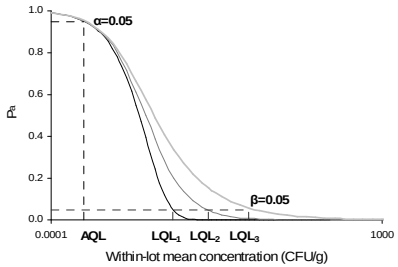
Derivation of a variables sampling plan based on a Poisson-gamma model representing within-batch and between-batch variability in low microbial counts



Dr. Ursula Gonzales-Barron
Prof. Francis Butler

UCD School of Biosystems Engineering, University College Dublin, Ireland

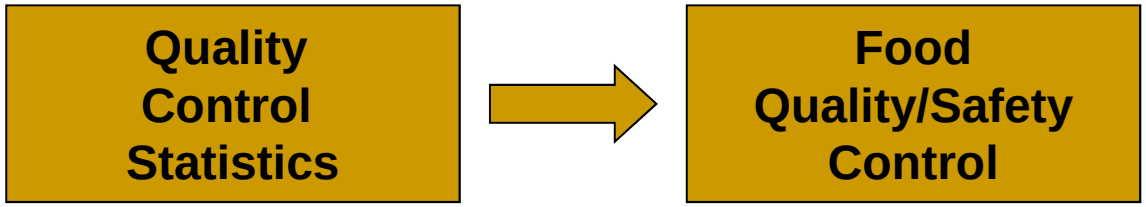
Background

Acceptance sampling plans

- Within-batch sampling
  - Compares the level of a microbial hazard detected in a food against a pre-specified limit or MC.
  - Regulation EC No.1441/2007
- Based on acceptance sampling plan theory
  - By attributes
    - Two-class: Ex. *L.m.* in RTE: n=5, c=0, m=100 CFU/g
    - Three-class: Ex. *E.coli* in minced meat: n=5, c=2, m=50, M=500
    - ICMSF spreadsheet: lognormal, constant variance batch to batch
  - By variables
    - Ex: *Enterobact.* on pig carcass: n=5, m=2 log, M=3 log CFU/g

Acceptance sampling plans

- Traditionally in acceptance sampling plans:
  - True concentration of microorganisms is **log-normally** distributed within a batch
  - Variance in microbial load is **constant batch to batch**



Recent findings (×4)

1) For data consisting of zero-counts (low microbial counts), Poisson-gamma (PG) distribution is far more suitable than PLN and lognormal

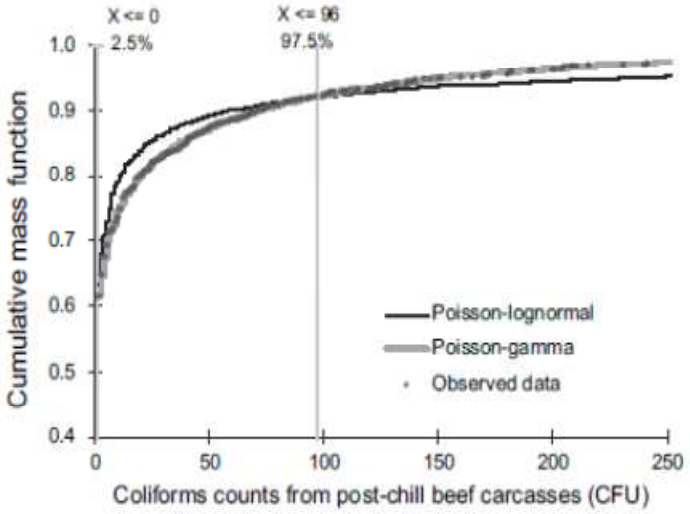


Fig. 2. Predictive distributions of the 'observed' coliforms plate counts *Y* from pre-chill (top) and post-chill (bottom) beef carcasses as modelled by the Poisson-lognormal and Poisson-gamma models. Point markers represent the ranked probabilities of the data.

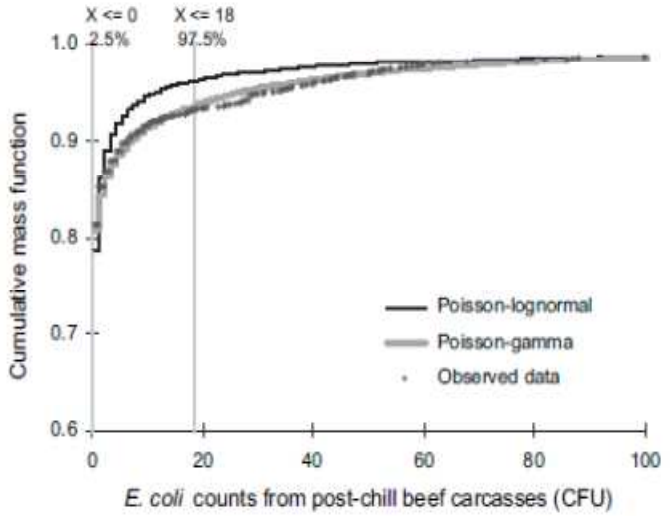


Fig. 3. Predictive distributions of the 'observed' *Escherichia coli* plate counts *Y* from pre-chill (top) and post-chill (bottom) beef carcasses as modelled by the Poisson-lognormal and Poisson-gamma models. Point markers represent the ranked probabilities of the data.

Recent findings (×4)

1) For data consisting of zero-counts (low microbial counts), Poisson-gamma (PG) distribution is far more suitable than PLN and lognormal

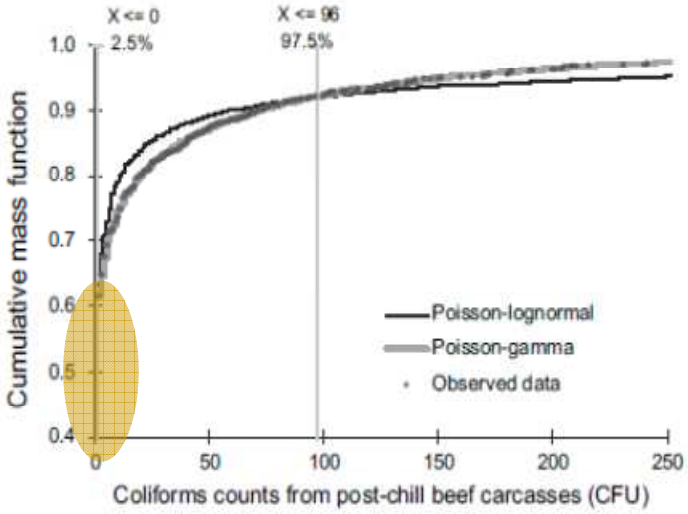


Fig. 2. Predictive distributions of the 'observed' coliforms plate counts *Y* from pre-chill (top) and post-chill (bottom) beef carcasses as modelled by the Poisson-lognormal and Poisson-gamma models. Point markers represent the ranked probabilities of the data.

