

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

# Food Control

journal homepage: [www.elsevier.com/locate/foodcont](http://www.elsevier.com/locate/foodcont)

## Conducting inferential statistics for low microbial counts in foods using the Poisson-gamma regression



Ursula Gonzales-Barron <sup>a,\*</sup>, Vasco Cadavez <sup>a</sup>, Francis Butler <sup>b</sup>

<sup>a</sup> CIMO Mountain Research Centre, ESA School of Agriculture, Polytechnic Institute of Braganza, Campus de Santa Apolónia, Apartado 1172, Braganza 5301-854, Portugal

<sup>b</sup> UCD School of Biosystems Engineering, University College Dublin, UCD Belfield, Dublin 4, Ireland

### ARTICLE INFO

#### Article history:

Received 22 May 2013

Received in revised form

24 August 2013

Accepted 14 September 2013

#### Keywords:

Carcass

Chilling

Coliforms

Negative binomial

Zero-inflated

Plate counts

### ABSTRACT

Mixed Poisson distributions have been shown to be able to represent low microbial counts more efficiently than the lognormal distribution because of its greater flexibility to model microbial clustering even when data consist of a large proportion of zero counts. The objective of this study was to develop an alternative modelling framework for low microbial counts based on heterogeneous Poisson regressions. As an illustration, Poisson-gamma regression models were used to assess the effect of chilling on the concentration of total coliforms from beef carcasses ( $n = 600$ ) sampled at eight large Irish abattoirs. Three Poisson-gamma and three zero-modified (hurdle and zero-inflated) models were appraised with a series of random-effects variants in order to extract any variability in microbial mean concentration, dispersion and/or proportion of zero counts. Models were compared and validated in their ability to predict the coliforms counts on carcasses after chilling. In all five test batches, the hurdle Poisson-gamma distributions predicted the observed post-chill counts closer than the Poisson-gamma distributions. This is justified by the better capacity of the hurdle model to represent a higher proportion of zero counts, which were in fact observed in the post-chill batches. Thus, with a coded variable (pre-chill/post-chill) as treatment, and extracting the significant variability of batches nested in abattoirs for the coliforms mean concentration ( $\sigma_u^2 = 2.68$ ), the dispersion measure ( $\sigma_v^2 = 2.39$ ) and the probability of zero counts ( $\sigma_w^2 = 0.89$ ), the validated hurdle Poisson-gamma model confirmed that chilling has a decreasing effect on the viability of coliforms from beef carcasses, and that the concentration is reduced by an average (pre-chill to post-chill) factor of 2.2 (95% CI: 2.15–2.24) at batch level. The model also indicated that chilling increases the odds of producing a zero count from a carcass swab in about 13.5 times, and that the higher the coliforms concentration in a batch, the weaker the effect that chilling has to reduce such contamination on the beef carcasses.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

In the evaluation of microbiological quality of foodstuffs, bacterial concentration is conventionally expressed in terms of log CFU cm<sup>-2</sup> or g<sup>-1</sup>. Logarithmic transformation is believed to approximate data normality, which is fundamental for the application of inferential statistical data analysis based on the Gaussian distribution. This assumption leads to the widely-held practice that whenever bacterial colonies are not observed (zero counts), a low log value corresponding to the limit of enumeration of the microbiological test can be inserted. This statistical practice for 'censored' observations is known as imputation; and, depending on the

proportion of zero counts or censored points, the mean values are overestimated. To avoid this, [Busschaert, Geeraerd, Uyttendaele, and Van Impe \(2011\)](#) developed a censored data method to fit a lognormal distribution to plate counting data with zero counts, which later on [Pouillot, Hoelzer, Chen, and Dennis \(2013\)](#) adapted for fitting lognormal and zero-inflated lognormal distributions to Most Probable Number (MPN) data. On the other hand, [Gonzales-Barron & Butler \(2011a,b\)](#) have demonstrated that the lognormal (and Poisson-lognormal) distributions are only appropriate for the representation of high microbial counts, while the Poisson-gamma distribution performs much better in the characterisation of low microbial counts and even for highly clustered microbial data consisting of a large proportion of zero counts (non-detections). Whereas the lognormal distribution cannot naturally take in zero values, the Poisson-lognormal distribution greatly overestimates the mean microbial concentration when it attempts to fit in a high

\* Corresponding author. Tel.: +351 273 303325.

E-mail address: [ubarron@ipb.pt](mailto:ubarron@ipb.pt) (U. Gonzales-Barron).

frequency of zero counts (Gonzales-Barron & Butler, 2011a). As a consequence, parametric statistical analyses, such as ANOVA or regression models, should only be performed on bacterial counts of high occurrence (such as mesophile or total viable counts), for which a logarithmic transformation can mostly bring about an approximate normal distribution. For bacterial counts of lower occurrence (such as some hygiene indicators and pathogens) that can be better described by a Poisson-gamma distribution, the alternative modelling framework of the Poisson-gamma regression should be explored.

The heterogeneous Poisson models and their zero-inflated variants correspond to a family of flexible count data distributions that can handle over-dispersion (i.e., variance of the observed count data exceeds the mean), which is a condition caused by microbial heterogeneity or clustering. A heterogeneous Poisson model loosens the Poisson restriction (whereby the nominal variance equals the mean) by allowing the expected number of counts ( $\lambda$ ) to be a function of some random variable. If this random variable follows a gamma distribution, the resulting heterogeneous Poisson will be a Poisson-gamma, also known as a negative binomial distribution. However, with some types of data, over-dispersion may also stem from a high percentage of zero counts, for which the variance function of the Poisson-gamma model may be insufficient. In this case, the zero-modified Poisson-gamma regression models, such as zero-inflated and hurdle, may be more convenient (Gonzales-Barron, Kerr, Sheridan, & Butler, 2010). Therefore, the objective of this work was to introduce a count data regression framework to conduct inferential statistics on microbial counts that do not approximate to a normal distribution after logarithmic transformation due to clustering and the relatively high proportion of zero counts. As an illustration, we will use the proposed Poisson-gamma and zero-modified Poisson-gamma regression models to assess whether there is an effect of chilling on the total coliforms recovered from beef carcasses; and, if so, to quantify the expected chilling effect at batch level. For this aim, a number of random-effects models were evaluated and compared in terms of goodness-of-fit and predictability.

## 2. Methodology

### 2.1. Sampling of pre-chill and post-chill beef carcasses and microbiological analysis

Eight beef export abattoirs, with a throughput of at least 30,000 cattle per annum each, located in the south, east and west of Ireland, were visited during 2007–2008 to obtain a representative sample of cattle being slaughtered throughout the country. Four of the abattoirs were each visited three times and the remaining four on two occasions. During each visit, 30 animals were randomly sampled throughout the day to obtain a representative sample of the cattle being presented for slaughter. After washing, at the end of the slaughter line, the two carcass sides of an animal to be sampled one before and one after chilling, were identified. Polyurethane sponges (Sydney Heath, Stoke on Trent, UK) were cut for use as carcass swabs (150 cm<sup>2</sup>), which were pre-soaked in 5 ml of Universal Quenching Agent (UQA). This solution contained 1.0 g peptone, 1.0 g sodium thiosulphate, 5.0 g Tween 80 and 0.7 g lecithin, in a litre of distilled water, at pH 7.0. The UQA solution was used because the levels of chlorine in the carcass wash water varied depending on the abattoir being visited. Swabs were placed in stomacher bags (Sarstedt, Numbrecht, Germany) and autoclaved at 115 °C for 10 min. Carcasses were swabbed using the method described by Lasta, Rodriguez, Zanelli, and Margaria (1992), which involved uniformly swabbing the entire outer surface of a carcass from the hindquarter to the forequarter. The extent of carcass

swabbed was estimated according to Lasta et al. (1992) and annotated for each animal. Areas swabbed ranged between 8900 and 11,200 cm<sup>2</sup> with a mean of ~10,000 cm<sup>2</sup>. A single sponge was used to swab each side of the carcass.

