



Bayesian Generalized Linear Model for Simulating Bacterial Inactivation/Growth Considering Variability and Uncertainty

Satoko Hiura, Hiroki Abe, Kento Koyama* and Shige Koseki

Graduate School of Agricultural Science, Hokkaido University, Sapporo, Japan

OPEN ACCESS

Edited by:

Bruce Michael Applegate,
Purdue University, United States

Reviewed by:

Alexandra Lianou,
University of Patras, Greece
Antonio Valero,
University of Córdoba, Spain

*Correspondence:

Kento Koyama
kento.koyama@frontier.hokudai.ac.jp

Specialty section:

This article was submitted to
Food Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 01 March 2021

Accepted: 17 May 2021

Published: 24 June 2021

Citation:

Hiura S, Abe H, Koyama K and
Koseki S (2021) Bayesian Generalized
Linear Model for Simulating Bacterial
Inactivation/Growth Considering
Variability and Uncertainty.
Front. Microbiol. 12:674364.
doi: 10.3389/fmicb.2021.674364

Conventional regression analysis using the least-squares method has been applied to describe bacterial behavior logarithmically. However, only the normal distribution is used as the error distribution in the least-squares method, and the variability and uncertainty related to bacterial behavior are not considered. In this paper, we propose Bayesian statistical modeling based on a generalized linear model (GLM) that considers variability and uncertainty while fitting the model to colony count data. We investigated the inactivation kinetic data of *Bacillus simplex* with an initial cell count of 10^5 and the growth kinetic data of *Listeria monocytogenes* with an initial cell count of 10^4 . The residual of the GLM was described using a Poisson distribution for the initial cell number and inactivation process and using a negative binomial distribution for the cell number variation during growth. The model parameters could be obtained considering the uncertainty by Bayesian inference. The Bayesian GLM successfully described the results of over 50 replications of bacterial inactivation with average of initial cell numbers of 10^1 , 10^2 , and 10^3 and growth with average of initial cell numbers of 10^{-1} , 10^0 , and 10^1 . The accuracy of the developed model revealed that more than 90% of the observed cell numbers except for growth with initial cell numbers of 10^1 were within the 95% prediction interval. In addition, parameter uncertainty could be expressed as an arbitrary probability distribution. The analysis procedures can be consistently applied to the simulation process through fitting. The Bayesian inference method based on the GLM clearly explains the variability and uncertainty in bacterial population behavior, which can serve as useful information for risk assessment related to food borne pathogens.

Keywords: parameter estimation, Bayesian inference, generalized linear model, poisson distribution, negative binomial distribution, model residual

INTRODUCTION

Predictive microbiology models explain bacterial number variations over time and how growth/inactivation rates are affected by environmental conditions (Lammerding and Fazil, 2000). In the development process of mathematical or statistical models, experimental data are collected, a model is selected, and curve fitting is applied to the data for parameter estimation. Least-squares estimation has been the most widely used curve fitting procedure (Gil et al., 2017). The least-squares

methods in frequentist statistics assume that the experimental error follows a normal distribution, and studies conducted thus far have described the experimental error using a normal distribution (van Boekel, 2020). In the case of bacterial growth or inactivation kinetics, the model residual with respect to the logarithmic number of cells has been assumed to follow a normal distribution (Ratkowsky et al., 1996), though the reason for this assumption is unclear given the variability and uncertainty in bacterial population behavior. The current model residual based on the normal distribution cannot clarify the origin of the error, which means that we need to identify what type of error is included, how the error can be separated, and how large the error is.

A predictive model has been employed for exposure assessment in risk assessment to quantify the changes in the number of bacteria along the farm-to-fork chain. Exposure assessment is necessary to qualitatively and/or quantitatively assess the likelihood of ingestion of pathogens (FAO/WHO, 2008). A quantitative exposure assessment requires the development of a model that mathematically describes all the relationships between the factors influencing the exposure (FAO/WHO, 2008). Since point estimation is performed using the mean values in the kinetic model (FAO/WHO, 2008), it is difficult to appropriately estimate the changes in bacterial behavior characterized by individual cell variation. To describe the variation in bacterial behavior considering the variability and uncertainty, the need to distinguish between variability and uncertainty has been pointed out (Nauta, 2000).