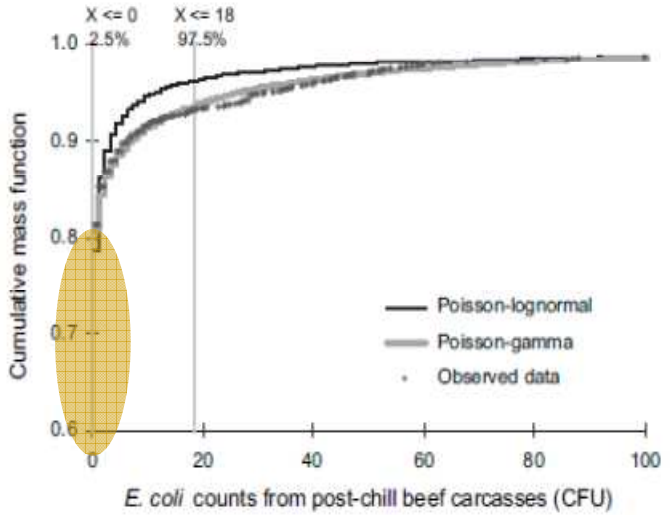
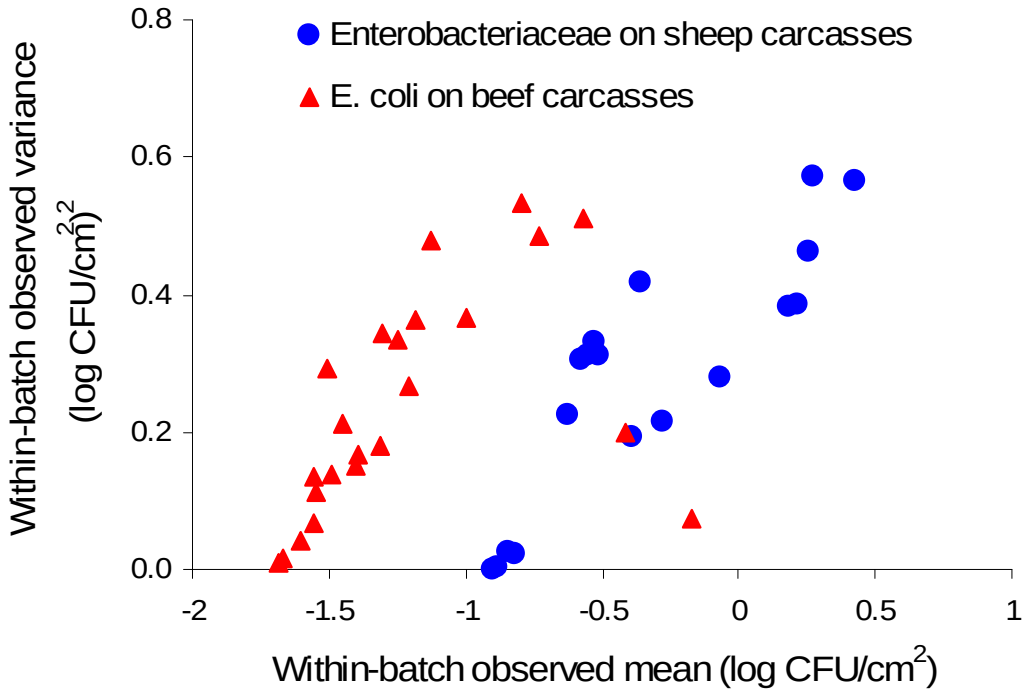


Fig. 3. Predictive distributions of the 'observed' *Escherichia coli* plate counts *Y* from pre-chill (top) and post-chill (bottom) beef carcasses as modelled by the Poisson-lognormal and Poisson-gamma models. Point markers represent the ranked probabilities of the data.

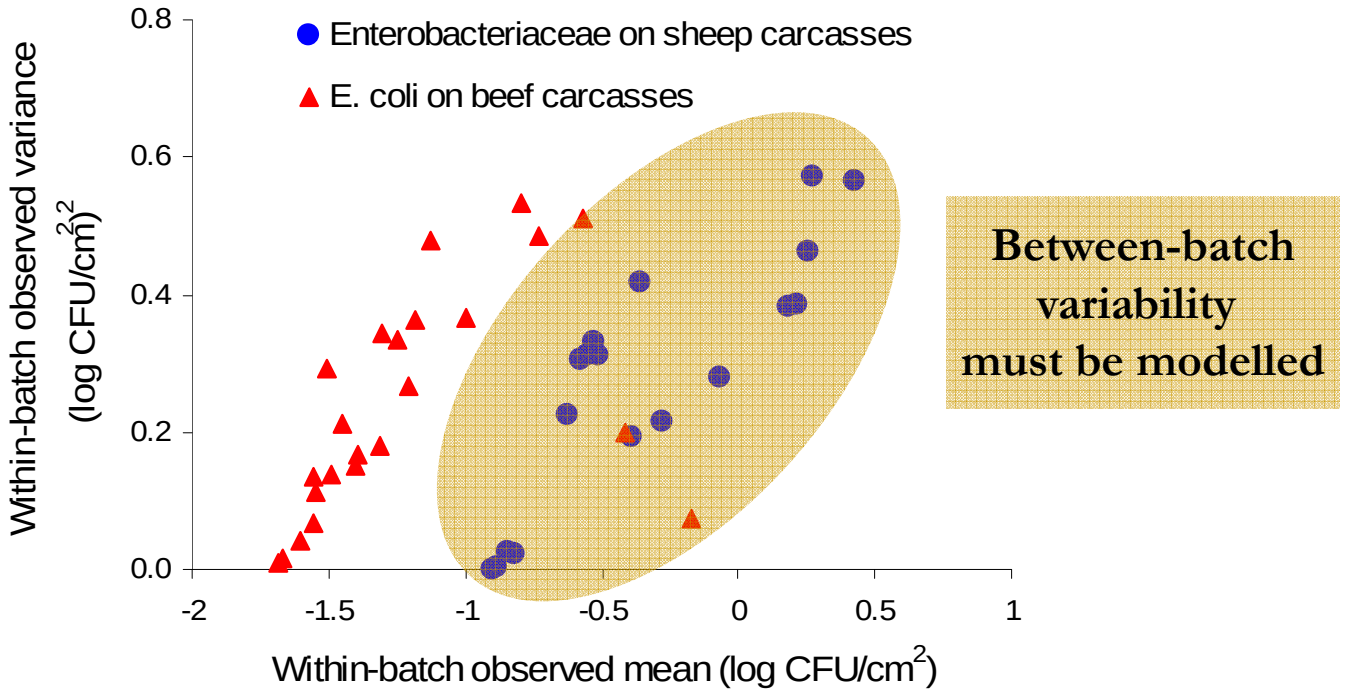
Recent findings (×4)

2) Within-batch variance is not constant and is associated to the within-batch mean



Recent findings (×4)

2) Within-batch variance is not constant and is associated to the within-batch mean



Recent findings (×4)

3) The association between WB mean and WB spread can be represented well by a Poisson-gamma regression model with correlated random effects

$$Y_{ij} \sim \text{Poisson}\left[\frac{A_i}{V} d_i t \times \lambda_{ij}\right]$$

