



**HAL**  
open science

# Parameter estimation for the distribution of single cell lag times

J. József Baranyi, Susan M. George, Z. Zoltán Kutalik

► **To cite this version:**

J. József Baranyi, Susan M. George, Z. Zoltán Kutalik. Parameter estimation for the distribution of single cell lag times. *Journal of Theoretical Biology*, 2009, 259 (1), pp.24. 10.1016/j.jtbi.2009.03.023 . hal-00554596

**HAL Id: hal-00554596**

**<https://hal.science/hal-00554596>**

Submitted on 11 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Author's Accepted Manuscript

Parameter estimation for the distribution of single cell lag times

József Baranyi, Susan M. George, Zoltán Kutalik

PII: S0022-5193(09)00132-5  
DOI: doi:10.1016/j.jtbi.2009.03.023  
Reference: YJTBI5508

To appear in: *Journal of Theoretical Biology*

Received date: 19 August 2008  
Revised date: 26 February 2009  
Accepted date: 13 March 2009

Cite this article as: József Baranyi, Susan M. George and Zoltán Kutalik, Parameter estimation for the distribution of single cell lag times, *Journal of Theoretical Biology* (2009), doi:[10.1016/j.jtbi.2009.03.023](https://doi.org/10.1016/j.jtbi.2009.03.023)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

**Parameter estimation for the distribution of single cell lag times**

József Baranyi<sup>1\*</sup>, Susan M. George<sup>1</sup>, Zoltán Kutalik<sup>2</sup>

5 <sup>1</sup> *Institute of Food Research, Norwich Research Park, Norwich NR4 7UA, United Kingdom;*

<sup>2</sup> *Department of Medical Genetics, Rue de Bugnon 27, CH-1005 Lausanne, Switzerland*

Running title: Distribution of single cell lag times

10 \* Corresponding author: [jozsef.baranyi@bbsrc.ac.uk](mailto:jozsef.baranyi@bbsrc.ac.uk)

Accepted manuscript

**ABSTRACT**

**In Quantitative Microbial Risk Assessment, it is vital to understand how lag times of individual cells are distributed over a bacterial population. Such identified distributions can be used to predict the time by which, in a growth-supporting environment, a few**  
5 **pathogenic cells can multiply to a poisoning concentration level.**

**We model the lag time of a single cell, inoculated into a new environment, by the delay of the growth function characterizing the generated subpopulation. We introduce an easy-to-implement procedure, based on the method of moments, to estimate the parameters of the distribution of single cell lag times. The advantage of the method is especially apparent for**  
10 **cases where the initial number of cells is small and random, and the culture is detectable only in the exponential growth phase.**

Keywords: single cell, lag time, stochastic process, parameter estimation, probability distribution

## Introduction

Food poisoning outbreaks can be triggered by contamination with even a few pathogenic cells, if they are able to grow in the food to reach an infective dose. Quantitative Microbial Risk Assessment studies frequently need to estimate the probability that a few contaminating cells multiply to a population level above a tolerance limit. To analyse this, one needs to identify the probability distribution of the lag time of single cells.

**Traditional modelling of lag.** If a homogeneous bacterial population is inoculated into a growth-supporting environment, then it will grow exponentially, possibly after a lag period. If  $t$  and  $N(t)$  denote the time and the number of cells, respectively, then the rate  $\frac{d \ln N(t)}{dt}$  is called the (instantaneous) specific growth rate of the population. The maximum of this rate, denoted by  $\mu$ , is characteristic of the species and the growth environment. If, from the onset of the exponential phase, the  $N(t)$  curve can be described by a delayed growth model of the form  $f(t-\lambda)$  (where  $f$  can be such as the Malthusian, logistic, etc. models, with  $\mu$  specific rate parameter) then  $\lambda$  is called the lag time (Baranyi and Roberts, 1994). Estimating the lag is vital for food safety modelling.

The lag phase has mainly been studied at the population level (Baty and Delignette-Muller, 2004), but lately more and more studies concentrate on single cell lag times (Augustin, 1999; Elfving *et al.* 2004; Francois *et al.* 2005; Kutalik *et al.* 2005; Prats *et al.* 2008; Pin and Baranyi, 2008). The reason for this is that, in practice, pathogenic contamination is frequently caused by a few cells only.

Apply the above definition to the case when the inoculum consists of  $N_0$  cells. The cell population,  $N(t)$ , sooner or later grows exponentially, so its logarithm can be modelled by the bi-phasic

$$y(t) = \ln N(t) = \begin{cases} \ln N_0 & t \leq \lambda \\ \ln N_0 + \mu \cdot (t - \lambda) & t > \lambda \end{cases}$$

5 function.

Suppose that a single cell when inoculated in a new environment first goes through a “physiological lag” period of the random length  $\tau$ . Suppose that the daughter cells are already in the exponential phase and the generation times ( $T_g$ : the time from birth to the next division) of all the future daughter cells are independent, identically distributed random variables. This means

10 that the first division happens at the time  $\tau + T_g$ , where the two independent variables follow two different distributions. Construct the above bi-phasic model for this stochastic birth process and denote the obtained lag by  $\lambda = L_g(N_0)$ , indicating that it depends on the number of initial cells, too. For shorthand, we will also use  $L_g$  for  $L_g(1)$ . Obviously  $L_g$  is also a random variable, depending primarily on  $\tau$ , but also on the subsequent generation times denoted by  $T_g$  in what

15 follows. We will call  $L_g$  the “geometrical” single cell lag time because of its definition (Fig.1). As  $N_0$  increases,  $L_g(N_0)$  converges to a limit value, a “population-lag” parameter. This convergence was analysed by Baranyi (1998).