Two hundred millilitre of Buffered Peptone Water (BPW) (Oxoid, Basinstoke, UK) was added to each sponge and homogenised in a Colworth stomacher (Model 400, Seward, London, UK) at 200 rpm for 1.0 min. A 15-ml aliquot of the homogenate was poured into 30 ml sterile polystyrene screw cap containers (Sterilin, Stone, UK). To determine total coliforms, 1-ml volumes of the neat homogenate were dispensed into each of two Petri dishes and over-poured with Chromocult coliform agar (Merck, Darmstadt, Germany) and incubated at 37 °C for 24 h. Total coliforms were identified as salmon-coloured colonies. Although the concept of microbiological limit of enumeration is not used in the Poisson-gamma modelling framework, it is worthy to mention that, for the present protocol, this value was approximately  $-2 \log \text{CFU/cm}^2$ , assuming a mean carcass swab area of 10,000 cm<sup>2</sup> (If one colony was counted in one of the two plates, the threshold would correspond to 1 CFU/(2 × 50 cm<sup>2</sup>)).

### 2.2. Poisson-gamma model

In its simplest form, the Poisson distribution models the number of events from a memory-less exponential process where the event rate  $\lambda$  is constant. Assuming that (i) there are no losses in the transfer of bacterial cells from the carcass surface to swabs and from the swabs to the homogenate (i.e., perfect recovery); (ii) the bacterial cells extracted from the swabs are randomly distributed in the 200-ml neat homogenate; and (iii) each of the plated cells will become a colony after incubation, let  $Y_i$  be the random variable for the number of bacterial colonies  $i$ , counted on a Petri dish. If  $Y_i$  follows a Poisson distribution, the probability mass function is,

$$\Pr(Y_i) = \frac{\exp(-\mu_i) \times \mu_i^{Y_i}}{Y_i!} \quad (1)$$

$$\mu_i = \frac{A}{V} dt \times \lambda_i \quad (2)$$

where  $A$  is the swabbed area of the carcass (cm<sup>2</sup>), which was measured for each of the sampled carcasses,  $V$  is the homogenate volume (200 ml),  $d$  is the dilution level at which the respective plate count  $Y_i$  was made (i.e., 10<sup>0</sup>, 10<sup>-1</sup>, 10<sup>-2</sup>, etc.),  $t$  represents the aliquot volume poured onto the Petri dish (1 ml), and  $\lambda_i$  is the unknown mean bacterial concentration of the carcass sample in cell/cm<sup>2</sup>.

In a Poisson regression model, the mean parameter  $\lambda_i$  would be a function of a vector of covariates  $X_n$ , where  $\beta$  is a parameter vector consisting of an intercept  $\beta_0$  and the chilling effect  $\beta_1$ . In our case, we wish to quantify the effect of the chilling treatment  $\beta_1$ ; thus, the only covariate was a coded variable  $X$  (0 as pre-chill, 1 as post-chill),

$$\lambda_i = \exp(\beta_0 + \beta_1 X) \quad (3)$$

Within a batch, carcasses do not share the same true unknown microbial concentration  $\lambda_i$  (CFU/cm<sup>2</sup>) due to the heterogeneity in contamination arising from systematic errors along the slaughter line and/or random sources of variability. Thus, the basic Poisson regression model was generalised by including a dispersion parameter to accommodate the heterogeneity in the count data. A generalised Poisson distribution lets the expected microbial concentration  $\lambda_i$  be a function also of some random variable  $e_i$  (Hinde & Demetrio, 1998),

$$\lambda_i = \exp(\beta_0 + \beta_1 X + e_i) = \exp(\beta_0 + \beta_1 X) \exp(e_i) \quad (4)$$

When  $\exp(e_i)$  follows a gamma distribution  $\Gamma(1/k, k)$  with expected value 1 and dispersion parameter  $k$ , the above modification to the Poisson distribution yields a more flexible distribution, the Poisson-gamma or negative binomial distribution. The discrete probability mass function for the microbial plate count  $Y$ , sampled from a batch of beef carcasses, can then be estimated as,

$$\Pr(Y_i) = \frac{\Gamma(Y_i + k^{-1})}{\Gamma(Y_i + 1)\Gamma(k^{-1})} \left[ \frac{k^{-1}}{k^{-1} + \lambda_i} \right]^{k^{-1}} \left[ \frac{\lambda_i}{k^{-1} + \lambda_i} \right]^{Y_i} \quad (5)$$

where  $\lambda_i = \exp(\beta_0 + \beta_1 X)$  and the gamma function  $\Gamma(a) = \int_0^\infty e^{-s} s^{a-1} ds$ . Now, the equi-dispersion implied by the Poisson distribution has been corrected, and clearly the variance is higher than the expected value.

$$\text{Var}(Y) = \lambda_i(1 + k\lambda_i) \quad (6)$$

The Poisson-gamma and Poisson models are considered to be nested because as  $k$  converges to zero, the Poisson-gamma converges to Poisson. For model fitting, the corresponding (rearranged) log-likelihood (LL) function is used.

$$LL = \sum_{i=1}^n \left[ \text{Log} \left[ \frac{\Gamma(Y_i + k^{-1})}{\Gamma(Y_i + 1)\Gamma(k^{-1})} \right] - (Y_i + k^{-1}) \text{Log}(1 + k\lambda_i) + Y_i \text{Log}(k\lambda_i) \right] \quad (7)$$

### 2.3. Zero-inflated Poisson-gamma model

With some types of data, a high percentage of zero counts is present, and may not be accounted for by the variance functions of the Poisson-gamma distribution. To model this excess of zeros, a mixture of two distributions: a degenerate distribution for the zero count and a standard count distribution, may be appropriate. This type of distribution is called zero-inflated, and it assumes that the individual counts can originate from two stochastic processes or groups. One group will always have a count of zero, while the other group will follow a count distribution, which is either a Poisson or a generalised Poisson, also responsible for generating zero counts. The probability mass function is given by,

$$\Pr(Y_i) = \begin{cases} p_0 + (1 - p_0)g(0) & \text{for } Y_i = 0 \\ (1 - p_0)g(Y_i) & \text{for } Y_i \geq 1 \end{cases} \quad (8)$$

where  $g$  is the probability mass function corresponding to the standard count distribution to be modified. Notice from the above equation, that a zero count can arise from the fixed-zero group with a probability  $p_0$  or from the count distribution  $g$  with a probability  $1 - p_0$ . The membership of the fixed-zero group is estimated by a probability  $p_0$ , which is calculated typically by a logit model with  $l_0$  as intercept.

$$\left[ \frac{p_0}{1 - p_0} \right] = \exp(l_0) \quad (9)$$

Substituting  $g$  in equation (8) by the Poisson-gamma yields the zero-inflated Poisson-gamma distribution (ZIPG), whose probability mass function becomes,

$$\Pr(Y_i) = \begin{cases} p_0 + (1 - p_0) \left( \frac{k^{-1}}{k^{-1} + \lambda} \right)^{k^{-1}} & \text{for } Y_i = 0 \\ (1 - p_0) \left[ \frac{\Gamma(Y_i + k^{-1})}{\Gamma(Y_i + 1)\Gamma(k^{-1})} \left[ \frac{k^{-1}}{k^{-1} + \lambda} \right]^{k^{-1}} \left[ \frac{\lambda}{k^{-1} + \lambda} \right]^{Y_i} \right] & \text{for } Y_i \geq 1 \end{cases} \quad (10)$$

The mean and variance of the ZIPG are, respectively,

$$E(Y) = \lambda_i(1 - p_0) \quad (11)$$

$$\text{Var}(Y) = \lambda_i(1 - p_0)(1 + \lambda_i(p_0 + k)) \quad (12)$$

The log-likelihood function of the ZIPG distribution is then computed as,

$$LL_{ZIPG} = \sum_{i=1}^n \left[ I(Y_i = 0) \log \left[ p_0 + (1 - p_0) \left( \frac{k^{-1}}{k^{-1} + \lambda} \right)^{k^{-1}} \right] + I(Y_i \geq 1) \left[ \log(1 - p_0) + \log \left[ \Gamma(Y_i + k^{-1}) \right] - \log[\Gamma(Y_i + 1)] - \log \left[ \Gamma(k^{-1}) \right] + Y_i \log(k\lambda) - (Y_i + k^{-1}) \log(1 + k\lambda) \right] \right] \quad (13)$$