$$\lambda_{ij} \sim \text{Gamma}(1/k_j \ , \ m_j)$$

$$\log(m_j) = \text{Int}_0 + u_j$$

$$\log(1/k_j) = \text{Int}_1 + v_j$$

$$[u_j, v_j] \sim \text{Multinormal}([0 \ , \ 0], \Sigma)$$

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$$

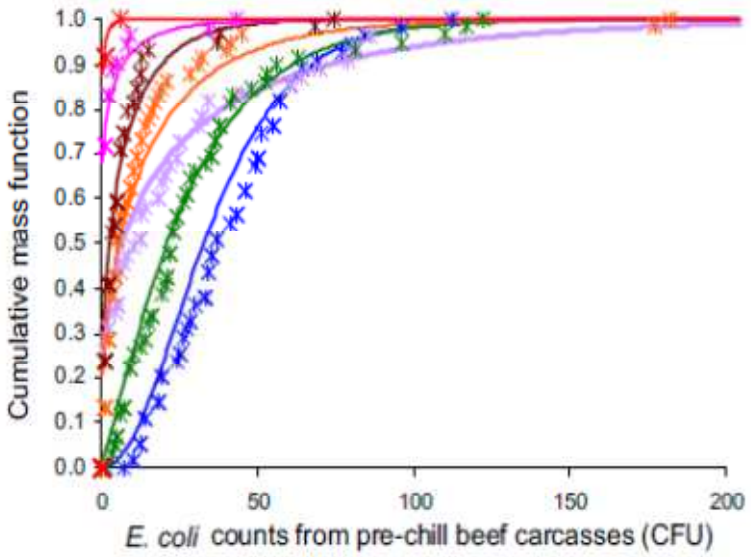
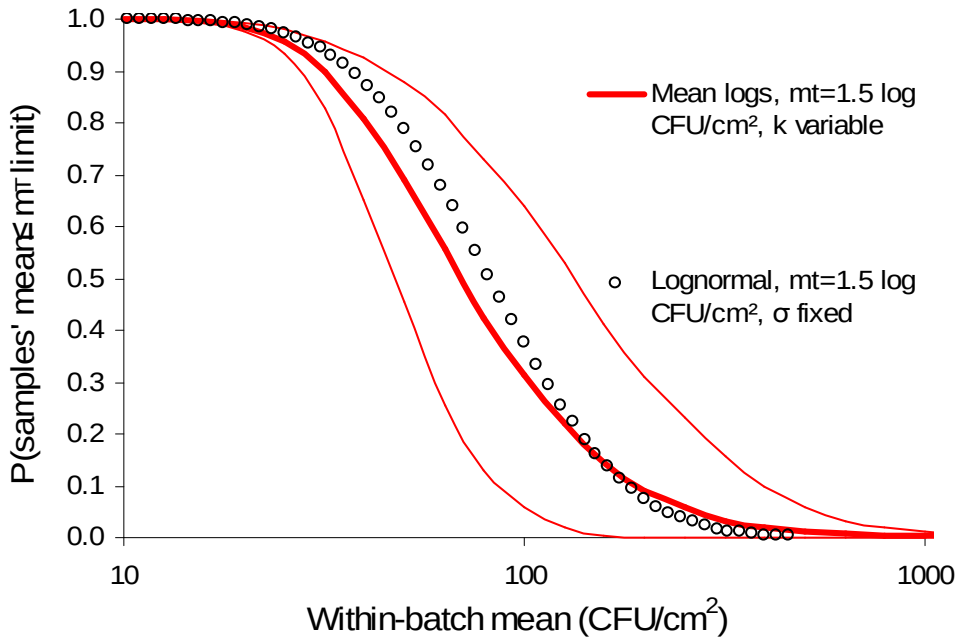


Fig. 5. Between-batch and within-batch variability in coliforms counts from post-chill beef carcasses (top) and in *Escherichia coli* counts from pre-chill beef carcasses (bottom) as modelled by the Poisson-gamma regression with correlated random effects for the mean and dispersion parameters. Asterisks of the same colour represent ranked observations from a batch, and for visual clarity, not all batches are shown.

Recent findings (×4)

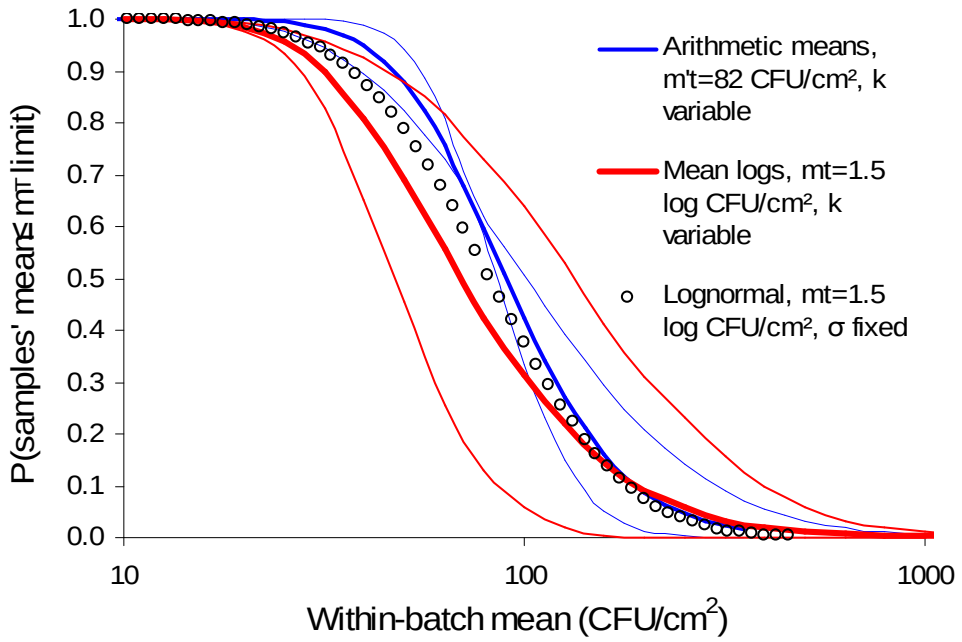
4) Sampling plans with a M_L expressed in arithmetic mean (CFU/g) are more effective than those expressed in mean log (log CFU/g)



MC: n=5, **m_L=1.5 log CFU/cm²**

Recent findings (×4)

4) Sampling plans with a M_L expressed in arithmetic mean (CFU/g) are more effective than those expressed in mean log (log CFU/g)



- The arithmetic mean scale generates less uncertainty in Pa, and hence causes a SP to be more effective.
- The arithmetic mean approach produce also steeper OC curve → higher discriminatory power of the SP

MC: n=5, **m_L= 82 CFU/cm²**

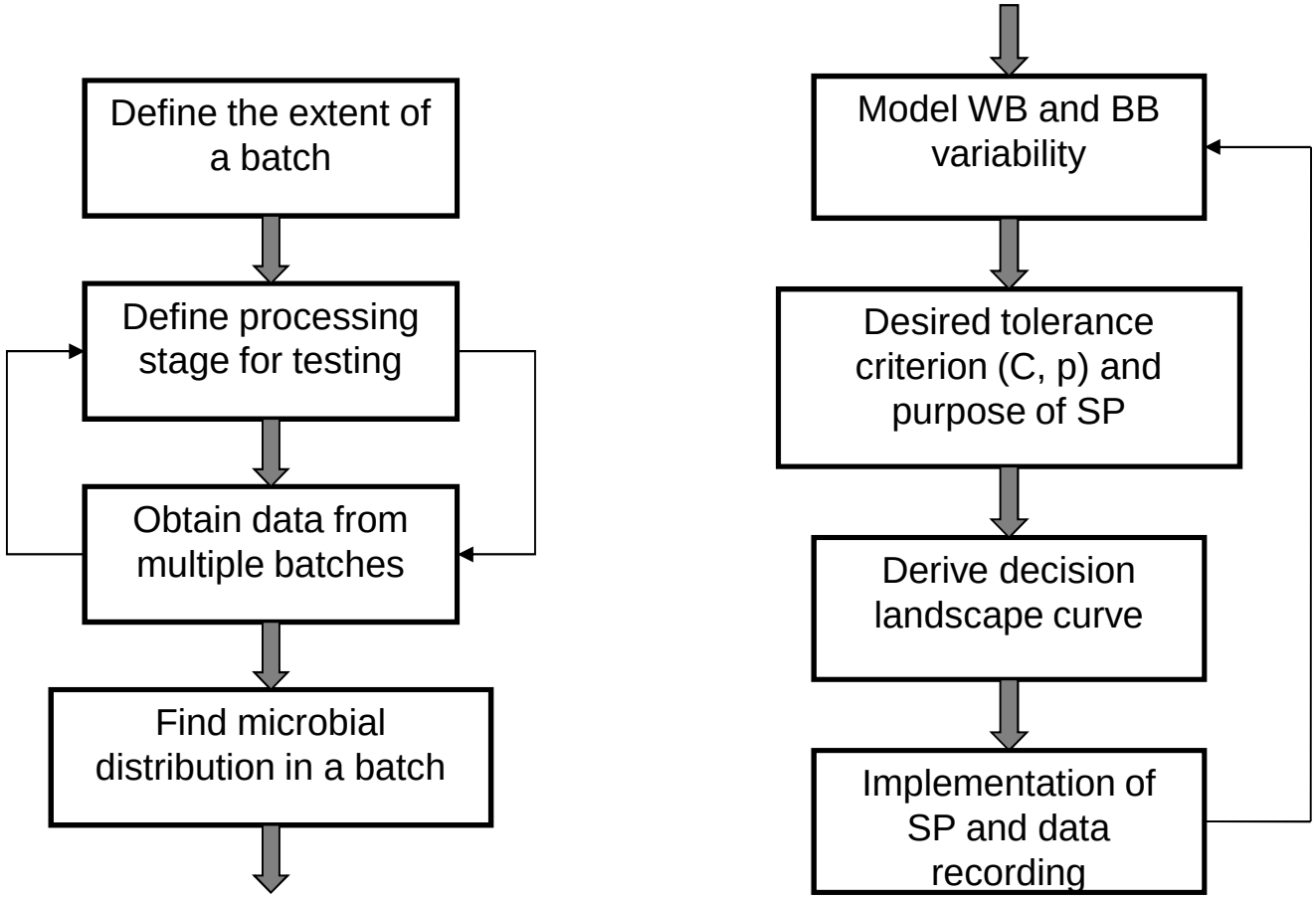
Objective

- This study proposes a novel methodology for the statistical derivation of a variable sampling plan for use in food production systems whose microbial counts are known to be low and can be represented by a Poisson-gamma model.