When interpreting the single cell lag time as above, it is important to see that  $L_g = L_g(1)$  is not a feature of the original single cell only, but that of the sub-population generated by the

20 single cell. The distribution of  $L_g$  is a function of the distributions of  $\tau$  and  $T_g$ . It is not easy (if possible at all) to calculate the distribution function of  $L_g$  by knowing those of the other two. One can expect that if  $\tau$  is dominantly bigger than  $T_g$ , then the distribution of  $L_g$  should be

close to that of  $\tau$ . However, if  $\tau$  is small, then the distributions of the two lag times can be very different. For demonstration, consider the simulated, single-cell-generated growth curves in Fig.2. The physiological lag was zero ( $\tau \equiv 0$ ), while the resulting  $L_g$  geometrical lag times are dispersed around zero, with a spread depending on the subsequent generation times.

5           The reason why we study  $L_g$  instead of  $\tau$  is that the latter variable, the physiological lag, is not observable for a single cell. For simulation studies, we need to use  $\tau$ , while for predictions it is more pragmatic to use  $L_g$ . Namely, by means of the  $L_g$ -delayed linear growth model, it is easy to predict the time by which the bacterial population would reach a given (e.g. harmful) level; this is exactly what risk assessors need.

10           **Detection time.** In practice, it is not feasible to follow the division of individual cells with automated measurements. Let  $y_{det} = \ln(N_{det})$  be the level at which a bacterial culture can be detected, for example, by measuring its turbidity in a liquid medium. Suppose that  $y_{det}$  is in the exponential phase and let  $T_{det}(N_0)$  denote the detection time at which the subpopulation reaches this  $y_{det}$  threshold value. Then the detection time is a shifted version of the  $L_g(N_0)$  random  
15           variable. However, as this value depends on the number of initial cells, which is unknown, it is not easy to infer to the  $L_g$  single cell lag times.

          Our main objective is to use detection time data to give an accurate estimation procedure for the parameters of the single cell lag time distribution, taking also into account that the initial  
20           count is random. Though maximum-likelihood principles serve as a standard for such parameter estimation problems, the use of these methods require more sophisticated software applications than what is commonly available on data logging computers in microbiology laboratories. Our

new moments-based method will be suitable for common applications, such as MS Excel spreadsheets.

We will also discuss some practical implications regarding Quantitative Microbial Risk Assessment.

5

### Practical considerations and measurements

**Simulation.** Suppose that the physiological lag time of a single cell,  $\tau$ , is Gamma distributed and the  $T_g$  generation times of the subsequent daughter cells are independent and identically distributed, also following a Gamma-distribution, but with parameters independent of the distribution for  $\tau$ . The original cell divides first at the  $\tau + T_g$  time; then the time from birth to division (single cell generation time), needed by any daughter cell of the subsequent generations, follow the same, the second Gamma distribution. Simulating the growth curves as above and fitting  $L_g$  for each case by the bi-phasic growth curve defined earlier, we could see that  $L_g$  was reasonably close to a shifted Gamma distribution, as demonstrated in Fig.3.

15 Another source of variability could be the specific growth rate of the single-cell generated population (the slope of the fitted bi-phasic function). As the simulation studies showed, the variance of the specific growth rate is negligible compared to the variance of  $L_g$  and the first few generation times (Métris et al. 2003). This can also be seen intuitively, since the specific rate is determined by many cells, therefore its variance is much less than that of the  
20 division time of a single cell. Hence we will consider the specific growth rate as a parameter, not as a random variable.

**Experimental data.** First (Case A), the detection times of a *Listeria innocua* culture (Métris et al. 2006) were used to test the new method. The detection times refer to the times

when the turbidity as measured by Bioscreen C plate reader (Labsystems, Helsinki, Finland) reached a fixed level (optical density, OD=0.11) which was shown to be equivalent to ca  $10^{7.7}$  cell/ml concentration.

In the second experiment (Case B), *L. innocua* was inoculated in the same way, but after receiving a mild heat shock, so the lag was significantly longer.

The maximum specific growth rate of the organism was measured in an independent experiment, by means of traditional plate count technique. The parameter, at 22°C, was estimated as  $\mu = 0.45 \text{ h}^{-1}$  from the plot of colony forming units against time, using the in-house curve-fitting program DMFit, available at [www.ifr.ac.uk/safety/DMFit](http://www.ifr.ac.uk/safety/DMFit).

**Random initial counts.** We evaluated the distribution of the "time to turbidity" values (detection times,  $T_{det}$ ) as measured by the Bioscreen C, which reads the OD of the wells of a microtitre plate. If the initial number of cells in a well is exactly one and the detection level is in the exponential phase, then the difference between the detection time and  $L_g$  depends only on the specific growth rate of the organism, which is reasonably constant. So  $T_{det}$  differs from the  $L_g$  single cell lag time only by an additive constant; otherwise their distribution is the same.

There is no easy way to sort single cells into the wells of the Bioscreen microtitre plates. Methods available include sorting by flow cytometer, which is expensive and can introduce bias during the process, e.g. if the cells are sorted according to size. Instead cultures are diluted to a level such that a sample in a well should contain only a few cells. With a sufficiently high dilution factor, most wells will receive zero or one cell. This technique has been applied to evaluate the variation of the lag times of single cells (Augustin (1999) and Francois (2005)). The disadvantage is that many wells will be empty and, for a statistically robust distribution

estimation, it is desirable to have as many positive wells as possible, a minimum of about one hundred (Bacanova, 2004).

### Estimation procedures

5 **Estimating the average initial counts per well.** Cell cultures were obtained after successive dilutions, and placed into  $W$  number of wells ( $W = 200$  for the microtitre plates of the Bioscreen). Hence, the initial number of cells ( $N_0$ ) in a well follows the Poisson distribution, i.e.

$$P_k = P(N_0 = k) = \exp(-\rho) \frac{\rho^k}{k!} \quad (k=0, 1, \dots) \quad (1)$$

where the expected value of  $N_0$  is  $E(N_0) = \rho$ . The fraction of empty wells can be used to estimate  $\rho$  by  $\hat{\rho} = -\ln(W_0 / W)$ , where  $W_0$  is the number of negative (empty) wells.