### 2.4. Hurdle Poisson-gamma model

Originally developed by Mullahy (1986), hurdle models are capable of dealing with excess zero counts. A hurdle-type model considers that a count outcome is generated by two systematically different statistical processes, a binomial distribution determining if a count outcome is zero or non-zero, and a truncated-at-zero distribution  $g(Y)$  for count data governing all positive counts. In its general form, a hurdle-at-zero distribution has probability mass function,

$$\Pr(Y_i) = \begin{cases} \omega_0 & \text{for } Y_i = 0 \\ (1 - \omega_0) \frac{g(Y_i)}{(1 - g(0))} = \Phi g(Y_i) & \text{for } Y_i \geq 1 \end{cases} \quad (14)$$

where the numerator of  $\Phi$  can be interpreted as the probability of crossing the hurdle (or more precisely in this case, the probability to have at least one colony count on the Petri dish) and the denominator gives a normalisation that accounts for the (purely technical) truncation of  $g(Y)$ . When the truncated count data distribution is a Poisson-gamma, the mean and variance of the hurdle Poisson-gamma (HPG) are given by,

$$E(Y) = \Phi \lambda \quad (15)$$

$$\text{Var}(Y) = \Pr(Y > 0) \text{Var}(Y > 0) + \Pr(Y = 0) E(Y > 0) \quad (16)$$

The probability mass function of the HPG is,

$$\Pr(Y_i) = \begin{cases} \omega_0 & \text{for } Y_i = 0 \\ (1 - \omega_0) \left[ \frac{\Gamma(Y_i + k^{-1})}{\Gamma(Y_i + 1)\Gamma(k^{-1})} (1 + k\lambda)^{-(Y_i + \alpha^{-1})} (\lambda k)^{Y_i} \right] & \text{for } Y_i \geq 1 \end{cases} \quad (17)$$

with the probability  $\omega_0$  calculated typically by a logit transformation with  $l_0$  as intercept.

$$\left[ \frac{\omega_0}{1 - \omega_0} \right] = \exp(l_0) \quad (18)$$

The log-likelihood function of the HPG distribution is computed as,

$$LL_{HPG} = \sum_{i=1}^n \left[ I(Y_i = 0) \log(\omega_0) + I(Y_i \geq 1) \left[ \log(1 - p_0) + \log \left[ \Gamma(Y_i + k^{-1}) \right] - (Y_i + k^{-1}) \log(1 + k\lambda) + Y_i \log k\lambda - \log \left[ 1 - \left( \frac{k^{-1}}{k^{-1} + \lambda} \right)^{k^{-1}} \right] \right] \right] \quad (19)$$

## 2.5. Fitted variants of Poisson-gamma regression and zero-modified models

A number of random-effects Poisson-gamma regressions variants were considered and they are described, as follows.

### 2.5.1. Model 1: Poisson-gamma with random-effects in the mean concentration

This model assumes that the measure of dispersion  $k$  is common for all the sampled batches of beef carcasses. The dispersion or variance is assumed to be the same for a low or a highly contaminated lot. Nevertheless, this model takes into account the between-batch heterogeneity (nested within abat-toirs) in the mean contamination. To extract this source of heterogeneity, the fitted function of the expected microbial concentration  $\lambda$  becomes,

$$\lambda = \exp\left([\beta_0 + u_{b(a)}] + \beta_1 X\right) \quad (20)$$

where the intercept  $\beta_0$  is allowed to take up different independent values batch to batch. The effect of chilling  $\beta_1$  is calculated as a fixed-effects term; this is, as an average effect of all production batches. The random effects  $u_{b(a)}$  of the production batches  $b$  nested in the abat-toirs  $a$  are assumed to have a normal distribution with mean zero and standard deviation  $\sigma_u$ .

### 2.5.2. Model 2: Poisson-gamma with two random-effects in the mean concentration

This model also estimates a common dispersion parameter  $k$  for all the batches. However, it does not only extract the batch-to-batch variability in the mean contamination level or intercept  $\beta_0$  but also in the chilling effect  $\beta_1$ ,

$$\lambda = \exp\left([\beta_0 + u_{b(a)}] + [\beta_1 + v_{b(a)}]X\right) \quad (21)$$

where the random effects  $v_{b(a)}$  model the shifts in the chilling effect  $\beta_1$  for each batch. The  $v_{b(a)}$  shifts are assumed to have a normal distribution with mean zero and standard deviation  $\sigma_v$ . The objective of this model was to assess if there is any correlation between the mean contamination level of a batch and the chilling effect. For this, the coefficient of correlation  $\rho$  between the two random effects was estimated from their covariance matrix.

### 2.5.3. Model 3: Poisson-gamma with random-effects in the mean concentration and in the dispersion measure

This is a flexible model in which now both the mean  $\lambda$  and the dispersion parameter  $k$  are affected by chilling. Because higher proportions of zero counts in a sampled batch lead to higher dispersion values  $k$ , it may be hypothesised that there is variability batch to batch in the dispersion parameter. The model then becomes,

$$\begin{aligned} \lambda &= \exp\left([\beta_0 + u_{b(a)}] + \beta_1 X\right) \\ k &= \exp\left([\alpha_0 + v_{b(a)}] + \alpha_1 X\right) \end{aligned} \quad (22)$$

The nested random-effects  $v_{b(a)}$  model the shifts in the intercept  $\alpha_0$  in the logarithm of the dispersion parameter in each batch, and they are assumed to have a normal distribution with mean zero and standard deviation  $\sigma_v$ . The correlation  $\rho$  between both random effects was assessed. The average effect of chilling on the dispersion parameter  $k$  is represented by  $\alpha_1$ .

### 2.5.4. Model 4: Zero-inflated Poisson-gamma with random-effects in the mean concentration and in the dispersion measure

This model is similar to Model 3, but assumes that there is an extra proportion of zero counts that cannot be described by the Poisson-gamma model. The logit transformation of the probability  $p_0$  (i.e., probability that a zero count arises from a fixed-zero group) is assumed to be affected by chilling, and its effect is represented by  $l_1$ .

$$\begin{aligned} \left[\frac{p_0}{1-p_0}\right] &= \exp(l_0 + l_1 X) \\ \lambda &= \exp\left([\beta_0 + u_{b(a)}] + \beta_1 X\right) \\ k &= \exp\left([\alpha_0 + v_{b(a)}] + \alpha_1 X\right) \end{aligned} \quad (23)$$

### 2.5.5. Model 5: Hurdle Poisson-gamma with random-effects in the mean concentration and in the dispersion measure

This model is analogous to Model 4, but governed by a hurdle-at-zero distribution.

$$\begin{aligned} \left[\frac{\omega_0}{1-\omega_0}\right] &= \exp(l_0 + l_1 X) \\ \lambda &= \exp\left([\beta_0 + u_{b(a)}] + \beta_1 X\right) \\ k &= \exp\left([\alpha_0 + v_{b(a)}] + \alpha_1 X\right) \end{aligned} \quad (24)$$

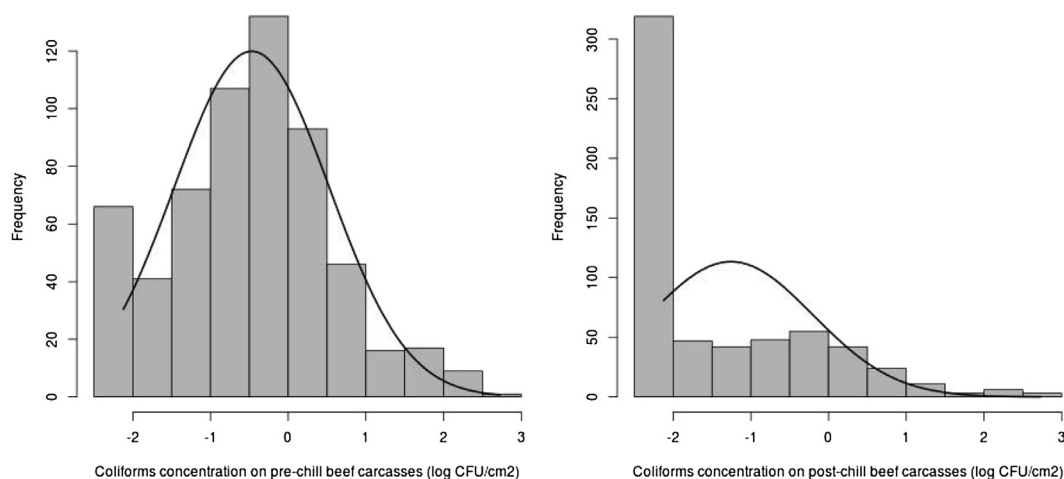
### 2.5.6. Model 6: Hurdle Poisson-gamma with random-effects in the mean concentration, in the dispersion measure, and in the proportion of zeros