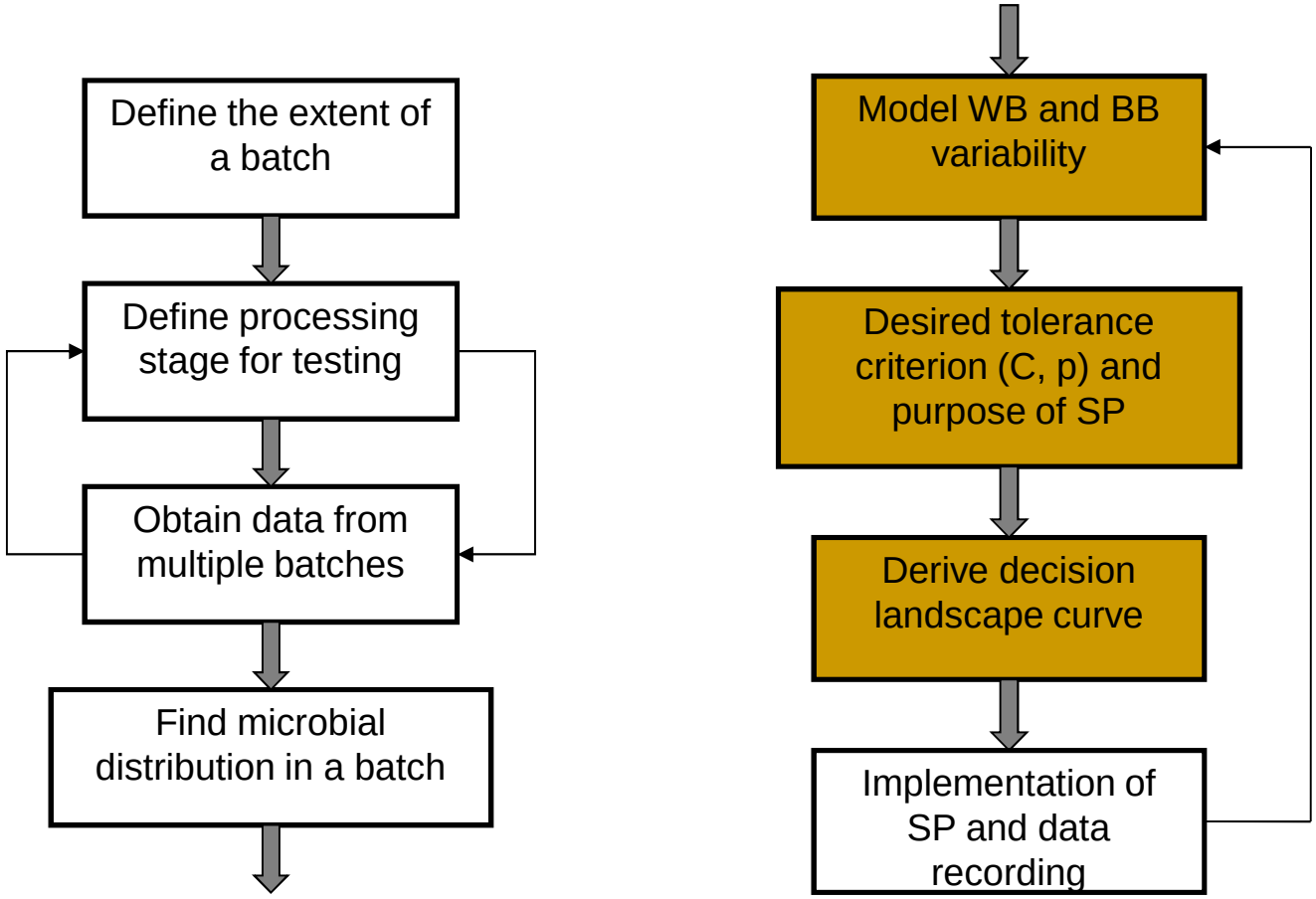
- Microbial data: *Enterobacteriaceae* counts on pre-chill sheep carcasses (n=400 carcasses, j=20 batches).

Methodology

Methodology



Methodology



1) Model WB and BW variability in microbial counts using Poisson-gamma model

- *Enterobacteriaceae* plate count data: i=400 carcasses, j=20 batches
- PG model was fitted

$$Y_{ij} \sim \text{Poisson}\left[\frac{A_i}{V} d_i t \times \lambda_{ij}\right]$$

$$\lambda_{ij} \sim \text{Gamma}(1/k_j \ , \ m_j)$$

$$\log(m_j) = \text{Int}_0 + u_j$$

$$\log(1/k_j) = \text{Int}_1 + v_j$$

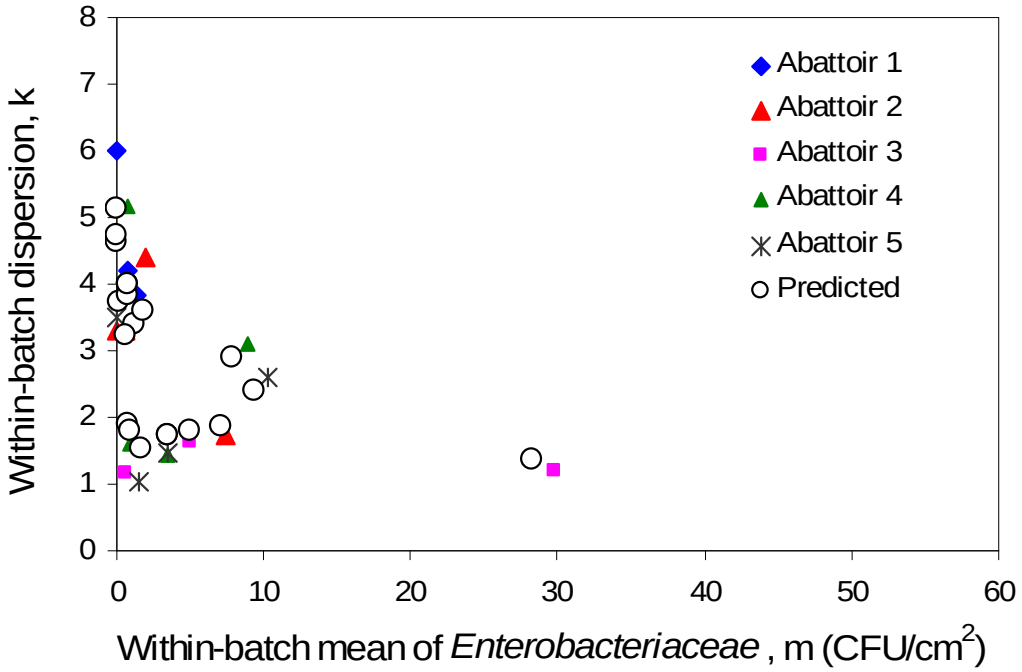
$$[u_j, v_j] \sim \text{Multinormal}([0 \ , \ 0], \Sigma)$$

$$\Sigma = \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}$$

1) Model WB and BW variability in microbial counts using Poisson-gamma model

Poisson-gamma models	**Mean**	**Standard error**	**Pr>|t|**	**BIC**
Correlated random effects for m, k				**2588**
Intercept [log(mean)], Int_0	-	-	-	
Random effects [log(mean)], σ_u^2	3.514	1.213	**0.009**	
Intercept [log(k)], Int_1	1.058	0.149	**<.001**	
Random effects [log(k)], σ_v^2	0.254	0.093	**0.048**	
Correlation coefficient, ϱ	-0.614	0.262	**0.031**	