The generalized linear model (GLM) is an approach incorporating various probability distributions into a fitting procedure to describe variability. The GLM, introduced by Nelder and Wedderburn (1972), has been used to describe variabilities, including discrete count data. Because a normal distribution is continuous and can take negative values, it is inappropriate for count data, which are discrete and can only take zero and positive integers, such as bacterial cell numbers. A discrete probability distribution integrated into the GLM would be suitable for expressing biological count data instead of a continuous distribution such as a normal distribution. The Poisson distribution is often used to describe death events at certain time intervals in survival analyses (Dickman et al., 2004; Dickman and Coviello, 2015). A negative binomial distribution is used for overdispersed count data in the field of ecology (Ver Hoef and Boveng, 2007). The GLM can be handled by both frequentist statistics and Bayesian inference. A GLM with Bayesian inference is often used to avoid overfitting (Dey et al., 2000). A Bayesian GLM can flexibly integrate various probability distributions as model residuals and parameter uncertainty.

Another problem in the current frequentist statistical fitting procedure is the point estimation of a model parameter. The model parameters in frequentist methods, such as the least-squares method and maximum likelihood, are often estimated by fitting the model to the data. The parameters are determined at one point in the estimation. However, because the experimental data are uncertain in a real situation, the obtained parameters are also uncertain (Garre et al., 2020; van Boekel, 2020). Therefore, the estimated parameters exhibit unexplained fluctuations (van

Boekel, 2020), and parameter estimation requires considering parameter errors (Dolan and Mishra, 2013). Furthermore, many types of uncertainties exist, such as model uncertainty and parameter uncertainty (FAO/WHO, 2008). In previous studies, model parameters were estimated using Bayesian inference (Jaloustre et al., 2011; Koyama et al., 2019; van Boekel, 2020). Bayesian inference has been used as a means to quantitatively estimate parameter uncertainty (Pouillot et al., 2003; Crépet et al., 2009; Koyama et al., 2019).

In the present study, a Bayesian GLM was introduced to fit observed bacterial inactivation data and growth data, and simulate bacterial behavior considering variability and uncertainty. Two types of bacteria were investigated to show applicability of the model to spoilage and pathogenic bacteria. For the inactivation data, we used datasets published in literature pertaining to the thermal inactivation of *Bacillus simplex*. As the growth data, the data obtained by investigating the growth of *Listeria monocytogenes* at 25°C were used. The data used contained three observed colony count replications for developing kinetic models, and over 50 observed colony count replications for validating bacterial behavior with small initial cell numbers. Individual cell heterogeneity and initial cell numbers were considered as variability and described using several probability distributions integrated into the model residual. The parameter uncertainty was obtained by Bayesian inference. From fitting to prediction, we consistently consider the variability in bacterial behavior. The modeling procedure considering the variability and uncertainty can contribute to improving risk-based processing design and exposure assessment.

MATERIALS AND METHODS

Dataset

Inactivation Dataset

The data reported by Abe et al. (2019) were used in this study. In their study, *Bacillus simplex*, which is a psychrophilic spore-forming bacterium, originating from pasteurized milk acquired from Hokkaido Research Organization (Japan). The strain was cultured in Nutrient Agar (Eiken, Tokyo, Japan) with some components and then in Nutrient Broth (Merck) with some components at 30°C for 24 h, respectively. *Bacillus simplex* spores were obtained by culturing in Spo8-agar (Faille et al., 2007; Helmond et al., 2017). The *Bacillus simplex* in the suspension with 10^5 cells was thermally inactivated at 94°C for kinetic evaluation. Viable counts were estimated by plating onto nutrient agar (Eiken, Tokyo, Japan) at 30°C after 2 days. Three independent trials were conducted. Furthermore, 60 replications of bacterial inactivation with an initial cell number of 10^n ($n = 1-3$) were used to observe the variation in bacterial inactivation.