This is the most complex model considered whereby batch-to-batch variability was assumed to be present in the mean concentration, the dispersion measure, and the logit of the probability of zero counts  $\omega_0$ . This model considers that the existing zero counts cannot be sufficiently modelled by the simple Poisson-gamma, and that the extra proportion of zero counts are variable among batches of production. The nested random-effects  $w_{b(a)}$  model the shifts in the intercept  $l_0$  predicting the logit of  $\omega_0$ , and they are assumed to have a normal distribution with mean zero and standard deviation  $\sigma_w$ .

$$\begin{aligned} \left[\frac{\omega_0}{1-\omega_0}\right] &= \exp\left([l_0 + w_{b(a)}] + l_1 X\right) \\ \lambda &= \exp\left([\beta_0 + u_{b(a)}] + \beta_1 X\right) \\ k &= \exp\left([\alpha_0 + v_{b(a)}] + \alpha_1 X\right) \end{aligned} \quad (25)$$

The six models were fitted to the coliforms counts data set using the PROC NLMIXED procedure for non-linear mixed models (SAS version 9.1, SAS Institute Inc., NC, USA). In this procedure, the log-likelihood functions were specified (Equations (7), (13) and (19)). Additionally, a measure of the reduction rate (RR) due to chilling in the mean microbial concentration in a batch was estimated as,

$$RR = 1/\exp(\beta_1) \quad (26)$$



**Fig. 1.** Histograms of frequency of the concentration of total coliforms on beef carcasses before (left) and after chilling (right) sampled from eight Irish abattoirs. First bar represents frequencies lower than or equal to the limit of enumeration ( $-2.0 \log \text{CFU}/\text{cm}^2$ ). Lack of fit to a normal distribution ( $p < 0.001$ ) is graphically shown.

Comparisons among non-nested models in terms of goodness-of-fit were performed through the Bayesian Information Criterion ( $BIC = -2LL + \log(n)p$ ), where  $LL$ ,  $p$  and  $n$  represent the fitted log-likelihood function, the number of parameters of the model, and the number of observations, respectively. The two best-fit models were then compared in their ability to predict the within-batch proportion of zero counts.

### 2.6. Model validation

Models were evaluated for their capacity to predict the effect of chilling (i.e., predict the coliforms counts in a post-chill batch from a pre-chill batch). The best-fit Poisson-gamma model was compared with the best-fit zero-modified Poisson-gamma model. As the aim of the models is to estimate the effect of chilling at batch level (and not at carcass level since the coliforms concentration of a post-chill carcass cannot be predicted from its pre-chill concentration), five production batches were randomly selected and their post-chill data excluded from the whole data set. Next, the Poisson-gamma and the zero-modified Poisson-gamma models were refitted using the smaller data set. Using the new model parameters and the random effects calculated for each of the five pre-chill batches, the parameters of their respective post-chill distributions were estimated. These distributions constitute the *predicted* distributions of the coliforms counts from post-chill carcasses for a given batch. In other words, the post-chill distributions were predicted from the fitted pre-chill distributions. The predicted Poisson-gamma and zero-modified distributions were then contrasted, in their cumulative mass probability form, against the ranked post-chill plate count observations for the five selected batches. Differences in the predictive ability of the models were assessed.

## 3. Results and discussion

### 3.1. Assessment of the coliforms data

Earlier, Gill, Deslandes, Rahn, Houde, and Bryant (1998) pointed out that, having a priori knowledge or assuming that the microbial concentration follows a normal distribution, for there to be a reasonable approximation, the bacteria of interest should be counted in at least 85% of the samples; otherwise, a censored model could be considered (Busschaert et al., 2011). In this work, the microbial data set of coliforms in beef, consisting of 67% of positive

counts, did not fit to a normal distribution. For plotting the histogram of frequencies in  $\log \text{CFU}/\text{cm}^2$ , the zero counts were replaced by the approximate microbiological limit of enumeration of  $-2.0 \log \text{CFU}/\text{cm}^2$ . As shown in Fig. 1, neither for the pre-chill nor for the post-chill group, the microbial concentration could be approached by a normal distribution. Here, it is worthy to mention that such practice of replacing zero counts by the limit of enumeration has been very common for statistical treatment either merely descriptive or inferential. Thus, it is clear that, while logarithmic transformation to induce data normality can be suitable for bacterial counts of high occurrence such as mesophile or total viable counts (Gonzales-Barron & Butler, 2011a), this approach may be inappropriate for bacterial counts of lower occurrence such as coliforms or pathogens.

If we still aimed to fit the low counts data to a lognormal distribution, due to the high level of zero counts (equivalent to less than  $-2.0 \log \text{CFU}/\text{cm}^2$ ), at least in the post-chill group (Fig. 1, right), the application of the Box-Cox method (Peltier, Wilcox, & Sharp, 1998) would fail to approximate the data to a normal distribution. On the other hand, the discrete heterogeneous Poisson distributions have been recently shown to closely describe microbial data sets containing zeros. While Jongenburger, Reij, Boer, Zwietering, and Boer (2012) indicated that both the Poisson-lognormal and the Poisson-gamma can be equally adequate for representing low microbial counts and a large proportion of zero counts; Gonzales-Barron, Lenahan, Sheridan, and Butler (2012), and Gonzales-Barron, Zwietering, and Butler (2013) pointed out that the Poisson-gamma distribution is more suitable than the Poisson-lognormal, as the latter tends to overestimate the mean concentration. In fact, Gonzales-Barron & Butler (2011a) have shown that the concentration of coliforms on beef carcasses is best described by the Poisson-gamma distribution. Thus, to carry out descriptive, inferential and predictive statistics on low microbial counts, the alternative approach proposed by the authors is to conduct count data regression models based on heterogeneous Poisson distributions; in this particular case, based on the Poisson-gamma. For the fitting of these Poisson-gamma models, the coliforms counts cannot be expressed in terms of a continuous concentration variable ( $\log \text{CFU}/\text{cm}^2$ ), but in terms of a discrete variable. The discrete response variable is defined as the number of colonies counted on a Petri dish originated from the 1-ml aliquot ( $Y$  in CFU). Thus, this type of count data models has the advantage of using the observation itself as the response variable and not a logarithmic transformation of a back-calculated variable ( $\log \text{CFU}/\text{cm}^2$ ; Table 1).

**Table 1**  
Differences between the inferential statistical methods based on the proposed Poisson-gamma distribution and the common data normality.

	Based on the Poisson-gamma and zero-modified variants	Based on normality
Over-dispersion Distribution characteristics	Capable to deal with highly over-dispersed and clustered microbial data. Discrete, takes in zero counts.	Deals with some over-dispersion. Continuous, does not take in zero counts.
Suitable data	Flexible enough to deal with low counts and/or presence of many zero counts (Mussida, Vose & Butler, 2013; Mussida, Gonzales-Barron & Butler, 2013; Gonzales-Barron & Butler, 2011a,b).	Only high counts and/or absence of zero counts (Gonzales-Barron & Butler, 2011a,b).
Variable	Observed number of colonies $Y$ counted on a petri-dish (CFU). Hence, the dilution level and the replicate values are both entered in the model.	Average logarithmic concentration $\lambda$ (log CFU/g or $\text{cm}^2$ ) back-calculated from $Y$ . Hence, the dilution level and the replicate values are not entered in the model.
Estimated mean values	Microbial mean concentrations are given in CFU/g or $\text{cm}^2$ . This is the arithmetic mean, which is a more suitable measure from the view point of risk-based metrics and sampling plans (Gonzales-Barron et al., 2012).	Microbial mean concentrations are given in log CFU/g or $\text{cm}^2$ , which represent the geometric mean. Estimation of the arithmetic mean can only be possible by knowing the standard deviation.
Advantages	The Poisson distribution considers the random sampling process of the bacteria taken from the homogenate solution. Models the zero counts in a natural way, and the zero-inflated variant allows for the distinction between true zeros and false non-detections due to sampling. Poisson-gamma can be applied equally well to microbial data with many, very few or no zero counts (Gonzales-Barron et al., 2012).	Very known and easy to fit.
Disadvantages	Less known, although most statistical packages have already built-in functions to fit negative binomial (Poisson-gamma) regression models.	Does not take into account any random sampling process; thus, there cannot be any distinction between true zeros and false non-detections. When applied to low counts data, it produces biased results (Mussida et al., 2013; Gonzales-Barron & Butler, 2011a,b). Cannot be used in all cases, rather the exception than the rule.