**A linear method relating detection time and single cell lag time.** Consider the delayed stochastic birth process model as defined above. At the time point  $t$  in the exponential phase, the natural logarithm of the population size in a well can be estimated from

$$y(t) = \ln N(t) = \ln(N_0) + \mu (t - L_g(N_0)) \quad \text{for } t > L_g(N_0) \quad (2)$$

15 where  $L_g(N_0)$  is the geometrical lag in the well, and  $\mu$  is the specific growth rate of the population. We assume that, for single cells,  $L_g = L_g(1)$  can be well described by a shifted Gamma distribution, *i.e.* for a  $T_{\text{shift}}$  parameter,  $L_g - T_{\text{shift}}$  is Gamma distributed (Fig.3). In order to create a simple bi-phasic model for the growth curve, the  $T_{\text{shift}}$  parameter compensates for the randomness of the first few division times. Alternatively, if  $T_{\text{shift}}$  is significantly different from  
20 zero, then it can be interpreted as a ‘compulsory repair time’ for a period after injury during which the probability of division is zero for any cell. An example will be shown for this later (Case B), with heat-shocked cells.

The linear method uses the above equation with  $y(t)=y_{det}$  and transforms the  $T_{det}(N_0)$  detection times into individual lag times by using the estimation for  $\rho$  :

$$L_g^{(LIN)} = T_{det}(N_0) - \frac{y_{det} - \ln(\hat{\rho})}{\mu} \quad (3)$$

Here,  $L_g^{(LIN)}$  denotes the linear estimation for the  $L_g$  single cell lag time. (As mentioned,  $y_{det}$  is the natural logarithm of the cell concentration at the detection level). This linear method substitutes the random variable,  $N_0$ , by an estimation of its mean, therefore the variability of  $L_g$  comes only from that of the observed detection times. If  $N_0$  was fixed at 1, then the difference between  $T_{det}$  and  $L_g$  would be only the additive constant  $y_{det}/\mu$ .

**The new ‘moments-based’ method.** Define

$$\alpha(N_0) = \exp(-\mu L_g(N_0)) \quad (4)$$

as the *physiological state* for the population generated by  $N_0$  initial cells in a well (Baranyi, 1998). This variable shows a certain affinity of the cells to the new environment. If it is 1, then no need for adjustment; if zero, then the lag is infinitely long.

Let  $S_\alpha(N_0)$  denote the sum of the physiological states of the one-cell generated subpopulations generated by the initial cells:

$$S_\alpha(N_0) = \sum_{j=1}^{N_0} \alpha_j(1) \quad (5)$$

As Baranyi (1998) proved it, the physiological state of a population is the arithmetical average of the physiological states of the constituent subpopulations:

$$\frac{S_\alpha(N_0)}{N_0} = \alpha(N_0) \quad (6)$$

Using the relationship

$$y_{det}(N_0) = \ln(N_0) + \mu [T_{det}(N_0) - L_g(N_0)] \quad (7)$$

we obtain

$$S_{\alpha}(N_0) = e^{y_{\text{det}}(N_0) - \mu T_{\text{det}}(N_0)} \quad (8)$$

This idea was used by Métris *et al.* (2006). Here we develop the method further. We utilise the fact that  $N_0$  is Poisson-distributed and we can get explicit formulae for the  $\mathbf{M}_1$ ,  $\mathbf{M}_2$ ,  $\mathbf{M}_3$  moments of  $S_{\alpha}(N_0)$ . The derivation can be found in the Appendix; here we show the main result only.

Let the  $L_g - T_{\text{shift}}$  Gamma distributed variable have the scale and shape parameters  $\theta$  and  $\beta$ , respectively. Then the moments of  $S_{\alpha}$  are given by the formulae A10-A12 of the Appendix. Make them equal to the empirical moments of  $S_{\alpha}$  :

$$e^{-\mu T_{\text{shift}}} \rho (1 + \mu \theta)^{-\beta} = \frac{\sum_{j=1}^n e^{y_{\text{det}}^{(j)} - \mu T_{\text{det}}^{(j)}}}{n} \quad (9)$$

$$10 \quad e^{-2\mu T_{\text{shift}}} (\rho (1 + 2\mu \theta)^{-\beta} + \rho^2 (1 + \mu \theta)^{-2\beta}) = \frac{\sum_{j=1}^n e^{2(y_{\text{det}}^{(j)} - \mu T_{\text{det}}^{(j)})}}{n} \quad (10)$$

$$e^{-3\mu T_{\text{shift}}} (\rho (1 + 3\mu \theta)^{-\beta} + 3\rho^2 (1 + \mu \theta)^{-\beta} (1 + 2\mu \theta)^{-\beta} + \rho^3 (1 + \mu \theta)^{-3\beta}) = \frac{\sum_{j=1}^n e^{3(y_{\text{det}}^{(j)} - \mu T_{\text{det}}^{(j)})}}{n} \quad (11)$$

where  $j=1 \dots n$  numbers the observations,

As mentioned,  $\rho$  can be estimated by the number of empty wells,  $W_0$ , out of the total wells,  $W$ . The specific rate is calculated from different data, so three equations remain with 3 parameters,  $T_{\text{shift}}$ ,  $\beta$  and  $\theta$ . The estimate for  $L_g$  will be

$$L_g^{(M)} = T_{\text{shift}} + \beta \theta \quad (12)$$

with the fitted parameters.