Growth Dataset

Bacterial Strain and Inoculum Preparation

Listeria monocytogenes (ATCC 19118) was used in the present study. The bacteria was maintained at -80°C in tryptic soy broth (TSB; Merck, Darmstadt, Germany) containing 10 vol/vol% glycerol. The strain was activated by incubating the cells at 37°C

for 24 h on tryptic soy agar (TSA; Merck) and twice at 37°C for 24 h in 5 mL of TSB to obtain a homogeneous and stable cell population. The cells were then collected by centrifugation (3,000 × g for 10 min). The resulting pellet was washed twice with TSB and re-suspended in 5 mL of TSB before the experiments.

Kinetic Evaluation of Bacterial Growth

Bacterial growth was investigated by colony counting methods. The inoculum [1×10^5 colony-forming units (CFU)/mL] was prepared by series 10-fold dilutions in TSB. Aliquots (100 μL) were dispensed into the wells of 8-well polymerase chain reaction (PCR) microplates for cell concentration of 10^4 CFU/100 μL per well. The high initial cell concentration was investigated for kinetic evaluation to avoid interference of variability derived from low cell concentrations. The microplates were incubated at 25°C. Samples were withdrawn at regular intervals to obtain kinetic data of microbial growth. At each sampling time, 8-well PCR microplates were incubated at 5°C to prevent further bacterial growth. The entire sample (100 μL) from each well was diluted by serial 10-fold dilution in TSB. The bacterial cell number was determined by plating 100 μL of the diluted suspensions on TSA plates, which were then incubated at 37°C for 24 h. The experiment was repeated independently three times.

Stochastic Evaluation of Bacterial Growth

Bacterial growth with a small number of initial cells ($n = 50$) were examined to evaluate the variation in cell growth. Suspensions with average of 10^n ($n = -1-1$) CFU/100 μL were prepared by 10-fold dilution in TSB. Aliquots (100 μL) from the same inoculum culture were dispensed into wells of an 8-well PCR microplate by using an 8-channel micropipette. Cell growth ($n = 50$ replicates) was independently assessed in 50 wells of multiple 8-well PCR microplates. The microplates were incubated at 25°C. Samples were withdrawn at regular intervals to obtain probabilistic data of microbial growth. At each sampling time, the 8-well PCR microplates were incubated at 5°C to prevent further bacterial growth. The bacterial cell numbers were determined by direct plating of 100 μL of the culture onto TSA plates without dilution (average of initial cell numbers of 0.1: examined after 0–10 h; 1: 0–6 h; and 10: 0–4 h) or after diluting (1 cells: 8 and 10 h; 10 cells: 6, 8, and 10 h). The plates were incubated at 37°C for 24 h. Fifty independent replicates were analyzed.

Modeling

We introduced a Bayesian GLM instead of the currently used least-squares method. The Bayesian GLM can flexibly integrate various probability distributions as model residuals and parameter uncertainties, unlike the least-squares method. **Figure 1** shows a conceptual diagram of the fitting procedure. Because the number of bacteria is count data, we used the Poisson distribution and negative binomial distribution as model residuals instead of the normal distribution.

Bayesian GLM for Inactivation Dataset

Figure 2A shows the conceptual diagram of the inactivation model. In the least-squares method used in frequentist statistics, the error in the logarithmic number of cells is assumed to follow a normal distribution (**Figure 2A**). In contrast, the error is assumed

to follow a Poisson distribution in the inactivation process when using the GLM (**Figure 2B**).

First, we show a kinetic model based on the least-squares method. The data with 10^5 inactivated cells were fitted to the Weibull model. The Weibull model was fitted to the inactivation data, and the Weibull model is described in Equation (1):

$$\log_{10} \frac{N_t}{N_0} = -\left(\frac{t}{\delta}\right)^p \quad (1)$$

where t , N_t , N_0 , p , and δ denote the inactivation time, bacterial population at time t , initial number of cells, shape parameter, and scale parameter, respectively. Curve regression to the Weibull model was conducted using a non-linear least-squares method.