The direct use of the discrete plate count raw data (as opposed to the concentration data) for low counts has been also recently supported by Commeau, Parent, Delignette-Muller, and Cornu (2012), who recommended it for microbial data consisting of 15%–75% zero counts, although such rule of thumb was derived based on an assumption that the data dispersion could be characterised by the lognormal distribution.

### 3.2. Assessment of the fitted models

The simplest Poisson-gamma regression model (Model 1), with a significant dispersion parameter  $k$  (Table 2), suggested that the over-dispersion of the plate count data is indeed accounted for by

**Table 2**  
Mean parameter estimates and standard errors (in brackets) of the random-effects Poisson-gamma regression models fitted to the plate count data of total coliforms on Irish beef carcasses. Value of parameters before and after chilling are also shown.

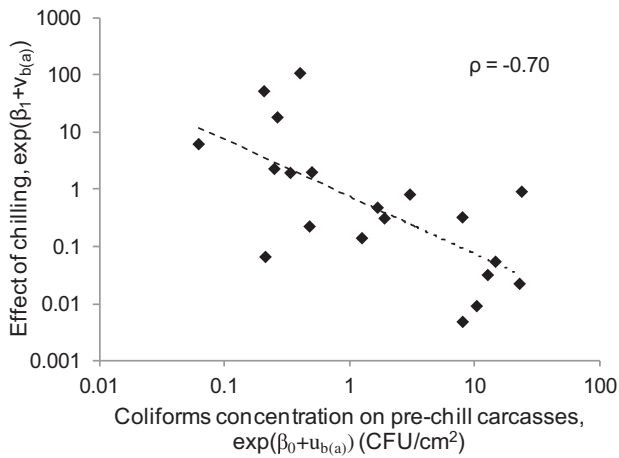
	Model 1	Model 2	Model 3
No. parameters	3	4	6
Parameters			
$\beta_0$	1.178*** (0.336)	0	0.800*** (0.008)
$\beta_1$	−0.834*** (0.205)	0	0.010 <sup>ns</sup> (0.009)
$k$	6.205*** (0.240)	5.121*** (0.204)	–
$\alpha_0$	–	–	1.016*** (0.003)
$\alpha_1$	–	–	1.493*** (0.002)
Random effects			
$\sigma^2_{u_i}$	2.367* (0.784)	3.626* (1.195)	2.595*** (0.013)
$\sigma^2_{v_i}$	–	7.904* (2.602)	0.363*** (0.003)
$\rho$	–	−0.698* (0.120)	0
Goodness-of-fit			
BIC	11229	11034	10817
Other estimates			
$\lambda$ before (CFU/ $\text{cm}^2$ )	3.246 (1.189)	–	2.225 (0.019)
$\lambda$ after (CFU/ $\text{cm}^2$ )	1.410 (0.517)	–	2.247 (0.022)
$k$ before	–	–	2.763 (0.008)
$k$ after	–	–	12.29 (0.046)
$RR = 1/\exp(\beta_1)$	2.302** (0.470)	–	0.990 <sup>ns</sup> (0.900)

Significance codes: 0 \*\*\*\* 0.001 \*\*\* 0.01 \*\* 0.05 \*<sup>ns</sup> Non-significant.

the Poisson-gamma model. It also suggests, although preliminarily, that there is a significant decreasing effect of chilling on the coliforms viability by an overall factor of 2.3 ( $RR$ : mean ratio of reduction in coliforms counts on carcasses from a pre-chill batch to a post-chill batch). There is also a significant heterogeneity in mean concentration among the batches nested in the abattoirs, represented by a standard deviation  $\sigma_{u_i}$  of 2.367 (Table 2). The fluctuation found batch to batch in the mean coliforms concentration is not unexpected as there are many factors contributing to their variability such as hygiene of the abattoirs, size of the abattoirs, practice of operators, type and capacity of the chilling systems, logistics, carcass surface dryness, proximity of contaminated carcasses, etc.

The objective of the second Poisson-gamma model was not to estimate the effect of chilling but to evaluate whether the extent of reduction in coliforms counts due to chilling depends on the initial level of contamination in a batch. Thus, the correlation between the random effects of the intercept  $\beta_0$  and the slope  $\beta_1$  of the logarithmic function of the mean concentration  $\lambda$  (Equation (9)) was estimated. Both random effects as well as their covariance were significant ( $p < 0.05$ ), with an also significant coefficient of correlation of  $\rho = -0.7$  (Table 2). This negative correlation indicated that as the initial (pre-chill) contamination of a batch increases ( $\exp(\beta_0)$ ), the reducing effect of chilling ( $\exp(\beta_1)$ ) on the viability of coliforms tends to be weaker. In other words, in batches of beef carcasses produced with good hygiene (i.e., low coliforms contamination), the chilling process will have a greater beneficial effect than in batches where contamination is high. The association can be clearly visualised in Fig. 2, where each of the markers represents the exponential values of the paired random effects ( $\exp(v_{b(a)})$ ,  $\exp(u_{b(a)})$ ) of a batch. Notice that in Model 2, both  $\beta_0$  and  $\beta_1$  were not different from zero ( $p > 0.05$ ), and hence they were dropped from the model.

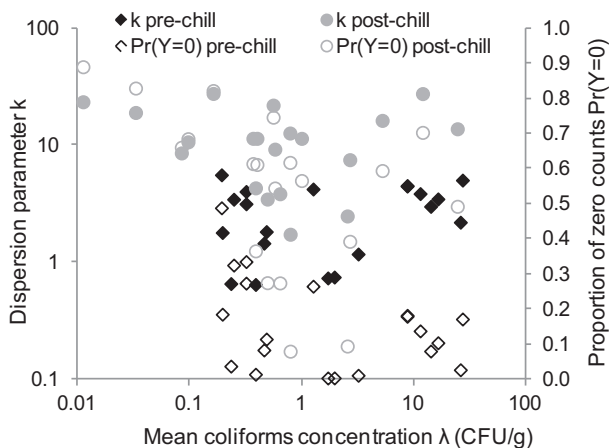
Further preliminary analysis of the coliforms data set aimed to assess whether there was another source of variability that needed to be accounted for by the model; specifically in the dispersion parameter  $k$ . To this effect, forty Poisson-gamma distributions were separately fitted to each of the twenty batches partitioned by



**Fig. 2.** Correlation between the fitted initial concentration of coliforms on a batch of pre-chill beef carcasses and its respective chilling effect  $\beta_1$  as modelled by a random-effects Poisson-gamma regression.

pre-chill and post-chill group. The scatter plot of their parameters showed that there was not only between-batch variability in the mean concentration  $\lambda$  but also in the dispersion parameter  $k$  (Fig. 3). Furthermore, it suggested that there was no correlation that needed to be accounted for between the mean and the measure of spread of the Poisson-gamma distribution, and that the dispersion values  $k$  of the post-chill group (1.7–30) were in general higher than the ones of their respective pre-chill group (0.6–6). Higher values of dispersion  $k$  produce higher probabilities of zero counts ( $\Pr(Y = 0)$ ). Thus, the next Poisson-gamma model fitted (Model 3) was formulated to collect the batch-to-batch variability in the mean dispersion parameter  $k$  as well as the effect of chilling on  $k$  (Equation (22)).

When a log-linear predictor for the dispersion parameter  $k$  was introduced (Model 3), the deviation measure  $BIC$  considerably decreased (in relation to Model 1). As expected, the between-batch variability  $\sigma^2_v$  in the (logarithm of) the dispersion parameter was significant, and the mean dispersion parameter for the pre-chill carcasses group was significantly lower (2.8,  $p < 0.05$ ) from the one for the post-chill group (12.3; Table 2). As explained in Gonzales-Barron & Butler (2011a), the higher the proportion of zeros



**Fig. 3.** Variability in the mean microbial concentration  $\lambda$  and in the dispersion parameter  $k$ , as observed by fitting the Poisson-gamma distribution separately for each sampled production batch. The fitted probabilities of zero counts ( $\Pr(Y = 0)$ ) present in every pre-chill and post-chill batch are also shown on the secondary y-axis.

**Table 3**

Mean parameter estimates and standard errors (in brackets) of the random-effects zero-inflated and hurdle Poisson-gamma regression models fitted to the plate count data of total coliforms on Irish beef carcasses.

