



**HAL**  
open science

## Quantitative risk assessment of Haemolytic and Uremic Syndrome (HUS) from consumption of raw milk soft cheese

Subhasish Basak, Janushan Christy, Laurent Guillier, Frédérique Audiat-Perrin, Moez Sanaa, Fanny Tenenhaus-Aziza, Julien Bect, Emmanuel Vazquez

### ► To cite this version:

Subhasish Basak, Janushan Christy, Laurent Guillier, Frédérique Audiat-Perrin, Moez Sanaa, et al.. Quantitative risk assessment of Haemolytic and Uremic Syndrome (HUS) from consumption of raw milk soft cheese. Food and Ecological Systems Modelling Journal, 2024, 5, pp.e109502. 10.3897/fmj.5.109502. anses-04578958

**HAL Id: anses-04578958**

**<https://anses.hal.science/anses-04578958v1>**

Submitted on 22 May 2024





**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Quantitative risk assessment of Haemolytic and Uremic Syndrome (HUS) from consumption of raw milk soft cheese

Subhasish Basak<sup>1,2</sup>, Janushan Christy<sup>3</sup>, Laurent Guillier<sup>1</sup>, Frederique Audiat-Perrin<sup>1</sup>, Moez Sanaa<sup>1</sup>, Fanny Tenenhaus-Aziza<sup>4</sup>, Julien Bect<sup>1</sup>, Emmanuel Vazquez<sup>2</sup>

<sup>1</sup> ANSES, Maison-Alfort, France

<sup>2</sup> L2S, CentraleSupélec, Gif-sur-Yvette, France

<sup>3</sup> ACTALIA, La Roche-sur-Foron, France

<sup>4</sup> CNIEL, Paris, France

Corresponding authors: Subhasish Basak ([subhasish.basak@centralesupelec.fr](mailto:subhasish.basak@centralesupelec.fr)); Janushan Christy ([janushan.christy@gmail.com](mailto:janushan.christy@gmail.com));

Fanny Tenenhaus-Aziza ([ftenenhaus@cniel.com](mailto:ftenenhaus@cniel.com)); Emmanuel Vazquez ([emmanuel.vazquez@centralesupelec.fr](mailto:emmanuel.vazquez@centralesupelec.fr))

## Abstract

The aim of this quantitative risk assessment model is to estimate the risk of Haemolytic Uremic Syndrome (HUS) caused by Shiga-toxin producing *Escherichia coli* (STEC) in raw milk soft cheese and explore intervention strategies to minimise this risk. Building upon previous work from literature, the model considers microbial contamination of raw milk at the farm level, as well as STEC growth and survival during cheese production, ripening and storage, along with intervention strategies in both pre- and post-harvest scenarios. It allows for the assessment of intervention steps at the farm level or during cheese production. Besides estimating the risk of HUS, it also assesses the production losses associated with interventions.

**Key words:** *E. coli*, QMRA model, raw milk soft cheese, R programming language



Academic editor: Hernán Gómez Redondo

Received: 13 July 2023

Accepted: 06 February 2024

Published: 6 March 2024

**Citation:** Basak S, Christy J, Guillier L, Audiat-Perrin F, Sanaa M, Tenenhaus-Aziza F, Bect J, Vazquez E (2024) Quantitative risk assessment of Haemolytic and Uremic Syndrome (HUS) from consumption of raw milk soft cheese. Food and Ecological Systems Modelling Journal 5: e109502. <https://doi.org/10.3897/fmj.5.109502>

**Copyright:** © Subhasish Basak et al. This is an open access article distributed under terms of the Creative Commons Attribution License ([Attribution 4.0 International – CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

## Model metadata

This section summarises the metadata of the concerning QMRA model.

### General metadata

**Source:** RISK ASSESSMENTS: Risk assessments models.

**Identifier:** d56222fe-149a-4f86-bf0f-1972bb22ce59.

**Rights:** GNU General Public License 3.0.

**Availability:** Open access.

**Language:** English.

**Software:** R.

**Language Written In:** R 4.

### Product/matrix

**Name:** Dairy products Cattle.

**Description:** Soft mould ripened cheese.

**Unit:** g.

**Method:** origin Country - France.

## Hazard

- **type:** Microorganisms; **Name:** *Escherichia coli*, pathogenic, VTEC; **Description:** -; **unit:** -; **Adverse Effect:** Haemolytic Uremic Syndrome (HUS).

## Data background

**Identifier:** QRA-STEC.

**Title:** Quantitative risk assessment of Haemolytic and Uremic Syndrome (HUS) from consumption of raw milk soft cheese.

## Introduction

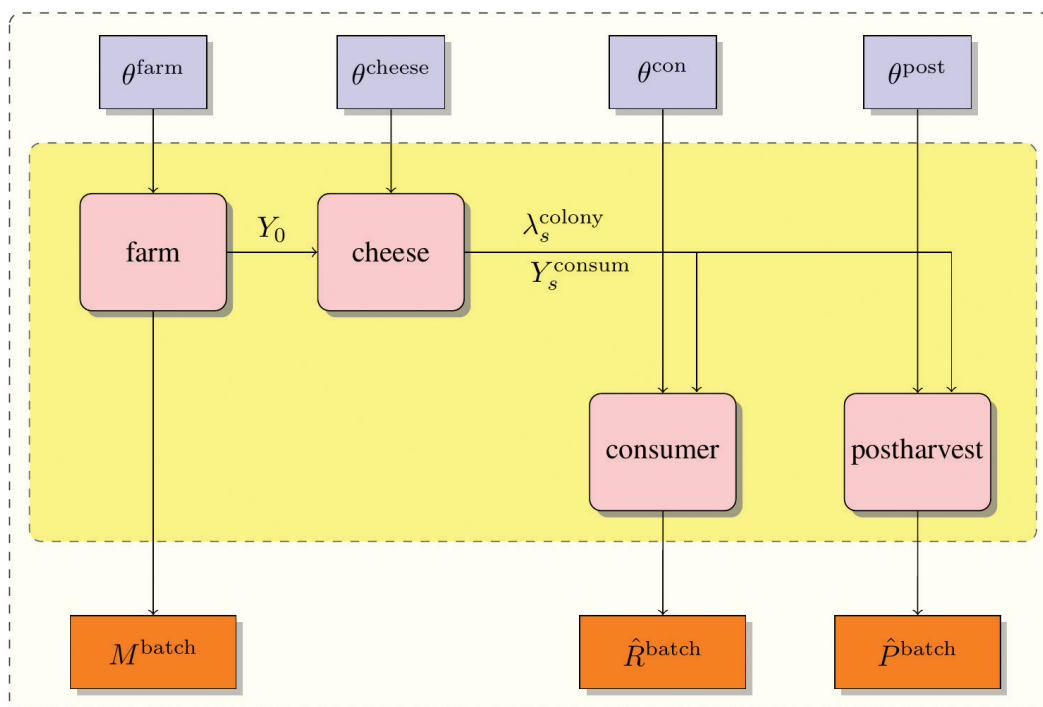
Microbiological food safety is a major challenge for the food sector (Plaza-Rodríguez et al. 2018). In this regard, the microbiological food safety community (i.e. food authorities, food industries and food research institutes) has invested great research efforts in the field of Quantitative Microbial Risk Assessment (QMRA). The framework for carrying out QMRA for food-borne pathogens is well established and relies on four components: hazard identification, hazard characterisation, exposure assessment and risk characterisation (F.A.O. and W.H.O. 2021). QMRAs have been employed to provide information for decision-making for the management of microbial risks (see, for example, Plaza-Rodríguez et al. (2018)).

The French partners of the ArtiSaneFood project, a European initiative aimed at improving the microbial safety of artisanal fermented foods in the Mediterranean Region, seek to optimise control measures to reduce the risk of food-borne illness from consuming soft cheese made from raw milk. While cheese is generally considered safe and nutritious, there are instances of food-borne illnesses related to its consumption, as noted by Sherkat et al. (1998). To address this issue, Perrin et al. (2014) investigated potential control measures to minimise the risk of Haemolytic and Uremic Syndrome (HUS) caused by Shiga-toxin producing *Escherichia coli* (STEC) in raw milk soft cheese. They developed a stochastic QMRA model to assess the risk of HUS associated with the five Main Pathogenic Serotypes<sup>1</sup> of STEC (MPS-STEC) in raw milk soft cheeses and a version of this model was implemented in SAS (SAS version 9.3 TS). In this work, we present an alternative implementation of the model using R language, along with several modifications to the original model.