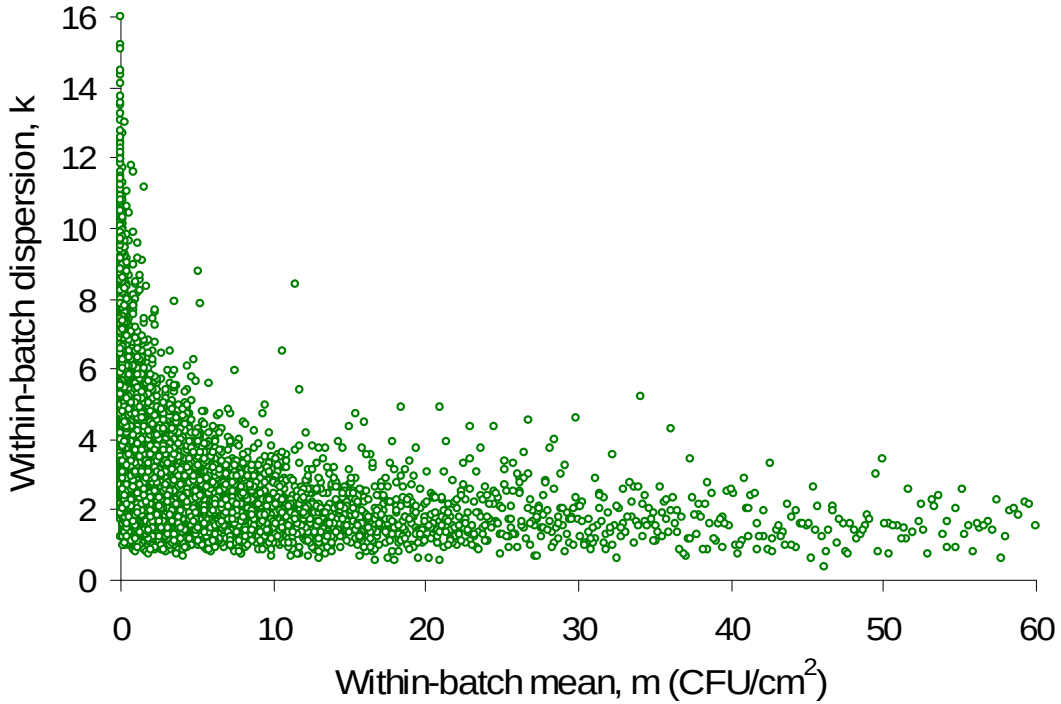
1) Model WB and BW variability in microbial counts using Poisson-gamma model



- The overall PG model represents well each of the batches

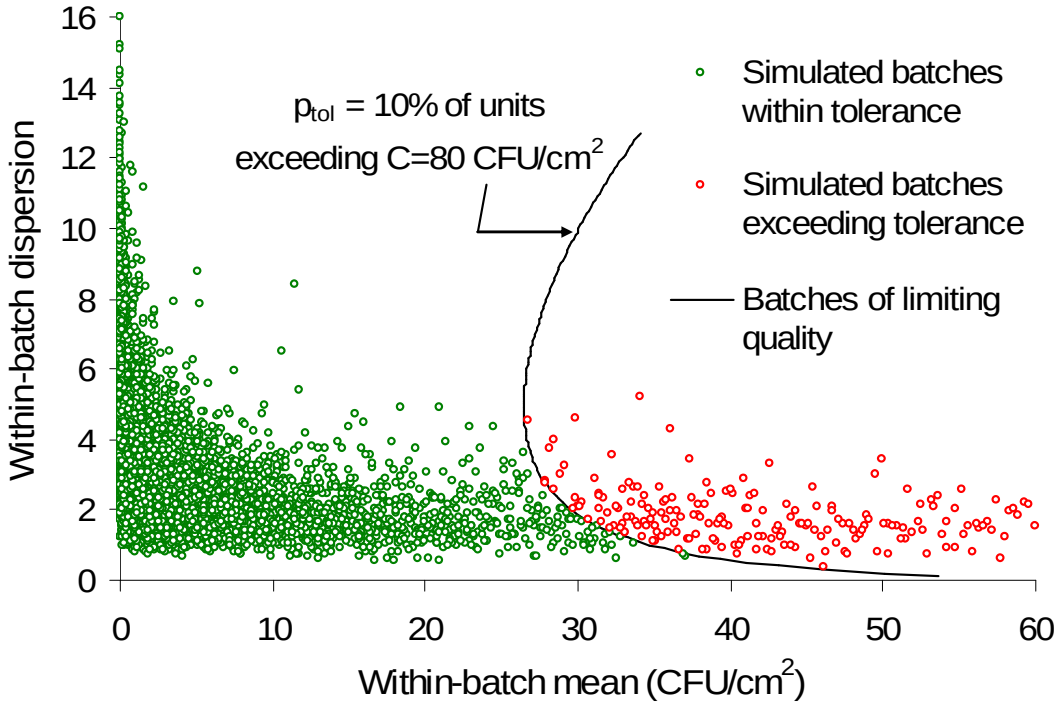
1) Model WB and BW variability in microbial counts using Poisson-gamma model

- It is possible to obtain a universe of contaminated batches through simulation
- Assumption: the PG model represents the level of *Enterobacteriaceae* under which Irish sheep abattoirs operate



2) Establish a tolerance criterion and purpose of the sampling plan

- Decide what makes a food unacceptable
- A batch is defined as "unacceptable" if more than a tolerance percentage (p_{tol}) of the food units has microbial concentrations exceeding a critical concentration (C)



2) Establish a tolerance criterion and purpose of the sampling plan

- Purpose of the SP should be known

Producer's side	Consumer's side
-For **control purposes** such as adherence to GMP limits or SPC -- Derived on the **α risk**: probability of taking action although the batch is of good quality - The producer requires some confidence that good batches will not be rejected -Critical concentration, **Cm**	-For **safety specifications** such as PO's. -- Derived on the **β risk**: probability of accepting a batch that in reality is defective - The consumer requires some confidence that bad batches will not be accepted -Critical concentration, **Cs>>Cm**

3) Derivation of decision landscape curves

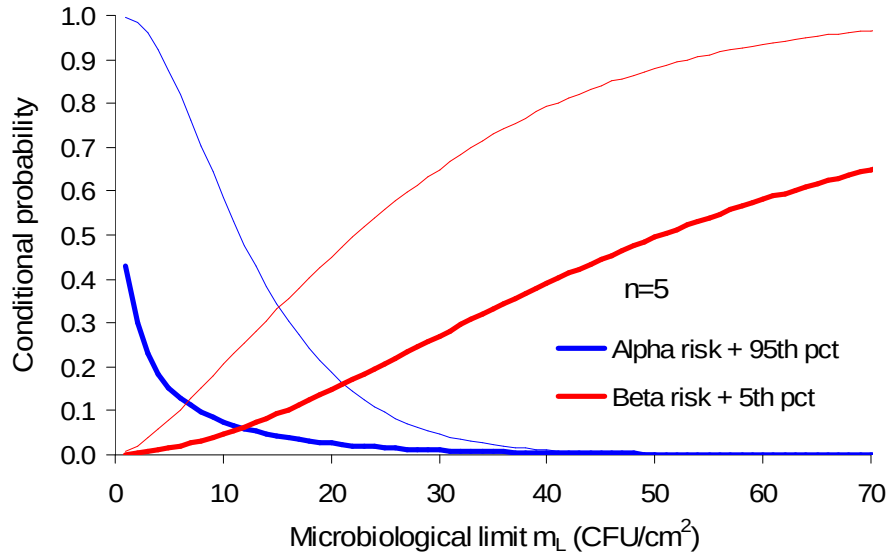
- In our case, SP will be derived on the producer's side
  - Because counts of *Enterobacteriaceae* are very low.
  - The food producer requires some confidence (1-α) that products of acceptable quality should not be rejected because of the imprecision of the sampling scheme.

