EuroReference
journal of Reference
anses
French agency for food, environmental and occupational health & safety

Journal No. 11
Winter 2013

Research

Modelling *Listeria monocytogenes* contamination to improve surveillance in the agri-food industry

Natalie Commeau (natalie.commeau@gmail.com)
INRA/AgroParisTech, UMR 518 MIA, Paris, France
ANSES, Laboratory for Food Safety, Maisons-Alfort, France

Agri-food companies are accountable for the quality of the products that they place on the market. One way to check this quality is to determine how contamination is distributed. A sampling plan would be a useful decision-support tool. To determine the optimal batch sample size, we used an approach based on Bayesian decision theory for finished food products that minimises the average cost incurred by the manufacturer. Here, we used data on the presence of *Listeria monocytogenes* during the production of diced bacon. We built models to describe the *L. monocytogenes* concentration by taking into account various factors, we estimated parameters using Bayesian inference and then compared our models with real data. Finally, we developed a model to determine how to minimise the average costs incurred by a meat-processing company in the case of *L. monocytogenes* contamination in diced bacon.

Within- and between-batch sampling

Knowledge on contamination by pathogens in a food-processing plant is necessary for agri-food companies so that they can take appropriate actions to reduce contamination. To acquire this knowledge, analyses are necessary (e.g. counting or screening) on food products or surfaces. How should samples be taken at a given point in the production process? Should a sample be randomly chosen from the production line? Or should it be chosen randomly from each batch? These questions are not trivial because, according to the finished food product and the processing method, variability in contamination within and between batches can be very different. **Figure 1** shows two hypothetical cases of distribution of contamination among several batches. In Figure 1a, the between-batch variability is much lower than the within-batch variability. In this case, randomly choosing a sample from the entire production line without considering batch identity is sufficient. However, if the distribution of contamination resembles that shown in Figure 1b, sampling by batch is essential for determining whether a given batch is contaminated or not. Within- and between-batch variability has been studied recently (ILSI, 2010; Gonzales-Barron and Butler, 2011).

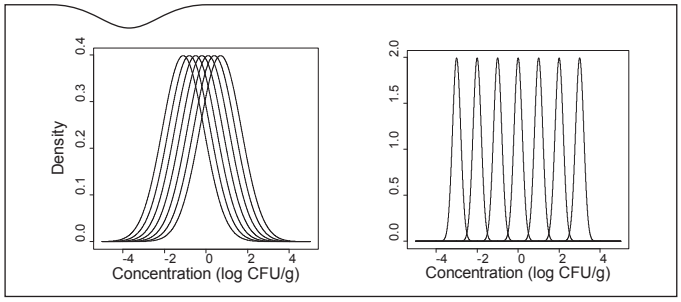

Figure 1: Representation of the variability within and between batches. Each curve shows the distribution of contamination in a given batch (log CFU/g). In Figure 1a (left), the standard deviation of contamination in a batch is 1 log CFU/g and the standard deviation between is 0.3 log CFU/g. For Figure 1b (right), the between-batches standard deviation is equal to 1 log CFU/g and the within-batch standard deviation is equal to 0.2 log CFU/g.

We begin by defining the term 'batch'. Although this term is used in everyday speech, it is not simple to define. According to European Commission Regulation (EC) No 2073/2005 (Article 2), a batch is "a group or set of identifiable products obtained from a given process under practically identical circumstances and produced in a given place within one defined production period." The definition given by the International Commission on Microbiological Specifications for Foods (ICMSF, 2002) begins by explaining that it is a quantity of food manufactured and handled in uniform conditions, but it goes further and indicates that this definition implies that the batch is homogeneous, e.g. the concentration of the contaminant follows a log-normal distribution. However, the ICMSF notes that batches do not always show homogeneous concentrations of microbial contaminants because microorganisms can be very heterogeneously distributed. The batch size should thus be adjusted according to the processing method. Nonetheless, statisticians modelling contamination must assume homogeneity to describe properly the distribution of contaminants in food production. Furthermore, the food business operator defines a 'batch' with respect to ensuring traceability and internal organisation.

Determining the structure of contamination in the production of diced bacon

To determine the structure of contamination, we sampled pork breast after the massaging process in a factory that produces fresh diced bacon and in which we analysed the presence and concentration of *Listeria monocytogenes*. A batch was defined as all the pork breasts contained in one tumbler, the first step in the production process. In total, eight or nine pork breasts were taken from 12 different batches. For each pork breast 100 cm² of meat was sampled and analysed to screen for detection and enumeration of *L. monocytogenes*. With the protocols used here, the limit of detection was 0.01 colony-forming units (CFU)/cm², while the limit of quantification was 0.2 CFU/cm². The raw data (presence or absence of detection, number of colonies counted) are shown in **Table 1**.