The QMRA model is implemented as a stochastic simulator that is composed of several hierarchical levels that will be called modules. The simulator

---

<sup>1</sup> The Shiga-toxin producing *Escherichia coli* a.k.a STEC is a special kind of *E. coli* bacteria. There are five main pathogenic stereotypes of STEC a.k.a MPS-STEC, that are identified so far in Europe. These are O157:H7, O26:H11, O103:H2, O111:H8 and O145:H28. In our study, we assume only MPS-STEC is responsible for the disease Haemolytic Uremic Syndrome (HUS).



**Figure 1.** Schematic diagram of the batch level simulator of the risk assessment model. Modules are denoted by pink coloured boxes with the blue boxes denoting the set of corresponding input parameters  $\theta = \{\theta^{\text{farm}}, \theta^{\text{cheese}}, \theta^{\text{con}}, \theta^{\text{post}}\}$  and the orange boxes denoting the outputs, namely, milk loss per batch  $M^{\text{batch}}$ , probability of rejecting a particular batch  $\hat{P}^{\text{batch}}$  and batch risk  $\hat{R}^{\text{batch}}$ .

is divided into two parts— namely, the batch-level simulator and the output module. In the batch level, the simulator is composed of a farm module followed by a pre-harvest intervention step, a cheese production module, a consumer module and a post-harvest sampling module. This corresponds to the fabrication of a single batch of cheeses, that is produced from a single batch of milk coming from a fixed number of farms. Fig. 1 provides a schematic diagram of the batch level simulator that generates three outputs corresponding to a particular batch to which pre- and post-harvest interventions are applied: milk loss per batch, probability of rejecting a particular batch and risk of illness if that particular batch of cheese is consumed. It also simulates internal variables that characterise a particular batch, namely the STEC concentration  $Y_0$  in the aggregated milk tank, the median concentration of bacteria at the time of consumption  $Y_s^{\text{consum}}$  and the average number of colonies  $\lambda_s^{\text{consum}}$  in a single cheese.

The model is made available in Food Safety Knowledge Markup Language (FSK-ML) format to facilitate its reuse. This open format is based on predefined terms, metadata and controlled vocabulary to harmonise annotations of risk assessment models (Haberbeck et al. 2018). This article details the construction of the modules, including the underlying assumptions and the methods employed to estimate various quantities of interest. The subsequent section presents specifics of the modules as they were implemented in R and are available in the FSK-ML format. We also provide a comparative summary of the model in relation to the initial model proposed by Perrin et al. (2014). The article concludes with a section that addresses the scope and applicability of the model.

## QRA simulator

In this section, we describe the different modules of the simulator that represent the farm- to-fork continuum in the cheese fabrication process. It starts with a farm module which computes the STEC concentration (CFU/ml) in the aggregated milk tank that is used for cheese-making. This module includes a pre-harvest intervention step which performs „farm milk sorting“ or, more precisely, rejecting the contaminated tankers of milk with a concentration in *E. coli* above a given threshold. Next, we have the cheese production module which describes the evolution of STEC over different stages of cheese fabrication – namely, milk storage, moulding, draining, salting, ripening and cheese storage. Then the consumer module computes the risk of HUS for a particular batch of cheese, based on the consumption behaviour for different age groups. The post-harvest intervention module describes the sampling of cheese after production and computes the probability of rejecting a particular batch of cheese. These four modules, all together, constitute the batch level simulator, which simulates the output corresponding to the fabrication of a particular batch of cheeses. In this model, with the default parameters, a batch is usually composed of 22,000 to 23,000 cheeses of 250 g. For the computation of ultimate quantities of interests, the output module simulates several batches and computes the overall risk of HUS  $R^{\text{HUS}}$ , the average batch rejection rate  $P^{\text{avg}}$  and the average milk loss due to sorting  $M^{\text{avg}}$ .

### Farm module

The outcome of the farm module is the STEC concentration (CFU/ml) in the aggregated milk tank that collects milk from all the farms after pre-harvest milk sorting. The inputs of this module are denoted by  $\theta^{\text{farm}}$ . The model considers a fixed number of farms ( $N_{\text{farms}}$ ) with given hygiene conditions (controlled by parameters  $\alpha$  and  $\sigma^2$ , characterising the *E.coli* concentration in milk) and each of the farms has a certain number of cows<sup>3</sup>. Milk from each of the farms is collected in a bulk tank and called Bulk Tank Milk (BTM), which is subjected to pre-harvest intervention (milk sorting). The milk from all the sorted farms is then collected in an aggregated milk tank. The farm module simulates the concentration  $Y_0$  of STEC (CFU/ml) in this aggregated milk tank.

### Module inputs

The set of inputs parameters of the farm module denoted by  $\theta^{\text{farm}}$ , are described in Table 1.

---

2 We have used a Bayesian approach to estimate the values of the hygiene parameters in the farm module: this uses a Gibbs sampler to estimate posterior distribution of alpha and sigma, based on the *E. coli* data provided by ACTALIA/CNIEL and used by Perrin et al. (2014). In this implementation, all the 31 farms considered are assumed to have same hygiene conditions with respect to these hygiene parameters, which are taken as, respectively, the means of the estimated posterior distributions.

3 Number of cows: In the current implementation, the number of cows per each farm is sampled from the cow distribution data provided by ACTALIA/CNIEL and used by Perrin et al. (2014).

**Table 1.** Inputs of farm module.

Symbol	Description	default values/references
N_farms	Number of farms	31 *
N_cow_i	Number of cows in i-th farm	see ***
alpha_i	Parameter for distribution of <i>E. coli</i> in milk (LogNormal distribution)	-1.3 **
sigma_i	Parameter for distribution of <i>E. coli</i> in milk (LogNormal distribution)	2.9 **
a_weibull	Parameter for distribution of STEC in faecal matter	0.264
b_weibull	Parameter for distribution of STEC in faecal matter	16.288
mu_ecoli	Parameter for distribution of <i>E. coli</i> in faecal matter	6
tau_ecoli	Parameter for distribution of <i>E. coli</i> in faecal matter	0.3
mu_u	Parameter for distribution of probability infected cows	-0.927
tau_u	Parameter for distribution of probability infected cows	1.47411
q_milk	Average quantity of milk from a cow	25 litres
sorting_frequency	Milk testing frequency	10
sorting_lim	Max. limit of <i>E. coli</i> concentration <sup>50</sup>	50
p_MPS_STEC	Proportion of pathogenic STEC in cows	0.025

\* ACTALIA SAS Script: The SAS script used by CNIEL and developed by ACTALIA uses a set of parameter values for the implementation. In our work, we have considered it as a reference for several parameter values.

\*\* We have used a Bayesian approach to estimate the values of the hygiene parameters in the farm module: this uses a Gibbs sampler to estimate posterior distribution of alpha and sigma, based on the *E. coli* data provided by ACTALIA/CNIEL and used by Perrin et al. (2014). In this implementation, all the 31 farms considered are assumed to have same hygiene conditions with respect to these hygiene parameters, which are taken as, respectively, the means of the estimated posterior distributions.

\*\*\* Number of cows: In the current implementation, the number of cows per each farm is sampled from the cow distribution data provided by ACTALIA/CNIEL and used by Perrin et al. (2014).

## Module description

The concentration of STEC in farm milk is usually too low to be assessed quantitatively through microbiological methods (because of their limit of detection). For this reason, we rely on the „relative“ approach proposed by Perrin et al. (2014), which uses collected data<sup>4</sup> on *Escherichia coli* (*E. coli*) to estimate the STEC concentration. This approach is based on the assumption that *E. coli* and STEC strains follow the same faecal routes in the cows.