Next, we construct the GLM. Equation (1) can be transformed into Equation (2):

$$N_t = N_0 \times 10^{-(t/\delta)^p} \quad (2)$$

Here, the bacterial cell number experimentally obtained via a dilution series was assumed to follow a Poisson distribution (Koyama et al., 2016). Therefore, it can be assumed that the initial cell data follow a Poisson distribution. In addition, we assumed that the inactivation rate of each cell was equal and that a cell inactivation event was independent of another event. The number of surviving cells can also follow a Poisson distribution under the assumption that the initial cells (which follow the Poisson distribution) die at random (Aguirre et al., 2009). Therefore, it can be assumed that the observed values of the number of surviving bacteria obtained at each time are taken from a Poisson distribution with an average of $N_0 \times 10^{-(t/\delta)^p}$, and the bacterial population at time t (N_t) can be described as in Equation (3):

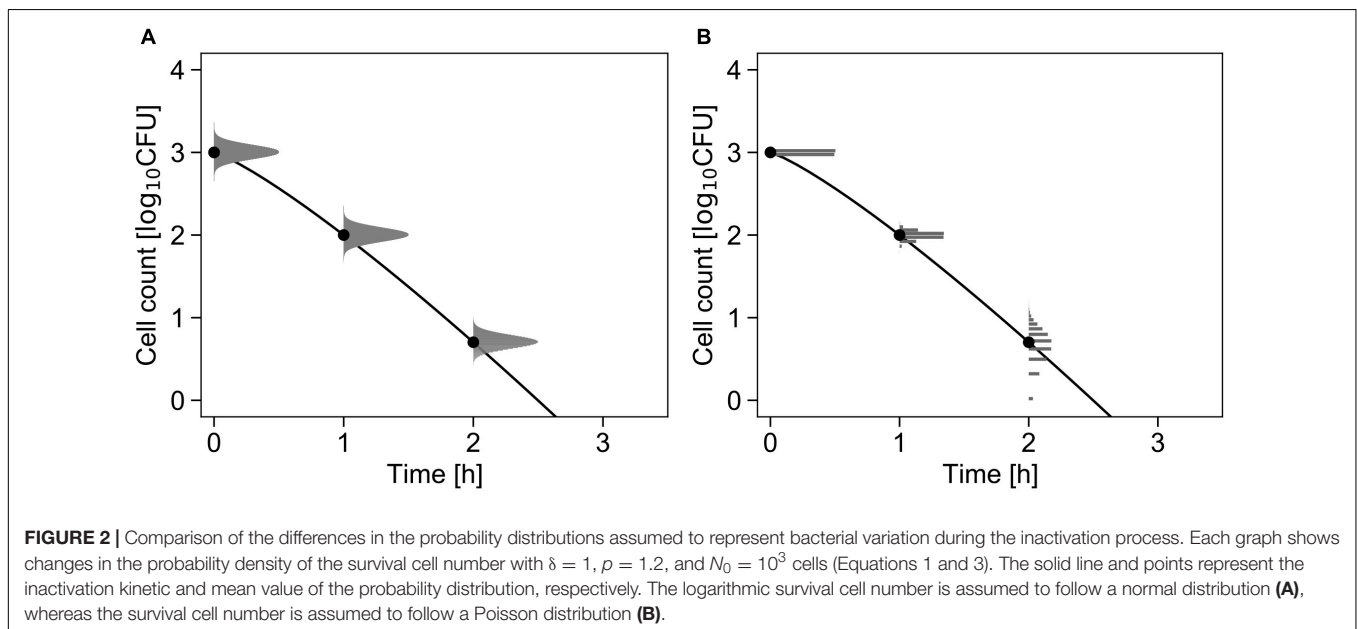
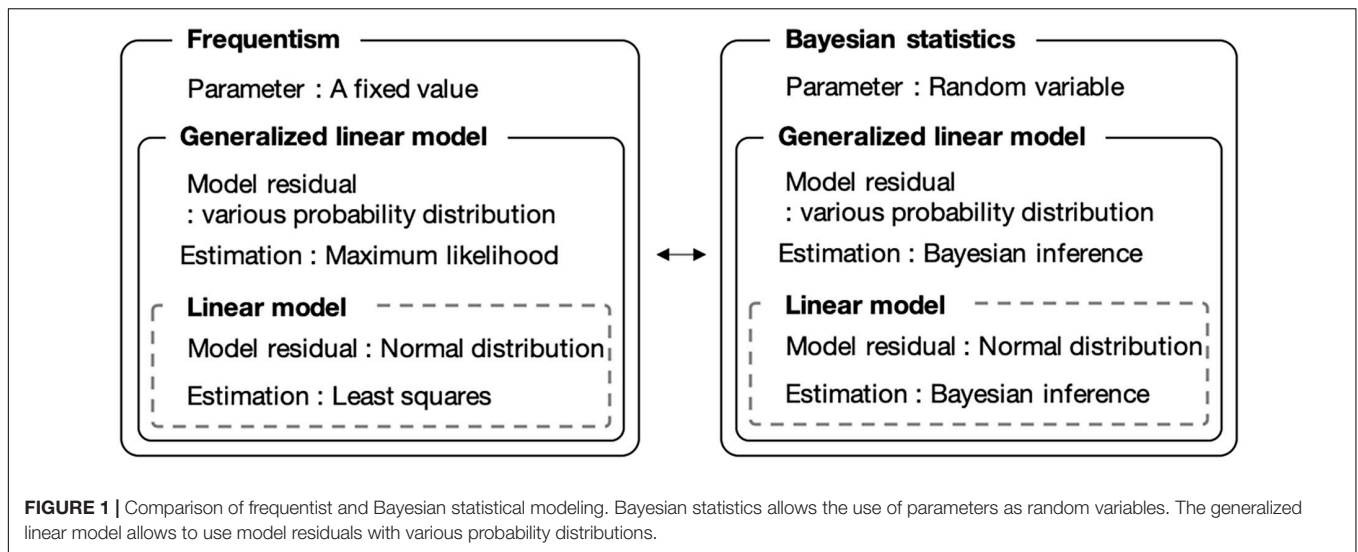
$$N_t \sim \text{Poisson} \left(N_0 \times 10^{-(t/\delta)^p} \right) \quad (3)$$

