Modelling contamination to build a sampling plan: application to French diced bacon industry and *Listeria monocytogenes*

N. Commeau[1,2,3], E. Parent[2], E. Billoir[4], V. Zuliani[5] and M. Cornu[1]

[1] Microbiologie quantitative et estimation des risques (MQER), Agence française de sécurité sanitaire des aliments (Afssa), 23 avenue du Général de Gaulle, F-94706 Maisons Alfort Cedex, France (n.commeau@afssa.fr)
[2] UMR 518, Mathématiques et Informatique Appliquées, AgroParisTech/INRA, 16 rue Claude Bernard F-75005 Paris, France
[3] AgroParisTech ENGREF, 19 avenue du Maine, F-75732 Paris, France
[4] UR341 Mathématiques et Informatique Appliquées, INRA, Domaine de Vilvert, F-78350 Jouy-en-Josas, France
[5] IFIP, Institut de la filière porcine, Pôle Viandes Fraîches et Produits Transformés, 7 avenue du Général de Gaulle, F-94704 Maisons Alfort Cedex, France

Abstract

We designed an experimental sampling plan to get knowledge on *L. monocytogenes* and on lactic acid bacteria (LAB) concentrations on pork breasts after tumbling in a French diced bacon plant. Then, we built a Bayesian model to estimate variability (including between batches-variability) and uncertainty. The results of the model are used to predict *L. monocytogenes* concentration on breasts after tumbling and could be useful to optimise a sampling plan for industrial own-checks.

Keywords

Bayesian modelling, sampling plan, *L. monocytogenes*, pork breasts

Introduction

Nowadays, food business operators who sell food products in Europe have to make sure the product they make is safe (regulation EC 2073/2005). In this context, a sampling plan a commonly tool used in order to assess contamination of a given food borne pathogen in the plant and to take decisions (throw away a batch, clean the plant better...). Performances of different sampling plans cannot be compared in situ in a plant for cost and feasibility reasons whereas modelling makes it possible. Among all the elements needed to model a sampling plan, it is necessary to get knowledge about pathogen concentration during the process.

We worked with a French plant producing diced bacon. After studying the process from pork breasts to package units of diced bacon (i.e. arrival, tumbling-and-brining, steaming, dicing and packaging (see Billoir et *al*, same conference for more details)), we designed an experimental sampling plan to get detection and enumeration data on (1) *L. monocytogenes* concentration and enumeration data on (2) LAB on breasts after tumbling and on diced bacon after packaging. To use concentration information, a Bayesian model, using both detection and enumeration data, was built to assess variability and uncertainty of *L. monocytogenes* concentration during the process. This paper is focused on the results concerning breasts after tumbling and their modelling.

Experimental sampling plan construction

To get data, the simplest way was to make a random sampling plan on the breasts. A random sampling is a method of selecting n units out of N such that every one of all the possible samples has an equal chance of being drawn. In that case, we thought breast population was quite heterogeneous and that dividing it into homogeneous subpopulations could produce a gain in the estimates. Breasts can be tumbled 3 times longer during weekend than during week. Thus, we chose a random sampling (sampling where population of N units is divided into $N_1,\ldots, N_S$ subunits ,here S=2, called strata) stratified on tumbling time. The drawing in each stratum is being made independently from the others. A simple random sampling is taken in each stratum.

To determine how many samples need to be taken in a stratified sampling, we used the method proposed by Cochran (1977) which consists of minimising the variance $V(\mu_{st})$ for a given cost C of taking the sample or of minimising the cost C for a given variance, where $V(\mu_{st})$ is the variance of the mean $\mu_{st} = \sum_{i=1}^{2} W_i \mu_i$ with $W_i = \frac{N_i}{N}$, where N is the size of the population and N_i the size of the stratum i.
The chosen cost function is $C = c_0 + \sum n_i c_i$, were c_0 represents a overhead cost and c_i the cost of taking one sample from the stratum i. Here, we made the assumption that the cost per unit is the same in all strata (c_i=c) and that the variance of μ_i are equal. In that case, the variance V is minimum for a specified cost C, and the cost is minimum for a specified variance V when $n_i = nW_i$. In our case, there are 2 strata: breasts tumbled during week and breasts tumbled during week-end tumbling.
Considering (*i*) that the sampling plan was constrained to ca. 100 breasts analysed, (*ii*) that in this plant 12% of breasts are tumbled during week-end and (*iii*) assuming that the breasts tumbled in one single tumbler compose a batch, the optimised plan was to sample 12 batches (10 tumbled during the week and 2 during week-end), and 9 breasts from each of them.

