

SCIENTIFIC REPORT OF EFSA

Quantitative risk assessment of *Salmonella* Enteritidis in shell eggs in Europe¹

Assessment Methodology Unit^{2,3}

European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

Prevalences in terms of the number of eggs contaminated (internally and externally) with *Salmonella* Enteritidis per million are presented based on a 2 stage Bayesian model. The first stage estimates the average flock prevalence over laying period in the production system and the second stage estimates the proportion of contaminated eggs for an infected flock. These are combined to give the expected number of infected eggs per million using the hen data from 2 European Union Member States (MSs). Interpretation and extrapolation of the results has to be done with caution as they are sensitive to the egg data used which are most likely biased due to selection effects. Working with the lower and upper limits of the 95% credible intervals for the parameter estimates is strongly recommended.

It is recommended to repeat a similar exercise with the current model with data from more EU MSs. These should be collected in the format received from the 2 MSs analysed in this report. Further, it is important to have hen and egg data from the same flocks (with flock sizes) for at least some Member States.

KEY WORDS

QMRA, Bayesian Model, Monte Carlo Markov Chain, laying hens, *Salmonella* Enteritidis in eggs, sensitivity, sensitivity analysis

1 On request of EFSA, Question No EFSA-Q-2009-00790 issued on the 19th of April 2010.

2 Correspondence: didier.verloo@efsa.europa.eu

3 Acknowledgement: This scientific report of EFSA is based on major contributions from: Jukka Ranta, Arie Havelaar, Saghir Bashir, Billy Amzal and Didier Verloo. This report was peer reviewed by Ivar Vågsholm.

SUMMARY

Impact on public health of *Salmonella* Enteritidis is typically believed to be proportional to the number of internally contaminated eggs distributed on the EU market for human consumption, although other food-stuff categories of public health relevance such as egg products and meat from spent hens could also contribute. In order to assess the impact on public health of various *Salmonella* flock prevalences (i.e. observed and targets), a quantitative model was developed.

Prevalences in terms of the number of eggs contaminated (internally and externally) with *Salmonella* Enteritidis per million are presented based on a 2 stage Bayesian model. The first stage estimates the average flock prevalence over laying period in the production system and the second stage estimates the proportion of contaminated eggs for an infected flock. These are combined to give the expected number of infected eggs per million using the hen data from 2 European Union Member States (MSs). Sensitivity analysis were conducted to assess the robustness of the results and to better quantify the risks. Further to assess the effects of potential targets different scenarios were considered based on the observed data.

Interpretation and extrapolation of the results has to be done with caution as they are sensitive to the egg data used which are most likely biased due to selection effects. Working with the lower and upper limits of the 95% credible intervals for the parameter estimates is strongly recommended.

For a similar exercise in the future data should be collected from more members states with some providing both hen and egg data. However, it is acknowledged that more detailed studies of infected flocks can be problematic due to legislation, costs or existing risk management interventions taking precedence.

TABLE OF CONTENTS

Abstract	1
Summary	2
Table of contents	3
List of tables	4
List of Figures	4
Background	5
Objectives.....	5
1. Introduction	6
2. Data availability and data management	7
2.1. Hen data from national monitoring programs	7
3. QMRA models and assumptions	8
3.1. Model the expected flock prevalence over laying period	8
3.1.1. Markov model of contamination in production flock.....	8
3.1.2. Sensitivity model and within flock prevalence.....	10
3.2. Model of number of contaminated eggs per contaminated flock.....	11
3.3. Summary of the Key Assumptions	12
4. Results	13
5. Discussion.....	15
Conclusions	16
References	17
Appendices	18
Sensitivity Outputs	18
Prior and posterior distributions	23
WinBUGS programs	28
R programs	36
Abbreviations	43

LIST OF TABLES

Table 1: Predicted prevalence (number of contaminated eggs per million) for the Member State A (MSA) and Member State B (MSB) hen data using the UK pooled and individual egg data with the lower and upper limits of the 95% credible interval.	13
Table 2: Predicted values for the intensity of infection (λ), intensity of recovery (μ) and initial infection probability (ν).	13
Table 3: Predicted eggs per million for MSA and MSB hen data using the UK pooled and individual egg data with the lower and upper limits of the 95% credible interval for different scenarios.	18
Table 4: Predicted values for the intensity of infection (λ), intensity of recovery (μ) and initial infection rate (ν) for different scenarios.	20

LIST OF FIGURES

Figure 1: Predicted number of contaminated eggs per million for the observed test sensitivity and a scenario where it is doubled.	15
Figure 2: Prior distributions for λ , μ and ν	23
Figure 3: Posterior Distributions for λ , μ , ν and number of contaminated eggs per million for MSA with the observed test sensitivity	24
Figure 4: Posterior Distributions for λ , μ , ν and number of contaminated eggs per million for MSA with double the observed test sensitivity.	25
Figure 5: Posterior Distributions for λ , μ , ν and number of contaminated eggs per million for MSB with the observed test sensitivity	26
Figure 6: Posterior Distributions for λ , μ , ν and number of contaminated eggs per million for MSB with double the observed test sensitivity.	27

BACKGROUND

As stated in mandate M-2008-0111, EFSA was asked to quantitatively estimate the impact on public health for various EU flock prevalence values, both theoretical and observed from the *Salmonella* national control programmes. This work should include the collection and summary of all relevant available data and the food chain modelling necessary to achieve a quantitative model-based risk assessment.

The BIOHAZ unit requested (July 2009) the Assessment Methodology unit to assist on this mandate and to base the modelling approach on a similar quantitative risk assessment for *Salmonella* in laying hens carried out by Ranta et al., 2002, published in the Finnish national report EELA, 2006. As this approach has been applied and tested for the Finnish data it was considered to be a suitable model to adapt to the EU data where similar data structure is available.

OBJECTIVES

Impact on public health of *Salmonella* Enteritidis is typically believed to be proportional to the number of internally contaminated eggs distributed on the EU market for human consumption, although other food-stuff categories of public health relevance could also be evaluated such as egg products and meat from spent hens.

The primary objective of the model-based QMRA is to build a quantitative link between the EU *Salmonella* in laying hens flock prevalence and the resulting number of internally and/or externally contaminated eggs per million (prevalence) that would be laid. The fact that the most up to date available data on *Salmonella* in eggs used was from eggs tested as 'whole eggs' (i.e. no differentiation possible between internally and/or externally contaminated), a correction factor should be applied to the final estimate of the number of only internally contaminated eggs. This correction factor is described in the BIOHAZ Opinion (EFSA 2010) and is not part of this Report.

For this purpose, the specific associated objectives should include:

- The identification, collection and summary of relevant data needed for the models and for the generation of realistic assumptions
- The review of the related statistical literature to investigate the published models and their adaptation to the current EFSA mandate.
- The building of a tailored food chain model, fitted to the most relevant available data
- A critical discussion of all assumptions made, including a quantitative measure of uncertainty and a related sensitivity analysis

1. Introduction

In order to assess the impact on public health of various *Salmonella* flock prevalences (i.e. observed and targets), it is necessary to build up a quantitative model that can describe the food chain from placement of flocks in the production environment to the eggs laid by those flocks.

Preferably such a food chain model should be stochastic, i.e. it incorporates the main sources of variability and uncertainty impacting on the final outcomes. Like every modelling exercise, the main characteristics of the model are based on:

- the most sensitive parameters to be accounted for
- the level of accuracy and precision required
- the data available

In the cases where some key parameters need to be included but no data are available, expert opinion is used to generate sensible assumptions. Once built, the model is fitted with the available data from the EU control program to infer on the model parameters. Subsequently the fitted model can be used to simulate the distribution of possible numbers of contaminated eggs laid, for various input scenarios, i.e. various 'what-if observed' flock prevalences and/or testing strategies.

The proposal for the modelling methodology and the data to be used is based on the work used by the similar Finnish assessment in EELA, 2006 because it is already applied and tested model with robust result that is currently updated to a more flexible continuous time model. A modification of this new version was used here. In the earlier EELA model, a statistical link is built between grandfather flock prevalence in breeding herds and the number of contaminated eggs. This accounts for:

- Vertical transmission, from grandparent flocks to parent flocks, and from parent flocks to production flocks.
- Horizontal transmission: within each flock (production or breeding flocks). The contamination can switch from the "*Salmonella*-free" status to the contaminated status and vice versa from one cycle (8 weeks) to another. The chances of such transitions are estimated from the data (Finnish National control program).

There are some major differences in objectives and context between the EFSA model needs and this EELA model. Major differences are:

- the EFSA model should focus only on the production flocks which is sufficient to meet the objectives, whereas the EELA model starts from the breeding phase
- the EELA model is tailored to the Finnish production systems, whereas the EFSA model should reflect the important heterogeneity of production systems and hygiene practices seen within EU. However, this is not included in the current model due to a lack of data.
- the EELA model does not account for the variation of (diagnostic) test sensitivity of the flock testing over time. This issue was addressed by setting values (best guess) of test sensitivity over age of the flock. The sensitivity cannot be estimated from current MS specific data but has to be given as an expert assumption. This will impact how the results can be generalized.

Finally, this model is a statistical model, meaning that it does not describe in details the mechanistic dynamics of the contamination. To estimate the distributions of the parameters Bayesian statistics using Monte Carlo Markov Chain are very appropriate to fit such models and make predictions from them. Bayesian statistics however require a prior distribution (See Figure 2 in the Appendix) reflecting the knowledge about the parameter before the analysis (which can be nil). Relevant estimates from the model should be driven mainly by MS specific data and other scientifically relevant data sources so that the result has sufficient empirical basis, instead of being driven excessively by assumptions or special case situations that do not apply generally. Therefore, a statistical model can have a limited number of parameters depending on the amount of data. In order to achieve this, excessive parameterizations and details are avoided.

2. Data availability and data management

2.1. Hen data from national monitoring programs

Eggs that are internally and externally contaminated with *Salmonella* Enteritidis are investigated in this work. Hereafter when the term contaminated is used it will refer to internally and externally contaminated eggs with *Salmonella* Enteritidis.

In order to fit the flock prevalence model over time, data from national monitoring programmes are needed. Ideally, they should come from as many Member States as possible to reflect the differences in production systems (e.g. outdoor vs. indoor, caged vs. floor production, vaccinated vs. non-vaccinated flocks) and hygiene practices. Those data should include test results at the flock level at each measurement time for *Salmonella* Enteritidis. The number of flocks by Member States is also required. This number could be approximated using data from the Zoonoses Baseline Study Report on *Salmonella* in Laying Hens.

Within the time frame available hen data was received from 6 EU member states. For the purposes of the modelling we only had sufficient data in the desirable format from two EU MSs, hereafter, referred to as Member State A (MSA) and Member State B (MSB). Nevertheless, EFSA wishes to acknowledge all the MSs that submitted data. A summary of the data from MSA and MSB is presented below:

- **MSA**
 - Approximately 3500 flocks (~houses)
 - Approximate testing times: 25, 40, 55 & 70 weeks
 - Infected flocks at each testing time: 4, 8, 5 & 7
 - Prevalence = $(4+8+5+7)/3500 = 0.0069$ (0.69%)
- **MSB**
 - Approximately 325 flocks (~houses)
 - Approximate testing times: 24, 39 & 54 weeks
 - Number of infected flocks at each testing time: 2, 1 & 4
 - Prevalence = $(2+1+4)/325 = 0.0215$ (2.15%)

Note that the prevalence calculations assume that the infected flocks are removed from production.

- **Data on egg prevalence (*Salmonella* contaminated) in positive flocks**

There were 3 possibilities for the egg data:

- US pooled egg data (Henzler 1998) 60 positive flocks.
- UK pooled test egg data⁴ 8 positive flocks.

⁴ The two UK studies used in the modelling work supporting this Scientific Opinion were carried out in the context of a wide-ranging study on the epidemiology and control of *Salmonella* in egg production (Ref. DEFRA project OZO325). Details of the studies are available in the project report, available on request from the Veterinary laboratories Agency (VLA) in UK.

- UK data⁴ on individually tested eggs 15 positive flocks.

The US pooled data was considered by the WG not to be suitable for this analysis, because it was believed that the prevalence was underestimated due to the use of insensitive test methods and therefore it was not used. However, US data does show results on contamination of egg contents, not whole eggs. The study protocol was also changed on the way, which affects the interpretation of results.

For the UK pooled egg data (UKp) 100 pooled samples were taken from each of 8 flocks, with a pool size of 40 eggs in each sample. There were 2, 2, 0, 0, 10, 1, 12 and 0 positive samples (egg shell and/or contents) from the 8 flocks. The mean prevalence for the samples is 3.4% ($100 \times 27/800$).

For the UK individual egg data (UKi) samples were taken from 15 flocks and the following data (positive egg shells and/or contents) was observed:

Positive	27	2	0	0	0	6	0	11	0	0	21	1	2	2	0
Size	180	200	200	200	200	200	200	200	200	200	300	300	300	300	300

The average prevalence for the individual eggs was 2.07% ($100 \times 72/3480$).

The UKi data contain some flocks with fairly large prevalence, particularly one (27/180 or 15%) could even be considered as an outlier, possibly leading to overestimate. Yet, all flocks were selected for egg sampling because they were known to be positive, which in general could lead to overestimate because we then do not have samples from false negative flocks (modestly infected flocks).

3. QMRA models and assumptions

3.1. Model the expected flock prevalence over laying period

3.1.1. Markov model of contamination in production flock

The main structure of the model describing the evolution of flock prevalence over the production time is similar to the one used in the EELA model, but using a continuous time variable as in Ranta et al. (paper on broilers). Using continuous time was necessary to easily handle the fact that different Member States have different sampling times. It allows a flexible and powerful description of the sampling schedules in EU. This is because the probability of infection at any time t can be solved and written as an expression that is a function of t . This function can then be used in computations for arbitrary number of time points t , without further structural changes to the model, in this respect. Moreover, the probability denotes the probability of true infection at time t , in contrast to the observed (detected) infection. Therefore, overall sensitivity of testing method is included as a parameter in the model, related to each testing time (age of flock). An overall sensitivity (either at a given age, or average over age) is defined as the probability to actually detect a flock to be SE infected when it is tested by the assumed sampling and laboratory methods and when there is any unknown level of within flock infection (i.e. not zero within flock prevalence).