Research

Table 1: Raw data for the detection (0=absence, 1=presence) and the counts (number of CFUs counted on Petri dishes) of *L. monocytogenes* conducted on 100 cm² pork breast samples after tumbling, i.e. the first step in the diced-bacon production process, during which bellies are tumbled with brine for several hours in a tumbler. The same sample was used for both detecting and counting *L. monocytogenes*.

Batch number	Detection results	Counting results (CFUs)
1	0-1-1-1-0-1-1-1-1	0-0-0-0-0-0-0-0-0
2, 8, 9 & 10	0-0-0-0-0-0-0-0-0	0-0-0-0-0-0-0-0-0
3	0-0-0-0-1-0-0-0-1	0-0-0-0-1-0-0-0-0
4	0-1-0-0-1-0-0-0-0	0-0-0-0-0-0-0-0-0
5	0-0-0-0-1-0-0-0-0	0-0-0-0-0-0-0-0-0
6	0-0-0-0-0-0-0-0	0-0-0-0-0-0-0-0
7	0-0-0-1-1-0-0-0-1	0-0-0-0-0-0-0-0-0
11	1-1-1-1-1-1-1-1	11-9-6-5-12-29-16-3
12	0-0-1-0-0-0-0-0-0	0-0-0-0-0-0-0-0-0

The results were used in four contamination models:
- contamination structured by units and batches (model REF);
- contamination structured by batch (model B);
- contamination structured by food unit (model U);
- contamination with no structure (model NS).

The food unit was the individual pork breast because we wished to determine whether there was within- and between-unit variability along with within and between-batch variability. We used a Bayesian approach, which allows the incorporation of information other than raw data into the model.

All models include a combination of binomial, Poisson and normal distributions. Models NS, B and U are nested in model REF.

In model REF, x_{ijk} is the detection result (1 if positive and 0 otherwise) of batch *i*, pork breast *j* and test portion *k*; y_{ijkl} is the enumeration of batch *i*, pork breast *j*, test portion *k* and fraction *l*. A test portion is the sample of meat on which the experiments were carried out (here 100 cm²). A fraction is the volume of the solution composed of the test portion diluted in an appropriate culture broth that is poured onto a Petri dish to count *L. monocytogenes* colonies. Variable x_{ijk} follows a binomial distribution and variable y_{ijk} follow a Poisson distribution:

$$x_{ijk} \sim Bin(1, 1 - exp(-10^{\theta_{ij}} S_k))$$
$$y_{ijkl} \sim P(10^{\theta_{ij}} S_k d_l)$$

where θ_{ij} is the logarithm to base 10 of the concentration of *L. monocytogenes* in pork breast *j* belonging to batch *i*; S_k is the surface of test portion *k*, and *d* is the dilution of the fraction *l*. The log concentration θ_{ij} follows a normal distribution:

$$\theta_{ij} \sim N(z_i, \lambda^2)$$

where z_i is the log concentration of *L. monocytogenes* in batch *i* and λ is the standard deviation of the log concentration in the food units. The log concentration z_i also follows a normal distribution:

$$z_i \sim N(\mu, \sigma^2)$$

where μ is the mean log concentration and σ is the standard deviation of the log concentration in batches. For the priors, parameter μ follows a normal distribution and σ^2 and λ^2 both follow an inverse gamma distribution.

There is no unit effect in model B, so $\lambda=0$. Conversely, there is no batch effect in model U, so $\sigma=0$. Model NS has neither of these effects, so $\lambda=\sigma=0$. Models B, U and NS are described in **Table 2**.

Table 2: Description of models B, U and NS. Subscripts *i, j, k* and *l* refer to a batch, a food unit (i.e. pork breast), a test portion and a fraction, respectively.

Model B	Model U	Model NS
$x_{ik} \sim Bin(1, 1 - exp(-10_i^z S_k))$ $y_{ikl} \sim P(10_i^z S_k d_l)$ $z_i \sim N(\mu, \sigma^2)$	$x_{jk} \sim Bin(1, 1 - exp(-10_j^\theta S_k))$ $y_{jkl} \sim P(10_j^\theta S_k d_l)$ $\theta_j \sim N(\mu, \lambda^2)$	$x_k \sim Bin(1, 1 - exp(-10^\mu S_k))$ $y_{kl} \sim P(10^\mu S_k d_l)$

To determine the parameters of the prior distributions, we used the self-inspection results that various companies carry out in the meat-processing industry. The posterior distributions of the parameters in the models were estimated using OpenBugs software (Thomas *et al.* 2006). According to the experimental protocol we carried out, S_k=100 cm² and d_l=0.05. Quantiles of the posterior distributions of the four models are shown in **Table 3**.