For each farm  $i = 1, 2, \dots, N^{\text{farms}}$ , the milk is collected into a bulk tank from  $N_i^{\text{cow}}$  cows, where  $N_i^{\text{cow}}$  is simulated from a cow distribution estimated from data<sup>5</sup>. The STEC concentration in the bulk tank milk (BTM) corresponding to farm  $i$  is denoted by  $Y_{0,i}$  and is computed using a proportion rule as:

$$Y_{0,i} = Y_i^{\text{EC}} \frac{\overline{F}_i^{\text{STEC}}}{\overline{F}_i^{\text{EC}}},$$

where  $Y_i^{\text{EC}}$  is the concentration of *E. coli* (CFU/ml) in bulk milk tank,  $\overline{F}_i^{\text{STEC}}$  is the average STEC concentration (CFU/gram) and  $\overline{F}_i^{\text{EC}}$  is the average *E. coli* concentration in faecal matter (CFU/gram) from all the cows.

- Pre-harvest intervention<sup>6</sup>

4 See the text of Footnote 2.

5 See the text of Footnote 3.

6 The pre-harvest intervention step does not implement the re-integration procedure of farms in the production process, once a particular farm is rejected. The typical process of re-integration involves conducting repeated tests on the milk from the farm over several days until it consistently shows no signs of contamination, ensuring the production of uncontaminated milk from the farm.

Bulk tank milk coming from each farm is tested for *E. coli* concentration and the farms with concentration  $Y_i^{EC}$  higher than a certain threshold  $I_{\text{sorting}}$  are rejected. Milk sorting is not performed for every batch of milk: instead, it is controlled by a parameter  $f^{\text{sorting}}$ , which denotes the frequency (in days) of testing the farms milk. Let  $S$  denote the set of farms that are qualified after milk sorting and let  $N^{\text{farms,sorted}} = |S|$ . The milk loss for a particular batch is given by  $M^{\text{batch}} = \sum_i^{N^{\text{farms}}} V_i \mathbf{1}_{\{i \in S\}}$ , where  $V_i = q_{\text{milk}} N_i^{\text{cow}}$  is the amount of milk produced by the  $i$ -th farm  $N_i^{\text{cow}}$ .<sup>7</sup> After milk sorting, milk from all the qualified bulk tanks is collected into a single aggregated milk tank. Hereafter, only the farms with accepted level of *E. coli* are considered.

For the  $i$ -th accepted farm,  $\overline{F}_i^{\text{STEC}}$  is the average of the individual STEC concentrations  $F_{ij}^{\text{STEC}}$  in infected cows  $j = 1, 2, \dots, k_i$ , where  $k_i$  is the number of infected cows in the  $i$ -th farm simulated according to a binomial distribution:

$$k_i \sim \text{Bin}(N_i^{\text{cow}}, p_i), \text{ where } \text{logit}(p_i) = u_i \text{ and } u_i \sim N(\mu_u, \tau_u).$$

The number of cows infected with MPS-STECS in the  $i$ -th farm is also simulated according to a binomial distribution:

$$k_i^{\text{MPS}} \sim \text{Bin}(k_i, p_{\text{STECS}}^{\text{MPS}}).$$

Remark: If we are interested to compute the concentration of MPS-STECS in the aggregated milk, we use  $k_i^{\text{MPS}}$  instead of  $k_i$  in the following computations:

$F_{ij}^{\text{STECS}}$  is simulated according to a Weibull distribution:

$$F_{ij}^{\text{STECS}} \sim \text{Weibull}(a^{\text{weibull}}, b^{\text{weibull}}),$$

$$\overline{F}_i^{\text{STECS}} = \frac{1}{N_i^{\text{cow}}} \sum_{j=1}^{k_i} F_{ij}^{\text{STECS}}.$$

The concentration of *E. coli* (CFU/ml) in BTM  $Y_i^{EC}$ , is modelled by a lognormal distribution:

$$Y_i^{EC} \sim \text{Lognormal}(a, \sigma).$$

$\overline{F}_i^{EC}$  is the average of individual *E. coli* concentration in faecal matter for each cow, denoted by  $F_{ij}^{EC}$ ,  $j = 1, 2, \dots, N_i^{\text{cow}}$  cow. It is simulated according to a Lognormal distribution:

$$\log_{10}(F_{ij}^{EC}) \sim N(\mu^{\text{ecoli}}, \tau^{\text{ecoli}}),$$

$$\overline{F}_i^{EC} = \frac{1}{N_i^{\text{cow}}} \sum_{j=1}^{N_i^{\text{cow}}} F_{ij}^{EC}.$$

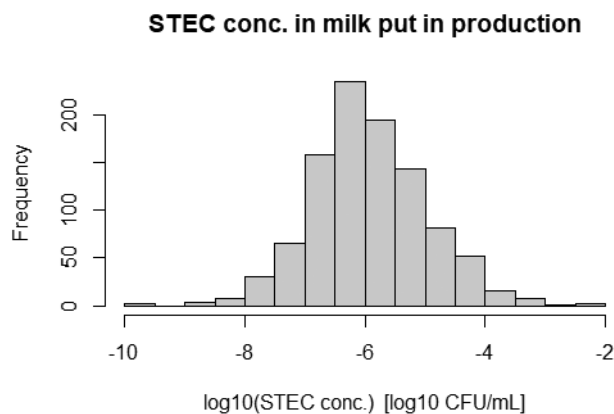
The STECS concentration  $Y_0$  (in CFU/ml) in this aggregated milk tank is then given by:

$$Y_0 = \sum_{i=1}^{N^{\text{farms}}} \left( Y_{0,i} \cdot \frac{V_i \mathbf{1}_{\{i \in S\}}}{\sum_{i=1}^{N^{\text{farms}}} V_i \mathbf{1}_{\{i \in S\}}} \right).$$

<sup>7</sup> See the text of Footnote 3.

## Outputs

The quantities of interest from the farm module are  $Y_0$ , the milk loss due to sorting, the  $M^{\text{batch}}$  total milk utilised and the number of farms discarded (due to sorting), i.e.  $N^{\text{farms}} - N^{\text{farms,sorted}}$ . Fig. 2 plots the histogram for the final STEC concentration in the aggregated milk tank without milk sorting testing.



**Figure 2.** Histogram of STEC (main pathogenic serotypes MPS-STECS) concentration ( $\log_{10}$  (CFU/ml)) in milk put into production.

## Cheese module

The cheese module begins with the input of the initial concentration  $Y_0$  of STEC in the aggregated milk tank and simulates the evolution of the STEC bacteria throughout the cheese-making process, which involves a transition from the liquid state of milk to the solid state of cheese.

### Module inputs

The inputs of the cheese module are the initial concentration  $Y_0$  of STEC and the parameters described in Table 2, which are denoted as  $\theta^{\text{cheese}}$ .

### Module description

The evolution of STEC involves six steps, with milk storage and moulding taking place during the liquid growth phase and draining and salting occurring during the solid growth phase. Ripening and cheese storage represent the (solid) decline phase of STEC.

In all the growth steps, the concentration  $y(t)$  for STEC at time  $t$ , is modelled using an ordinary differential equation:

$$\frac{dy}{dt} = \mu^{\text{max}}(t)y(t) \left(1 - \frac{y(t)}{y^{\text{max}}}\right)$$

where  $\mu^{\text{max}}(t)$  stands for the maximum growth rate (in  $\text{h}^{-1}$ ) and  $y^{\text{max}}$  is a parameter that represents the hypothetical maximum population of STEC strain in milk or cheese. For each step, the physico-chemical parameters  $\{d, pH, T, a_w\}$  are computed from Table III in Perrin et al. (2014). The maximum growth rate  $\mu^{\text{max}}(t)$  for a particular step, at time  $t$  is according to model 5 in Augustin et al. (2005).



**Table 2.** Parameters of cheese module  $\theta^{\text{cheese}}$ . Unless specified, the parameter values are taken from Perrin et al. (2014).