The model version used here is focused on the production phase (layers). It includes a vertical transmission parameter which accounts for the contamination transmitted from the breeding phase but also the remaining contamination inherited from previous productions in the same premises. The 'vertical' transmission probability in this modified model simply denotes the probability of a flock being infected at the beginning of laying.

For a given flock, the contamination status at time t is described by a Markov process. The probability of this flock to be truly infected, given that it was not at time zero, is given by the following formula:

$$P(I_t = 1 | \lambda, \mu, I_0 = 0) = \frac{\lambda}{\lambda + \mu} - \frac{\lambda}{\lambda + \mu} \exp(-(\lambda + \mu)t)$$

where λ is the infection intensity (rate) of flocks and μ is the recovery intensity (rate) of flocks. Likewise, a modified expression can be written for the situation starting from infected state. (The last term is $+\mu/(\lambda+\mu)\exp(-(\lambda+\mu)t)$). Hence, as a combination of the two possible initial states, the probability of infection at time t is

$$P(I_t = 1 | v, \lambda, \mu) = vP(I_t = 1 | \lambda, \mu, I_0 = 1) + (1-v)P(I_t = 1 | \lambda, \mu, I_0 = 0).$$

This shows the role of parameter v . Parameters λ and μ were assumed to be constant for all flocks of a same production system type and over time in a MS, as determined by the choice of data set. The interpretation of λ is that if there is a large number of uninfected flocks at some time point, then after time t , the proportion of infected flocks is $1 - \exp(-\lambda t)$. Likewise, if starting with a large number of infected flocks, $1 - \exp(-\mu t)$ would be the proportion of flocks that have recovered after time t . Also, $1/\lambda$ is the expected waiting time until an uninfected flock is infected, and $1/\mu$ is the expected waiting time until an infected flock recovers. Hence, *the parameters are rates for flocks, in a population of flocks, not rates for hens*. No prior information is assumed about the infection intensity, but the prior distribution of recovery rate is based on expert assumption about the proportion of flocks that could be expected to have recovered after a given time. It is generally assumed that the recovery is slow. Therefore, an informative beta-distribution Beta(49,1), mean 0.98, was chosen to describe the uncertainty about the proportion η of infected flocks that would *not* have recovered (spontaneous clean up) in 2 months (1/6 years). The values can be transformed back to the intensity by the equation $\mu = -6 \ln(\eta)$ $E(\mu) = 0.121$ /year ($1/E(\mu) = 8.25$ years).). For example: if the proportion of infected flocks that have recovered in 2 months is $1 - 0.98 = 0.02$, in a year it would be $1 - \exp(-\mu) = 1 - \exp(-6 \ln(0.98)) = 0.11$. Uninformative uniform prior over a wide range was assumed for infection intensity λ .

As a result, the model gives an expression of probability for hidden infection (i.e. the biological idea of plain existence of salmonella in a flock, regardless of whether we test or not, or what the test result would be) at a specific age. This depends on the model parameters and what is known about the history of a flock up to that time point. From a specified set of observed testing results, the parameters can be estimated taking into account imperfect test sensitivity as explained below. Generally, the methodological approach is related to estimating infection dynamics when the detectability is imperfect (Nagelkerke N J et al, 1990).

In addition to parameter estimates, some further results can be drawn, for example the expected average infection prevalence of flocks over laying period (defined as EQ). This means calculating the expected proportion of flocks that would be infected at a given time t of the production period, and then taking the average of this value over a time span as a summary of that whole time span. This expected value can either be computed assuming (1) no interventions take place regardless of the testing results, or (2) it may be assumed that test positive flocks are removed immediately. The latter approach thus involves *survival probabilities* (the probability of a flock showing negative test results – which may be correct negatives or false negatives). Due to intervention, the survival probabilities have an effect on the expected prevalence at a later time, since the flock population then consists of those who were not detected positive previously. Since the model here is defined over the laying period only, it is assumed that all flocks enter laying period (infected or uninfected) and the removals are then possible only at subsequent testing times. The expected value over laying period from T_1 to T_2 , with testing times t_1, \dots, t_n , accounting for the survival probability is of the form below.

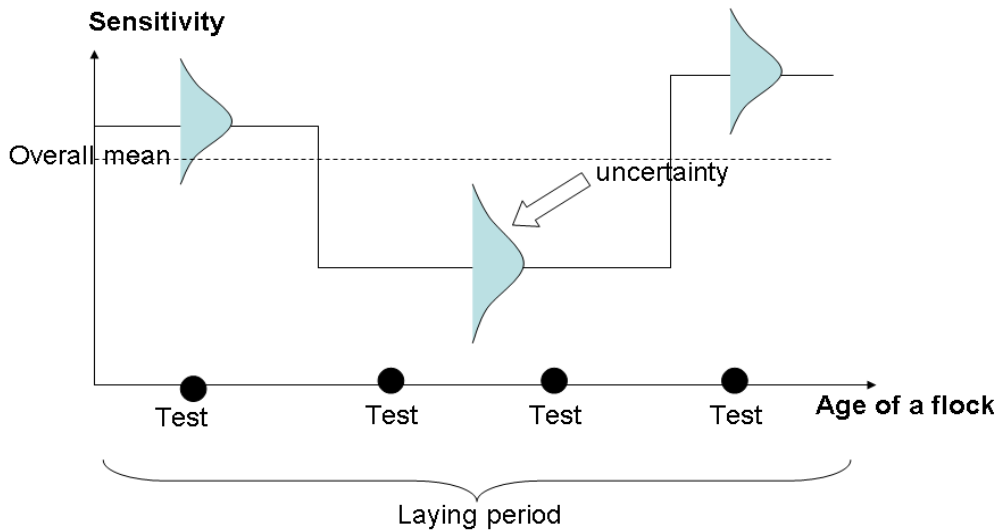
$$\begin{aligned}
 EQ &= Q(T_1, t_1 | H) \times P(t_1^{\oplus}) \\
 &+ Q(T_1, t_2 | H) \times P(t_1^-, t_2^{\oplus}) + \dots \\
 &+ Q(T_1, t_k | H) \times P(t_1^-, \dots, t_{k-1}^-, t_k^{\oplus}) \\
 &+ Q(T_1, T_2 | H) \times P(t_1^-, \dots, t_n^-)
 \end{aligned}$$

The expected average flock prevalence, Q , between beginning of laying T_1 and a testing time t , becomes weighted by the survival probability for all previous tests and detection probability at the latter point t . For a specific time interval, Q can be calculated as an integral of the infection probability $P(I_i=1 | v, \lambda, \mu, H)$, divided by length of the interval. Infection probability H is calculated conditionally to the past history H of testing results (all negatives). Finally, the expected average prevalence over the whole laying period becomes weighted by the survival probability over all tests. (A flock may be detected at the 1st test, or the 2nd test, etc, or not at all, hence the summation over all these possibilities). If a flock is tested positive then it is assumed that the flock is removed. All probabilities are conditional to the history of the past testing results (all negatives) before the time point in question.

3.1.2. Sensitivity model and within flock prevalence

The within flock dynamics is expected to have a strong impact on the sensitivity of detection and consequently on model parameter estimates. This is because the within flock dynamics affect the within flock prevalence which is related to test sensitivity for a given flock (sensitivity is assumed to be <100%, but specificity is assumed 100%). Depending on the within flock situation (e.g. caged versus free range), the chances of detection can be different in different infected flocks. This effect could be accounted for by modelling the sensitivity parameter as a random effect so that it is allowed to take different values at different times for different flocks. However, this would introduce $T \times N$ unknown parameters with T testing times and N flocks. An approach with similar effect is to replace the sensitivity parameter p_{sen} with its expected value $ep_{\text{sen}} = E(p_{\text{sen}})$ and treat that as a parameter to be estimated. In simple Bernoulli experiments, if the success probability is different in each experiment and independently drawn from a same distribution, the only parameter that can be estimated is the 'average probability' which will yield the same results when taken as a binomial probability.

A model that takes within flock prevalence into account could provide a function describing p_{sen} as a function of T and would therefore allow a more detailed approach, but it would still represent only a prediction unless the MS specific data could include within flock data linked to the flock testing times at which the flocks are now reported simply as negative or positive. In such approach it should be noted that the actual sensitivity of a specific flock depends on the time since infection. While this might be constructed in a functional form, based on expert opinion, it would depend on the unknown time of infection. The mathematical calculations would then require integrating over all possible infection times. In a same manner, any sensitivity parameter can be seen as a weighted average for a reference population when integrating over more detailed unknown events. In other words, sensitivity is a population quantity with an interpretation that depends on the adopted population description. Hence, one approach is to define all the tested flocks at all testing times as a population, in which case sensitivity would refer to that population as a whole. However, the testing times are characterized by the testing ages, according to the national testing scheme. Therefore, this makes a natural stratification of all testing occasions by testing age, and it is possible to define an age related sensitivity parameter. This describes the chance of detecting an infected flock that is of specific age (the reference population is all those flocks that have been infected before the testing, and still are infected, when tested at given age). It may then be assumed that there is some variation of test sensitivity over age as such, that can be roughly accommodated as an expert opinion. Alternatively, the overall mean is given over age. In each case, an informative assumption is needed about its value.



For the 2 available data sets (MSA and MSB) for one year there were approximately 3500 and 325 flocks under the testing scheme with 4 and 3 testing times over the laying period at specified ages of flocks, respectively. The number of positive flocks were reported at each of these testing times. In reality, the times are not exactly these for all flocks but an approximation of the testing age was used as data for the model. Also, in some cases the number of testing times per flock could have been different from the planned. Again, fixed testing plan with given number of testing times was assumed in the model. In the demonstrating calculation it was assumed that the sensitivity of flock testing was higher for the beginning and end of laying period, but possibly lower in the middle. These values were taken as expert opinions, based on empirical data from the UK, giving best estimates of 0.30, 0.25 and 0.35. The sensitivity should describe how big the chances were to actually detect a positive flock given these data, considering it was an operator sampling, as an average under all possible circumstances that may have taken place in these data. At each testing time for a flock, the outcome (pos/neg) was modelled as Bernoulli-variable with probability $p_{\text{sen}}(t) \cdot P(I_t=1 \mid H, v, \lambda, \mu)$ i.e. the product of the sensitivity at that age, and the probability of true infection, given the history of previous results. The Bernoulli probabilities then build up the complete model for all observed sampling results.

3.2. Model of number of contaminated eggs per contaminated flock

The proportion of contaminated eggs, produced by an infected flock, depends on the within flock prevalence. In a detailed approach, a within flock model would describe the hen level dynamics which depend on the time since first infection and the interactions between hens and the transmission from hens to eggs. In a statistical analysis of empirical data on reported positive eggs (or pools of eggs) in samples from infected flocks, it cannot be known when the infection started in the flock or what the within flock prevalence was at the time of sampling. Therefore, the empirical egg prevalence can only be used to describe 'infected flocks', without more detailed reference to the underlying situation. There is only one recent experiment where both egg prevalence and hen prevalence were investigated at the same time from the same flocks (UK report).

Empirical samples from various infected flocks can be used to estimate the proportion of contaminated eggs in each such flock. Moreover, the results can be combined to estimate the average proportion of contaminated eggs in such infected flocks in general. However, the estimates may be biased since the time (age) of sampling and the selection of infected flocks is typically not formally randomized within the group of all infected flocks and over all ages. Furthermore, such studies are only rarely reported from few countries only, and their applicability requires assuming reasonable similarity between countries, production systems, and seasons.

No information in the egg-model has an effect on the hen-model and vice versa (i.e. the 2 models are fitted independently of each other). One model estimates flock prevalence (and finally, the average EQ) and the other deals with egg prevalence for infected flocks in general (p_{μ} parameter). Both parameters are then described by posterior distribution. The final overall estimate of the egg prevalence (for idealized large population, describing the production system) is the product of EQ and p_{μ} . The uncertainty about this product is obtained as the distribution of the product of two random parameters, each with their own distribution. When the member state is changed, the flock level data are changed (positive & negative flocks at specific testing times) so that the distribution of EQ becomes different, but it is still assumed that the overall egg prevalence in infected flocks is p_{μ} (based on UKi and/or UKp data, but applied to each MS). The final result is a product of these two uncertain averages, summarized e.g. by the median of that product.

Following the latter approach, i.e. estimating egg prevalence in infected flocks, 2 data sources were combined to estimate the egg prevalence: UK data on pooled samples (8 flocks) and UK data on individual eggs (15 flocks). In all cases, samples were collected from flocks known to be infected. As a sensitivity of the egg testing method, expert opinion was used in the Finnish RA (2006) which assumed it to be 0.3-0.7. In a recent experimental setting (UK) pooled eggs were inoculated and tested to study the sensitivity. Information from this experiment provides $8/12 = 0.67$ as a point estimate. Both sources of information were combined to give a distribution for sensitivity (see WinBUGS programs in appendix) which is then used in estimating the true egg prevalence in infected flocks. It is assumed that infected flocks have the same contaminated egg prevalence distributions. Sensitivity parameter was assumed to be applicable to both UK individual egg data and UK pooled eggs data. Inclusion of US pooled egg data would need a specification of another level of test sensitivity, in comparison to the one used.

3.3. Summary of the Key Assumptions

- The sensitivity cannot be estimated from current MS specific data but has to be given as an expert assumption (sensitivity is assumed to be <100%, but specificity is assumed 100%).
 - Fixed testing plan with given number of testing times was assumed in the model. It was assumed that the sensitivity of flock testing was higher at the beginning and end of laying period, but possibly lower in the middle. These values were taken as expert opinions, based on empirical data from the UK, giving best estimates of 0.30, 0.25 and 0.35.
- Sensitivity parameter was assumed to be applicable to both UK individual egg data and UK pooled eggs data.
- Parameters λ (infection intensity) and μ (recovery rate) were assumed to be constant for all flocks of a same production system type and over time in a MS, as determined by the choice of data set.
 - No prior information is assumed about the λ and therefore an uninformative uniform prior over a wide range was assumed.