Table 3: Descriptive statistics of the posterior distributions of models REF, B, U and NS.

Model	Parameter	Descriptive statistics of the posterior distributions				
		Mean	S.D.	2.5th perc.	50th perc.	97.5th perc.
REF	μ	-3.09	0.53	-4.25	-3.05	-2.15
	σ	1.55	0.49	0.89	1.45	2.77
	λ	0.38	0.08	0.25	0.36	0.57
B	μ	-3.12	0.51	-4.21	-3.09	-2.18
	σ	1.72	0.47	1.06	1.63	2.86
U	μ	-3.51	0.15	-3.81	-3.51	-3.21
	λ	1.99	0.24	1.59	1.97	2.51
NS	μ	-0.94	0.005	-0.95	-0.94	-0.93

S.D., standard deviation; perc., percentile

We investigated the ability of the models to replicate real data with a visual criterion based on data simulations: detection data were simulated using the posterior distributions of the parameters (same number of datasets per batch and same number of batches as for the observed data), then the proportions of batches with (1) only presences, (2) only absences, or (3) a mixture of presences and absences, were counted. This process was repeated *n* times to calculate the median and the credibility intervals at 50% and 95%. A credibility interval at *x*% indicates that there is an *x*% probability that a value is within the interval. The same process was then repeated for counting. The results are shown in **Figure 2**. The model that best replicated the data was model B. The model REF performed only slightly worse (not shown). Model B is the best of the four studied models. 3.

Research

Example illustrating the determination of the optimal sample size that minimises the costs for the company

Knowing the distribution of contamination helps to define a sampling strategy. However, sampling strategies must also consider the specific processing practices used in a given factory and the reasons for sampling. Several types of sampling plans are used in the agri-food industry. A widely used model is the two-class sampling plan: *n* products are sampled and screened (generally 25 g of finished product); if the number of positive results *y* exceeds a certain number *c*, then the batch is rejected (destroyed or sold for a different use); if not, the batch is delivered. A two-class plan assumes that the product is still in the factory when the results are made available, which is not always the case. To adapt this type of plan to bacon processing, we modified the definition of the sampling plan slightly. After discussion with an industry expert, we developed the following sampling plan:

- sampling is not based on a production batch but on a certain production time period (e.g. 1 week or 1 month);
- according to the number of positive results (x), three possible decisions are made by the processing plant: (1) do nothing; (2) take minor corrective actions because the prevalence of *L. monocytogenes* during the production period is intermediate; (3) take major corrective actions because contaminant prevalence is high.

Our goal was to determine the optimal sample size *n* as well as the thresholds c_1 and c_2, the values of *x* beyond which minor or major corrective actions, respectively, are taken. To achieve this goal, we used Bayesian decision theory. This theory was used to determine the best solution for an operator *in situ*ation of uncertainty. Application of this theory involves several steps:

- determine the set $\mathcal{D}$ of all the possible decisions (here, the three decisions described above);
- determine all the values $\mathcal{S}$ of the states of nature (here, contamination of pork breasts by *L. monocytogenes*) and the prior distributions;
- determine the set of all the observations $\mathcal{O}$ (here, bacterial detection and counts) and their distributions;
- define a so-called loss function *L* defined for $\mathcal{D} \times \mathcal{S} \times \mathcal{O}$ in $\mathbb{R}^+$ (see below);
- determine the best decision rule (function which associates a decision *d* with a set of observations), obtained by minimising the expected loss over the states of nature and the observables.

For more information on this theory, see Berger (1985), Parent (2007) or Robert (2006).

According to contaminant prevalence in the batches of finished product sampled during the chosen period, the customer (distributor) can apply a penalty for non-compliance with specifications and order additional tests over a given period of time. The cost of the penalties depends on the level of prevalence (i.e. the higher the prevalence, the higher the cost of the penalty), but can be adjusted according to any corrective actions taken by the meat-processing company (i.e. if the company applies a corrective action, the penalty decreases). To keep the model simple, prevalence was divided into three classes: low, intermediate and high. We asked our expert to estimate the cost of these penalties and the corrective actions. These are summarised in **Table 4**.