Symbol	Description	Values/reference
Parameters for mu_max	Parameters to compute mu_max	table I in Perrin et al. (2014)
mu_opt	Optimal growth rate	1.85 (average from Table II in Perrin et al. (2014))
y_max_milk	Hypothetical maximum population in milk	$10^9$ CFU/ml
y_max_cheese	Hypothetical maximum population in cheese	$10^5$ CFU/g
p_0157H7	Class probability of O157:H7	0.76 (taken from *)
p_MPS_STEC	Proportion of main pathogenic STEC	0.025
rho_0157H7	Parameter for decline phase of O157:H7	$0.14 \log_{10}$ CFU/day
rho_otherMPS	Parameter for decline phase of other MPS	$0.033 \log_{10}$ CFU/day
a_w	Water activity parameter	0.99
w_loss	Proportion of water loss	0.9
v_cheese	Milk used in a single cheese	2200
t_consum	Consumption time	Triangular (22,30,60)
d	Duration of a step	table III in Perrin et al. (2014)
pH	pH of a step	table III in Perrin et al. (2014)
T	Temperature of a step	table III in Perrin et al. (2014)
NA	Absolute tolerance of ode solver	$10^{-6}$
NA	Maximal step size of ode solver	0.01

\* ACTALIA SAS Script: The SAS script used by CNIEL and developed by ACTALIA uses a set of parameter values for the implementation. In our work, we have considered it as a reference for several parameter values.

$\mu^{\text{max}}(t)$  is dependent on the values of physico-chemical parameters  $\{d, pH, T, a_w\}$  at time  $t$  and the optimal growth rate parameter  $\mu^{\text{opt}}$ . The optimal growth rate is taken to be the average of the growth rates corresponding to different strains as provided in Table II in Perrin et al. (2014).

The milk storage step starts with initial concentration  $Y_0$  and outputs the final concentration  $Y^{\text{storage}}$  after growth at the end of storage. The moulding step takes  $Y^{\text{storage}}$  as input to model the corresponding growth of the bacteria during moulding and outputs the final concentration  $Y^{\text{molding}}$ . After moulding, the milk is converted into solid cheese and the STEC bacteria from colonies inside a single cheese made from a volume  $v^{\text{cheese}}$  (ml) of milk. The number of colonies in a cheese is modelled as a Poisson variable  $N_s^{\text{colony}} \sim \text{Poisson}(\lambda_s^{\text{colony}})$ , for each strain  $s$  of MPS-STECS  $s \in \{0157, \overline{0157}\}$ , with the following mean parameters:

$$\lambda_{0157}^{\text{colony}} = Y_{\text{molding}} \cdot v_{\text{cheese}} \cdot w_{\text{loss}} \cdot p_{0157} \cdot p_{\text{MPS}}^{\text{STEC}}$$

$$\lambda_{\overline{0157}}^{\text{colony}} = Y_{\text{molding}} \cdot v_{\text{cheese}} \cdot w_{\text{loss}} \cdot (1 - p_{0157}) \cdot p_{\text{MPS}}^{\text{STEC}}$$

Remark: The factor  $p_{\text{MPS}}^{\text{STEC}}$  takes into account only the concentration of pathogenic STEC bacteria (MPS-STECS). Alternatively, this can be taken into account by directly computing the concentration of MPS-STECS in the aggregated milk tank as an output of the farm module using  $k_i^{\text{MPS}}$  instead of  $k_i$ . The FSKX implementation provides a flag **flag\_MPS = FALSE** that allows the user to directly compute the concentration of MPS-STECS in the farm module using the random proportion of MPS-STECS infected cows in a farm.

Starting from the draining phase, the evolution of the size of colony, stemming from one immobilised STEC cell, is studied. The draining phase commences

with an initial colony size of 1 CFU and growth continues until the salting phase. It is assumed that the evolution of each colony inside each cheese (weighing 250 g) is identical during these phases, since they have the same environmental conditions. The output of the draining and salting steps are called  $Y^{\text{draining}}$  and  $Y^{\text{salting}}$ , respectively.

The growth in colony size stops after the salting phase. Then starts the decline phase, which is composed of two steps—namely, ripening and cheese storage. The ripening phase lasts until the 14<sup>th</sup> day of cheese production and the cheese storage phase duration depends on the consumption time  $t^{\text{consum}}$  of the cheese. The rate of decline is different for different strains—namely, MPS (O157, O157) and non-MPS STEC. The median number of bacteria per colony, for a particular strain after decline, at the end of the ripening phase and at the time of consumption, is computed as:

$$Y_s^{\text{ripening}} = Y_s^{\text{salting}} \cdot 10^{-\rho_s \times (14 \times 24 - t) / 24},$$

$$Y_s^{\text{consum}} = Y_s^{\text{salting}} \cdot 10^{-\rho_s \times (t^{\text{consum}} \times 24 - t) / 24}.$$

where  $t$  the time (in hours) taken until the salting step.

The expected number of colonies  $\lambda_s^{\text{colony}}$  is adjusted taking the inactivation during the ripening and cheese storage phases. For instance if the median number of bacteria for a particular strain  $Y_s^{\text{consum}}$  falls below 1, it signifies that the colony might have disappeared with probability  $Y_s^{\text{consum}}$  and the corresponding expected number of colonies is obtained by multiplying  $\lambda_s^{\text{colony}}$  by  $Y_s^{\text{consum}}$ . In case of  $Y_s^{\text{consum}} > 1$ , the expected number of colonies remains unchanged.

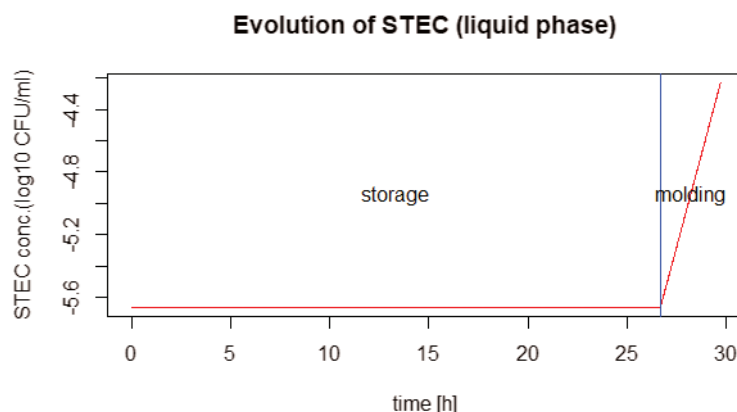
$$\lambda_s^{\text{colony}} \leftarrow \lambda_s^{\text{colony}} \cdot \min(1, Y_s^{\text{consum}}).$$

## Outputs

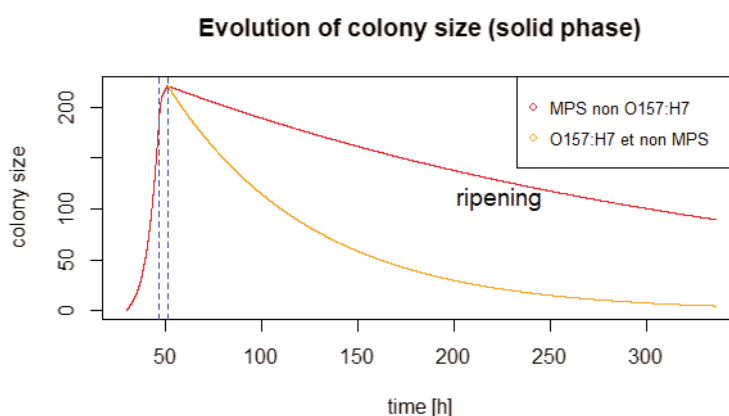
The outputs of interest for the cheese module are the average number of colonies  $\lambda_s^{\text{colony}}$  and the median colony size at the time of consumption  $Y_s^{\text{consum}}$ , where the subscript  $s$  denotes the strain {O157, O157}. Fig. 3 shows the growth of STEC concentration (CFU/ml) starting from  $Y_0 = 10^{-3.8}$  CFU/ml, during the storage and moulding steps, for the baseline scenario (with no interventions and default values of the parameters). After the transition from liquid to solid state, the evolution of a single colony is shown in Fig. 4. The colonies grow during the draining and salting step and then decline during the ripening and cheese storage steps.

## Consumer module

In this section, we describe the module of the batch level simulator that computes the risk of HUS for a particular batch. Given the outputs of the cheese module, i.e. the average number of colonies  $\lambda_s^{\text{colony}}$  and the median number of bacteria at the end of the decline phase  $Y_s^{\text{consum}}$ , the consumer module estimates the risk of HUS, based on the raw milk cheese consumption behaviour of people in different age groups  $\theta^{\text{con}}$ .