Sampling plan results

Material and methods

From May to June 2009, 104 breasts (8 to 9 from 12 batches) were sampled just after tumbling, following the sampling plan justified above.
From each sample, 100 cm² of the lean surface were excised for *L. monocytogenes* analyses (detection and enumeration) and 100 cm² of the lean surface for LAB enumeration.
Detection and enumeration of *L. monocytogenes* were carried out on the same test portion. Each 100cm^2-test portion was stomached with 100 mL of half Fraser broth, stored 1h at 20°C, and 5mL of this first dilution were pour-plated into 5 plates of ALOA (agar for *Listeria* according to Ottaviani and Agosti). Antibiotics were added to the remaining first dilution for the enrichment procedure of the standard ISO 11290-1/A1. For the analyses of LAB, enumeration was performed onto MRS (Mann-Rogosa-Sharpe), according to the standard ISO 15 214.

Results

LAB concentration was always greater than 10 cfu/cm² and under 3200 cfu/cm². Breast from the same batch have very similar LAB concentration, as it is shown on Figure 1. Usually, differences between the smaller and the higher concentration within batch were less than one log, clearly suggesting that cells concentration is homogenised during tumbling. Concentration for batches 2 and 5 are the highest ones and these two batches are the only ones tumbled during week-end. An ANOVA concluded to a significant week-end effect (p-value<10^{-16}). The extended length of time breasts spend in the tumbler during week-end (3 times longer than the week) may explain this higher concentration: LAB has more time to grow.
L. monocytogenes was detected in 25 of the 104 analysed test portions (24%), with prevalences per batch ranging from 0% to 100%. Here, absence means absence in 100 cm², so the detection threshold is equal to 0.01 cfu/cm², which is much lower than in standard own-checks. In the 11th batch, all tested breasts were positive for detection and enumeration (0.6 to 5.8 cfu./cm²). Among the 11 other batches, only one other sample was positive for enumeration (at the threshold: 0.2 cfu/cm^2). This between-batches difference and within-batch homogeneity is consistent with the results obtained with LAB: cells distribution among breasts taken from a single tumbler appears then very homogenous. No week end effect was observed on *L. monocytogenes* data.

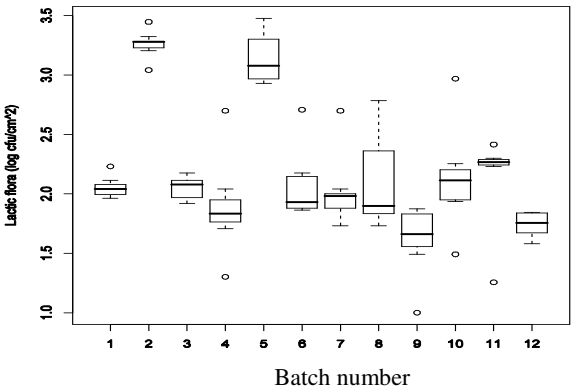

Figure 1: Boxplots of the LAB concentration in cfu/cm² per batch for pork breasts after tumbling

Modelling the *L. monocytogenes* contamination in breasts

Material and methods (model, priors, computation)

The Bayesian model used here to describe *L. monocytogenes* results is composed of two parts: the latent part and the data one. For the latent part, we considered every batch i has a mean concentration z_i which follows a lognormal distribution with mean μ and standard deviation σ.

$$z_i \sim LogN(\mu; \sigma)$$

Here, we used enumeration data expressed in colony counts (raw data). For the i^{th} batch, we considered that v_{ij}, the number of cells in the test portion j, follows a Poisson distribution with parameter $z_i * S$, where S is the surface of the test portion (S=100 cm^2). The colony count y_{ij} (ie the number of cells in the volume plated onto the 5 ALOA plates for the j^{th} breast of the i^{th} batch) follows a Poisson distribution with parameter $v_{ij}*d$, where d is the ratio between the volume spread onto the 5 plates and the volume of the first dilution (d=5/100).

$$v_{ij} \sim P(z_i * S) \qquad y_{ij} \sim P(v_{ij} * d)$$

Concerning detection results (expressed in a binary way: 'absence' or 'presence'), we considered that 'absence' meant there was indeed no *L. monocytogenes* in the test portion (but the latter could have been elsewhere on the breast). Considering cells dispersion on the breast follows a Poisson distribution, then the detection result x_{ij} for the batch i and the test portion j follows a Bernoulli distribution with parameter 1-exp(-z_i*M).

$$x_{ij} \sim Ber(1 - e^{-(z_i * M)})$$

The directed acyclic graph (DAG) of the model is shown on figure 2.