- The problem consists of finding a decision criterion
  - m_L
  - n

  that satisfies the pre-defined minimum confidence (1-α) measured on the samples' mean distributions.

3) Derivation of decision landscape curves

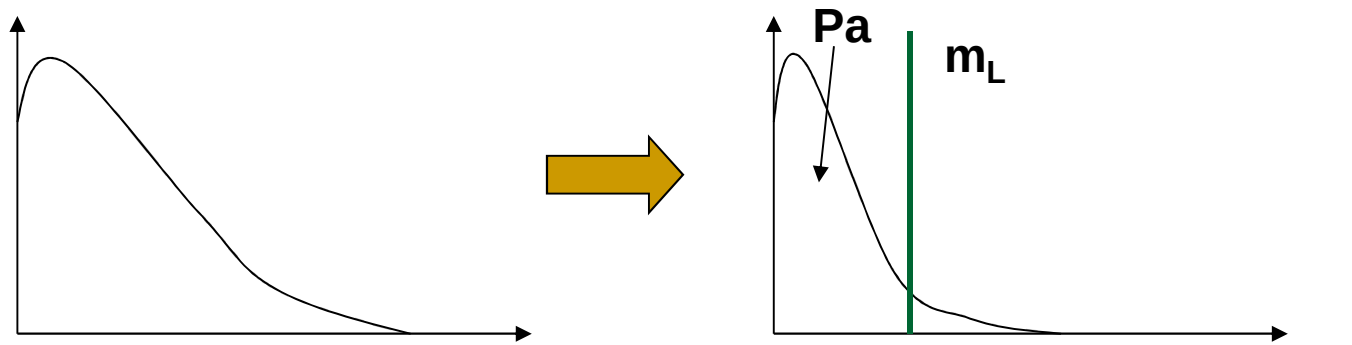
- A decision landscape curve plots the values of α and β risks for different microbiological limits $\mathbf{m_L}$ at a constant sample size **n**.



- It is constructed by Monte Carlo simulation so as to propagate the BB variability to the α and β risks.
- α and β risks are measured on a 'samples' mean distribution'

3) Derivation of decision landscape curves

- Every ‘true within-batch distribution’ from the universe should be converted to ‘sample mean’s distribution’ in order to calculate the probability of accepting the batch

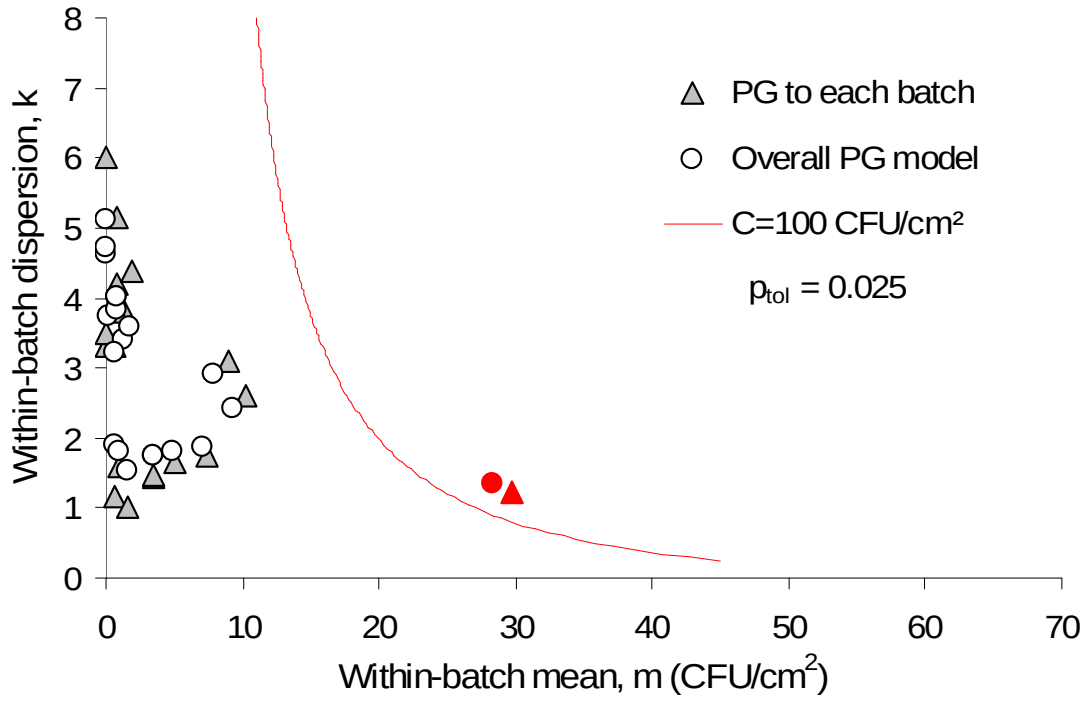


True WB distribution
Gamma (k, m)

Samples’ mean distribution
Gamma (k/n, m)