**Figure 3.** Evolution of STEC (main pathogenic serotypes MPS-STECS) in  $\log_{10}$  CFU/ml during the storage and moulding step. The blue vertical line shows the end of the storage phase.



**Figure 4.** Evolution of STEC colony size during draining, salting and ripening of cheese fabrication. The decline rate for the MPS O157:H7 strain and non-MPS strains are equal (orange line) and significantly higher than the decline rate of MPS non-O157:H7 strain (red line). The three phases, namely, draining, salting and ripening are separated by vertical blue dotted lines.

## Module description

The dose  $\Gamma$ , corresponding to the concentration of MPS-STECS per 25 g of cheese serving, is computed as:

$$\Gamma = \sum_S N_{s,\text{sample}}^{\text{colony}} Y_S^{\text{colony}}$$

where  $N_{s,\text{sample}}^{\text{colony}}$  denotes the number of colonies in a cheese serving of weight  $w^{\text{serving}}$  g, which is a Poisson random variable with mean  $\lambda_s^{\text{colony}} \times w^{\text{serving}} / w^{\text{cheese}}$ .  $Y_S^{\text{colony}}$  denotes the size of all colonies in a cheese, which follows a Log-Normal distribution:

$$Y_S^{\text{colony}} = Y_S^{\text{consum}} \cdot 10^{\epsilon_s},$$

where  $\epsilon_s \sim N(0, \tau_{\epsilon_s})$  represents inter-batch variability (i.e. variability between different cheeses in a single batch). Note that here we assume all the colonies inside a single cheese, for a particular batch and a given strain  $s$ , have identical size  $Y_S^{\text{colony}}$ .

The average number of colonies  $\lambda_s^{\text{colony}}$  depends on several random quantities, such as the initial STEC concentration  $Y_0$ ,  $t^{\text{storage}}$ ,  $d^{\text{storage}}$  and  $t^{\text{consum}}$  (see

Table 2). The median colony size  $Y_s^{\text{consum}}$  depends on the random variable  $t^{\text{consum}}$ . In summary, the dose  $\Gamma$  is a combination of several Poisson and Log-Normal random variables corresponding to different strains and their parameters are dependent on the random quantities  $\xi^{\text{dose}} = \{Y_0, t^{\text{storage}}, d^{\text{storage}}, t^{\text{consum}}\}$ . The probability of HUS (risk) associated with the ingestion of a dose  $\gamma$ , for age  $a$  is:

$$P(\text{HUS}|a,\gamma) = 1 - (1 - r_a)^\gamma,$$

where  $r_a = r_0 \times e^{-ka}$  for age  $a$  varying from 1 to 15. The risk of HUS for a particular batch of milk, conditional on  $\xi^{\text{dose}}$ , can be derived by integrating w.r.t  $\Gamma|\xi^{\text{dose}}$ :

$$R^{\text{batch}} = \sum_{a=1}^{15} g(a) \int_0^\infty [1 - (1 - r_a)^\gamma] p(\gamma|\xi^{\text{dose}}) d\gamma,$$

where  $p(\gamma|\xi^{\text{dose}})$  is the conditional probability density function of  $\Gamma$  given  $\xi^{\text{dose}}$  and  $g(a)$  is the proportion of cheese consumed by the age group  $a$ .

### Module inputs

The inputs of the consumer module are the average number of colonies  $\lambda_s^{\text{colony}}$ , the median colony size  $Y_s^{\text{consum}}$  and other parameters denoted by  $\theta^{\text{con}}$ , which are listed in Table 3.

**Table 3.** Parameters of the consumer module  $\theta^{\text{con}}$ . Unless specified, the parameter values are taken from Perrin et al. (2014).

Symbol	Description	Values/references
k	Parameter for risk computation	0.38
r_0	Parameter for risk computation	$10^{-2.33}$
wt_cheese	weight of a single cheese	250 gm
wt_serving	weight of a single serving	25 gm
g(a)	Proportion of cheese consumed by age group $a$	taken from *
tau_eps_0157H7	Parameter for inter cheese variability	0.000279659 (taken from *)
tau_eps_otherMPS	Parameter for inter cheese variability	0.000065399 (taken from *)
a_max	Maximum age group	15
N_dose	Monte Carlo sample size	0

\* ACTALIA SAS Script: The SAS script used by CNIEL and developed by ACTALIA uses a set of parameter values for the implementation. In our work, we have considered it as a reference for several parameter values.

### Numerical methods for the estimation of the batch risk

For a given set of input parameters  $\xi^{\text{dose}}$  the conditional risk  $R^{\text{batch}} = P(\text{HUS}|\xi^{\text{dose}})$  can be computed using simple Monte Carlo using i.i.d. samples from  $p(\Gamma|\xi^{\text{dose}})$ . The number of Monte Carlo samples is given by the numerical parameter  $N^{\text{dose}}$ . Note that the randomness in  $\Gamma|\xi^{\text{dose}}$  is only due to the fact that the dose is a combination of several pairs of a Poisson and Lognormal random variables.

Alternatively, the batch risk can be computed by approximation of the integral  $\mathbb{E}[(1 - r_a)^\Gamma]$ . Since the two classes of MPS-STECS are independent, we can write the dose as a sum  $\Gamma = \Gamma_1 + \Gamma_2$  with:

$$\Gamma_s = Y_s^{\text{consum}} \cdot N_s^{\text{colony}} \cdot 10^{t_{\text{es}} v_s} = d_s \cdot N_s^{\text{colony}} \cdot b_s^{v_s},$$

where  $v_s$  is a standard normal variable. Now the expectation can be written as  $\mathbb{E}[(1 - r_a)^r] = \prod_s \mathbb{E}[\mathbb{E}[(1 - r_a)^{r_s} | v_s]]$ . Using the result in <sup>8</sup>, the expectation can be further reduced to:

$$\mathbb{E}[(1 - r_a)^r] = \exp\left(\sum_s \lambda_s^{\text{colony}}\right) \prod_s \mathbb{E}\left[c_{s,1}^{c_{s,3}^v} c_{s,2}^{c_{s,2}^v}\right]$$

where  $c_{s,1} = \exp(\lambda_s^{\text{colony}})$ ,  $c_{s,2} = (1 - r_a)^{d_s}$  and  $c_{s,3} = b_s$ . Since  $c_{s,1} > 1$ ,  $c_{s,2} < 1$  and  $c_{s,3} > 1$ , the function

$$(c_{s,1})^{(c_{s,2})^{c_{s,3}^v}}$$

is monotone (non-increasing) in  $v_s$ . For such functions, deterministic quadrature methods (such as the trapezoidal rule) have convergence rate  $O(n^{-1})$ , which is better than simple Monte Carlo (see, for example, Novak (1992)).

The choice of the method for computing the batch risk is determined through the parameter value  $N^{\text{dose}}$ . For a non-zero integral value, the simulator uses the simple Monte Carlo method and, if  $N^{\text{dose}}$  is set to zero, it used the integral approximation method.