## Results

The linear method is reliable only if the average number of initial cells is very low, so the majority of the wells contain one cell only. We demonstrate below that there can already be a significant difference between the true  $L_g$  values and the linear estimate produced by Eq. (3)

5 when the average initial number of cells in a well is between 2 and 3.

We simulated two cases for the Gamma parameters of the physiological single cell lag times:

Simulation\_1:  $\tau \sim \Gamma(2,1)$  [  $E(\tau) = 2h$ ,  $\sigma(\tau) = 1.4h$  ]

Simulation\_2:  $\tau \sim \Gamma(2,10)$  [  $E(\tau) = 20h$ ,  $\sigma(\tau) = 14h$  ]

10 The coefficient of variation values of these Gamma distributed variables are the same and, from observed data, realistic (Kutalik et al. 2005). The difference between them is the ratio of their scale parameters to their means and to the subsequent specific growth rate, which is fixed at  $\mu = 0.45 \text{ (h}^{-1}\text{)}$ .

Assume that the cells are inoculated into 200 wells, with  $E(N_0) = \rho = 2.5$  cells per well.

15 Simulate the theoretical outcome considering only wells with less than 5 cells (88% of the non-empty wells). As mentioned, the  $\sigma$  standard deviation of the  $T_{det}(N_0)$  detection times will be the same as that of the  $L_g(N_0)$  population lag times. The first four rows of Table 2 refers to  $L_g(k)$ , for wells containing exactly  $k = 1, 2, 3, 4$  cells (in  $P_1 P_2 P_3 P_4$  proportions, respectively). The last row refers to their convoluted distribution, which would be the result from the linear method,

20 apart from the slight under-representation because of the missing 12% of the wells with  $>4$  initial counts. With Simulation\_1,  $\sigma(L_g(N_0)) > \sigma(L_g(1))$ , while with Simulation\_2, the result is the opposite. In reality, the detection time variability is equal to the variability of the  $L_g(N_0)$  lag time, where  $N_0$  is random. As the table show, it can significantly differ from the variability of

single cell lag times. If the subsequent generation times in the exponential phase are comparable to the single cell lag times (Simulation\_1), then the difference is around 10%; however if the lag time is much longer than the generation times (Simulation\_2), then the relative difference between the two standard deviations can be as high as 40-50%.

5

It is evident that the higher the Poisson parameter,  $\rho$ , the less it is true that the linear method estimates the distribution of the single cell lag time. However, we demonstrate below that the new method is also suitable for  $\rho > 1$  cases. We applied it to two experimental data sets obtained with *L. innocua*.

10 **Case A.** The first example is the same as used by Métris *et al.* (2006). Out of the  $W=200$ , 146 were positive, so the average initial number of cells per well was  $\hat{\rho} = -\ln(54/200) = 1.31$ . The specific rate was fixed at the value  $\mu=0.15 \text{ h}^{-1}$  as reported in the original paper. By means of the formulae (6) and (8), the physiological state of the cell population could be calculated by using the estimate for the mean of  $N_0$  :

$$15 \quad \alpha = \frac{\sum_{j=1}^n e^{\gamma_{\text{det}}^{(j)} - \mu \cdot T_{\text{det}}^{(j)}}}{\rho \cdot n} = 0.019$$

The standard deviation of the detection times was  $\sigma(T_{\text{det}}(N_0))=12.5 \text{ h}$  while our method resulted in  $\sigma(L_g(1))=32.23 \text{ h}$  for the single cell lag times.  $T_{\text{shift}}$  was insignificant, and its value could be fixed at 0. After fitting the  $\beta$ ,  $\theta$  parameters, the average single cell lag time was calculated as  $T_{\text{shift}} + \beta\theta = 55.5 \text{ h}$ . Note that the lag of the population was  $-\ln(\alpha)/\mu = 26.5$ , about half of the mean single cell lag time. These results coincide with those of Métris *et al.* (2006), because the  $T_{\text{shift}}$  parameter was zero.

20

**Case B.** In the second example, after a mild heat shock, cells were inoculated into wells as described earlier. In this case, 189 wells out of  $W=198$  were positive, so the average initial counts per well was  $\hat{\rho} = -\ln(9/198)=3.09$ . Equipped with the linear method only, experiments with such a high average initial count would normally be discarded as unsuitable to measure the variability of single cell lag times. Using our method, we were able to estimate that the standard deviation of the detection times was  $\sigma(L_g(N_0)) = \sigma(T_{det}(N_0))=15.52$  h and that of the individual lag times was  $\sigma(L_g(1))=23.6$  h. Here, the  $T_{shift}$  parameter was significant, its estimate was 6.6 hours. The physiological state and the lag of the population was ca.  $10^{-4}$  and 20.6 h, respectively, the latter being less than half of the mean of the single cell lag times that was estimated as 49h. These result could not have been obtained without introducing the  $T_{shift}$  parameter.

## Discussion

**Developing the new method.** Automated readers are used increasingly to follow the physico-chemical properties of bacterial cultures. They are not suitable to detect low cell concentrations, but able to provide for example the times by which the cultures reach a given concentration level. It is desirable to process such data with methods that can easily be attached to the data acquisition software. We gave such a method in this paper, in order to identify the distribution of single cell lag times. The novelty of our method is due to the explicit expressions that can be obtained for the moments of the “physiological state” random variable.

**Estimating initial counts by the number of negative wells.** There is no practical alternative to estimating the Poisson parameter more efficiently than by means of the number of negative wells. The level of initial counts in a well is much too low for direct detection. Another way would be by means of the original cell concentration before the dilution series. However,

estimating the *per*-well cell concentration this way is much more inaccurate than the one obtained by the number of negative wells. Nonetheless the calculation is useful to estimate the number of necessary dilutions to arrive at the concentration of ca 1-3 cells per well.