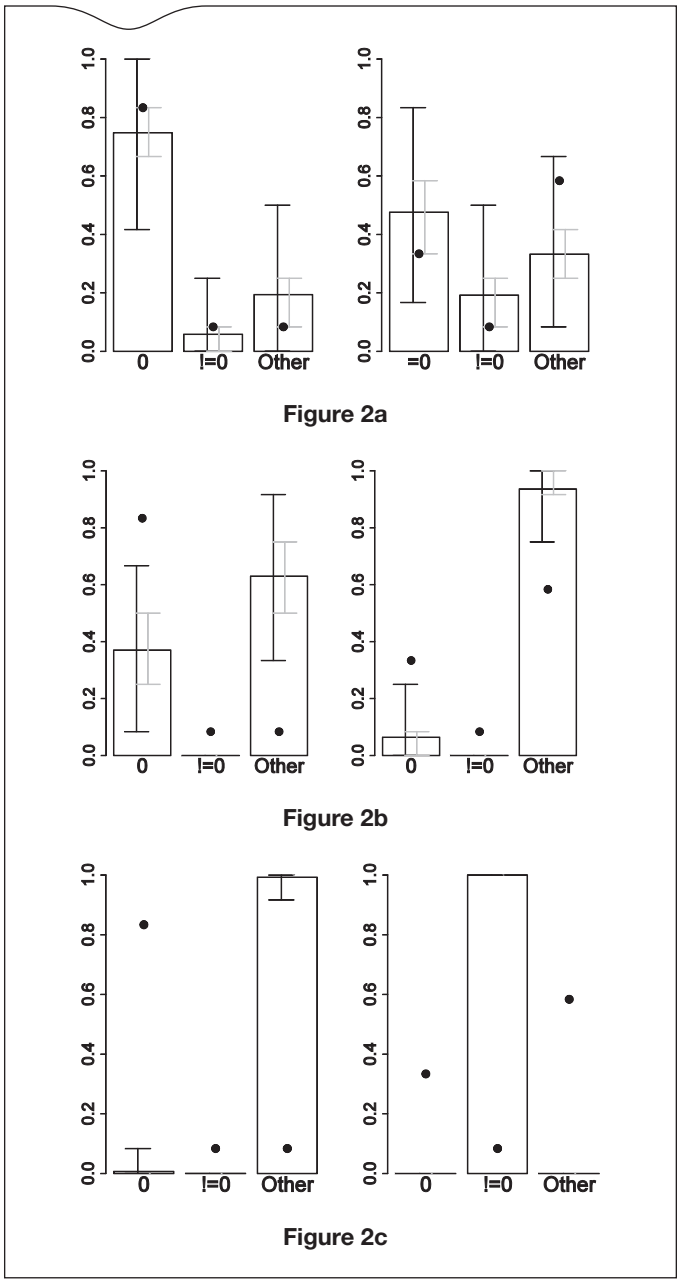

Figure 2a

Figure 2b

Figure 2c

Figure 2: Observed and simulated data for three models: (a) model B, (b) model U, and (c) model NS. The left-hand panel shows data for the detection method, and the right-hand panel data for the counting method. The histograms represent the average data for each group (0: proportion of batches with only null data, !=0: proportion of batches with only non-null data, and Other: all other batches). The grey error bar represents the credibility interval at 50% and the black error bar the credibility interval at 95%. The black dots indicate observed data.

Research

Table 4: Costs incurred (in euros) by the meat-processing company according to contaminant prevalence in the finished product and the decision made.

		Decision made		
		No action taken	Minor corrective action taken	Major corrective action taken
Actual results (Prevalence of contaminant)	Low prevalence	€0	€4 250	€14 000
	Intermediate prevalence	€6 200	€6 110	€14 930
	High prevalence	€92 050	€31 900	€27 800

The cost of sampling, estimated at €20 by the expert, must be added to each of these costs. To complete the calculation, we determined the thresholds of prevalence: below 0.2, prevalence is considered to be low and above 0.6, it is considered to be high. Finally, the beta distribution of parameters 2 and 3 was used to describe prevalence (see **Figure 3**). A beta distribution on parameters a and β has a probability distribution function equal to $\frac{\Gamma(\alpha+\beta)}{\Gamma\alpha+\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta}$, where $\Gamma t = \int_0^{\infty} z^{t-1}e^{-z}dz$. With this information, we can calculate the loss function which, for a given set of observations and value of contamination, there is an associated cost. The decision depends on the value of the observations; therefore knowing the observations automatically determines the decision to take.