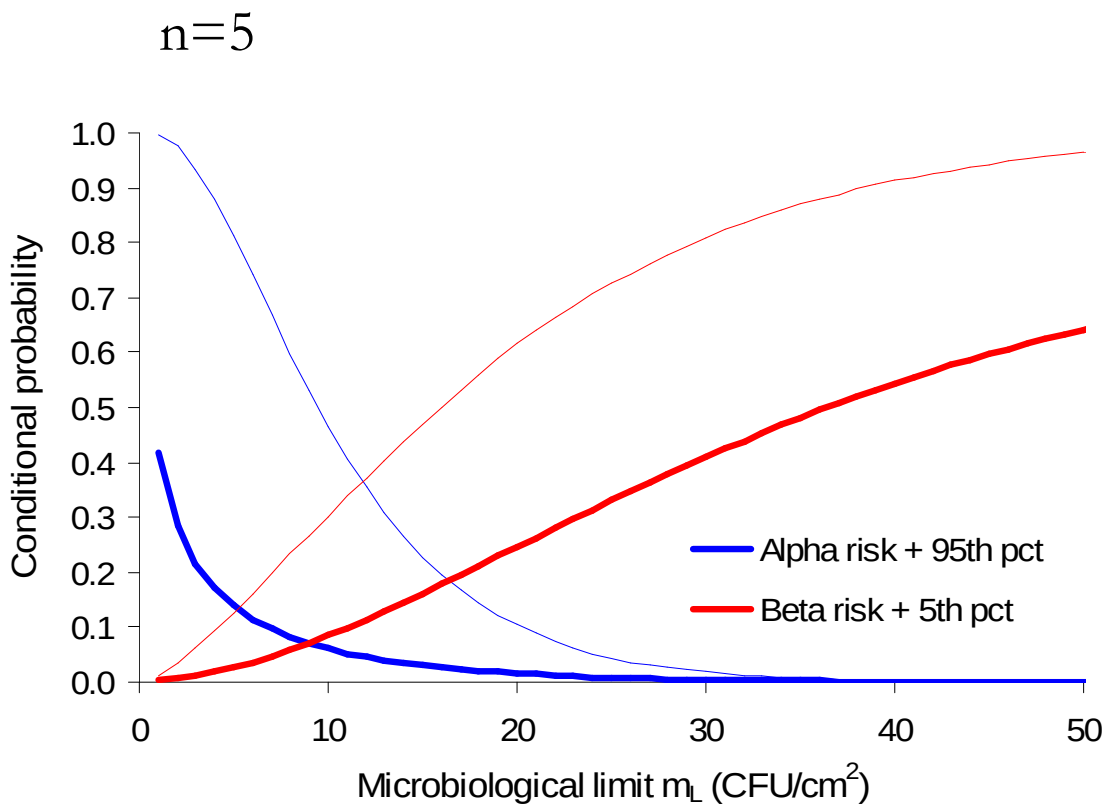
Results

Results: Definition of tolerance criterion



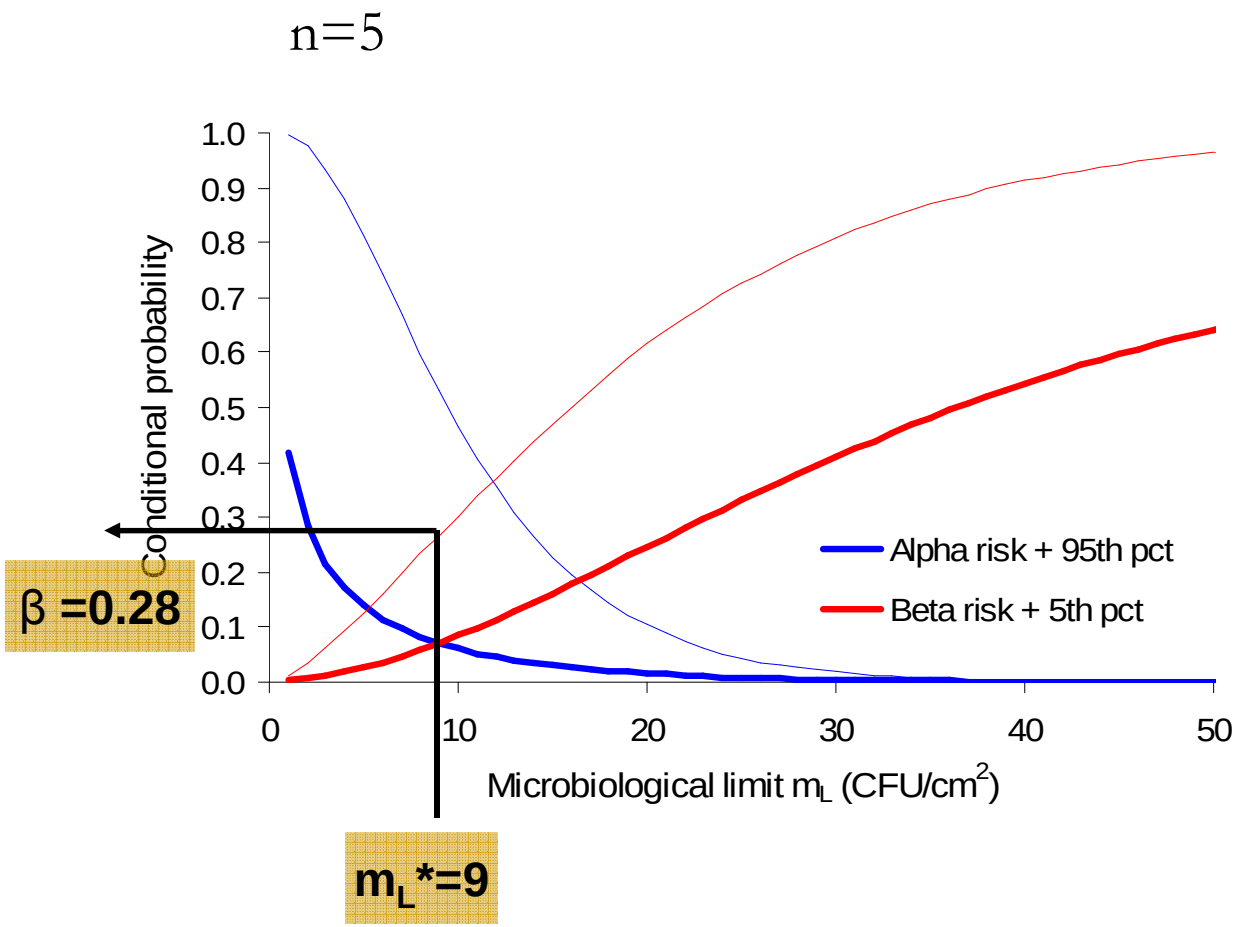
- Tolerance criterion: A batch is considered unsafe if more than 2.5% of carcasses have microbial load beyond 100 CFU/cm^2.
- The tolerance criterion defines the limiting quality contour.

Results: Decision landscape curve



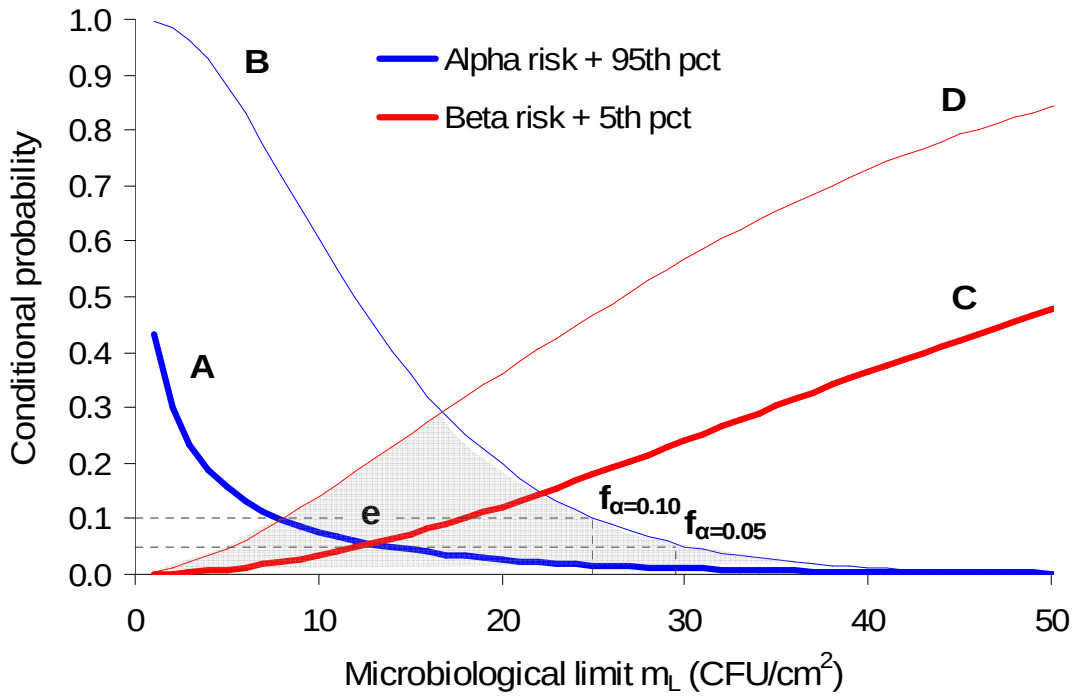
- As m_L increases, α risk decreases, and β risk increases.
- At every m_L, we have distributions of uncertainty of α and β, originating from the BB variability.