**Numerical stability.** The estimation procedure requires solving the three equations (9)-(11) for  $T_{\text{shift}}$ ,  $\theta$  and  $\beta$ . The system of equations is highly non-linear and a common solver tends to find extreme values of the parameters, when all terms disappear. To avoid this, we instead minimize the sum of squares of the differences between the two sides of the equations. If that minimum is sufficiently small, then we accept the obtained parameters as solutions.

Even in this case, a global minimisation algorithm can still cause problems because of the extreme parameter values. A local minimum is needed which requires good initial estimations of the parameters. This is quite demanding, and probably no better way exists than to examine the sum of squares on a feasible grid of the parameters. If no local minimum is found then the data are not compatible with the basic assumptions of the method (Poisson distribution for  $N_0$  and shifted Gamma distribution for  $L_g$ ).

As mentioned, and shown in Table 2, the distribution of the single cell lag times can be either wider or narrower than the distribution of the  $T_{\text{det}}(N_0)$  detection times, depending mainly on  $N_0$  and the relation between the lag and the growth rate. With  $N_0$  increasing, the estimation for  $\rho$  is less accurate, as are the estimates for the moments. However, when cells grow together in a well, then their population lag has a smaller variability than that of a single cell lag time. This increases the accuracy of the estimation procedure.

In Simulation\_2,  $\sigma(L_g(1))$  was bigger than  $\sigma(L_g(N_0))$ . This relatively wide distribution of the single cell lag times means that if cells grow together then the population lag of  $N_0 > 2$  cells is dominantly determined by the subpopulation of the first dividing initial cells (by the time the

slowest cells divide, the others are in the exponential phase). In such cases, it is crucial to know the specific growth rate *a-priori*. Here we use viable count growth curves for its accurate estimation, but Baranyi and Pin (1999) showed a method to also measure the specific rate as well as determining the lag time distribution using the Bioscreen.

5 An obvious alternative to our moments-based method is one based on the Maximum Likelihood principle. We compared our results with those generated by an ML program and they were practically identical. However, the computational cost (time, resources) of the ML method is much higher than that of the method of moments. Note that even if software packages are available and can provide ML-results, our method is useful to give initial parameter values for  
10 the ML-iteration.

**Application of single cell lag variation in Microbial Risk Assessment.** Our results can be used in simulations for Quantitative Microbial Risk Assessment. A framework for this is demonstrated in Fig. 4: Suppose that the initial number of cells on a portion of food is a random variable following a distribution that we call "vertical" for obvious reasons. This distribution is  
15 the result of the history of the cells. Let  $y_0 = \ln N_0$  denote the natural logarithm of the initial counts. Each cell goes through a lag period,  $L_g$ , which follows a distribution we call "horizontal", then, during the exponential phase, the log number increases linearly. The question a risk assessor is most interested in is the probability that the contamination level at the time of consumption,  $t_c$ , is above a certain threshold. The "vertical" distribution of this contamination level,  $y(t_c)$  is affected  
20 by the "vertical" distribution of  $y_0$ , the (horizontal) distribution of  $L_g$  and the specific growth rate that we consider constant here. The distribution of the bacterial population level can seldom be calculated explicitly by algebraic operations but Monte-Carlo simulations provide a straightforward solution to study them. For those simulations, the underlying distributions should

be measured as accurately as possible. Unfortunately, the region of interest is at relatively low cell concentrations, where automated measurements, which are necessary for sufficient data to obtain robust distribution estimates, are not available. At high concentrations, however, assuming that the detection level  $y_{det}$  is still in the exponential phase, one can use the "horizontal" distribution of the  $T_{det}$  detection times instead of the  $L_g$  lag times, as we did in this paper.

Generally, only vertical distributions are measurable in food, while broth cultures are suitable to determine both horizontal and vertical distributions. The transformation between them can be made by means of the specific growth rate, since the two types of distributions are rotated by  $90^\circ$ .

The lag time distribution has significance in predicting contamination levels only if the estimation of the specific growth rate is accurate. When this is the case, the ideas shown in this paper can be built into Monte-Carlo simulations as the "computational engine" behind Microbial Risk Assessment programs.

## ACKNOWLEDGMENTS

This work was partially funded by the EU Programme Quality of Life and Management of Living Resources, project no. QLK1-CT-2001-01145 (BACANOVA) and Biotechnology and Biological Research Council (grant number 42266A).

**Appendix**

Calculate the first three moments of the random variable  $e^{-\mu\tau}$  around zero, where  $\tau$  is Gamma distributed with scale parameter  $\theta$  and shape parameter  $\beta$ :

$$m_1 = (1 + \mu\theta)^{-\beta} \quad (\text{A1})$$

$$5 \quad m_2 = (1 + 2\mu\theta)^{-\beta} \quad (\text{A2})$$

$$m_3 = (1 + 3\mu\theta)^{-\beta} \quad (\text{A3})$$

Suppose that the  $L_g = L_g(1)$  geometrical lag has a (constant) time shift parameter, i.e.  $L_g - T_{\text{shift}}$  follows the above Gamma distribution. Therefore, the respective moments of the physiological

10 state  $\alpha = e^{-\mu L_g}$  are

$$M_1 = \mathbf{E}(e^{-\mu L_g}) = e^{-\mu T_{\text{shift}}} (1 + \mu\theta)^{-\beta} \quad (\text{A4})$$

$$M_2 = \mathbf{E}(e^{-2\mu L_g}) = e^{-2\mu T_{\text{shift}}} (1 + 2\mu\theta)^{-\beta} \quad (\text{A5})$$

$$M_3 = \mathbf{E}(e^{-3\mu L_g}) = e^{-3\mu T_{\text{shift}}} (1 + 3\mu\theta)^{-\beta} \quad (\text{A6})$$

The moments of the compound random variable  $S_\alpha(N_0)$  can be calculated by means of the