- The prior distribution of recovery rate is based on expert assumption about the proportion of flocks that could be expected to have recovered after a given time. It is generally assumed that the recovery is slow.
- It is assumed that all flocks enter laying period (infected or uninfected) and the removals are then possible only at subsequent testing times.

4. Results

Note that the estimates presented below refer to the median of the posterior distributions.

From the model, a posterior predictive distribution was computed for the flock testing results (observable prevalence in all flocks, i.e. N_{pos}/NF , denominator given by NF flocks) and for the egg prevalence per million. Table 1 below presents the number of predicted *S. Enteritidis* contaminated eggs per million for the hen data of member states A and B using both the UK pooled and individual egg data (all 3 possible combinations). It was assumed that all infected flocks are removed directly after taking the sample that will ultimately result in a positive result. The lower and upper columns best reflect the uncertainty that there is and perhaps best indicators of the risks involved.

Table 1: Predicted prevalence (number of contaminated eggs per million) for the Member State A (MSA) and Member State B (MSB) hen data using the UK pooled and individual egg data with the lower and upper limits of the 95% credible interval.

Hen Data	Egg Data	Eggs / Million		
		Estimate	Lower	Upper
MSA	UKi	28.2	1.1	158.9
	UKp	2.3	0.0	20.7
	UKp-UKi	9.5	1.1	43.6
MSB	UKi	151.7	7.1	978.6
	UKp	13.9	0.2	132.2
	UKp-UKi	56.8	5.9	300.8

Table 2 presents the posterior distribution summaries for the intensity of infection (λ), intensity of recovery (μ) and initial infection probability (v). Note that these are calculated from the MSA and MSB hen data only and therefore are independent of the egg data. The prior and posterior distributions for λ , μ , v and the egg prevalence per million eggs are presented **below** (see Appendix: From Figure 2 to Figure 6: Prior and Posterior Distributions).

Table 2: Predicted values for the intensity of infection (λ), intensity of recovery (μ) and initial infection probability (v).

Hen Data	Parameter	Estimate	Lower	Upper
MSA	lambda (λ)	0.010	0.003	0.019
	mu (μ)	0.093	0.003	0.480
	v	0.004	0.001	0.009
MSB	lambda (λ)	0.068	0.010	0.172
	mu (μ)	0.083	0.003	0.450
	v	0.017	0.001	0.051

The posterior distributions of μ are similar to the prior distribution (and it has a long tail), but the posterior distributions of λ and ν are strongly based on the data. The distribution of the egg prevalence have a very long tail too.

At the request of the WG the 6 scenarios were also modelled where the actual observed prevalence was “reset” according to the following flock prevalence for the hen data:

- 0.88%, 1.00%, 1.76%, 2.00%, 3.10% and 13.6%

The apparent observed data was recalculated for each of the scenario prevalences using

$$\underline{X}_s = (p_s/NF) * (\underline{X}/\text{sum}(\underline{X}))$$

where \underline{X} is a vector of the observed number of infected flocks for each time point, p_s is the scenario proportion (percentage), NF is the number of flocks. \underline{X}_s was rounded to the nearest integer which could be problem for small p_s . Therefore, the scenario is “what if we had observed this different prevalence, but the number of positives over age would still have (nearly) the same age-profile”.

Furthermore, the scenario of doubling the observed test sensitivity was also investigated to calculate the impact this would have on the estimated egg prevalence per million. Note that the scenarios represent alternative data that are taken to replace the actual data we had, to rerun the estimation. Therefore, the results describe what our estimates would be *if* we had observed such data. For example, if, together with a given ‘observed’ prevalence, we assume test sensitivity to be higher than before, then the true prevalence will be estimated lower than before, which means that either ν or λ , or both, become estimated lower than before, and all this has an effect on what the estimate of EQ will be (it depends on ν , λ , μ and p_{sen}). In every case, a new posterior distribution is computed and it will be conditional to each given data and prior assumption. Note: change in p_{sen} will therefore lead to a different estimate of ν and λ . This scenario (‘seeing’) is not the same as the scenario (‘doing’) of assuming p_{sen} could be changed independently of the other parameters.

The full results of all the combinations of the sensitivity analyses are presented in Table 3 and Table 4 (see Appendix) and Figure 1 presents a summary of the results graphically. The same set of “observed” scenarios lead to different egg prevalences for the 2 member states. This is not unexpected as the sampling points for the hen data and the distribution of the infected flocks differ between member states.

For reference purposes all the programs used are presented at the end of this report (see WinBUGS and R Programs in appendix).

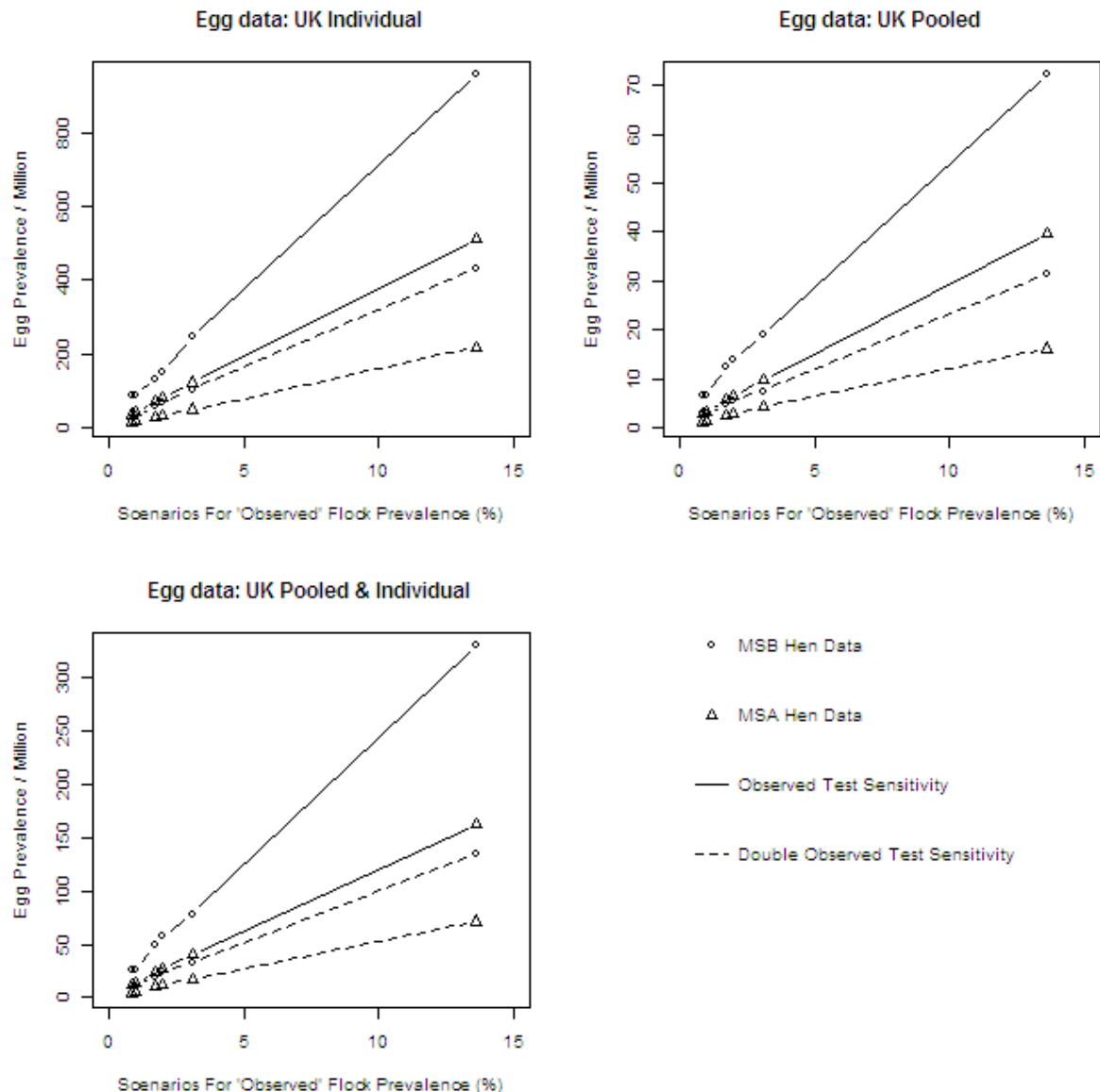


Figure 1: Predicted number of contaminated eggs per million for the observed test sensitivity and a scenario where it is doubled.

5. Discussion

Model with overall mean sensitivity (though unknown) and a model with age-dependent sensitivity as a step function (though unknown) provide same results if the overall sensitivity matches the mean of the age specific sensitivity. In all cases, realistic results may only be obtained if the sensitivity in general is assumed to be greater than 15-20%. Otherwise, assuming very small sensitivity, the testing would not be informative about the prevalence of infected flocks which would effectively deny any useful connection between the true flock status and testing results. In other words, one needs to assume that the test is usually working reasonably well, and assign this level either as a fixed value or as an informative distribution, excluding extremely small values.

Although baseline data can provide within flock information, they would only be available at the end of the production phase. Hence it would not be able to link it with time or age of the flock over the whole laying period. The collection/search for additional complementary data as well as the additional

modelling required would need substantial resources and the added value for such an approach is uncertain.

The relationship between hen prevalence and egg prevalence in an infected flock cannot be estimated accurately so that even if hen prevalence would be known exactly (which is not the case), the egg prevalence would still be quite uncertain to predict from the hen prevalence although correlation exists. It is advisable to quote lower and upper limits of the credible intervals.

The statistical approach for estimating egg prevalence in infected flocks, based on direct samples from infected flocks would ideally need a well randomized study over all infected flocks (not just some that were selected) representing all ages at all production systems during all seasons in several MSs. Since the data is limited and selection effects cannot be controlled, the generalizability of the estimates cannot be fully guaranteed. Hence, the estimates may represent the circumstances in these data sets only (i.e the results are likely to be biased and therefore difficult to generalise).

The flock size distribution was not included in the model and its inclusion could cause some additional uncertainty. In further work, if flock size data were available, it should be MS specific and ideally linked directly with the hen data (positive and negative sampling results of laying flocks). Then, predictive results could be computed for each flock in production system, and a MS summary computed from those.

The total number of flocks tested was roughly estimated and does not represent an exact number, from which to calculate the total number of flocks that tested negative at all times. This is difficult to get exact due to annual reporting practice and the fact that flocks usually fall partly over two consecutive calendar years, whereas it would be ideal to have complete data on a set of flocks as a cohort. Also, positive results are typically better documented than negative results, which make it difficult to estimate the total number of tested flocks, for a flock specific model.

CONCLUSIONS

The results should be quoted and interpreted with caution as they are sensitive to the egg data used and most likely biased. Working with the lower and upper limits of the 95% credible intervals for the parameter estimates is strongly recommended.

In order to repeat a similar exercise with the current model, data from more EU MSs should be collected in the format received from the 2 MSs analysed in this report. Further, it is important to have hen and egg data from the same flocks (with flock sizes) for at least some Member States. However, it is acknowledged that more detailed studies of infected flocks can be problematic due to legislation, costs or existing risk management interventions taking precedence. A statistician should be consulted in advance to help find the best compromise to achieve this and the added information value must be assessed prior to data collection.

REFERENCES

- Chriél M, Stryhn H, Dauphin G. Generalized linear mixed models analysis of risk factors for contamination of Danish broiler flocks with Salmonella Typhimurium. Preventive Veterinary Medicine, 40:1-17, 1999.
- EELA, Salmonella in Egg Production in Finland – a Quantitative Risk Assessment, Eelan Julkaisu, EELA report, April 2006
- EFSA Panel on Biological Hazards (BIOHAZ), 2010. Scientific Opinion on a quantitative estimate of the public health impact of setting a new target for the reduction of Salmonella in laying hens. EFSA Journal, 8(4):1546, 87 pp.
- Gast RK, Detection of Salmonella Enteritidis in experimentally infected laying hens by culturing pools of egg contents. Poultry Science, 72(2):267-274, 1993
- Gormley FJ et al. Microbiological Study on Salmonella Contamination of Pooled Raw Shelled Egg Mix and Environmental Samples from Catering Establishments, LACORS/HPA Coordinated Food Liaison Group Studies, 2009
- Lindley D V, Seeing and doing: the concept of causation (with discussion). Internat. Statist. Rev, 70, 191-214. 2002.
- Nagelkerke N J; Chung R N; Kinoti S N, Estimation of parasitic infection dynamics when detectability is imperfect. Statistics in medicine, 9(10):1211-9. 1990
- Ranta J, Maijala R, A probabilistic transmission model of salmonella in the primary broiler production chain. Risk Analysis, 22(1):47-58, 2002
- Ranta J, Tuominen P, Mikkilä A, Wahlström H, Bayesian hidden Markov process model of salmonella over three generations of flocks in broiler production chain. Abstract. S.Co. 2009, Complex Data Modeling and Computationally Intensive Statistical Methods for Estimation and Prediction. Politecnico di Milano, Milan, September 14-18.2009. <http://mox.polimi.it/sco2009/>
- Valentin-Bon IE, Brackett RE, Seo KH, Hammack TS, Andrews WH; Preenrichment versus direct selective agar plating for the detection of Salmonella Enteritidis in shell eggs, Journal of food protection 66(9):1670-1674, 2003
- WHO/FAO, Development of Practical Risk Management Strategies based on Microbiological Risk Assessment Outputs – Case Study: Salmonella Enteritidis in Table Eggs, Background Paper 6, March 2006

APPENDICES

SENSITIVITY OUTPUTS

Table 3: Predicted eggs per million for MSA and MSB hen data using the UK pooled and individual egg data with the lower and upper limits of the 95% credible interval for different scenarios.