In this study, the parameters (δ and p) and the initial bacterial count (N_0) were estimated from the heating time (t) and the number of surviving bacteria at each time (N_t). The random variable in the number of cells is Poisson-distributed, which is equivalent to the model residual of the dependent variable in the GLM. The Bayesian GLM is constructed using Equation (3).

Bayesian GLM for the Growth Dataset

Figure 3 shows a conceptual diagram of the growth model. In the least-squares method used in frequentist statistics, the error in the logarithmic number of cells is assumed to follow a normal distribution (**Figure 3A**). In contrast, the error in the initial cell number is assumed to follow a Poisson distribution, and the error in the number of divisions during the exponential phase is assumed to follow a negative binomial distribution (**Figure 3B**).

First, we show a kinetic model based on the least-squares method. The growth model used in this study was based on a three-phase linear model (Buchanan et al., 1997; McKellar and Lu, 2003) without a stationary phase. In this study, to simplify the calculation, the stationary phase was not included in the data



used for the analysis. The kinetic model is described in Equation (4):

$$\log_{10}N_t = \log_{10}(N_0 \times \exp(\mu \times (t - \lambda))) (t > \lambda) \quad (4)$$

$$\log_{10}N_t = \log_{10}N_0 (t \leq \lambda)$$

where t , N_0 , N_t , μ , and λ denote the incubation time, initial cell number, number of bacteria, maximum growth rate, and lag time, respectively. Curve regression to the growth data was conducted using the least-squares method, and μ and λ were estimated.