15 moments of  $N_0$  and  $\alpha_j$ :

$$\mathbf{M}_1(S_\alpha) = \mathbf{E}\left(\sum_{j=1}^{N_0} \alpha_j\right) = \mathbf{E}(N_0) \cdot M_1 = \rho M_1 \quad (\text{A7})$$

$$\mathbf{M}_2(S_\alpha) = \mathbf{E}\left(\sum_{j=1}^{N_0} \alpha_j^2\right) = \mathbf{E}(N_0) \cdot M_2 + (\mathbf{E}(N_0) M_1)^2 = \rho M_2 + \rho^2 M_1^2 \quad (\text{A8})$$

$$\mathbf{M}_3(S_\alpha) = \mathbf{E}\left(\sum_{j=1}^{N_0} \alpha_j^3\right) = \rho M_3 + 3\rho^2 M_1 M_2 + \rho^3 M_1^3 \quad (\text{A9})$$

Combining A1-A3 and A7-A9:

$$\mathbf{M}_1(S_\alpha) = e^{-\mu T_{\text{shift}}} \rho (1 + \mu\theta)^{-\beta} \quad (\text{A10})$$

$$\mathbf{M}_2(S_\alpha) = e^{-2\mu T_{\text{shift}}} (\rho (1 + 2\mu\theta)^{-\beta} + \rho^2 (1 + \mu\theta)^{-2\beta}) \quad (\text{A11})$$

$$\mathbf{M}_3(S_\alpha) = e^{-3\mu T_{\text{shift}}} (\rho (1 + 3\mu\theta)^{-\beta} + 3\rho^2 (1 + \mu\theta)^{-\beta} (1 + 2\mu\theta)^{-\beta} + \rho^3 (1 + \mu\theta)^{-3\beta}) \quad (\text{A12})$$

5

## REFERENCES

- Augustin, J.-C., Rosso, L., Carlier, V.,** 1999. Estimation of temperature dependent growth rate and lag time of *Listeria monocytogenes* by optical density measurements. *J. Microbiol. Methods*, **38**: 137-146.
- BACANOVA,** 2005. Final report on the EU-QLRT-2000-01145: Optimisation of safe food processing methods based on accurate characterisation of bacterial lag time using analysis of variance techniques, BACANOVA.
- Baranyi, J. and Roberts, T.,** 1994. A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol.*, **23**: 277–294.
- Baranyi, J.,** 1998. Comparison of stochastic and deterministic concepts of bacterial lag. *J. Theor. Biol.* **192**, 403–408.
- Baranyi, J. and Pin, C.,** 1999. Estimating bacterial growth parameters by means of detection times. *Appl. Environ. Microbiol.* **65**: 732–736.
- Baty, F. and Delignette-Muller, M.-L.,** 2004. Estimating the bacterial lag time: which model, which precision? *Int. J. Food Microbiol.* **91**: 261-277.
- Brehm-Stecher, B.F., and Johnson, E.A.,** 2004. Single-cell microbiology: tools, technologies, and applications. *Microbiol. and Mol. Biol. Reviews*, **68**: 528–559.

- Elfwing, A., Le Marc, Y., Baranyi, J., and Ballagi, A., 2004. Observing the growth and division of large number of individual bacteria using image analysis. *Appl. Environ. Microbiol.*, **70**: 675–678.
- 5 Francois, K., Devlieghere, F., Smet, K., Standaert, A.R., Geeraerd, A.H., Van Impe, J.F., and Debevere, J., 2005. Modelling the individual cell lag phase: effect of temperature and pH on the individual cell lag distribution of *Listeria monocytogenes* *Int. J. Food Microbiol.*
- 10 Kotalik, Z., Razaz, M., Elfwing, A., Ballagi, A. and Baranyi, J., 2005. Stochastic modelling of individual cell growth using flow chamber microscopy images. *Int. J. Food Microbiol.*, **105**: 177-190.
- Métris A., George S.A. Peck M.W. and Baranyi J., 2003. Distribution of turbidity detection times produced by single cell-generated bacterial populations. *Journal of Microbiological Methods*, 55: 821-827.
- 15 Métris, A., George, S.M., and Baranyi, J., 2006. Using optical density detection times to assess the effect of acetic acid on single cell kinetics *Appl. Env. Microbiol.* **72**: 10.
- Pin C, Baranyi J., 2008. Single-cell and population lag times as a function of cell age. *Appl. Env. Microbiol.* 74, 2534-2536.
- 20 Prats, C., Giró, A., Ferrer, J., López, D. and Vives-Rego, J., 2008. Analysis and IbM simulation of the stages in bacterial lag phase: basis for an updated definition. *Journal of Theoretical Biology* 252, 56-68.

## Legends

### Table 1.

Summary of the key notations used in the paper.

5

**Table 2.** Standard deviation of the  $L_g$  geometrical lag times for exactly  $k$  and for  $N_0$  cell/well where  $N_0$  is random, in case of Simulation\_1 and Simulation\_2.

### Figure 1.

10 Simulated growth curve of a one-cell-generated population growing as a result of subsequent divisions. The horizontal "steps" represent the intervals when no division occurred. The physiological lag,  $\tau$ , for the initial cell was a random number generated to follow the Gamma distribution with mean=5.4 h,  $\sigma$ =1.8 h. After the lag, the subsequent individual generation times for the daughter cells were also Gamma distributed, with mean=1 h and  $\sigma$ =0.26 h. The time to  
15 the first division (equal to the "detection time" when the detection level is  $y_{det}=\ln 2$ ) is the sum of  $\tau$  and the first generation time. The  $L_g$  geometrical lag is not marked by a measurable event but depends on the subsequent growth of the population. Its distribution can be measured by that of the detection times,  $T_{det}$ . The parameters of the simulation were taken from flow-chamber experiments presented in Elfwing (2004).