Hen Data	Egg Data	Test Sensitivity Weight	Scenario	Eggs / Million		
				Estimate	Lower	Upper
MSB	UKi	1	Observed	151.7	7.1	978.6
			0.88%	87.1	2.0	630.2
			1.00%	87.1	2.0	630.2
			1.76%	132.8	5.4	889.7
			2.00%	151.7	7.1	978.6
			3.10%	245.2	13.4	1423.0
			13.60%	957.2	49.0	5386.3
MSB	UKi	2	Observed	68.0	2.7	443.5
			0.88%	32.6	1.3	241.5
			1.00%	32.6	1.3	241.5
			1.76%	56.7	1.7	375.1
			2.00%	68.0	2.7	443.5
			3.10%	102.8	4.6	619.4
			13.60%	430.5	20.2	2314.0
MSB	UKp	1	Observed	13.9	0.2	132.2
			0.88%	6.6	0.0	74.9
			1.00%	6.6	0.0	74.9
			1.76%	12.2	0.2	113.9
			2.00%	13.9	0.2	132.2
			3.10%	18.8	0.1	204.0
			13.60%	72.2	0.4	640.9
MSB	UKp	2	Observed	5.5	0.0	60.7
			0.88%	2.7	0.0	29.1
			1.00%	2.7	0.0	29.1
			1.76%	4.7	0.1	44.1
			2.00%	5.5	0.0	60.7
			3.10%	7.4	0.0	80.7
			13.60%	31.5	0.2	297.1
MSB	UKp-UKi	1	Observed	56.8	5.9	300.8
			0.88%	26.6	2.1	161.3
			1.00%	26.6	2.1	161.3
			1.76%	48.7	5.0	261.7
			2.00%	56.8	5.9	300.8

			3.10%	78.3	8.6	401.1
			13.60%	329.4	42.3	1509.1
MSB	UKp-UKi	2	Observed	23.7	2.5	125.6
			0.88%	11.5	1.2	67.8
			1.00%	11.5	1.2	67.8
			1.76%	19.7	1.8	107.7
			2.00%	23.7	2.5	125.6
			3.10%	32.4	3.8	167.5
			13.60%	134.9	15.3	631.8
MSA	UKi	1	Observed	28.2	1.1	158.9
			0.88%	34.4	1.5	186.8
			1.00%	41.2	1.7	226.8
			1.76%	70.0	2.9	385.0
			2.00%	80.2	3.8	440.3
			3.10%	121.9	5.6	655.9
			13.60%	510.6	23.2	2719.0
MSA	UKi	2	Observed	12.2	0.6	70.0
			0.88%	16.0	0.8	89.1
			1.00%	16.5	0.6	97.9
			1.76%	27.9	1.1	165.3
			2.00%	32.9	1.4	184.6
			3.10%	49.6	2.1	274.1
			13.60%	217.3	9.7	1155.0
MSA	UKp	1	Observed	2.3	0.0	20.7
			0.88%	2.6	0.0	25.1
			1.00%	3.0	0.0	29.8
			1.76%	5.7	0.1	46.5
			2.00%	6.4	0.0	50.0
			3.10%	9.7	0.1	85.0
			13.60%	39.7	0.4	324.2
MSA	UKp	2	Observed	1.0	0.0	8.4
			0.88%	1.1	0.0	10.9
			1.00%	1.4	0.0	12.6
			1.76%	2.3	0.0	20.6
			2.00%	2.8	0.0	23.6
			3.10%	4.1	0.1	36.8
			13.60%	16.1	0.1	138.8
MSA	UKp-UKi	1	Observed	9.5	1.1	43.6
			0.88%	12.1	1.2	53.6
			1.00%	13.5	1.8	62.3

			1.76%	23.9	3.2	102.9
			2.00%	27.0	3.0	120.4
			3.10%	40.4	4.6	190.0
			13.60%	162.7	23.5	713.6
MSA	UKp-UKi	2	Observed	3.9	0.5	18.1
			0.88%	4.5	0.5	21.8
			1.00%	5.5	0.6	26.1
			1.76%	10.1	1.3	43.5
			2.00%	11.8	1.5	52.9
			3.10%	16.9	2.0	75.0
			13.60%	71.1	8.8	314.6

Table 4: Predicted values for the intensity of infection (λ), intensity of recovery (μ) and initial infection rate (ν) for different scenarios

Hen Data	Test Sensitivity		Parameter	Scenarios	Estimate	Lower	Upper
	Weight						
MSB	1		λ	Observed	0.068	0.010	0.172
				0.88%	0.032	0.003	0.106
				1.00%	0.032	0.003	0.106
				1.76%	0.052	0.006	0.149
				2.00%	0.068	0.010	0.172
				3.10%	0.096	0.020	0.224
				13.60%	0.461	0.237	0.792
MSB	2		λ	Observed	0.044	0.013	0.098
				0.88%	0.021	0.003	0.061
				1.00%	0.021	0.003	0.061
				1.76%	0.035	0.008	0.087
				2.00%	0.044	0.013	0.098
				3.10%	0.062	0.024	0.126
				13.60%	0.278	0.175	0.426
MSB	1		λ	Observed	0.083	0.003	0.450
				0.88%	0.090	0.004	0.460
				1.00%	0.090	0.004	0.460
				1.76%	0.088	0.004	0.465
				2.00%	0.083	0.003	0.450
				3.10%	0.088	0.003	0.478
				13.60%	0.081	0.003	0.425
MSB	2		λ	Observed	0.084	0.003	0.438
				0.88%	0.086	0.003	0.451
				1.00%	0.086	0.003	0.451

			1.76%	0.088	0.003	0.468
			2.00%	0.084	0.003	0.438
			3.10%	0.085	0.003	0.455
			13.60%	0.081	0.003	0.427
MSB	1	v	Observed	0.017	0.001	0.051
			0.88%	0.010	0.001	0.034
			1.00%	0.010	0.001	0.034
			1.76%	0.017	0.002	0.049
			2.00%	0.017	0.001	0.051
			3.10%	0.021	0.002	0.060
			13.60%	0.079	0.025	0.153
MSB	2	v	Observed	0.009	0.001	0.028
			0.88%	0.005	0.000	0.020
			1.00%	0.005	0.000	0.020
			1.76%	0.009	0.001	0.028
			2.00%	0.009	0.001	0.028
			3.10%	0.011	0.001	0.032
			13.60%	0.039	0.012	0.075
MSA	1	lambda (λ)	Observed	0.010	0.003	0.019
			0.88%	0.013	0.005	0.022
			1.00%	0.014	0.006	0.025
			1.76%	0.027	0.014	0.041
			2.00%	0.029	0.015	0.045
			3.10%	0.046	0.028	0.067
			13.60%	0.191	0.137	0.261
MSA	2	lambda (λ)	Observed	0.008	0.004	0.013
			0.88%	0.010	0.006	0.015
			1.00%	0.012	0.007	0.017
			1.76%	0.021	0.014	0.029
			2.00%	0.023	0.016	0.031
			3.10%	0.036	0.027	0.047
			13.60%	0.149	0.122	0.182
MSA	1	mu (μ)	Observed	0.093	0.003	0.480
			0.88%	0.094	0.003	0.473
			1.00%	0.091	0.003	0.484
			1.76%	0.096	0.004	0.507
			2.00%	0.091	0.003	0.495
			3.10%	0.097	0.004	0.517
			13.60%	0.108	0.004	0.549

MSA	2	mu (μ)	Observed	0.091	0.003	0.481
			0.88%	0.091	0.003	0.467
			1.00%	0.089	0.004	0.484
			1.76%	0.095	0.003	0.486
			2.00%	0.095	0.003	0.484
			3.10%	0.096	0.004	0.508
			13.60%	0.104	0.004	0.547
MSA	1	v	Observed	0.004	0.001	0.009
			0.88%	0.005	0.001	0.010
			1.00%	0.006	0.001	0.012
			1.76%	0.009	0.003	0.017
			2.00%	0.011	0.004	0.019
			3.10%	0.016	0.007	0.027
			13.60%	0.073	0.050	0.102
MSA	2	v	Observed	0.002	0.000	0.004
			0.88%	0.002	0.000	0.005
			1.00%	0.002	0.000	0.005
			1.76%	0.003	0.001	0.007
			2.00%	0.004	0.001	0.008
			3.10%	0.006	0.002	0.011
			13.60%	0.030	0.018	0.046

PRIOR AND POSTERIOR DISTRIBUTIONS

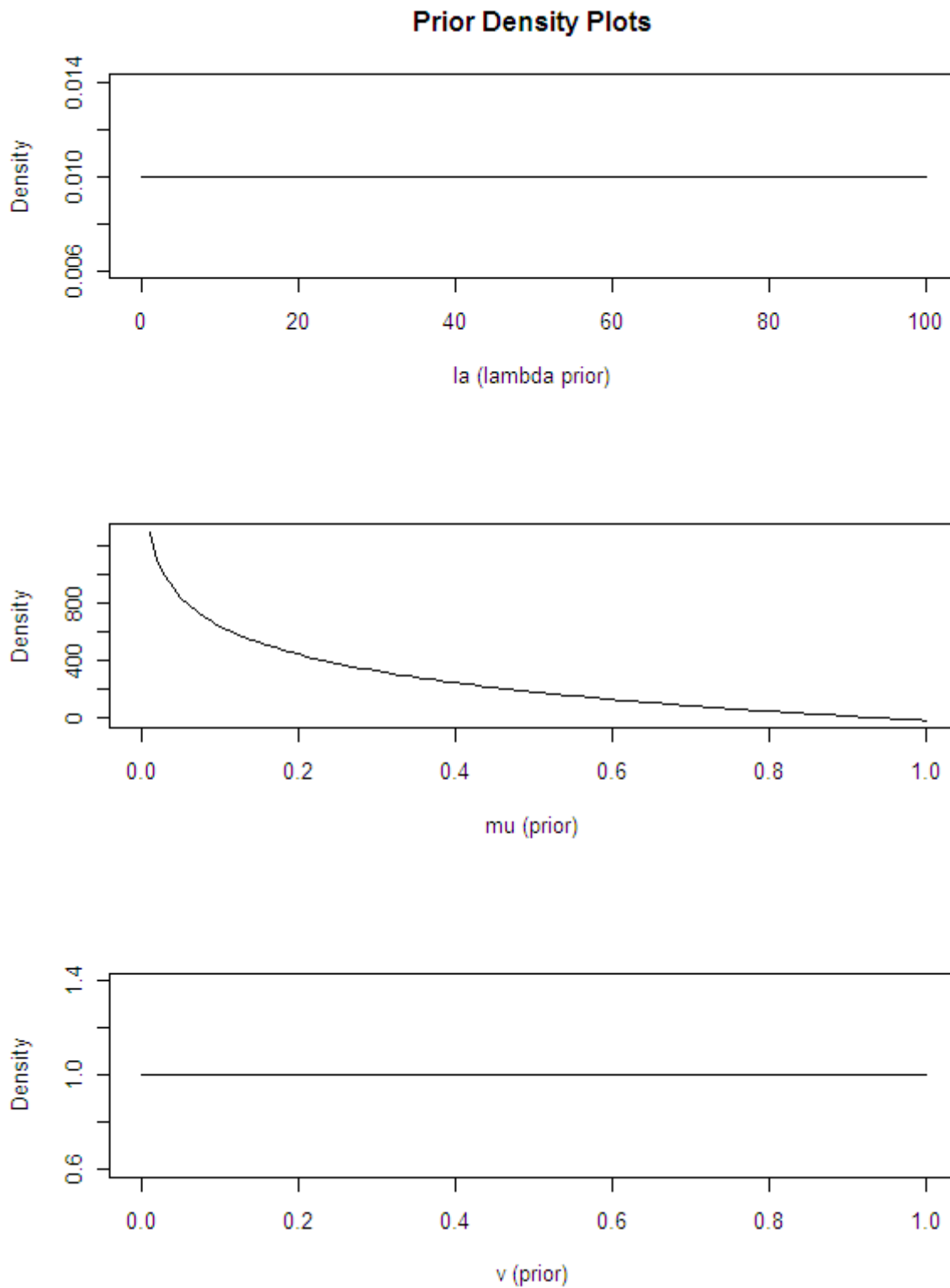


Figure 2: Prior distributions for lambda, mu and v

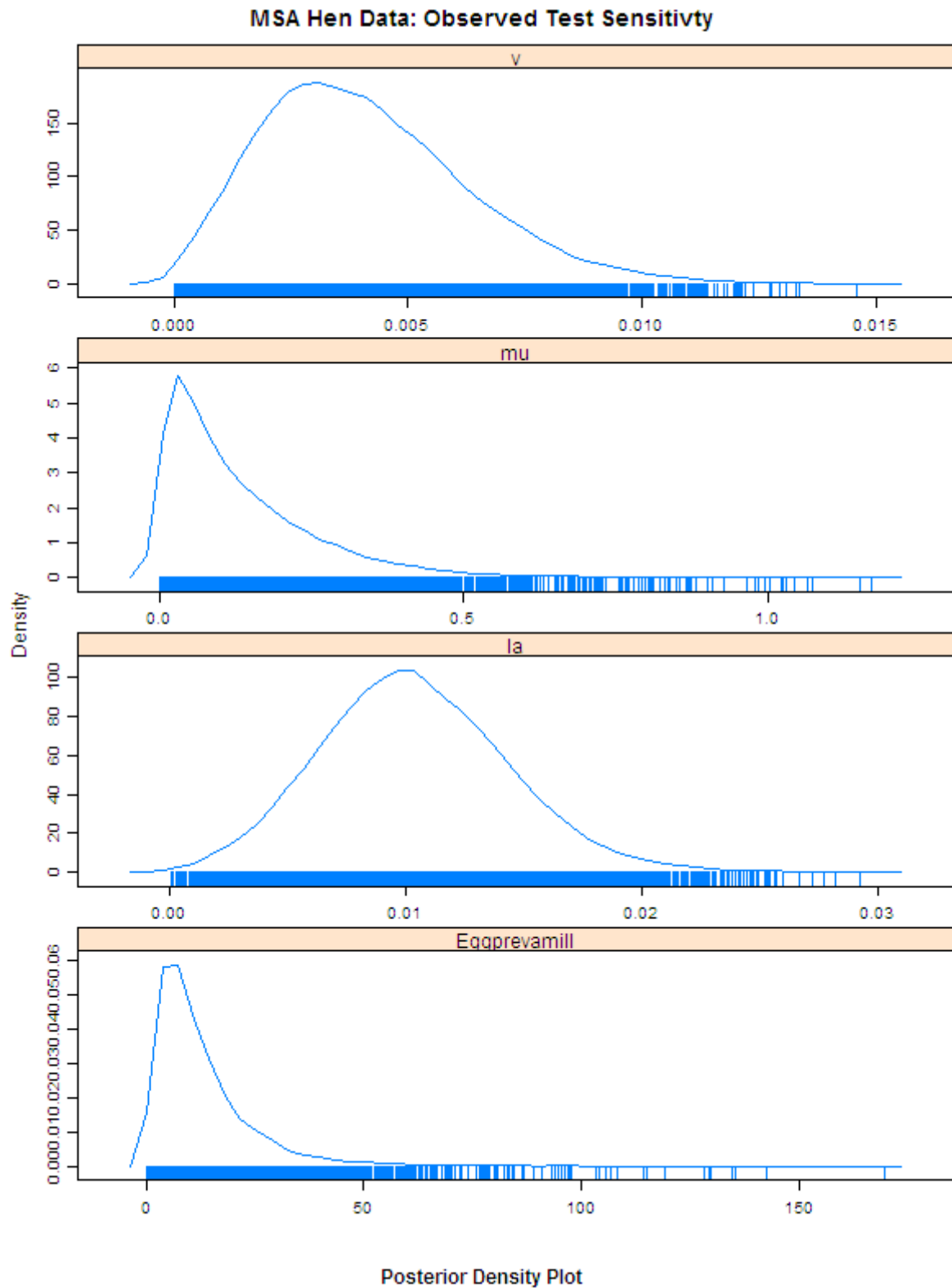


Figure 3: Posterior Distributions for lambda, mu, v and number of contaminated eggs per million for MSA with the observed test sensitivity