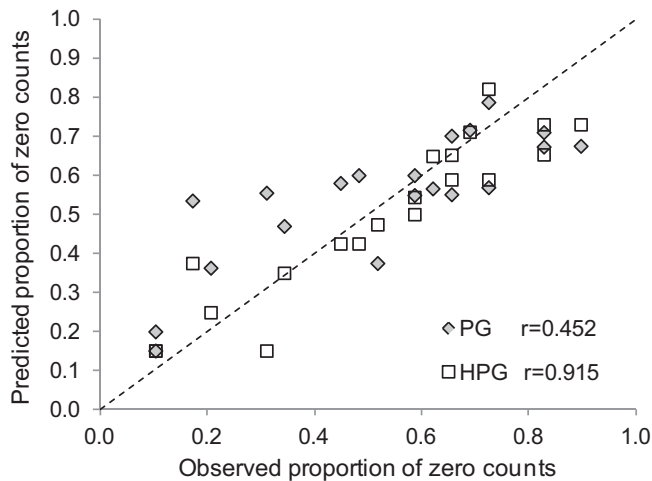
	Model 4	Model 5	Model 6
No. parameters	8	8	9
Parameters			
$\beta_0$	0.801*** (0.013)	0.040** (0.011)	0.040 <sup>ns</sup> (0.142)
$\beta_1$	0.009** (0.001)	-0.792*** (0.010)	-0.786*** (0.009)
$\alpha_0$	1.017*** (0.004)	2.015*** (0.002)	2.025*** (0.148)
$\alpha_1$	1.493*** (0.001)	2.287*** (0.006)	2.290*** (0.017)
$l_0$	-15.24 <sup>ns</sup> (14.22)	-2.313*** (0.071)	-2.630*** (0.295)
$l_1$	2.061 <sup>ns</sup> (2.032)	2.307*** (0.010)	2.600*** (0.325)
Random effects			
$\sigma^2_u$	2.596*** (0.063)	2.679*** (0.028)	2.682*** (0.288)
$\sigma^2_v$	0.365*** (0.001)	2.391*** (0.011)	2.391*** (0.051)
$\sigma^2_w$	–	–	0.886*** (0.006)
Goodness-of-fit			
$BIC$	10821	10773	10683
Other estimates			
$\lambda$ before (CFU/cm <sup>2</sup> )	2.228 (0.029)	1.042 (0.011)	1.000 (–)
$\lambda$ after (CFU/cm <sup>2</sup> )	2.249 (0.032)	0.477 (0.010)	0.475 (0.072)
$k$ before	2.765 (0.012)	7.506 (0.016)	7.576 (1.125)
$k$ after	12.303 (0.055)	34.77 (0.190)	34.83 (9.809)
$p_0, \omega_0$ before	2.40E-7 (2.26E-5)	0.090 (0.006)	0.067 (0.018)
$p_0, \omega_0$ after	1.88E-6 (0.0002)	0.498 (0.017)	0.493 (0.017)
OR treatment	–	10.04*** (0.100)	13.46** (4.379)
$RR = 1/\exp(\beta_1)$	–	2.182*** (0.022)	2.194*** (0.022)

Significance codes: 0 \*\*\*\* 0.001 \*\*\* 0.01 \*\* 0.05 \*<sup>ns</sup> Non-significant.

in a data set, the more skewed the Poisson-gamma distribution, and hence, the higher the dispersion parameter. However, according to this model, the effect of chilling  $\beta_1$  on the mean concentration was not significant, as opposed to what was found when the fixed dispersion parameter was assumed as fixed (i.e., Model 1; Table 2). Instead, according to Model 3, chilling had an effect only in the dispersion parameter  $k$ , which is driven by the higher proportion of zeros in the post-chill batches. No significant correlation ( $p > 0.05$ ) was found between the random effects for the mean microbial concentration ( $u_{b(a)}$ ) and the random effects for the dispersion parameter ( $v_{b(a)}$ ) (Table 2), as preliminary suggested by Fig. 3.

It was also noticed that some of the batches presented a high proportion of zero-count observations (Fig. 3). The probabilities of zero counts ( $\Pr(Y = 0)$ ) in every pre-chill and post-chill batch were estimated from the forty Poisson-gamma distributions separately fitted. Fig. 3 shows that many batches presented high  $\Pr(Y = 0)$  values, above 60%, which posed the question as to whether the variance function of the Poisson-gamma model may be insufficient to accommodate all the zero counts. This led to the evaluation of count data regressions based on the zero-modified Poisson-model; namely, the zero-inflated and hurdle variants. The zero-inflated Poisson-gamma model (Model 4) did not offer any improvement over the Poisson-gamma model (Model 3) since the parameters  $l_0$  and  $l_1$  defining the probability  $p_0$  were not significant (Table 3). Notice in Equation (23), that the non-significance ( $p > 0.05$ ) of the parameters  $l_0$  and  $l_1$  makes the zero-inflated variant collapse into the simpler Poisson-gamma (Equation (22)). The estimated  $p_0$  for the pre-chill and post-chill groups were practically null (2.40E-7 and 1.88E-6, respectively), suggesting that there is no need to allow for zero counts additional to those already modelled by the Poisson-gamma. The convergence of the zero-inflated Poisson-gamma to the simpler Poisson-gamma is also evidenced by the similarity of their parameter estimates and the insignificant improvement in  $BIC$  (Tables 2 and 3).

In contrast to the zero-inflated Poisson-gamma, the analogous hurdle Poisson-gamma (Model 5;  $BIC = 10,773$ ) was superior in



**Fig. 4.** Probability of zero counts predicted by the Poisson-gamma (PG) and the hurdle Poisson-gamma (HPG) models in comparison with the observed proportion of zero counts of coliforms for the twenty post-chill batches of beef carcasses. Coefficient of correlations ( $r$ ) between the observed and predicted values are shown for both models.

goodness-of-fit to the simpler Poisson-gamma (Model 3;  $BIC = 10,817$ ). In this case, the parameters  $l_0$  and  $l_1$  defining the probability  $w_0$  of zero counts were both significant (Table 3). Thus, according to Model 5, chilling has a significant effect on all the parameters: the microbial concentration  $\lambda$ , the dispersion parameter  $k$  and the proportion of zero counts  $w_0$ . For the microbial data analysed, chilling decreased the mean coliforms concentration from an average of 1.04 to an average of 0.48 CFU/cm<sup>2</sup>, and increased the probability of zero counts from 0.09 to 0.50 (Table 3). However, in Model 5, the logit parameters of  $w_0$  consist only of fixed effects, while Fig. 3 clearly shows that there is some variability in the proportion of zero counts among the pre-chill and post-chill groups. Thus, Model 6 was formulated to include such between-batch variability by means of a nested random-effects term  $w_{b(a)}$  added to the intercept  $l_0$  predicting the logit of  $w_0$ . With this complex model, a considerably better fit was obtained ( $BIC = 10,683$  in Table 3). No correlations were found among the three random effects, and basically all the model parameters were significant, except for the intercept  $\beta_0$ . The parameter estimates of Model 6 (with three random effects) were quite close to those of Model 5 (with two random effects); and likewise, Model 6 pointed out a significant decreasing effect of chilling on the concentration of coliforms ( $\beta_1 = -0.786$ ) at batch level. Hence, the reduction rate ( $RR$ ) was significant and estimated at 2.2, meaning that chilling reduces the mean concentration of the total coliforms (CFU/g) in a batch approximately by half. Interestingly, this chilling effect estimate was very similar to that from the simpler Model 1 ( $RR = 2.3$  in Table 2). Model 6 suggested that, in the data set analysed, chilling decreased the mean coliforms concentration from 1.00 to 0.48 CFU/cm<sup>2</sup>, and whereas the probability of encountering a zero count ( $w_0$ ) from a pre-chill carcass was on average 0.07 (95% CI: 0.03–0.11), in the case of a post-chill carcass it was significantly higher at 0.49 (95% CI: 0.46–0.52). Explained in other terms, the odds ratio (OR) for the treatment covariate was significant, and indicated that, chilling increases the odds of producing a zero count from a carcass swab in about 13.5 times (Table 3).