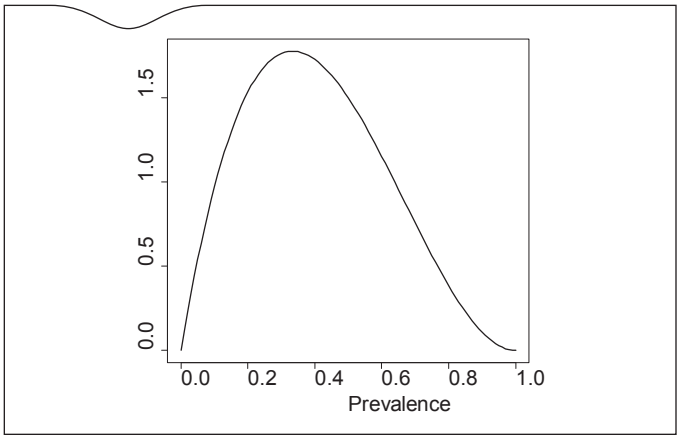

Figure 3: Distribution of prevalence between the different production periods.

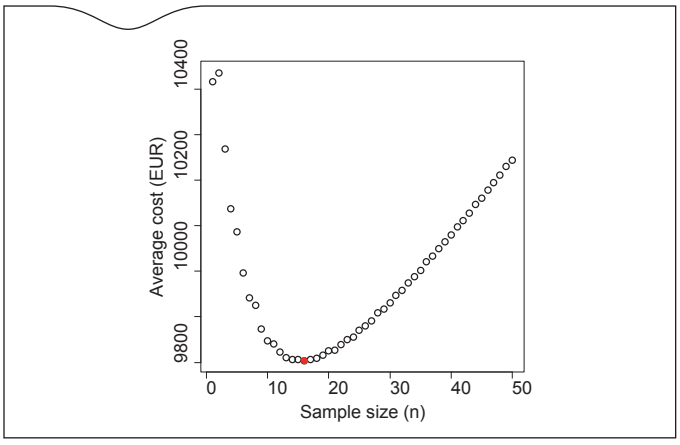

Figure 4: Average cost (in euros) incurred by the meat-processing company according to sample size n. The minimal cost is reached at n=16, c_1=4 and c_2=11 (red dot).

Based on this information and according to prevalence and analysis results, we calculated the average cost per production period for the meat-processing company with respect to sample size. By varying sample size, we can determine the value that minimises the average cost. The average costs based on the chosen numerical values are shown in Figure 4 and depend only on sample size. The minimum value was found for n=16, c_1=4 and c_2=11.

The distribution and thresholds of prevalence were set to complete the exercise. Obviously, when they vary this leads to a change in the optimal sampling plan: with a prevalence following a beta distribution of parameters 2 and 20 and thresholds of prevalence of 0.05 and 0.1, the average minimum cost for the company is reached at n=48, c_1=1 and c_2=6, which differs greatly from the previous result. Similarly, if costs vary then so does the sampling plan.

The application of Bayesian decision theory provides additional support for the decision-maker particularly *in situ*ations with many unknowns. This approach requires defining the population to which the method will be applied (e.g. here, we defined 'batch'), modelling prevalence, defining the set of decisions and their possible consequences, determining the costs, and, finally, carrying out probabilistic calculations. The final values depend strongly on the model used and current costs, which means that they must be defined carefully for each application. For more information on this work, see Commeau (2012).

References

Berger J. 1985. Statistical decision theory and Bayesian analysis. *Springer Verlag, New*

York, second edition: 617 pp.

Commeau N, (2012). Modélisation de la contamination par *Listeria monocytogenes* pour l'amélioration de la surveillance dans les industries agro-alimentaires. PhD thesis, AgroParisTech. http://tel.archives-ouvertes.fr/index.php?alsid=3euus0iet62q4s4jjf39p0e935&view_this_doc=pastel-00770790&version=1

Gonzales-Barron U, Butler F. 2011. Characterisation of within-batch and between-batch variability in microbial counts in foods using Poisson-gamma and Poisson-lognormal regression models. *Food Control*, 22:1268–1278.

International Commission on Microbiological Specifications for Foods (ICMSF). 2002. Microorganisms in Foods 7: microbiological testing in food safety management. *New-York, Kluwer academic/Plenum Publisher:* 375 pp.

International Life Science Institute (ILSI). 2010. Impact of microbial distribution on food safety. 1-68.

Parent E, Bernier J. 2007. Le raisonnement bayésien. *Springer Verlag, France:* 327 pp.

Regulation (EC) no 2073/2005 of the European Commission on microbiological criteria for foodstuffs. *Official Journal of the European Union* L338:1-26.

Robert C. 2006. Le choix bayésien. *Springer-Verlag, France*: 639 pp.

Thomas, A., O'Hara, B., Ligges, U. and Sturtz, S. (2006). Making BUGS open. *R News*, 6:12–17.