Results: Decision landscape curve



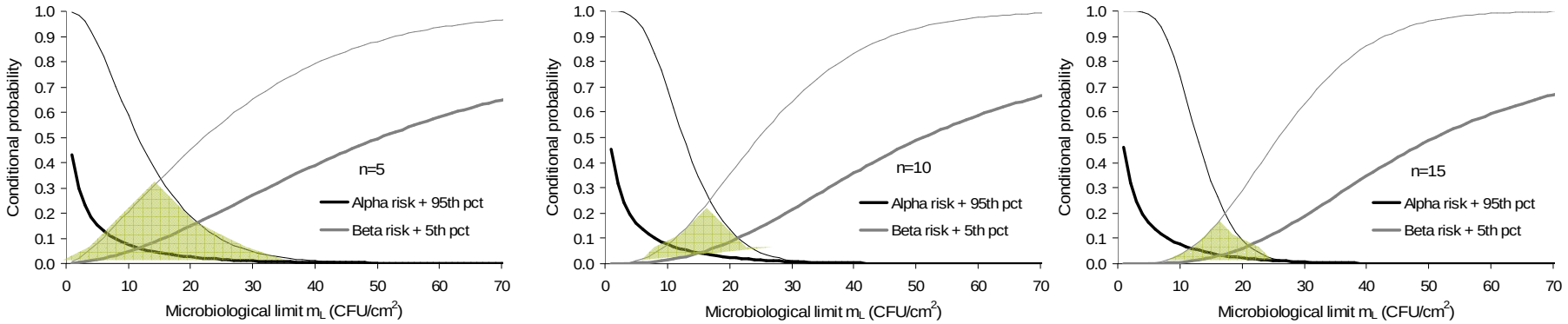
- As m_L increases, α risk decreases, and β risk increases.
- At every m_L, we have distributions of uncertainty of α and β, originating from the BB variability.
- Even if we minimise both risks at m_L*=9 CFU/cm², β risk may be still as high as 28%
- Therefore, assess higher sample size n.

Results: Relevant descriptors from the decision landscape curve



- When safety of production system is under control, m_L can be found at a desirable maximum α.
- Ex: reasonable quality batches should be accepted with a minimum confidence of 90% (α=0.10)

Results: Effect of sample size



- When sample size is higher, the misclassification errors are lower - the overlap area is smaller

Results: Possible sampling plans

- Derived under the assumption that good batches should be accepted with a minimum confidence of 90%

n	$m_L \mid (\alpha_{95pct}=0.10)$ (CFU/cm^2)	α (Mean + 90% CI)	$\beta \mid m_L$ (Mean + 90% CI)
5	20.0	0.0160 [0 – 0.10]	0.246 [0 – 0.615]
6	19.5	0.0166 [0 – 0.10]	0.217 [0 – 0.595]
8	18.5	0.0172 [0 – 0.10]	0.172 [0 – 0.537]
10	17.5	0.0188 [0 – 0.10]	0.134 [0 – 0.496]
12	17.0	0.0192 [0 – 0.10]	0.115 [0 – 0.461]
14	16.8	0.0198 [0 – 0.10]	0.101 [0 – 0.447]
16	16.5	0.0212 [0 – 0.10]	0.087 [0 – 0.413]
18	16.2	0.0223 [0 – 0.10]	0.077 [0 – 0.381]
20	15.9	0.0231 [0 – 0.10]	0.066 [0 – 0.366]

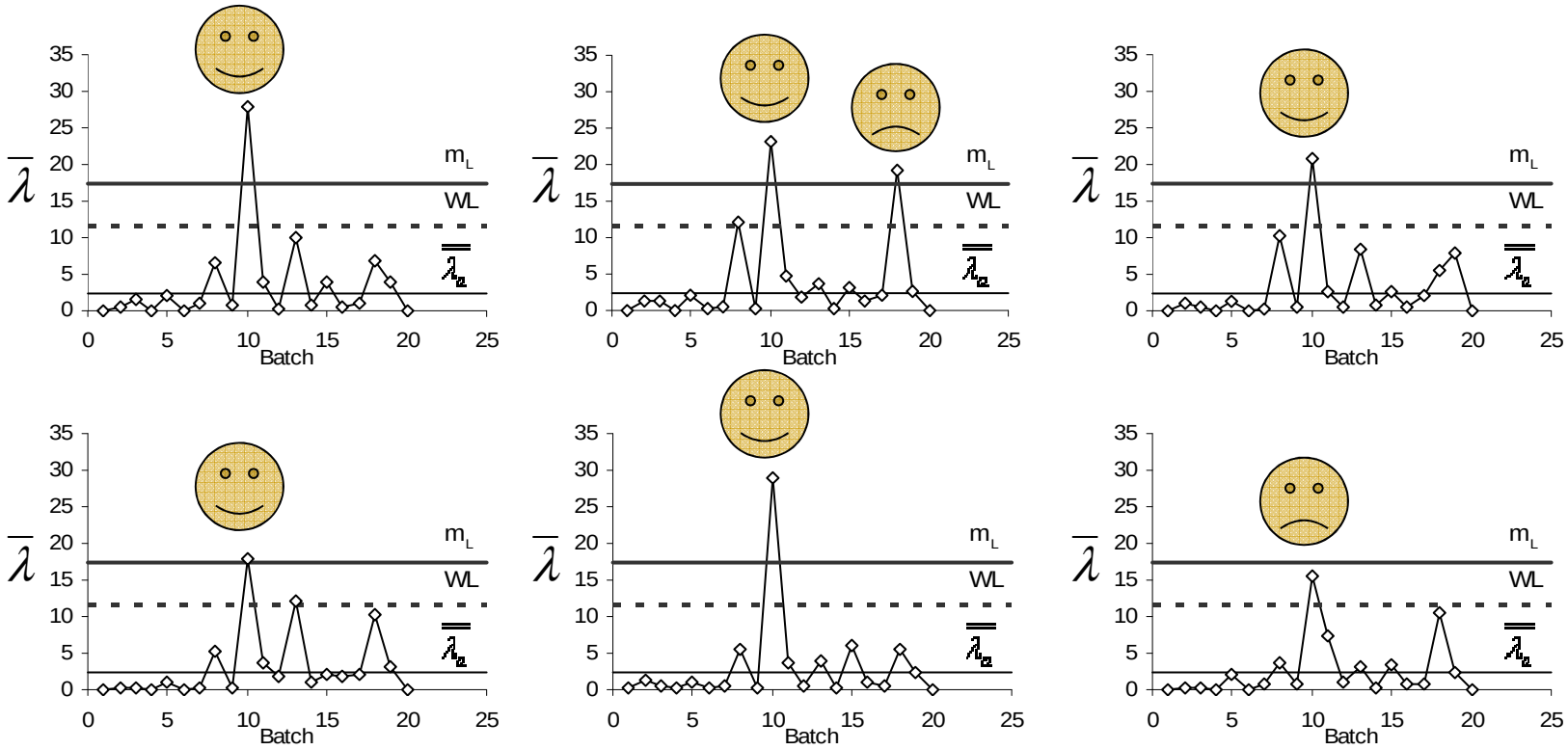
Results: Possible sampling plans

- We may wish to establish that the consumer's risk should on average be 0.10.

n	$m_L \mid (\alpha_{95pct}=0.10)$ (CFU/cm^2)	α (Mean + 90% CI)	$\beta \mid m_L$ (Mean + 90% CI)
5	20.0	0.0160 [0 – 0.10]	0.246 [0 – 0.615]
6	19.5	0.0166 [0 – 0.10]	0.217 [0 – 0.595]
8	18.5	0.0172 [0 – 0.10]	0.172 [0 – 0.537]
10	17.5	0.0188 [0 – 0.10]	0.134 [0 – 0.496]
12	17.0	0.0192 [0 – 0.10]	0.115 [0 – 0.461]
14	**16.8**	**0.0198 [0 – 0.10]**	**0.101 [0 – 0.447]**
16	16.5	0.0212 [0 – 0.10]	0.087 [0 – 0.413]
18	16.2	0.0223 [0 – 0.10]	0.077 [0 – 0.381]
20	15.9	0.0231 [0 – 0.10]	0.066 [0 – 0.366]

Results: Performance of chosen sampling plan

- n=14, m_L=17 CFU/cm^2



Conclusions

- The proposed methodology proved to be statistically sound:
  - It is the first to address the derivation of a sampling plan as a classification problem capable of propagating the between-batch variability
  - Adequate for microbial data consisting of many zero counts
  - Uses the more effective arithmetic means, compatible with the Poisson-gamma

Thank you !!!

Merci beaucoup !!!

Gracias !!!