Another important aspect to take into consideration in the comparison of the Poisson-gamma regression variants is their ability to represent or predict the proportion of zeros; especially in the present work where we deal with *low counts*. To this effect, the probability of zero counts for the Poisson-gamma model ( $\Pr(Y = 0)$ ) and for the hurdle Poisson-gamma model ( $w_0$ ) were calculated

with the model parameters and random effects from Model 3 and Model 6 (Equations (5) and (17), were employed, respectively). When the predicted probabilities of zero counts ( $\Pr(Y = 0)$ ,  $w_0$ ) were compared against the observed proportion of zero counts in every batch, it was noticed that there was not a significant difference in the predictions between the Poisson-gamma and the hurdle Poisson-gamma models for the pre-chill batches of beef carcasses (results not shown). Nonetheless, for the post-chill batches, a considerable difference was observed between the Poisson-gamma and the hurdle Poisson-gamma in their capacity to describe the observed within-batch proportion of zero counts. Fig. 4 shows that the hurdle Poisson-gamma model yielded predicted probabilities of zero counts in post-chill batches that had a high correlation with the observed proportions ( $r = 0.915$ ), in contrast to the poorer zero counts predictions produces by the Poisson-gamma model ( $r = 0.452$ ). The capacity of the hurdle Poisson-gamma to predict the proportion of zero counts is outstanding, bearing in mind that the predicted probabilities plotted in Fig. 4 were not obtained by fitting separate distributions to each of the twenty batches, but by fitting a *single model* to the entire data set. This single model produces different probabilities of zero counts arising from the different random-effects values marginally estimated for each of the batches.

### 3.3. Model validation

In terms of goodness-of-fit ( $BIC$ ) and the ability to represent the zero counts of the coliforms data on beef carcasses, the hurdle Poisson-gamma (Model 6) was superior to the Poisson-gamma (Model 3). However, as the objective of this work is to assess the accuracy of the Poisson-gamma regression to estimate the effect of chilling on the coliforms counts, it is also necessary to test the model's ability to predict the post-chill observations of a batch (i.e., a predicted post-chill distribution) using the coliforms counts of the pre-chill carcasses from that batch (i.e., a fitted pre-chill distribution). The model validation was performed using five batches, and by means of a two-stage process that we proceed to explain for the hurdle Poisson-gamma model. The procedure used for the Poisson-gamma model is analogous. Firstly, using the model parameters  $\beta_0$ ,  $\alpha_0$  and  $l_0$ , and the random effects  $u_{b(a)}$ ,  $v_{b(a)}$  and  $w_{b(a)}$  of a given test batch, the parameters of the hurdle Poisson-gamma distribution representing the coliforms concentration of that pre-chill batch were calculated as:  $\lambda = \exp(\beta_0 + u_{b(a)})$ ,  $k = \exp(\alpha_0 + v_{b(a)})$  and  $w_0 = \exp(l_0 + w_{b(a)}) / (1 + \exp(l_0 + w_{b(a)}))$ . This hurdle Poisson-gamma distribution ( $\lambda$ ,  $k$ ,  $w_0$ ; Equation (17)) is the *fitted* pre-chill distribution of coliforms counts that arises from the single hurdle Poisson-gamma model. For comparison, this distribution was plotted in its cumulative form with the ranked pre-chill observations. Secondly, the parameters of the hurdle Poisson-gamma distribution for the corresponding post-chill batch were calculated as:  $\lambda' = \exp(\beta_0 + u_{b(a)} + \beta_1)$ ,  $k' = \exp(\alpha_0 + v_{b(a)} + \alpha_1)$  and  $w_0' = \exp(l_0 + w_{b(a)} + l_1) / (1 + \exp(l_0 + w_{b(a)} + l_1))$ . This hurdle Poisson-gamma distribution ( $\lambda'$ ,  $k'$ ,  $w_0'$ ) represents the *predicted* distribution of coliforms counts on the post-chill carcasses from the same test batch. This post-chill predicted distribution was compared in its cumulative form with the ranked post-chill observations corresponding to the pre-chill test batch. Notice that, while the pre-chill distribution is derived using the raw pre-chill observations by the single hurdle Poisson-gamma model (reason why the pre-chill distribution is referred to as *fitted*), the post-chill distribution is not derived using the post-chill observations but using the chilling effect estimates  $\beta_1$ ,  $\alpha_1$  and  $l_1$  (reason why this distribution is referred to as *predicted*).

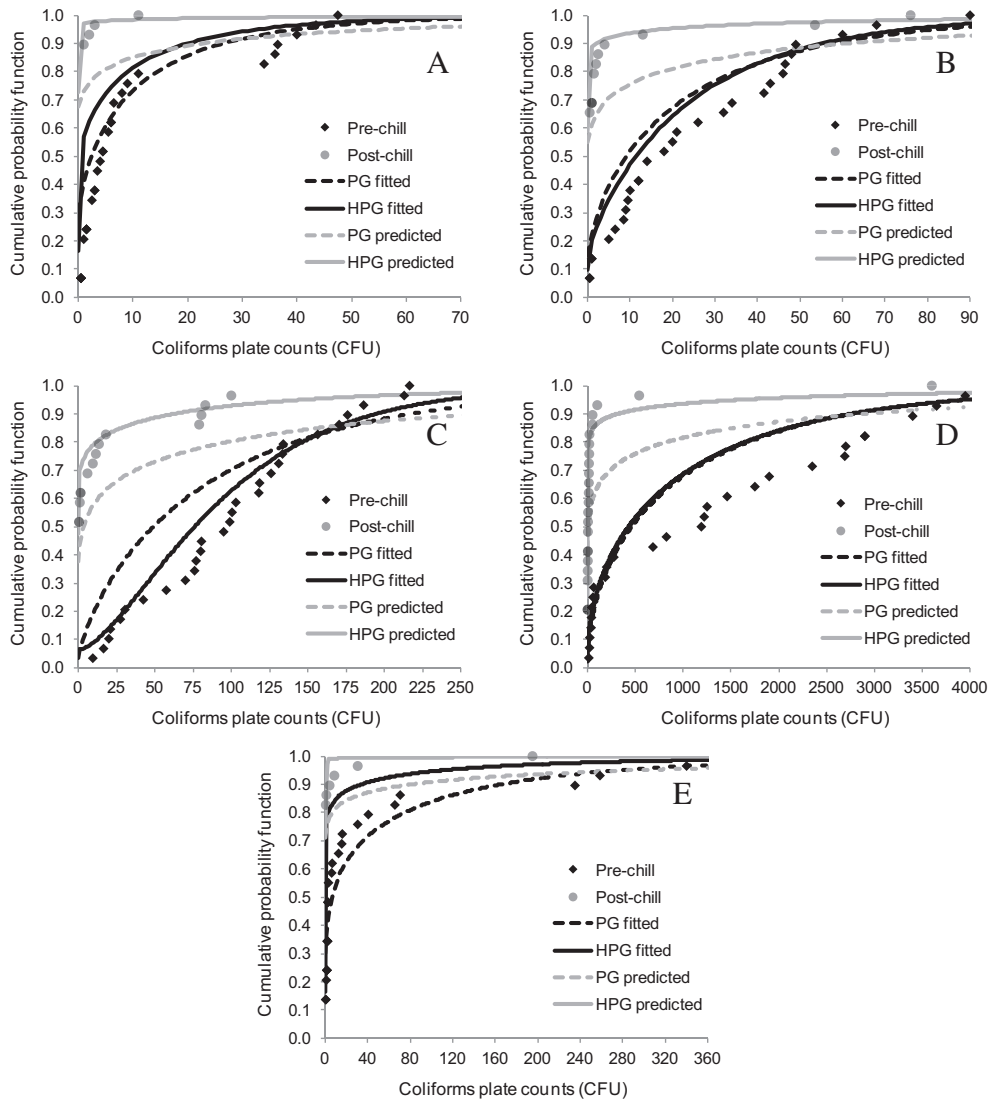
The results of the model validation for the fitted pre-chill and predicted post-chill distributions are shown in Fig. 5 for the five test



batches. In batches A and E, the fitted Poisson-gamma distributions were closer to the pre-chill observations, while in batches B and C, the hurdle Poisson-gamma distributions depicted better the observations. Interestingly, in the highly contaminated batch D, there was no observable difference between the Poisson-gamma and the hurdle Poisson-gamma distributions. Thus, we can say that, regarding the fitted distributions, the Poisson-gamma and the hurdle Poisson-gamma models had overall a comparable capacity to describe the pre-chill batches, which are batches of lower proportion of zero counts. This reinforces the previous finding that for the pre-chill data, the Poisson-gamma model could predict equally well the probability of zero counts as the hurdle Poisson-gamma model. The explanation for this is that the lower proportion of zeros found in the pre-chill batches (Fig. 3) produces a within-batch variance magnitude sufficient for the Poisson-gamma model to deal with. However, this does not always occur with the post-chill batches, where the hurdle Poisson-gamma distribution with its increased variance is by definition better equipped to deal with higher proportions of zero counts. This explains why in the post-chill batches (Fig. 5), the hurdle Poisson-gamma model predicted

more closely the observed counts than the Poisson-gamma model for the five test batches. In all test batches, the predicted (post-chill) Poisson-gamma distributions underestimated the coliforms plate counts; and for batches B, C, D and E, the within-batch probability of zero counts predicted by the Poisson-gamma was lower than the one predicted by the hurdle Poisson-gamma model (Fig. 5). Furthermore, it was generally observed that, despite such differences, both the Poisson-gamma and hurdle Poisson-gamma models can characterise batches of higher contamination (Fig. 5D; Table 1).