The current implementation also allows the user to compute the conditional batch risk  $\xi^{\text{dose}} \setminus t^{\text{consum}}$  averaged out with respect to the consumption time. This is done by using the flag **flag\_consum = TRUE** that computes the integration of  $R^{\text{batch}}$  with respect to  $t^{\text{consum}}$ . The integral  $I = \mathbb{E}_{t^{\text{consum}}}[R^{\text{batch}}(t^{\text{consum}})]$  is approximated using a piecewise constant function on a regular grid  $\{t_1^{\text{consum}}, t_2^{\text{consum}}, \dots, t_n^{\text{consum}}\}$  of size  $n$ , that spans the support of the triangular distribution of the variable  $t^{\text{consum}}$ . For each interval, the value of the piecewise constant function is taken equal to the value of  $R^{\text{batch}}$  at the left endpoint of that interval and the approximation is computed as:

$$\hat{I} = \sum_{i=1}^{n-1} P[t_i^{\text{consum}} \leq t^{\text{consum}} \leq t_{i+1}^{\text{consum}}] \cdot R^{\text{batch}}(t_i^{\text{consum}}).$$

For STEC, due to the inactivation during the post-ripening storage phase,  $R^{\text{batch}}$  turns out to be a decreasing function of  $t^{\text{consum}}$ , which slightly overestimates the risk with this approach. It is worth noting that the computation of

<sup>8</sup> The integral inside  $R^{\text{batch}}(y_0)$  can be written as  $\mathbb{E}[1 - (1 - r_a)^{N_s^{\text{colony}} \cdot C}]$ .

$$\begin{aligned} &= \sum_{N_s^{\text{colony}}=0}^{\infty} \left[1 - (1 - r_a)^{N_s^{\text{colony}} \cdot C}\right] e^{-\lambda_s^{\text{colony}}} \frac{(\lambda_s^{\text{colony}})^{N_s^{\text{colony}}}}{N_s^{\text{colony}}!} \\ &= \sum_{N_s^{\text{colony}}=0}^{\infty} e^{-\lambda_s^{\text{colony}}} \frac{(\lambda_s^{\text{colony}})^{N_s^{\text{colony}}}}{N_s^{\text{colony}}!} - \sum_{N_s^{\text{colony}}=0}^{\infty} (1 - r_a)^{N_s^{\text{colony}} \cdot C} e^{-\lambda_s^{\text{colony}}} \frac{(\lambda_s^{\text{colony}})^{N_s^{\text{colony}}}}{N_s^{\text{colony}}!} \\ &= 1 - \sum_{N_s^{\text{colony}}=0}^{\infty} \frac{\{(1 - r_a)^C \lambda_s^{\text{colony}}\}^{N_s^{\text{colony}}}}{N_s^{\text{colony}}!} \frac{e^{-\{(1 - r_a)^C \lambda_s^{\text{colony}}\}} \cdot e^{-\lambda_s^{\text{colony}}}}{e^{-\{(1 - r_a)^C \lambda_s^{\text{colony}}\}}} \\ &= 1 - e^{-\lambda_s^{\text{colony}} [1 - (1 - r_a)^C]} \sum_{N_s^{\text{colony}}=0}^{\infty} e^{-\{(1 - r_a)^C \lambda_s^{\text{colony}}\}} \frac{\{(1 - r_a)^C \lambda_s^{\text{colony}}\}^{N_s^{\text{colony}}}}{N_s^{\text{colony}}!} \\ &= 1 - e^{-\lambda_s^{\text{colony}} [1 - (1 - r_a)^C]} \end{aligned}$$

$R^{\text{batch}}$  for a grid of values is not very expensive compared to the computation of the same for a single value of  $t^{\text{consum}}$ , since both of them require a single run to the cheese module that can compute all the necessary outputs,  $\lambda_s^{\text{colony}}$  and  $Y_s^{\text{consum}}$  corresponding to the consumption time points.

## Output

The output of the consumer module is the estimated batch risk of HUS. For a given set of input parameters, Fig. 6 plots the relative risk.

## Post-harvest: Cheese sampling

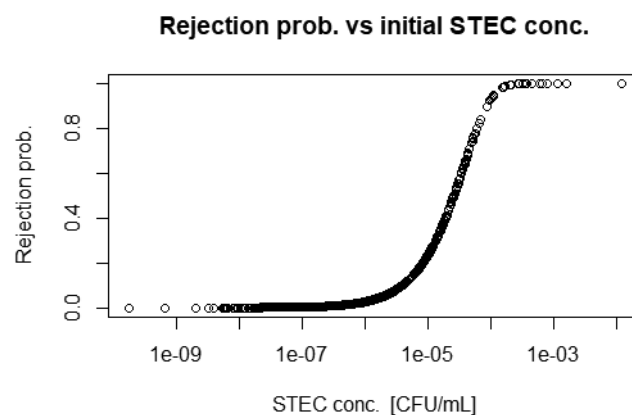
The post-harvest step, also known as the sampling step, can be carried out at various stages of cheese production, depending on the type of bacteria. For STEC, this step is conducted at the end of the ripening phase, more precisely at the 14<sup>th</sup> day of production by default. However the current implementation allows us to change this parameter with a minimum value of 3 days and a maximum value of 14 days. During this step, the batch of cheese produced is examined for MPS-STE $C$  contamination by taking small portions of cheese samples from the batch. Once a single sample unit tests positive, the entire batch of cheese is rejected, meaning that the specific batch does not enter into the calculation of the overall risk of HUS.

## Module inputs

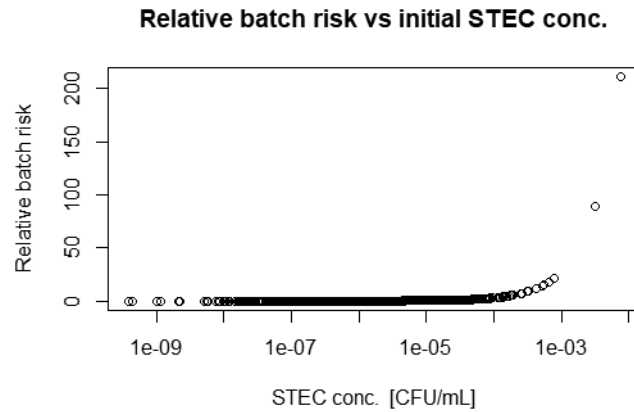
The inputs of the post-harvest module are initial STEC concentration in milk  $Y_0$ , average number of colonies  $\lambda_s^{\text{colony}}$  and the parameters denoted by  $\theta^{\text{post}}$ , listed in Table 6.

## Module description

We assume that the colonies are homogeneously distributed inside a cheese, a colony contains at least one MPS-STE $C$  cell and the test is accurate enough to detect a colony with a single MPS-STE $C$  cell. We observe that probability of a sample unit testing positive is  $P^{\text{sample}} = P[N_{\text{sample}}^{\text{colony}} > 0]$ , where  $N_{\text{sample}}^{\text{colony}}$  is the total number of colonies from all strains. Since the number of cells corresponding



**Figure 5.** Batch rejection probability as a function of the initial STEC (main pathogenic serotypes MPS- STE $C$ ) concentration (CFU/ml).



**Figure 6.** The relative batch risk (with respect to a baseline risk value) is plotted as a function of the initial STEC (main pathogenic serotypes MPS-STECS) concentration (CFU/ml).

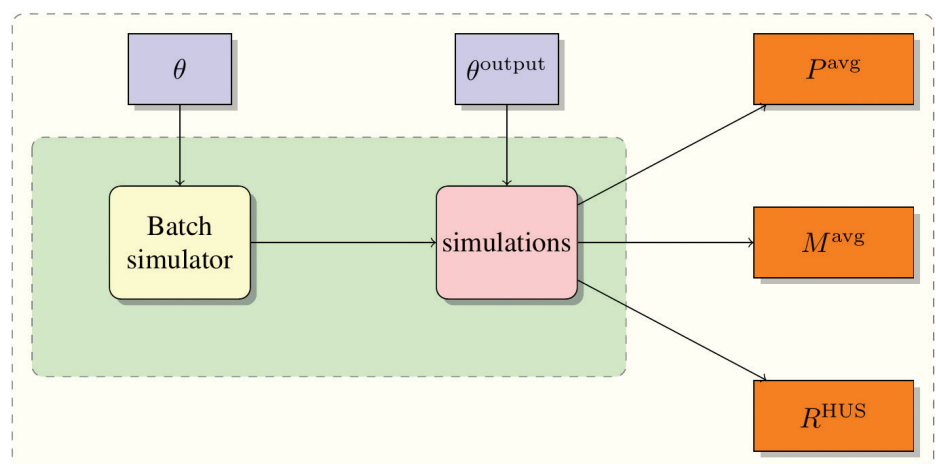
to two different strains grow independently, the total number of colonies is also a Poisson random variable,  $N_{\text{sample}}^{\text{colony}} \sim \text{Poisson}((\lambda_1^{\text{colony}} + \lambda_2^{\text{colony}}) \cdot m_{\text{sample}}/\text{wt}^{\text{cheese}})$ . For given values of  $\xi^{\text{dose}}$  using this Poisson distribution, we compute the probability of rejecting a particular batch  $P^{\text{reject}} = 1 - (1 - P_{\text{sample}})^{n^{\text{sample}}}$ .