20

### Figure 2.

Two simulated examples for a one-cell-generated population **with no lag**. The distribution of the generation times is as in Fig. 1. The distribution of  $L_g=L_g(1)$  (see Fig. 1) comes exclusively from

the distribution of the subsequent generation times, since the physiological lag time is zero. The two panels demonstrate that both (i)  $L_g(1) < 0$  and (ii)  $L_g(1) > 0$  are possible for the population growth.

5 **Figure 3.**

Distribution of the  $L_g=L_g(1)$  single cell lag times with the same simulation parameters as in Fig. 2, based on 1000 replicates of a Monte-Carlo simulation. The age of the initial cell was also picked randomly. The cells had no physiological lag time, so this distribution of the *geometrical lag time* is a sole consequence of that of the subsequent generation times. It is compared with a zero-centered, shifted Gamma distribution (continuous line), where the shift is equal to the mean of the generation times (1 h) but the deviation is higher (0.4 h vs. 0.26) than that of the generation times.

**Figure 4.** Schematic representation of the horizontal and vertical distributions for a Monte-Carlo simulation for Quantitative Microbial Risk Assessment. The vertical distribution of the  $\ln N(t_c)$  log cell number, where  $t_c$  is the time of consumption, depends both on the vertically distributed initial load,  $y_0$ , and the horizontally distributed  $L_g$  geometrical lag time. The horizontal distribution of the detection time is measurable at higher cell concentration. The vertical and horizontal distributions can be transformed into each other by means of the specific growth rate,  $\mu$ , which is assumed to be constant.

Table 1.

Notation	Definition
$\alpha(N_0)$	Physiological state of the initial population consisting of $N_0$ cells (a compound random variable): $\alpha(N_0) = \exp(-\mu \cdot L_g(N_0))$
$E(X)$	Expected value of a random variable $X$
$\Gamma(\theta, \beta)$	Gamma distribution, as a generalisation of that obtained by adding $\beta$ exponentially distributed random variables each of which has $\theta$ mean value. Then the mean of the distribution is $\beta\theta$ , its variance is $\beta\theta^2$ .
$M_i(X)$	$i$ -th moment of the random variable $X$ ( $E(X) = M_1(X)$ )
$L_g(N_0)$	Geometrical lag time produced by $N_0$ initial cells in a well (a compound random variable; it depends on both $\tau$ and $T_g$ )
$L_g^{(M)}$	Estimator for $L_g=L_g(1)$ , based on the moments-method
$L_g^{(LIN)}$	Estimator of $L_g=L_g(1)$ , based on the linear method
$\mu$	Maximum specific growth rate (constant)
$N_0$	Initial cell number in a well (random variable)
$\rho$	Parameter of the Poisson distribution; average cell number per well: $\rho = E(N_0)$
$\hat{\rho}$	Estimate for the unknown $\rho = E(N_0)$ parameter: $\hat{\rho} = -\ln(W_0/W)$
$S_\alpha(N_0)$	Sum of the individual physiological states in the initial population consisting of $N_0$ cells: $S_\alpha(N_0) = \alpha_1(1) + \dots + \alpha_{N_0}(1)$ . It can be proven that $S_\alpha(N_0) = N_0 \cdot \alpha(N_0)$
$\sigma(X)$	Standard deviation of the random variable $X$
$T_{det}(N_0)$	Detection time, a compound random variable: $\mu \cdot (T_{det}(N_0) - L_g(N_0)) = y_{det} - \ln N_0$
$T_{shift}$	Shift parameter for $L_g=L_g(1)$ : $L_g - T_{shift}$ is Gamma distributed.
$\tau$	Physiological lag time of a single cell of the initial population (its first division happens at $\tau + T_g$ )
$y_{det}$	Natural logarithm of the detection level (constant)
$W$	Total number of wells in the Bioscreen
$W_0$	Number of empty wells

Table 2.

Number of initial cells in a well.	Proportion of wells with $k$ cells among the positive wells	Simulation_1 $E(\tau) = 2$ h $\sigma(\tau) = 1.4$ h	Simulation_2 $E(\tau) = 20$ h $\sigma(\tau) = 14$ h
$N_0 \sim \text{Poisson}(2.5)$	$P_k = P(N_0=k) =$	(lag $\approx$ gen.time)	(lag $\gg$ gen.time)
$N_0 = k \quad (1 \leq k \leq 4)$	$(\rho^k e^{-\rho} / k!) / (1 - e^{-\rho})$	$\sigma(L_g(k))$	$\sigma(L_g(k))$
$k = 1$	$P_1 = 0.223564$	<b>1.41</b>	<b>14.15</b>
$k = 2$	$P_2 = 0.279455$	0.86	8.25
$k = 3$	$P_3 = 0.232879$	0.67	6.12
$k = 4$	$P_4 = 0.145549$	0.56	5
<b>Convolution (<math>k &lt; 5</math>)</b>	<b>0.881447</b>	$\sigma(L_g(N_0)) \approx$ <b>1.54</b>	$\sigma(L_g(N_0)) \approx$ <b>10.36</b>

Fig 1.

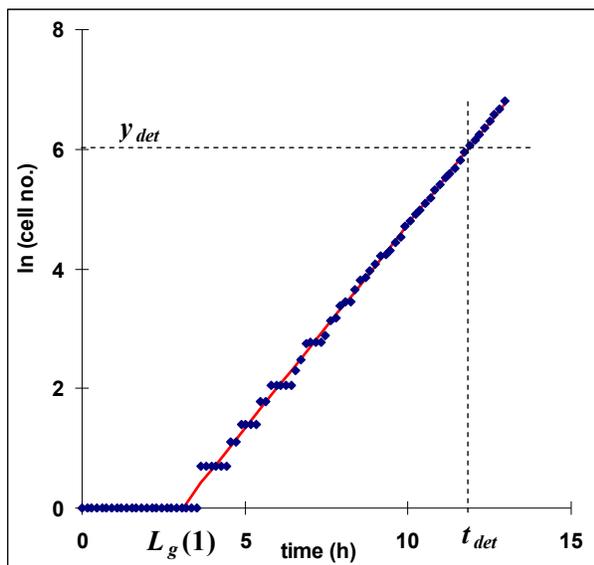


Fig 2.