After this analysis of validation, it can be stated that the hurdle Poisson-gamma distribution is more suitable than the simpler Poisson-gamma to characterise the coliforms counts sampled from beef carcasses, given not only its better goodness-of-fit measure but also its better capacity to predict the effect of chilling in a batch as well as the within-batch proportion of zero counts. Although the common ANOVA based on normality constitutes a more straightforward procedure to analyse microbial data, special care should be taken as it cannot be used in all cases, certainly not when microbial counts are low and with many non-detections. In such cases, the Poisson-gamma model and its zero-modified variants offer an



**Fig. 5.** Cumulative distributions of the coliforms plate counts from post-chill beef carcasses as predicted by the Poisson-gamma (PG) and the hurdle Poisson-gamma (HPG) model for the five batches selected for model validation. The fitted distributions for the respective pre-chill group are also shown. Markers indicate ranked observations from the pre-chill and post-chill carcasses.

alternative framework to perform descriptive and inferential statistics. Although it is apparently more complex, the count data regressions can be easily performed using most statistical packages. The advantages of the Poisson-gamma approach are many (Table 1), namely: (i) the random sampling process of the bacteria taken from the homogenate is considered; (ii) although it is more adequate for low counts, still it can represent equally well batches that have occasionally higher microbial concentrations; (iii) it takes in the zero counts in a natural way, so there is no need to replace them with limit of quantification, nor fitting censored normal distributions without the certainty that the distribution is really normal; (iv) it produces estimates of microbial concentration in arithmetic mean, and not in geometric mean (for a detailed explanation on the suitability of arithmetic means in sampling plans, see Gonzales-Barron et al., 2012); (v) there is a distinction between the observed plate count data and the unknown true microbial concentration; whereas in the common microbial data treatment, the true concentration is directly (back-) calculated from the observation; and (vi) the zero-inflated variant, although not demonstrated in this study, has the capability to distinguish between true absence of contamination and false non-detections due to sampling.

#### 4. Conclusions

This work introduced an alternative Poisson-gamma regression to conduct inferential statistics on plate count data from microorganisms of low counts in foods. This represents a neater approach as it directly uses the observed number of colony forming units and the zero counts, making a distinction between observed distribution (Poisson-gamma for the plate counts) and true distribution (gamma for the true microbial concentration); and therefore there is no need to employ either the ordinary log-transformed data (log CFU per g or cm<sup>2</sup>) or a censored model. As an illustration, to estimate the effect of chilling on the concentration of total coliforms on beef carcasses, Poisson-gamma and zero-modified Poisson-gamma models were appraised with a series of random-effects variants in order to extract any variability in microbial mean concentration, dispersion and/or proportion of zero counts among batches. A hurdle Poisson-gamma model with three random effects fitted the coliforms data better than a Poisson-gamma model with two random effects. However while their performance for the batches of pre-chill carcasses was comparable, it was mostly for the batches of post-chill carcasses that the superiority of the hurdle model was noticeable, as evidenced by its better capacity to predict both the higher proportion of zero counts and the effect of chilling at batch level. This occurs because the hurdle Poisson-gamma is mathematically equipped to deal with greater variance than the Poisson-gamma model. The hurdle model proved that the chilling operation decreases significantly the recovery of the total coliforms, and the pre-chill to post-chill counts reduction in a batch is by an average factor of 2.2 (95% CI: 2.15–2.24). The model also estimated that, on average, the beef carcasses sampled from the Irish abattoirs before chilling presented a concentration of coliforms of 1.00 CFU/cm<sup>2</sup> with an expected probability of zero counts of 7%, and after chilling, this concentration dropped significantly to 0.48 CFU/cm<sup>2</sup> with an increased expected probability of non-detections of 49%. By assessing the random effects, the Poisson-gamma model also suggested that the lower the coliforms concentration (i.e., the better the hygiene), the greater the beneficial effect that can be attained

by chilling to reduce such contamination. The Poisson-gamma framework offers the additional advantage of estimating the mean microbial concentrations straight on a CFU arithmetic scale, which is of more significance for low microbial counts. Finally, these results can find applications in the derivation of risk management measures such as the design of more efficient sampling plans for hygiene indicators that take into account the batch-to-batch variability.

#### Acknowledgements

Dr. Gonzales-Barron wishes to acknowledge the financial support provided by the Portuguese Foundation for Science and Technology (FCT), first through the postdoctoral fellowship SFRH/BPD/87670/2012, and later on, through the award of a five-year Investigator Fellowship (IF) in the mode of Development Grants (IF/00570).

#### References

- Busschaert, P., Geeraerd, A. H., Uyttendaele, M., & Van Impe, J. F. (2011). Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. *Food Microbiology*, 28, 712–719.
- Commeau, N., Parent, E., Delignette-Muller, M. L., & Cornu, M. (2012). Fitting a lognormal distribution to enumeration and absence/presence data. *International Journal of Food Microbiology*, 155, 146–152.
- Gill, C., Deslandes, B., Rahn, K., Houde, A., & Bryant, J. (1998). Evaluation of the hygienic performances of the processes for beef carcass dressing at 10 packing plants. *Journal of Applied Microbiology*, 84, 1050–1058.
- Gonzales-Barron, U., Zwietering, M. H., & Butler, F. (2013). A novel derivation of a within-batch testing regime based on a Poisson-gamma model characterising low microbial counts in foods. *International Journal of Food Microbiology*, 161(2), 84–96.
- Gonzales-Barron, U., Lenahan, M., Sheridan, J. J., & Butler, F. (2012). Use of a Poisson-gamma model to assess the performance of the EC process hygiene criterion for *Enterobacteriaceae* on Irish sheep carcasses. *Food Control*, 25, 172–183.
- Gonzales-Barron, U., & Butler, F. (2011a). A comparison between the discrete Poisson-Gamma and Poisson-Lognormal distributions to characterise microbial counts in foods. *Food Control*, 22, 1279–1286.
- Gonzales-Barron, U., & Butler, F. (2011b). 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.
- Gonzales Barron, U., Kerr, M., Sheridan, J. J., & Butler, F. (2010). Count data distributions and their zero-modified equivalents as a framework for modeling microbial data with a relatively high occurrence of zero counts. *International Journal of Food Microbiology*, 136, 268–277.
- Hinde, J., & Demetrio, C. (1998). Overdispersion: models and estimation. *Computational Statistics and Data Analysis*, 27, 151–170.
- Jongenburger, I., Reij, M. W., Boer, E. P. J., Zwietering, M. H., & Gorris, L. G. M. (2012). Modelling homogenous and heterogeneous microbial contaminations in a powdered food product. *International Journal of Food Microbiology*, 157, 35–44.
- Lasta, J., Rodriguez, R., Zanelli, M., & Margaria, C. A. (1992). Bacterial count from bovine carcasses as an indicator of hygiene at slaughtering places: a proposal for sampling. *Journal of Food Protection*, 54, 271–278.
- Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics*, 33, 341–365.
- Mussida, A., Vose, D., & Butler, F. (2013). Efficiency of the sampling plan for *Cronobacter* spp. assuming a Poisson-lognormal distribution of the bacteria in powder infant formula and the implications of assuming a fixed within- and between-lot variability. *Food Control*, 33, 174–185.
- Mussida, A., Gonzales-Barron, U., & Butler, F. (2013). Effectiveness of sampling plans by attributes based on mixture distributions characterising microbial clustering in food. *Food Control*, 34(1), 50–60.
- Peltier, M. R., Wilcox, C. J., & Sharp, D. C. (1998). Application of the Box–Cox data transformation to animal science experiments. *Journal of Animal Science*, 76, 847–849.
- Pouillot, R., Hoelzer, K., Chen, J., & Dennis, S. (2013). Estimating probability distributions of bacterial concentrations in food based on data generated using the most probable number (MPN) method for use in risk assessment. *Food Control*, 29, 350–357.