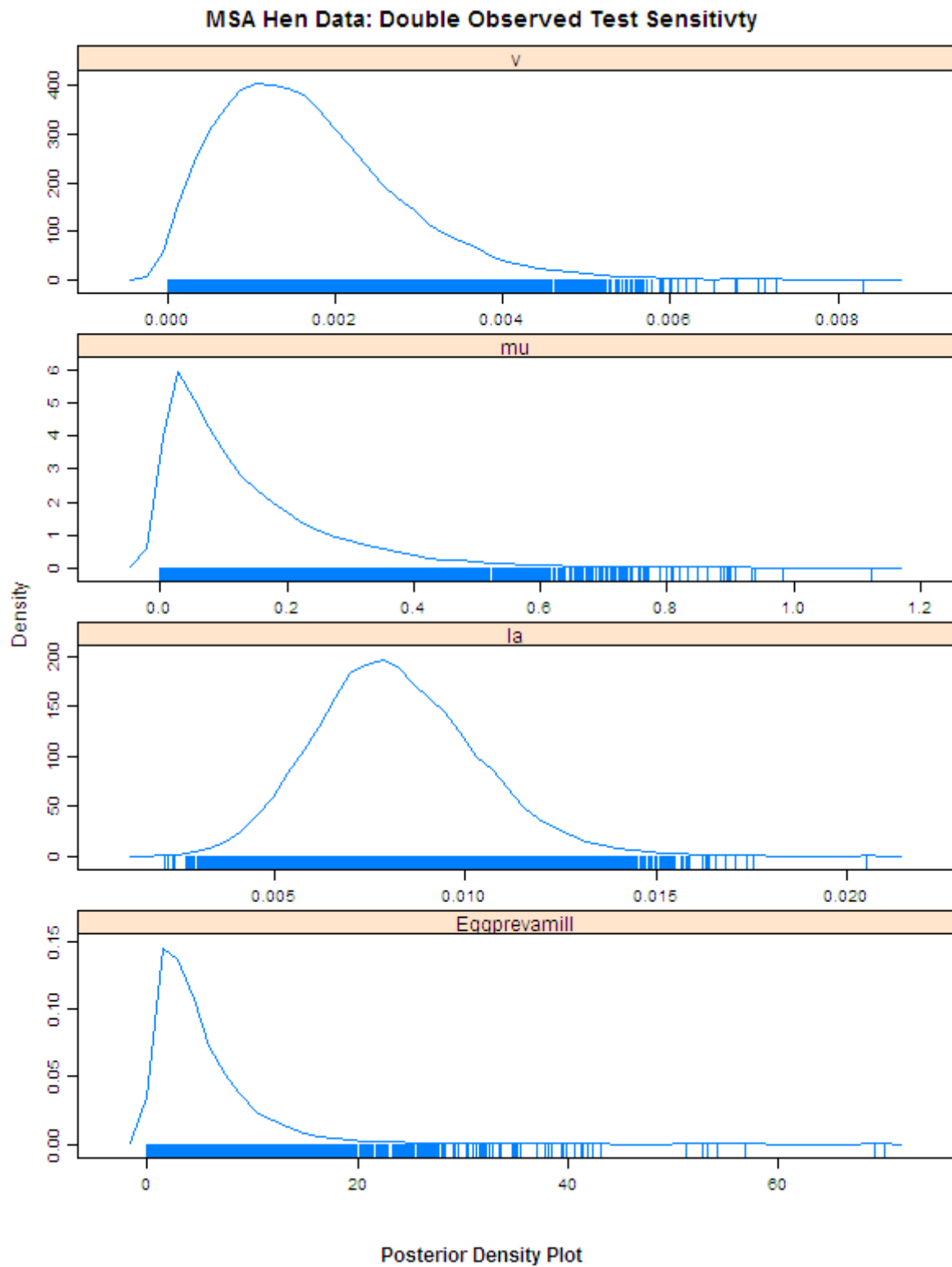


Figure 4: Posterior Distributions for lambda, mu, v and number of contaminated eggs per million for MSA with double the observed test sensitivity

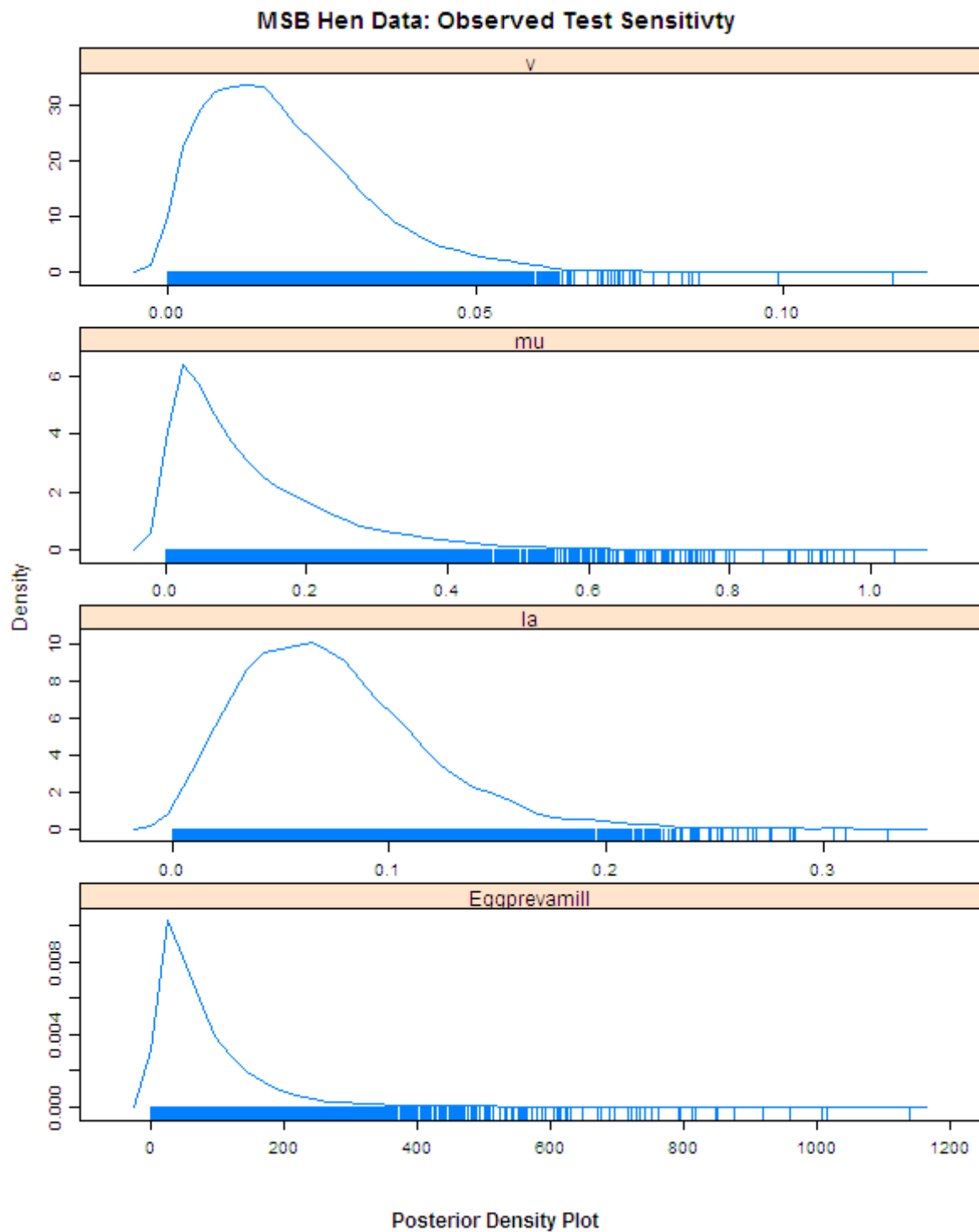


Figure 5: Posterior Distributions for lambda, mu, v and number of contaminated eggs per million for MSB with the observed test sensitivity

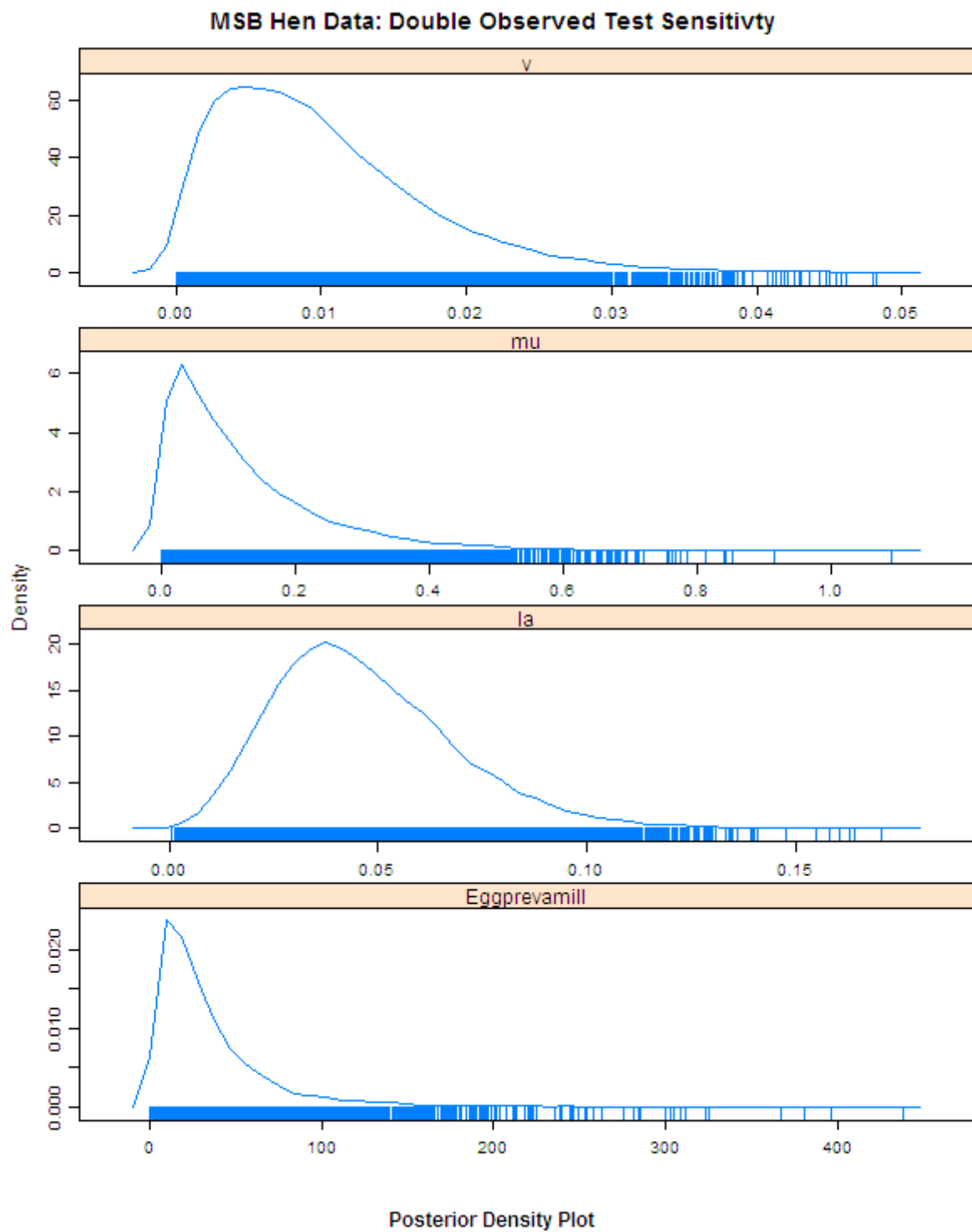


Figure 6: Posterior Distributions for lambda, mu, v and number of contaminated eggs per million for MSB with double the observed test sensitivity

WINBUGS PROGRAMS

The following are the key WinBUGS programs for the analysis (which have to be run under the WBDev environment within Black Box):