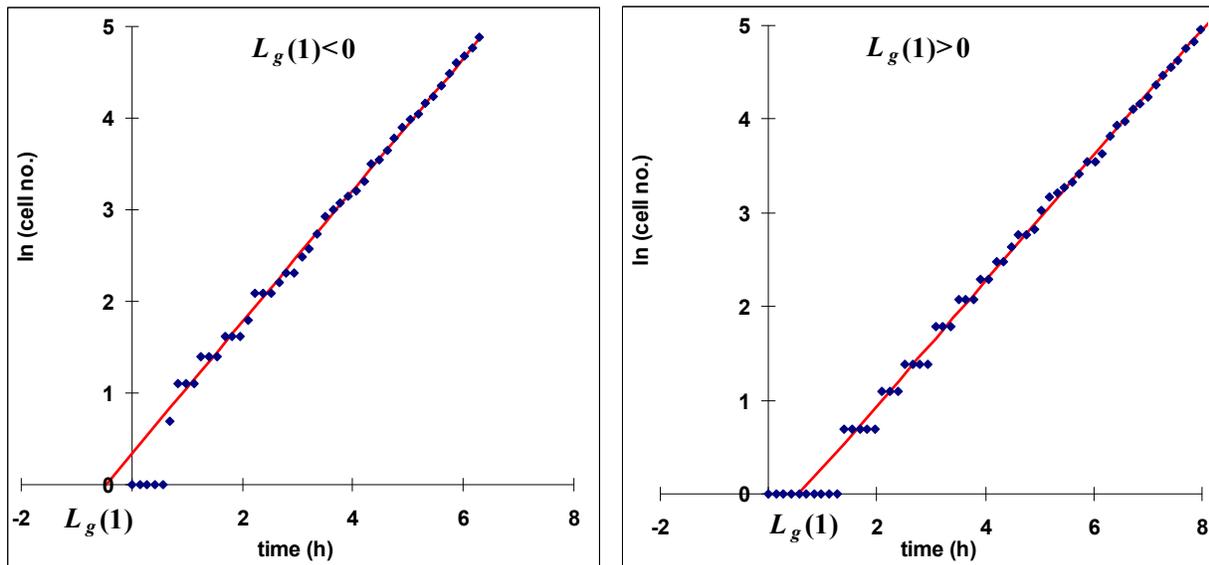


Fig 3.

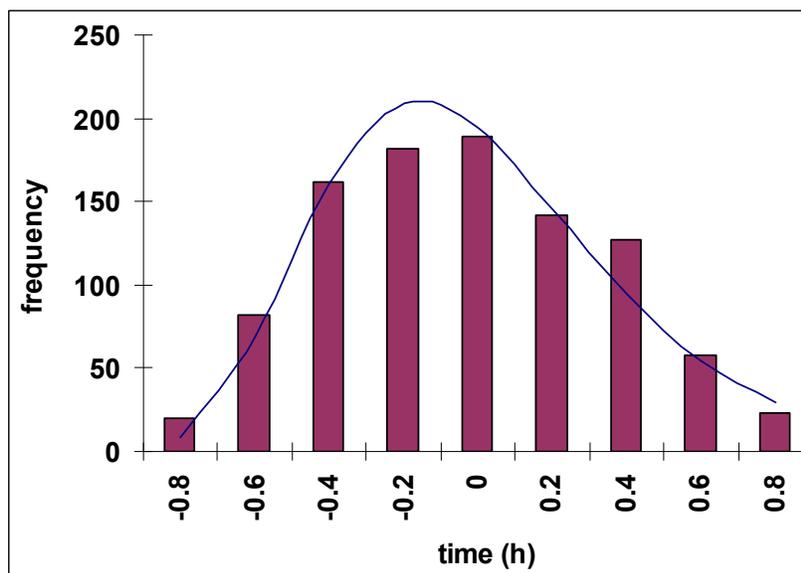
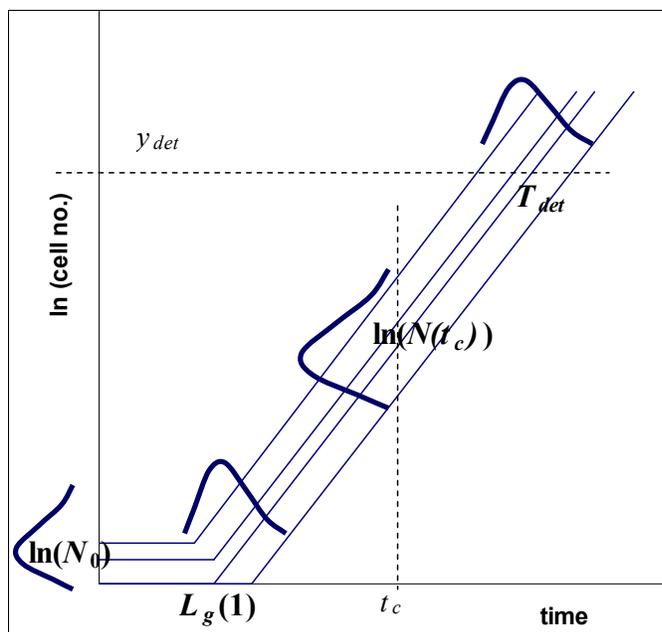


Fig 4.

5



10

15

Accepted manuscript