### Output

The output of the post-harvest module is the probability of rejection. Fig. 5 plots the probability of rejecting  $P^{\text{reject}}$  with fixed  $m_{\text{sample}} = 25$  and  $n_{\text{sample}} = 5$ , for  $10^3$  batches corresponding values of  $y_0$ .

### Output module

The output module computes the overall risk of HUS from MPS-STECS and other quantities required to assess the analytical cost corresponding to the intervention steps. This module is outside the batch-level simulator that computes the quantities of interest corresponding to the fabrication of a particular batch of cheese (see Fig. 7).



**Figure 7.** Output module.

## Module input

The inputs of the output module are the outputs of the farm, consumer and post-harvest module along with the parameters denoted by  $\theta^{\text{output}}$ , listed in Table 4.

**Table 4.** Inputs of output module  $\theta^{\text{output}}$ .

Symbol	Description	Values/reference
N_batch	Monte Carlo sample size	1
p_test	Proportion of cheese batch tested	0.5

## Model description

The output module simulates several batches and produces estimates of the overall risk  $R^{\text{HUS}}$ , the probability of rejection  $P^{\text{avg}}$  and the milk loss  $M^{\text{avg}}$ . We define the average milk loss  $M^{\text{avg}} = \mathbb{E}[M^{\text{batch}}]$ , average batch rejection rate:

$$P^{\text{avg}} = \mathbb{E}[P^{\text{batch}} \cdot p^{\text{test}}] = \int_{\xi^{\text{dose}}} P^{\text{batch}} \cdot p^{\text{test}} \cdot p(\xi^{\text{dose}}) d\xi^{\text{dose}},$$

with  $p^{\text{test}}$  being the proportion of batch tested and the average over batch risk:

$$R^{\text{avg}} = \mathbb{E}[R^{\text{batch}} \cdot (1 - P^{\text{batch}} \cdot p^{\text{test}})],$$

where the batch risk is set to zero for rejected batches.

The overall risk of HUS is conditional on the event that the batch actually goes into the market, i.e. not rejected. It is computed by dividing  $R^{\text{avg}}$  by the probability that a produced batch is not rejected:

$$R^{\text{HUS}} = \frac{\int_{\xi^{\text{dose}}} R^{\text{batch}} \cdot (1 - P^{\text{batch}} \cdot p^{\text{test}}) \cdot p(\xi^{\text{dose}}) d\xi^{\text{dose}}}{1 - P^{\text{avg}}}.$$

The quantities are estimated using simple Monte Carlo with sample size  $N^{\text{batch}}$ . We use i.i.d. samples  $\{\xi_1^{\text{dose}}, \xi_2^{\text{dose}}, \dots, \xi_{N^{\text{batch}}}^{\text{dose}}\}$ , drawn from the conditional distribution  $p(\xi^{\text{dose}})$  to construct the unbiased simple Monte Carlo estimator:

$$\hat{P}^{\text{avg}} = \frac{1}{N^{\text{batch}}} \sum_{l=1}^{N^{\text{batch}}} \hat{P}^{\text{batch}}(\xi_l^{\text{dose}}) \cdot p^{\text{test}},$$

$$\hat{R}^{\text{avg}} = \frac{1}{N^{\text{batch}}} \sum_{l=1}^{N^{\text{batch}}} \hat{R}^{\text{batch}}(\xi_l^{\text{dose}}) \cdot (1 - \hat{P}^{\text{batch}}(\xi_l^{\text{dose}}) \cdot p^{\text{test}}).$$

The current implementation of the output module computes the relative risk of HUS with respect to a baseline scenario with no intervention steps. This quantity is obtained by dividing  $R^{\text{HUS}}$  by the risk of HUS in the baseline scenario (the baseline risk is also estimated by simple Monte Carlo with sample size  $N^{\text{batch}}$ ).

## Output

The output module returns the relative risk of HUS, average proportion of rejected batches  $P^{\text{avg}}$  and average milk loss  $M^{\text{avg}}$ .



## Comparison with the previous implementation

The R implementation proposed in this article differs in several respects from the QMRA model originally proposed by Perrin et al. (2014). The modifications have been validated by experts from ANSES, CNIEL and L2S. In this section, we summarise the differences.

### Modifications in the farm module

In the farm module, the hyper-parameters of the distribution of the concentration of *E. coli* (CFU/ml) in bulk milk tank have been estimated using a Bayesian approach, based on a hierarchical Poisson mixed model as described by Equation 4 in Perrin et al. (2014). The approach is based on the *E. coli* data provided by<sup>9</sup> and uses a Gibbs sampler to derive the posterior distribution of  $\alpha$  and  $\sigma$ .

The proportion of MPS infected cows is considered to be  $p_{MPS}^{STEC}$ . Previously, this proportion was multiplied with the average colony size of STEC after the moulding step in order to obtain the average colony size of MPS-STEC. In the current implementation, this proportion is used directly inside the farm module to simulate the number of MPS infected cows in each farm, that allows us to simulate directly the concentration of MPS-STEC as an output of the farm module.

As an additional metric of cost of intervention, this model computes the average quantity of milk lost due to pre-harvest milk sorting.

### Modifications in the cheese and consumer module

The variable  $u_i$  in Equation (7) of Perrin et al. (2014), denoting intra-cheese variability is set to zero. In other words, all the colonies inside a particular cheese have an identical colony size.

The volume of milk used to produce a single cheese is taken (default value) as 2.2 litres instead of 2.5 litres. However, this value can be changed depending on the production scenario.

In the current implementation, the batch risk is computed at the time of consumption which includes the inactivation (decline in concentration) during the cheese storage phase. This is different from the batch risk computed at the end of production which does not take into account the decline in bacteria population during the cheese storage. The duration of this phase is modelled as a Triangular distribution with more recent and updated values of the parameters as shown in Table 2.

## Usage and applicability

The FSKX implementation (Suppl. material 1) allows the user to execute the simulator on KNIME, using a set of input parameters listed in Table 5. By suitably adjusting the input parameter **cm\_n\_batch**, the user can run the FSKX implementation to either simulate a single batch (by setting **cm\_n\_batch** = 1) or

<sup>9</sup> ACTALIA SAS Script: The SAS script used by CNIEL and developed by ACTALIA uses a set of parameter values for the implementation. In our work, we have considered it as a reference for several parameter values.

**Table 5.** Default simulation settings.

<b>defaultSimulation</b>	
fm_N_farms	31
fm_q_milk	25
fm_sorting_freq	10
fm_sorting_lim	50
fm_mu_u	-0.927
fm_tau_u	1.47411
fm_a_weibull	0.264
fm_b_weibull	16.288
fm_mu_ecoli	6
fm_tau_ecoli	0.3
cm_mu_max_T_min	5.5
cm_mu_max_T_opt	40.6
cm_mu_max_T_max	48.1
cm_mu_max_pH_min	3.9
cm_mu_max_pH_opt	6.25
cm_mu_max_pH_max	14
cm_mu_max_aw_min	0.9533
cm_mu_max_aw_opt	0.999
cm_mu_max_mu_opt	2.03
cm_w_activity	0.99
cm_rho_0157H7	0.14
cm_rho_otherMPS	0.033
cm_y_max_milk	1e+09
cm_y_max_cheese	1e+05
cm_storage_duration	12
cm_storage_duration_min	1
cm_storage_duration_max	40
cm_storage_duration_mode	12
cm_storage_temperature	5
cm_storage_temperature_min	1
cm_storage_temperature_max	6
cm_p_0157H7	0.76
cm_p_MPS_STEC or fm_p_MPS_STEC	0.025
cm_mu_eps_0157H7	0
cm_tau_eps_0157H7	0.000279659
cm_mu_eps_otherMPS	0
cm_tau_eps_otherMPS	6.5399e-05
cm_molding_duration	3
cm_draining_duration	17
cm_salting_duration	4.5
cm_consumption_time_min	22
cm_consumption_time_max	60
cm_consumption_time_mode	30
cm_v_cheese	2200
cm_w_loss	0.9
cm_wt_cheese	250
cm_wt_serving	25
cm_m_sample	25
cm_n_sample	5
cm_k	0.38
cm_r0	1e-2.33
cm_age_max	14
cm_p_test	0.5

defaultSimulation	
cm_d_test	14
cm_n_dose	0
cm_n_batch	1
flag_consum	TRUE
flag_MPS	FALSE

**Table 6.** Parameters of the post-harvest module  $\theta^{\text{post}}$ . Unless specified, the parameter values are taken from \*.