EFSA_Model_Laying_Hens.odc

```
#####
##
##   Program name: EFSA_Model_Laying_Hens.odc
##   AMU Directory: R:\Egg QMRA\Analysis\WinBUGS
##   Author: Ranta Jukka [Jukka.Ranta@evira.fi]
##   Updated by Saghir Bashir (EFSA)
##   Date Finalised: 18 FEB 2010
##
##   Mandate Number: M-2008-0111
##   Question Number: EFSA-Q-2008-292
##
## Useful publications:
##
## Original model (discrete time) in:
##
## Ranta J, Maijala R: A Probabilistic Transmission Model of
## Salmonella in the Primary Broiler Production Chain. Risk
## Analysis, Vol. 22, No. 1, 2002. 47-58.
##
## New model version (continuous time) in:
##
## Ranta J, Tuominen P, Mikkela A, Wahlstrom H: Bayesian hidden
## Markov process model of Salmonella over three generations of
## flocks in broiler production chain. Conference oral
## presentation abstract. S.co. 2009.
## Complex data modeling and computationally intensive
## statistical methods for estimation and prediction. Milan.
## http://mox.polimi.it/sco2009/
##
#####

##
### WinBUGS CODE: Start
##
### WinBUGS MODEL starts here
##

model{

# intensity of infection; uninformative, 'flat', prior here:
  la ~ dunif(0,100)
# Or: probability of no infection in two months time (theta),
# if expert opinion formulated about this:
  # theta ~ dbeta(?,?)
  # la <- -6*log(theta);

# intensity of recovery:
  mu <- -6*log(eta);
# probability of no recovery in two months time (eta),
# if expert opinion formulated about this:
  eta ~ dbeta(49,1)

# probability of initial infection. Assume no prior knowledge
# (uniform density) so that we rely in the model and data for estimation:
  v ~ dbeta(1,1)
# alternatively, could assume a fixed small value as a simplification.

# expected waiting time of a flock to get infected:
  ewt[1] <- 1/la

# expected waiting time of a flock to recover from infection:
  ewt[2] <- 1/mu
# Note: realistic values for both la and mu are small if we assume the
# infections to be rare, and the recovery to be difficult
# (persistent infection). Therefore, the inverse value can be large
```

```

# and its distributional shape (posterior distribution), especially its
# upper bound, difficult to estimate with Monte Carlo. The waiting time
# can exceed the lifetime of a flock.
# Note: the long run probability of infection would be  $la/(la+\mu)$ ,
# so that if this stationary prevalence would be known as "pstat", then:
#  $la = \mu * pstat / (1 - pstat)$ 
#-----
# Model for overall flock test sensitivity (as a function of age):
# (1) Simple functional approach: define it as the curve of gamma density
# function whose maximum value corresponds to a chosen peak sensitivity, # and this
# peak may be located at an unknown time point
# between 0 and AgeMax.
# OR (2) Simple approach defining the overall sensitivity (one parameter) # over all
# ages.
# OR (3) Step function approach defining the sensitivity as a step
# function over age, with specified levels and numbers of steps,
# with fixed or unknown change points.
# OR (4) assign sensitivity values directly at each testing age-point:
psen[1] <- psensit[1];
psen[2] <- psensit[2];
psen[3] <- psensit[2];
psen[4] <- psensit[3]
# Note: sensitivity values outside the actual testing ages are
# redundant and not possible to link with the actual data.
# Alternatives 1-3 are thus mostly superfluous, i.e. relevant only
# in terms of what the values at the given testing ages would be.

# After setting the time origin at the beginning of laying period,
# the testing times during laying period are,
# for example: 0.4615385-0.4 = 0.0615385 and so on. (time unit = 1 year).

# Specify the values of sensitivity as three possible levels,
# to be applied in a given order at all testing ages:
psensit[1] ~ dbeta(as[1],bs[1])
psensit[2] ~ dbeta(as[2],bs[2])
psensit[3] ~ dbeta(as[3],bs[3])

# losens[i] = test sensitivity lower bound (~95% in prior)
# upsens[i] = test sensitivity upper bound (~95% in prior)
# these bounds are given as data, for the 3 sensitivity levels (3
# parameters) applied to the different testing ages

Ms[1] <- (losens[1]+upsens[1])/2;
Vars[1] <- ((upsens[1]-losens[1])/4)*((upsens[1]-losens[1])/4)

as[1] <- -Ms[1]*(Ms[1]*Ms[1]-Ms[1]+Vars[1])/Vars[1];
bs[1] <- (Ms[1]*Ms[1]-Ms[1]+Vars[1])*(Ms[1]-1)/Vars[1]

Ms[2] <- (losens[2]+upsens[2])/2;
Vars[2] <- ((upsens[2]-losens[2])/4)*((upsens[2]-losens[2])/4)

as[2] <- -Ms[2]*(Ms[2]*Ms[2]-Ms[2]+Vars[2])/Vars[2];
bs[2] <- (Ms[2]*Ms[2]-Ms[2]+Vars[2])*(Ms[2]-1)/Vars[2]

Ms[3] <- (losens[3]+upsens[3])/2;
Vars[3] <- ((upsens[3]-losens[3])/4)*((upsens[3]-losens[3])/4)

as[3] <- -Ms[3]*(Ms[3]*Ms[3]-Ms[3]+Vars[3])/Vars[3];
bs[3] <- (Ms[3]*Ms[3]-Ms[3]+Vars[3])*(Ms[3]-1)/Vars[3]
#-----
meanpsen <- mean(psen[]) #average sensitivity over all testing ages
params[1] <- la;
params[2] <- mu;
params[3] <- v
rho[1:NT] <- Rhofun2(NT,params[1:3],psen[1:NT],T[1:NT])

# testing times i.e. observations
for(t in 2:NT){

# probability of observed data X, conditionally on previous observed
# history:
X[t] ~dbin(proba[t],NF) # flock with zero positives history X[1:t-1]
proba[t] <- psen[t]*ptrue[t]
ptrue[t] <- ((1-p10[t])*rho[t-1]+p01[t]*(1-rho[t-1]))
# Note: ptrue[t] is the "true hidden prevalence" of infected flocks

```

```

# in the population of flocks of age t
# (given that previous samplings were negative).
# proba[t] is the corresponding apparent prevalence.

# transition probabilities for the hidden state:
p01[t] <- (la/(la+mu))*(1-exp(-(la+mu)*(T[t]-T[t-1])));
p10[t]<- (mu/(la+mu))*(1-exp(-(la+mu)*(T[t]-T[t-1]))
)

# probability of observed data X[1]:
X[1] ~ dbin(proba[1],NF)
proba[1] <- psen[1]*ptrue[1]
ptrue[1]<- ((1-p10[1])*v+p01[1]*(1-v))

# transition probability for the hidden state:
p01[1] <- (la/(la+mu))*(1-exp(-(la+mu)*T[1]));
p10[1]<- (mu/(la+mu))*(1-exp(-(la+mu)*T[1]))

# calculating P(detection at t | survival t-1):
for(tt in 2:NT){

# calculate P(detection at tt | survival t-1 i.e. X[1]=0,...,X[t-1]=0)
pdet[tt] <- psen[tt]*(p01[tt]*(1-rho[tt-1])+(1-p10[tt])*rho[tt-1])
}

# First time point is now the 1st test during laying
pdet[1] <- psen[1]*(p01[1]*(1-v)+(1-p10[1])*v)

for(i in 1:NT){
  T[i]<-TT[i]-StartLayingAge;
  TX[i+1]<-T[i];
  rhoX[i+1]<-rho[i]; pdetX[i+1]<-pdet[i]
}

# Set the time origin at the beginning of laying (=StartLayingAge)
# so that v is now the probability to be infected at the beginning of
# laying

# time at the beginning of laying
TX[1] <- 0

# probability of infection at the beginning of laying
rhoX[1] <- v

# probability of detection at the beginning of laying
pdetX[1] <- 0

# In this function call:
# TimeIdx-1 time points during laying + one time point
# before that (at origin = start of laying).
# Note: indexing of rho, pdet and T has to match each other.
# (Same index refers to same time point).
# And the first element in pdet is left out and the first time point (in # T) is now
set to beginning of laying. (Time origin).
# EQ = expected average flock prevalence over whole production period
EQ <-EQfun2(TimeIdx,la,mu,rhoX[1:TimeIdx],pdetX[2:TimeIdx],TX[1:TimeIdx])
#-----

# Submodel for the egg prevalence in infected flocks:
# Proportion of internally contaminated eggs in infected flocks:
# combine several flock specific results
tau <- 1/(sig*sig); sig~dunif(0,100)

# muu,tau = hyper parameters for all flocks
muu ~dnorm(0,0.001)

for (i in 1:NUK){
  logitp[i] ~ dnorm(muu,tau)

# prevalence in flock i
p[i] <- exp(logitp[i])/(1+exp(logitp[i]))

# P(at least one pos egg in a pool)
ppool[i] <- 1-pow(1-p[i],PoolsizeUK)
psentest[i] <- ppool[i]*ps

```

```

# model of pooled test results
  XUK[i] ~ dbin(psentest[i],NpooledSamplesUK)
}

for (i in NUK+1:NUS+NUK){
  logitp[i] ~ dnorm(muu,tau)

# prevalence in flock i
  p[i] <- exp(logitp[i])/(1+exp(logitp[i]))

# P(at least one pos egg in a pool)
  ppool[i] <- 1-pow(1-p[i],PoolsizeUS)
  psentest[i] <- ppool[i]*ps

# model of pooled test results
  XUS[i-NUK] ~ dbin(psentest[i],NpooledSamplesUS)
}

for (i in NUS+NUK+1:NUS+NUK+NNUK){
  logitp[i] ~ dnorm(muu,tau)

# prevalence in flock i
  p[i] <- exp(logitp[i])/(1+exp(logitp[i]))
  psentest[i] <- p[i]*ps

# model for single tested eggs (UK)
  peggsUK[i-NUK-NUS] ~ dbin(psentest[i],neggsUK[i-NUK-NUS])
}

# posterior predictive distribution for the "next" flock:
# logitpNext ~ dnorm(muu,tau)
# pnext <- exp(logitpNext)/(1+exp(logitpNext))

# overall egg prevalence for infected flocks:
pmuu <- exp(muu)/(1+exp(muu))
pmuumill <- pmuu* 1000000 # per million eggs

# test sensitivity prior originally used with Henzler data:

# test sensitivity lower bound (~95%)
lo <- 0.3
# test sensitivity upper bound (~95%)
up <- 0.7

ms <- (lo+up)/2; vars <- ((up-lo)/4)*((up-lo)/4)
alpha_s <- -ms*(ms*ms-ms+vars)/vars;
beta_s <- (ms*ms-ms+vars)*(ms-1)/vars
ps ~ dbeta(alpha_s,beta_s)

# test sensitivity prior based on UK study (pools of 40, with 1
# inoculated egg): 8 detected in 12 trials
posits ~ dbin(ps,12); posits <- 8

#-----
# Resulting expected proportion of contaminated eggs in whole production:
Eggpreva <- EQ*pmuu
# per million eggs
Eggprevamill <- Eggpreva*1000000
}

##
### WinBUGS MODEL ends here
##

##
### WinBUGS Initial values
##

# (Some) initial values for MCMC:
list(eta=0.9,la=0.9,muu=-9,sig=2) #v=0.01

##
### Country A data (tested flocks and ages)
## (for egg prevalence, US and UK data are combined)

# Country A data: 2008, operator sampling

```



```

(*1*) MODULE WBDevEQfun3;
(* INPUT k: INTEGER; lambda, mu: REAL; rho, T: ARRAY k OF REAL; pdet: ARRAY k-1 OF REAL *)
(* This is for computing expected infection prevalence over laying phase *)
(* (i.e. flock population prevalence for breeding flocks), *)
(* given that the flock has survived over previous testing times. *)
(* It is assumed that a flock is destroyed whenever a testing result is positive. *)
(* Hence, the detection probabilities at any given testing time are *)
(* conditional to negative testing results at all previous testing times. *)
(* Flock is assumed to be born at time 0, its laying phase is assumed to start at time T1=0.4, *)
*)
(* But now T1=0 is set as a new origin of time *)
(* and the laying phase is assumed to end at time T2=1.2 (now 0.8 due to new origin) *)
(* when the flock is slaughtered. *)
(* The first testing time of input T should be the last testing time before laying phase (<T1)
-now set to 0 *)
(* The last testing time of input T should be the last testing time before slaughter (<T2)
*)

IMPORT
    WBDevScalar,
(*2*)    Math;

TYPE
    Function = POINTER TO RECORD (WBDevScalar.Node) END;
    Factory = POINTER TO RECORD (WBDevScalar.Factory) END;

VAR
    fact-: WBDevScalar.Factory;

(*3*) PROCEDURE (func: Function) DeclareArgTypes (OUT args: ARRAY OF CHAR);
(*4*) BEGIN
(*5*)     args := "sssVVV";
(*6*) END DeclareArgTypes;

(*7*) PROCEDURE (func: Function) Evaluate (OUT value: REAL);
CONST
    input = 0;
    lainput = 1;
    muinput = 2;
    rhoinput = 3;
    pdetinput = 4;
    Tinput = 5;
    T1 = 0; (* beginning of laying phase 0.4 -> change origin to T1=0*)
    T2 = 1.0; (* end of laying phase 1.2 -> 1.4 (73 weeks) -> 1.0, due to change of
origin *)
(*10*) VAR
(*11*)     i, k: INTEGER;
    lambda, mu: REAL;
    rho, pdet, Psurvlay, integral, Q, Q0, Q1, T, QQ, Qsum, Qsumw: POINTER TO ARRAY
OF REAL;
(*12*) BEGIN
    k := SHORT( ENTIER (func.arguments[input][0].Value() ));
    lambda := func.arguments[lainput][0].Value();
    mu := func.arguments[muinput][0].Value();
    NEW(rho, k); (* probabilities of true infection at given testing times *)
    NEW(pdet, k-1); (*probability of detection at given testing times *)
    NEW(integral, k); (* expected average infection prevalence over laying phase, *)
    (* accounting for possible destruction of flock whenever a positive
test result *)
    NEW(Psurvlay, k-1); (* probability of survival over consecutive testing times
during laying phase*)
    NEW(T, k);
    NEW(Q, k); (* weighted average of two average infection probabilities Q0 and
Q1*)
    NEW(Q0, k); (* average infection probability over a time interval, started as not
infected *)
    NEW(Q1, k); (* average infection probability over a time interval, started as
infected *)
    NEW(QQ, k); (* weighted Q, weighting by length of time interval*)
    NEW(Qsum, k); (*cumulative sum of Q*)
    NEW(Qsumw, k); (*weighted average Q over several intervals from the 1st to ith,
i=1,...,k*)
    FOR i := 0 TO k-1 DO
        rho[i] := func.arguments[rhoinput][i].Value();
        T[i] := func.arguments[Tinput][i].Value();
    END;