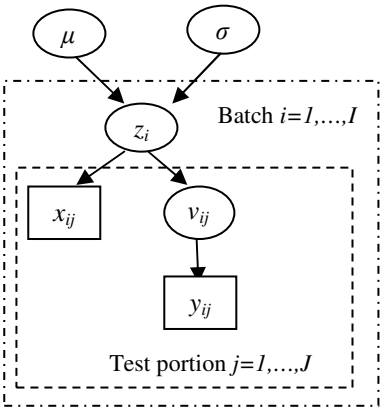

Figure 2: DAG of the model. The mean concentration of the batch i is noted z_i. For a test portion j, there are 2 kinds of data: colony count (y_{ij}) and detection result (x_{ij}). All the parameters are represented in ellipses and data are in rectangles. Arrows indicate a probabilistic dependence between parents and children. For example, z_i has two parents (μ and σ) and 2*J children (all the y_{ij} and the x_{ij}). Given all its parents, a node is independent from all the other nodes, except its children. Nodes without parents are called hyperparameters.

Informative priors were based on outputs of Billoir et *al*. (same conference):

$$\mu \sim N(-2.94; 2) \quad \text{and} \quad \sigma^2 \sim IG(5.10^{-3}; 7.10^{-2})$$

Bayesian estimation of the posterior distributions were performed using OpenBugs (http://mathstat.helsinki.fi/openbugs/, 2007). 200,000 iterations were performed for each model. 1 value out of 10 was taken to avoid autocorrelation.

Results

Table 1: Descriptive statistics of empirical distributions of hyperparameters of the model expressed in log cfu/cm² (S.D. stands for standard deviation and perc. for percentile)

Parameter	Mean	S.D.	2.5^{th} perc.	50^{th} perc.	97.5^{th} perc.
μ	-2.95	0.41	-3.80	-2.93	-2.19
σ	1.52	0.48	0.86	1.42	2.74

Descriptive empirical distributions of the hyperparameters are shown on table 1.

The posterior distribution of μ represents the remaining uncertainty on the mean log concentration of the breasts after tumbling and the posterior distribution of σ represents the remaining uncertainty on the between-batches variability. To assess the capacity of the model to replicate data, we chose 50 values of μ and σ from their posterior distributions and replicated 30,000 data y_{rep} for every observed datum y_{obs}. Then, we calculated the mean and the credibility interval at 95% of y_{rep}=0, $1 \leq y_{rep} < 10$ and $y_{rep} \geq 10$ colony counts. The result is shown on figure 3. Each time, observed data (black dots) are very close to the mean or, at least are within the credibility interval.

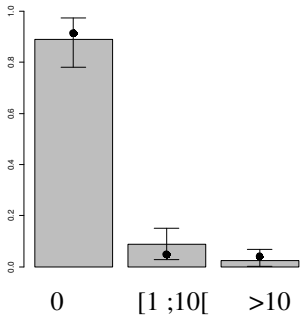

Figure 3: Mean (barplots) and 95% credibility interval (segments) of the replicated data, and real data (black points). Data are colony counts of *L. monocytogenes* measured in breasts after tumbling. Barplots for 0, [1;10[and >10 colony counts.

Conclusion

To optimise an industrial sampling plan for own-checks, the first step is to get knowledge about contamination of the considered hazard during the process. To do that, a way is to select an interesting step and to study the pathogen concentration. In the process of the diced bacon, tumbling is the chosen step. Both results of LAB and *L. monocytogenes* support the conclusion that cells are homogenized during tumbling.

In the considered plant, sampling of breasts is currently performed at the really start of the process. We suggest that sampling after the tumbling step could be more appropriate.

Our modelling of the concentration of *L. monocytogens* can then be used to optimise the industrial sampling plan but also to calibrate the model describing the process of Billoir et *al.*.

Acknowledgments

This work was supported by a grant from the Agence National de la Recherche (ANR) (France), as a part of the Quant'HACCP project.

References

Cochran W. (1977) *Sampling techniques*, 3rd edition, John Wiley & Sons, New York, USA, 428pp. (ISBN: 047116240X)

Billoir E., Denis J-B., Commeau N., Cornu M. and Zuliani V. Probabilistic modelling of L. monocytogenes behaviour in diced bacon along the manufacture process chain. *Submitted to the same conference.*