Symbol	Description	Values/reference
n_sample	Number of test portions	5
m_sample	Mass of each test portion	25 gm
d_test	Post-harvest sampling day	14 days

\* ACTALIA SAS Script: The SAS script used by CNIEL and developed by ACTALIA uses a set of parameter values for the implementation. In our work, we have considered it as a reference for several parameter values.

multiple independent batches (by setting **cm\_n\_batch** > 1) to estimate the ultimate quantities of interest. Simulation of a single batch produces three numerical outputs, namely, the STEC concentration (CFU/ml) in milk put in production, the amount of milk loss (in litres) due to testing and the probability of rejecting the cheese batch. This also produces graphical representation of the evolution of STEC and colonies during cheese fabrication (both solid and liquid phase) and the evolution of the bacteria growth rate over the different phases of production. On the other hand, when multiple batches are simulated, it produces the estimates of the ultimate quantities of interest, averaged over these batches, namely, the relative risk of HUS computed with respect to a baseline scenario (with no intervention steps), the average milk loss (in litres) and the average probability of rejecting a cheese batch after production. The corresponding graphical outputs show the distribution of STEC concentration (CFU/ml) in aggregated milk tank, the relative batch risk of HUS (computed with respect to the baseline risk) and the relative batch risk and batch rejection probability as a function of initial STEC concentration, as shown in Figs 5, 6. The baseline scenario—i.e. the cheese production without any intervention step—can be simulated by appropriately choosing the parameters  $p^{\text{test}} = 0$  and  $f^{\text{sorting}} = \infty$ .

This article presents a QMRA model that offers a scientific approach to simulate the real-life scenarios encountered during the production of raw milk soft cheese. The model builds upon the work of Perrin et al. (2014), as well as inputs from ANSES, CNIEL, ACTALIA and L2S. The primary goal of this model is to conduct optimisation studies for the process intervention parameters and to make recommendations to cheese producers. It is essential to note that the outputs obtained using the simulator, such as the batch risk, loss of milk and proportion of rejected cheese batches, are just the estimates of a hypothetical scenario simulated with a state-of-the-art<sup>10</sup> QMRA model available for raw milk cheese. These quantities can be used to improve the intervention processes and also study the effects of different input parameters in the production process, but should not be interpreted as the actual prevalence of HUS observed in reality.

<sup>10</sup> See Footnote 6.

**Table 7.** QMRA model assumptions.

Assumptions	Significance	Comments
Homogeneous distribution of colonies inside a cheese.	The distribution of colonies inside a cheese impacts the post-harvest sampling step. This assumption is used to simplify the cheese testing step, which assures that, if the cheese is contaminated, it is always tested positive.	This overestimates the detection probability when colonies are clustered.
Identifying MPS-STEC as the unique HUS-causing hazard.	It was not taken into account that certain non-MPS-STEC strains can also cause HUS and that, within MPS-STEC, some of the strains maybe less virulent.	Some recent publications, see, for example, Auvray et al. (2023) suggest that MPS could be based on <i>stx</i> subtypes rather than serotypes. The current level of knowledge does not allow us to determine the prevalence, based on this new definition.
STEC and <i>E. coli</i> follow the same intestinal route in the farm animal.	The pre-harvest intervention step is based on this assumption. The milk sorting is carried out using the <i>E. coli</i> concentration in the farm milk.	This assumption is based on Perrin et al. (2014).
No intra-cheese variability.	All the colonies inside a single cheese are of same colony size.	

### Model assumptions

According to F.A.O. and W.H.O. 2021 „*Models are always incomplete representations of the system they are intended to model, but they can still be useful*“. The QMRA model for STEC is based on several assumptions which are quite common in microbiological risk modeling, however some of the important assumptions are listed below in Table 7.

### Additional information

#### Conflict of interest

The authors have declared that no competing interests exist.

#### Ethical statement

No ethical statement was reported.

#### Funding

This work is part of the ArtiSaneFood project (grant number: ANR-18-PRIM-0015) which is part of the PRIMA programme supported by the European Union.

#### Author contributions

Subhasish Basak: Writing – Original Draft, Writing – Review & Editing; Janushan Christy, Laurent Guillier, Fanny Tenenhaus-Aziza, Julien Bect, Emmanuel Vazquez: Methodology, Writing – Reviewing and Editing; Fanny Tenenhaus-Aziza: Project coordinator; Moez Sanaa, Frédérique Audiat-Perrin: Writing Reviewing and Editing.

#### Author ORCIDs

Subhasish Basak  <https://orcid.org/0000-0003-3522-1236>

Janushan Christy  <https://orcid.org/0000-0003-1504-4316>

Laurent Guillier  <https://orcid.org/0000-0002-7867-2937>

Julien Bect  <https://orcid.org/0000-0002-0867-0215>

#### Data availability

All of the data that support the findings of this study are available in the main text or Supplementary Information.

## References

- Augustin JC, Zuliani V, Cornu M, Guillier L (2005) Growth rate and growth probability of *Listeria monocytogenes* in dairy, meat and seafood products in suboptimal conditions. *Journal of Applied Microbiology* 99(5): 10191042. <https://doi.org/10.1111/j.1365-2672.2005.02710.x>
- Auvray F, Cointe A, Desvaux M, Jones G, Kurkdjian PM, Oswald E, Kooh P, Guillier L, Audiat-Perrin F (2023) Avis de l'Anses relatif à la définition des souches pathogènes d'*Escherichia coli* productrices de shigatoxines. Maisons-Alfort, Anses, 63 pp. <https://anses.hal.science/anses-04171498>
- F.A.O., W.H.O. (2021) Microbiological risk assessment - Guidance for food. Guidance. Microbiological Risk Assessment Series (36).
- Haberbeck LU, Plaza-Rodríguez C, Desvignes V, Dalgaard P, Sanaa M, Guillier L, Nauta M, Filter M (2018) Harmonized terms, concepts and metadata for microbiological risk assessment models: The basis for knowledge integration and exchange. *Microbial Risk Analysis* 10: 312. [In English] <https://doi.org/10.1016/j.mran.2018.06.001>
- Novak E (1992) Quadrature formulas for monotone functions. *Proceedings of the American Mathematical Society* 115 (1): 5968. <https://doi.org/10.1090/s0002-9939-1992-1086337-x>
- Perrin F, Tenenhaus-Aziza F, Michel V, Miszczyscha S, Bel N, Sanaa M (2014) Quantitative risk assessment of Haemolytic and Uremic Syndrome linked to O157:H7 and Non-O157:H7 Shiga-toxin producing *Escherichia coli* strains in raw milk soft cheeses. *Risk Analysis* 35(1): 109128. <https://doi.org/10.1111/risa.12267>
- Plaza-Rodríguez C, Haberbeck LU, Desvignes V, Dalgaard P, Sanaa M, Nauta M, Filter M, Guillier L (2018) Towards transparent and consistent exchange of knowledge for improved microbiological food safety. *Current Opinion in Food Science* 19: 129137. <https://doi.org/10.1016/j.cofs.2017.12.002>
- Sherkat F, Pollard A, Tan SP (1998) Effects of yield increase on texture of fetta cheese. *Australian Journal of Dairy Technology; Melbourne* 53(2): 138.

## Supplementary material 1

### QRA simulator

Authors: Subhasish Basak, Janushan Christy, Laurent Guillier, Frederique Audiat-Perrin, Moez Sanaa, Fanny Tenenhaus-Aziza, Julien Bect, Emmanuel Vazquez

Data type: fsk

Copyright notice: This dataset is made available under the Open Database License (<http://opendatacommons.org/licenses/odbl/1.0/>). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: <https://doi.org/https://doi.org/10.3897/fmj.5.109502.suppl1>