```

```

FOR i := 0 TO k-2 DO (*pdet[0] is prob of detection at T[1], NOT at T[0] *)
  pdet[i] := func.arguments[pdetinput][i].Value();
END;

(* First time interval starts at T1=T[0]=0 either as 0 or 1, and ends at time
T[1], the 1st. test during laying*)
Q0[0] := lambda/(lambda+mu) + ( lambda/((lambda+mu)*(lambda+mu)) )*( Math.Exp(-
(lambda+mu)*T[1])-Math.Exp(-(lambda+mu)*T[0]))/(T[1]-T[0]);
Q1[0] := lambda/(lambda+mu) - ( mu/((lambda+mu)*(lambda+mu)) )*( Math.Exp(-
(lambda+mu)*T[1])-Math.Exp(-(lambda+mu)*T[0]))/(T[1]-T[0]);
Q[0] := rho[0]*Q1[0] + (1-rho[0])*Q0[0]; (* rho = probability to start as 1*)
QQ[0] := Q[0]*(T[1]-T1); (* QQ=weighted Q, weighting by length of time,
T1=T[0]=0 *)

FOR i := 1 TO k-2 DO (* here every interval, length T[i+1]-T[i] starts either
as 0 or 1 at T[i] *)
  Q0[i] := lambda/(lambda+mu) + ( lambda/((lambda+mu)*(lambda+mu)) )*(
Math.Exp(-(lambda+mu)*(T[i+1]-T[i]))-Math.Exp(-(lambda+mu)*0))/(T[i+1]-T[i]);
  Q1[i] := lambda/(lambda+mu) - ( mu/((lambda+mu)*(lambda+mu)) )*(
Math.Exp(-(lambda+mu)*(T[i+1]-T[i]))-Math.Exp(-(lambda+mu)*0))/(T[i+1]-T[i]);
  Q[i] := rho[i]*Q1[i] + (1-rho[i])*Q0[i];
  QQ[i] := Q[i]*(T[i+1]-T[i]); (* QQ=weighted Q *)
END;
(* The last interval, length T2-T[k-1] starts either as 0 or 1 at T[k-1]*)
Q0[k-1] := lambda/(lambda+mu) + ( lambda/((lambda+mu)*(lambda+mu))
)*( Math.Exp(-(lambda+mu)*(T2-T[k-1]))-Math.Exp(-(lambda+mu)*0))/(T2-T[k-1]);
Q1[k-1] := lambda/(lambda+mu) - ( mu/((lambda+mu)*(lambda+mu)) )*(
Math.Exp(-(lambda+mu)*(T2-T[k-1]))-Math.Exp(-(lambda+mu)*0))/(T2-T[k-1]);
Q[k-1] := rho[k-1]*Q1[k-1] + (1-rho[k-1])*Q0[k-1];
QQ[k-1] := Q[k-1]*(T2-T[k-1]); (* QQ=weighted Q *)

Qsum[0]:=QQ[0]; (* weighted sum of Q, over time periods, weighting by length of
time *)
FOR i := 1 TO k-1 DO
  Qsum[i] := Qsum[i-1]+QQ[i];
END;
FOR i:= 0 TO k-2 DO (*weighted average of Q, over time periods, weighting by
length of time *)
  Qsumw[i] := Qsum[i]/(T[i+1]-T[0]); (* note T[0]=T1=start of laying*)
END;
Qsumw[k-1]:=Qsum[k-1]/(T2-T[0]); (* note T2=end of laying*)

Psurvlay[0]:= (1-pdet[0]);
integral[0]:= Qsumw[0]*pdet[0];
FOR i := 1 TO k-2 DO
  integral[i]:= integral[i-1]+Qsumw[i]*Psurvlay[i-1]*pdet[i];
  Psurvlay[i]:=Psurvlay[i-1]*(1-pdet[i]);
END;
integral[k-1] := integral[k-2] + Qsumw[k-1]*Psurvlay[k-2];

value := integral[k-1];
END Evaluate;
(*26*)

PROCEDURE (f: Factory) New (option: INTEGER): Function;
VAR
  func: Function;
BEGIN
  NEW(func); func.Initialize; RETURN func;
END New;

PROCEDURE Install*;
BEGIN
  WBDevScalar.Install(fact);
END Install;

PROCEDURE Init;
VAR
  f: Factory;
BEGIN
  NEW(f); fact := f;
END Init;

BEGIN
  Init;
(*1*) END WBDevEQfun3.

```

Rhofun2.odc

```
(*1*) MODULE WDevRhofun2;
(* INPUT k: INTEGER; lambda, mu, ver: REAL; psen ARRAY k OF REAL; T ARRAY k OF REAL *)
(* This is the recursive formula for the true infection probability at times T[1:k], *)
(* conditional to observed negative results (test sensitivity psen[i]) at the same time point *)
(* and *)
(* all previous observation times, based on Markov process model (intensities lambda,mu), *)
(* starting as infected at time 0 with probability ver. *)
IMPORT
    WDevVector,
(*2*)    Math;

TYPE
    Function = POINTER TO RECORD (WDevVector.Node) END;
    Factory = POINTER TO RECORD (WDevVector.Factory) END;

VAR
    fact-: WDevVector.Factory;

(*3*) PROCEDURE (func: Function) DeclareArgTypes (OUT args: ARRAY OF CHAR);
(*4*) BEGIN
(*5*)     args := "svvv";
(*6*) END DeclareArgTypes;

(*7*) PROCEDURE (func: Function) Evaluate (OUT values: ARRAY OF REAL);
CONST
    input = 0;
    paraminput = 1;
    pseninput = 2;
    Tinput = 3;
(*10*) VAR
(*11*)     i, k: INTEGER;
    lambda, mu, ver: REAL;
    rho, T, p01, p10, psen: POINTER TO ARRAY OF REAL;
(*12*) BEGIN
    k := SHORT( ENTIER (func.arguments[input][0].Value() )); (*number of time
points*)
    lambda :=func.arguments[paraminput][0].Value(); (*infection intensity*)
    mu :=func.arguments[paraminput][1].Value(); (*recovery intensity*)
    ver :=func.arguments[paraminput][2].Value(); (*vertical transmission
probability for initial state*)
    NEW(rho, k); (*probability of true infection at given time points*)
    NEW(p01,k); (*transition probability from 0 to 1 in a given time interval*)
    NEW(p10,k); (*transition probability from 1 to 0 in a given time interval*)
    NEW(psen,k); (*sensitivity at given time points*)
    NEW(T,k);

    FOR i := 0 TO k-1 DO
        T[i] :=func.arguments[Tinput][i].Value(); (* given time points*)
        psen[i] :=func.arguments[pseninput][i].Value(); (* sensitivity of
testing method *)
    END;

    p01[0] := Math.Exp(
    Math.Ln(lambda)-Math.Ln(lambda+mu)+Math.Ln(1-Math.Exp( -(lambda+mu)*T[0] ))
    );
    p10[0] := Math.Exp(
    Math.Ln(mu)-Math.Ln(lambda+mu)+Math.Ln(1-Math.Exp( -(lambda+mu)*T[0] )) );
    rho[0] := Math.Exp(
    Math.Ln(1-psen[0])+Math.Ln( (1-ver)*p01[0]+ver*(1-p10[0]) ) -
    Math.Ln( (1-psen[0])*((1-ver)*p01[0]+ver*(1-p10[0])) + (1-ver)*(1-
p01[0])+ver*p10[0] )
    );
    values[0] := rho[0];
    FOR i:= 1 TO k-1 DO
        p01[i] :=Math.Exp(
        Math.Ln(lambda)-Math.Ln(lambda+mu)+Math.Ln(1-Math.Exp( -(lambda+mu)*(T[i]-T[i-1])
))
        );
        p10[i] := Math.Exp(
        Math.Ln(mu)-Math.Ln(lambda+mu)+Math.Ln(1-Math.Exp( -(lambda+mu)*(T[i]-T[i-1]) ))
        );
        rho[i] := Math.Exp(
```

```

Math.Ln(1-psen[i])+Math.Ln((1-p01[i]-p10[i])*rho[i-1]+p01[i]) -
Math.Ln(1-psen[i]*((1-p01[i]-p10[i])*rho[i-1]+p01[i]) )
);
  values[i] := rho[i];
END;

(*26*)      END Evaluate;

PROCEDURE (f: Factory) New (option: INTEGER): Function;
VAR
  func: Function;
BEGIN
  NEW(func); func.Initialize; RETURN func;
END New;

PROCEDURE Install*;
BEGIN
  WBDevVector.Install(fact);
END Install;

PROCEDURE Init;
VAR
  f: Factory;
BEGIN
  NEW(f); fact := f;
END Init;

BEGIN
  Init;
(*1*)      END WBDevRhofun2.

```

R PROGRAMS

The following 2 R programs were used to automate the running of the analysis especially given the number of scenarios that had to be investigated.

EFSA_Write_HenData.R

```

#####
##
##   Program name: EFSA_Write_HenData.R
##   AMU Directory: R:\Egg_QMRA\Analysis\WinBUGS\Final\
##   Author: Saghir Bashir
##   Date Finalised: 18 FEB 2010
##
##   Mandate Number: M-2008-0111
##   Question Number: EFSA-Q-2008-292
##
##   Description: An R function to recalculate the observed hen data for the
##               Member State A (MSA) & Member State B (MSB)
##
##               Write the data files for the WinBUGS data that are needed.
##
##               Write the script file for the full analysis
##
#####

##
## Reading in all the libraries that are needed by this program and setting the
## "global" parameters that will be used by this program.
##

# Key parameters that could be useful later
workdir = "R:/Egg_QMRA/Analysis/WinBUGS/Final/"
outdir  = "R:/Egg_QMRA/Analysis/WinBUGS/Final/"
programe = "EFSA_Write_HenData.R"

# stem for the output files from coda.
hstem = "EFSA_Data_Hens_"

```

```

## A macro to run this program from within R without needing to copy & paste each time
qrun = function(){
  source(paste(workdir,progname,sep=""))
}

# Note any negative number corresponds to the original data
scenarios = c(-9, 0.0088, 0.01, 0.0176, 0.02, 0.031, 0.136)

## A function to recalculate the "observed" data based on the original data
recalc.obs = function(X, NF, sprp) {
  # X = observed infected flocks
  # NF = Number of flocks
  # NT = Number of measured time points
  # sprp = Scenario for hypothetical observed propotion of flocks infected

# Total prop of flocks infected
tprp = sum(X)/NF

# Calculate the total number of infections under scenario
# Note the negative proportions mean that we should keep the original X
tnum = ifelse(sprp<0, sum(X), sprp*Nf)
cat("\n tnum = ", tnum, " tprp = ", tprp, "sprp = ", sprp, "\n")

NewX = tnum*(X/sum(X))
cat(" X = c(", X, ") & NewX = c(", NewX, ")\n")
round(NewX)
}

write.hens = function(outloc, country, sX, sNF, scen, index){

# File where the data will stored:
fstem = "EFSA_Data_Hens_"
filename = paste(outloc, fstem, country, "-", index, ".txt", sep="")

# Header for the output data
head = paste("## Hen data \n## Country:", country,"\n## Scenario", index, ":", scen, "\n")
# Recalculating the "observed" data
NewObs = recalc.obs(sX, sNF, sprp=scen)
start = paste("\nlist(\n X=c(", toString(NewObs), ")\n NF=", toString(sNF), ",", sep="")

## The hen data
BBtxt = "\n TimeIdx=4,\n NT=3,\n TT=c(0.4615385, 0.7500000, 1.0384615)\n )\n"
AAtxt = "\n NT=4,\n TimeIdx=5,\n TT=c(0.4807692, 0.7692308, 1.0576923, 1.3461538)\n )\n"

full = paste(head, start, eval(parse(text=paste(country, "txt", sep="))))
write(full, file=filename)
full
}

sendat.wrt = function(senwgt){
  # Writing the scenario data out

  # Losens and upsens will depend on the senwgt but the psen[] does not need to multiplied
  # by this factor as the losens and upsens will double the sensitivity by changing the
  # priors.
  e1 = "\n StartLayingAge=0.4,\n losens=c(0.3,0.2,0.25),\n upsens=c(0.4,0.3,0.35)\n)"
  e2 = "\n StartLayingAge=0.4,\n losens=c(0.6,0.4,0.5),\n upsens=c(0.8,0.6,0.7)\n)"
  othdat = eval(parse(text=paste("e", senwgt, sep="")))

  top = paste("## Automatically generated\n##\n## Data for the different scenarios:",
    "\n## senwgt = weight applied to the sensitivity\n##\n", sep="")

  sendat = paste("list(\n senwgt = 1,", othdat, "\n")
  senfile = paste(workdir, "EFSA_Data_Psen_", senwgt, ".txt", sep="")
  write(paste(top, sendat, sep=""), file=senfile)
}

##
## Writing a Script file automatically for the different scenarios for psens too.
##

script.bugs = function(fname, hen, egg, suf, scen, senwgt){
  # hen = hen data (AA, BB)
  # egg = egg data (UKp, UKi, UKp-UKi)

```

```

# suf = Scenario suffix (1 to 7)
# senwgt = Weight for testing sensitivity (1 or 2)

codafile = paste(workdir, "codaR/EFSA_coda_H", hen, "-", suf, "_", egg, "_Psen", senwgt,
sep="")
senfile = paste("EFSA_Data_Psen_", senwgt, ".txt", sep="")

head1 = paste("\n##\n## Hen Data:", hen, "\n## Egg data:", egg, "\n##")
head2 = paste(" Scenario:", suf, "\n## Prevalence:", scen, "\n##")
head3 = paste(" Psen wgt:", senwgt, "\n##\n\n")

check = paste(" check('", workdir, "EFSA_Model_", egg, ".txt')\n", sep="")
hdat = paste(" data('", workdir, "EFSA_Data_Hens_", hen, "-", suf, ".txt')\n", sep="")
sdat = paste(" data('", workdir, senfile, ")\n", sep="")
euki = ifelse(grep("UKi", egg), paste(" data('", workdir, "Egg_UKi.txt')\n", sep=""), "")
eukp = ifelse(grep("UKp", egg), paste(" data('", workdir, "Egg_UKp.txt')\n", sep=""), "")

brun = " compile(1)
inits(1, 'R:/Egg_QMRA/Analysis/WinBUGS/Final/EFSA_Inits_Laying_Hens.txt')
gen.inits()
update(10000)
set(Eggprevamill)
set(la)
set(mu)
set(v)
set(psen)
update(10000)
history(*)
stats(*)
density(*)\n"
coda = paste(" coda(*, "", codafile, ")\n\n", sep="")

fullscript = paste(head1, head2, head3, check, hdat, sdat, euki, eukp, brun, coda)
write(fullscript, fname, append=TRUE)
}

###
### Writing the hen data files and the automatic script file.
###

fullhead =
quote("#####
##
## Program name: Laying_Hens_Script_All.odc
## AMU Directory: R:\\Egg_QMRA\\Analysis\\WinBUGS\\Final
## Author: Saghir Bashir
## Date Finalised: 19 FEB 2010
##
## Mandate Number: M-2008-0111
## Question Number: EFSA-Q-2008-292
##
## Description: Running the laying hen analysis automatically via this script
## file. We will run 6 different models based on the different
## published egg data that we have.
##
## We will also look at the different scenarios for the initial
## flock infection (v).
##
## NOTE: Run this script from WinBUGS WITHIN Blackbox
##
#####

#_#
#_# Abbreviations:
#_#
#_# LHA: Lay hen data for MSA
#_# LHB: Lay hen data for MSB
#_# UKp: UK pooled egg data
#_# UKi: UK individual egg data
#_# USp: US pooled egg data
#_#

display('log')\n\n")

```

```
##
## Automatically generate the script file to append too.
##
scrfile = paste(workdir, "EFSA_Script_Hens_AUTO.txt", sep="")
atxt1 = "##\n## This file is automatically generated by the following program:\n##\n## "
atxt2 = paste(workdir, progname, "\n##", sep="")
atxt3 = paste("\n## Generated:", date(), "\n##")
write(paste(fullhead, atxt1, atxt2, atxt3), file=scrfile, append=FALSE)

iX = list(BB=c(2,1,4), AA=c(4,8,5,7))
iNF = list(BB=325, AA=3500)

for (dhen in c("BB", "AA")){
  write(paste("\n#_#\n#_# (1) Country START:", dhen, "\n#_#\n"), file=scrfile, append=TRUE)
  for (psen in c(1,2)){
    write(paste("\n#_#\n#_# (2) Psen START: psen = ", psen, "\n#_#\n"), file=scrfile,
append=TRUE)
    if(dhen=="BB"){sendat.wrt(psen)}
    for (degg in c("UKi", "UKp", "UKp-UKi")){
      suffix = 1
      for (i in scenarios){
        script.bugs(scrfile, dhen, degg, suffix, i, psen)

        # The ifelse is just to prevent files being written multiple times.
        if(degg=="UKi"){write.hens(outdir, dhen, iX[[dhen]], iNF[[dhen]], i, suffix)}

        suffix <- suffix+1
      }
    }
    write(paste("\n#_#\n#_# (2) Psen END: psen = ", psen, "\n#_#\n"), file=scrfile,
append=TRUE)
  }
  write(paste("\n#_#\n#_# (1) Country END:", dhen, date(), "\n#_#\n"), file=scrfile,
append=TRUE)
}

## Adding the place to save the log file at the end

save = paste("##\n## Save the log file\n##\n save(' ", workdir, "EFSA_Results_Hens_Final')\n",
sep="")
write(save, file=scrfile, append=TRUE)
```

EFSA_Output_Final.R

```
#####
##
## Program name: EFSA_Output_Final.R
## AMU Directory: R:\Egg_QMRA\Analysis\WinBUGS\Final\codar
## Author: Saghir Bashir
## Date Finalised: 18 FEB 2010
##
## Mandate Number: M-2008-0111
## Question Number: EFSA-Q-2008-292
##
## Description: Reading the output files from Final analysis carried out in
## WinBUGS (via Blackbox) and producing tables for the reports.
##
## There are many scenarios that will be considered
##
#####

##
## Reading in all the libraries that are needed by this program and setting the
## "global" parameters that will be used by this program.
##

library(coda)

# Key parameters that could be useful later
workdir = "R:/Egg_QMRA/Analysis/WinBUGS/Final/"
codadir = "R:/Egg_QMRA/Analysis/WinBUGS/Final/codar/"
outdir = "R:/Egg_QMRA/Analysis/WinBUGS/Final/SharedResults/"
progname = "EFSA_Output_Final.R"
```

```

# stem for the output files from coda.
hstem = "EFSA_coda_"

#
# Defining the working directory where the winBUGS outputs are stored and the
# results will be saved.
#

setwd(codadir)

# A macro to run this program from within R without needing to copy & paste each time
qrun = function(){
  source(paste(workdir,progname,sep=""))
}

##
## Useful abbreviations to know for later:
##
##   HAA: MSA hen data
##   HBB: MSB hen data
##   UKi: UK individual egg data
##   UKp: UK pooled egg data
##   nscen: Number of scenarios
##

##
## A function to produce some "pretty" tables from the summary statistics
##
##   hens.dat = Hen data (HAA or HBB)
##   eggs.dat = Egg data (UKi, UKp or UKp-UKi)
##   chains = Chain from WinBUGS
##

hens.read = function(stem, hendata, eggdata, scenar, senwgt){
  ## Read in the data to summarise
  hens.stem = paste(stem, hendata, "-", scenar, "_", eggdata, "_Psen", senwgt, sep="")
  # There is only one chain hence the 1.txt
  hens.out = paste(hens.stem, "1.txt", sep="")
  hens.ind = paste(hens.stem, "Index.txt", sep="")
  hens.dat = read.coda(hens.out, hens.ind)
  hens.dat
}

# Summarize the data
hens.tab = function(stem, hendata, eggdata, scenar, senwgt){

  hens.dat = hens.read(stem, hendata, eggdata, scenar, senwgt)
  ## Summarys statistics for the WinBUGS outputs

  hens.sum = summary(hens.dat)
  hens.pct = hens.sum[[2]]

  ## Some variables to add to the dataframe that will be given at the end of this
  ## function
  pnum = dim(hens.pct)[1]
  cat("\n pnum:", pnum, "\n")
  base.dat = data.frame(hen=rep(hendata, pnum), egg=rep(eggdata, pnum),
                        scenario=rep(scenar, pnum), psenwgt=rep(senwgt, pnum))

  hens.tab = data.frame(base.dat, data.frame( param = row.names(hens.pct),
                                             lower = hens.pct[,1], median = hens.pct[,3], upper = hens.pct[,5]),
                        row.names = NULL)

  hens.tab
}

Hen.prep = NULL
for (dhen in c("HBB", "HAA")){
  for (psen in c(1,2)){
    for (degg in c("UKi", "UKp", "UKp-UKi")){
      negg = ifelse(degg=="UKp-UKi", "UKpUKi", degg)
      for (i in 1:7){
        ndat = paste(dhen, i, negg, sep="")
        cat("\n -->", ndat, "\n")
        ndat = hens.tab(hstem, dhen, degg, i, psen)
      }
    }
  }
}

```



```

        Hen.prep <- rbind(Hen.prep, ndat)
    }
}
}

##
## Summarize the results
##

scendat = data.frame(scenario=1:7,
                    Scenario.text = c("Observed", "0.88%", "1.00%", "1.76%", "2.00%", "3.10%",
"13.60%"))
snumdat = data.frame(scenario=1:7,
                    Scenario.num = c(NA, 0.88, 1.00, 1.76, 2.00, 3.10, 13.60))
country = data.frame(hen=c("HBB", "HAA"), HenData=c("MSB", "MSA"))
eggdat = data.frame(egg=c("UKi", "UKp", "UKp-UKi")
                    , EggData=c("UK Individual", "UK Pooled", "UK Pooled \& Individual"))

Hen.scn = merge(Hen.prep, scendat, by.x="scenario", by.y="scenario")
Hen.num = merge(Hen.scn, snumdat, by.x="scenario", by.y="scenario")
Hen.egg = merge(Hen.num, eggdat, by.x="egg", by.y="egg")
Hen.cty = merge(Hen.egg, country, by.x="hen", by.y="hen")
Hen.Sum = Hen.cty[order(Hen.cty$hen, Hen.cty$egg, Hen.cty$param, Hen.cty$psenwgt,
Hen.cty$scenario),]

## Write summary statistics to comma separated file
write.csv(Hen.Sum, file=paste(outdir,"EFSA_Hen_Analysis_Summary.csv", sep=""))

##
## Some plots for the WG
##

# Egg prevalence per million

eprev = subset(subset(Hen.Sum, param=="Eggprevamill"), scenario>1)
summary(eprev)

##
## Plotting the scenario versus the eggs/million. There will be 4 lines on the graph:
##
## * 2 for each country (MSA & MSB) for each sensitivity weight
## * for each of the different egg models (UKi UKp, UKp-UKi)
##

plot.eprev = function(prevedf){
## 4 plots on one page.
    par(mfrow=c(2,2))

    for (degg in c("UKi", "UKp", "UKp-UKi")){
        pdat = subset(prevedf, egg==degg)
        ## Plot all the points in white just to create the axis as automatically as possible.
        plot(pdat$Scenario.num, pdat$median, xlim=c(0,15), col="white",
            , main= paste("Egg data:", unique(pdat$EggData[pdat$egg==degg]))
            , xlab="Scenarios For 'Observed' Flock Prevalence (%)"
            , ylab="Egg Prevalence / Million")

        ## Plot the lines
        for (dhen in c("HBB", "HAA")){
            # Symbol for the county
            csym = ifelse(dhen=="HBB", 1, 2)
            for (psen in c(1,2)){
                lines(pdat$Scenario.num[pdat$hen==dhen & pdat$psenwgt==psen]
                    , pdat$median[pdat$hen==dhen & pdat$psenwgt==psen]
                    , lty=psen, pch=csym, type="b")
            }
        }
    }
}

## Legend in the final plot region
plot(c(0,25), c(0,5) , type="n", axes=F, xlab="", ylab="")
points(2, 5, pch=1)
text(3, 5, "MSB Hen Data", pos=4, pch=1)
points(2, 4, pch=2)
text(3, 4, "MSA Hen Data", pos=4, pch=1)
lines(c(1,3), c(3,3), lty=1)

```

```

text(3, 3, "Observed Test Sensitivity", pos=4)
lines(c(1,3), c(2, 2), lty=2)
text(3, 2, "Double Observed Test Sensitivity", pos=4)
}

#Plotting all the 3 models with a legend on one page and outputting to pdf and png formats:
pdf(file=paste(outdir,"EFSA_Hen_Scenario_Plot.pdf", sep=""), paper="a4")
  plot.eprev(eprev)
dev.off()

png(file=paste(outdir,"EFSA_Hen_Scenario_Plot.png", sep="")
    , width = 640, height = 640)
  plot.eprev(eprev)
dev.off()

##
## Some posterior and prior distribution density plots for lambda, v and mu.
##

# lambda, mu & v for the MSB model
HBB_S1_P1 = hens.read(hstem, "HBB", "UKp-UKi", 1, 1)
HBB_S1_P2 = hens.read(hstem, "HBB", "UKp-UKi", 1, 2)
HAA_S1_P1 = hens.read(hstem, "HAA", "UKp-UKi", 1, 1)
HAA_S1_P2 = hens.read(hstem, "HAA", "UKp-UKi", 1, 2)

# Density Plots in a single PDF file
pdf(file=paste(outdir,"EFSA_Plot_Densities.pdf", sep=""), paper="a4")
  densityplot(HBB_S1_P1[,c(1:3,8)], main="MSB Hen Data: Observed Test Sensitivity")
  densityplot(HBB_S1_P2[,c(1:3,8)], main="MSB Hen Data: Double Observed Test Sensitivity")
  densityplot(HAA_S1_P1[,c(1:3,8)], main="MSA Hen Data: Observed Test Sensitivity")
  densityplot(HAA_S1_P2[,c(1:3,8)], main="MSA Hen Data: Double Observed Test Sensitivity")
dev.off()

# Posterior Density plots in PNG format
png(file=paste(outdir,"EFSA_Plot_Posterior_Density_BB-P1.png", sep=""), width=600,
height=800)
  densityplot(HBB_S1_P1[,c(1:3,8)], main="MSB Hen Data: Observed Test Sensitivity"
    , sub="Posterior Density Plot")
dev.off()

png(file=paste(outdir,"EFSA_Plot_Posterior_Density_BB-P2.png", sep=""), width=600,
height=800)
  densityplot(HBB_S1_P2[,c(1:3,8)], main="MSB Hen Data: Double Observed Test Sensitivity"
    , sub="Posterior Density Plot")
dev.off()

png(file=paste(outdir,"EFSA_Plot_Posterior_Density_AA-P1.png", sep=""), width=600,
height=800)
  densityplot(HAA_S1_P1[,c(1:3,8)], main="MSA Hen Data: Observed Test Sensitivity"
    , sub="Posterior Density Plot")
dev.off()

png(file=paste(outdir,"EFSA_Plot_Posterior_Density_AA-P2.png", sep=""), width=600,
height=800)
  densityplot(HAA_S1_P2[,c(1:3,8)], main="MSA Hen Data: Double Observed Test Sensitivity"
    , sub="Posterior Density Plot")
dev.off()

# Prior Density plots for lambda, mu and v
png(file=paste(outdir,"EFSA_Plot_Prior_Density.png", sep=""), width=600, height=800)
  par(mfrow=c(3,1), cex=1.1)
  x <- seq(0, 1, by = 0.01)
  muprior <- -6*log(dbeta(x, 49, 1))
  vprior = dbeta(x, 1, 1)
  plot(c(0,100), c(0.01, 0.01), ylab="Density", xlab="la (lambda prior)", type='l',
    main="Prior Density Plots")
  plot(x, muprior, ylab="Density", xlab="mu (prior)", type='l')
  plot(x, vprior, ylab="Density", xlab="v (prior)", type='l')
dev.off()

```

ABBREVIATIONS

EELA	National Veterinary and Food Research Institute. After 2006: Finnish Food Safety Authority, Evira.
MSA	EU Member State A
MSB	EU Member State A
UKi	UK individual egg data
UKp	UK pooled egg data
WG	Working Group