Next, we construct the GLM. Here, as with the inactivation model, since the initial cells were experimentally obtained using a dilution series, the initial cells were assumed to follow a Poisson distribution. In addition, we assumed that the exponential growth rate of the individual cells is equal and that cell division is

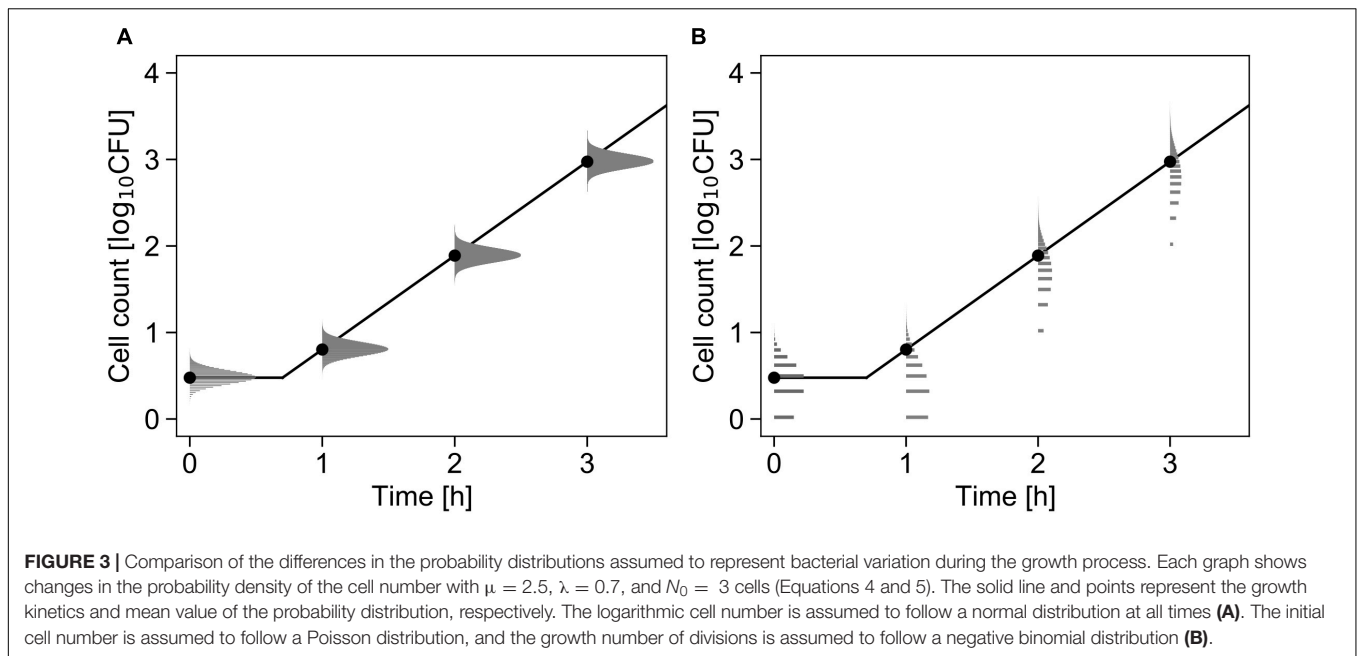
independent of another event (Coleman and Marks, 1999). Under these assumptions, a pure birth process is used to calculate the stochastic growth of bacteria (Renshaw, 1993; Coleman and Marks, 1999). In the pure birth process, the number of divisions can be described as a negative binomial distribution, as in Equation (5):

$$D_t \sim \text{Negbin}(N_0, \exp(-\mu \times (t - \lambda))) (t > \lambda) \quad (5)$$

$$N_0 \sim \text{Poisson}(N_0)$$

$$D_t = 0 (t \leq \lambda)$$

$$N_t = N_0 + D_t$$



where t , D_t , N_0 , N_t , μ , and λ denote the incubation time, a total number of cell divisions in bacterial population up to time t , number of initial cells, number of bacteria, maximum growth rate, and lag time, respectively. In this study, Bayesian inference was conducted using the growth dataset in both the lag and

exponential phases. The parameters (μ and λ) were estimated from the incubation time (t) and the number of bacteria at each time point (N_t). The Bayesian GLM is constructed using Equation (5).

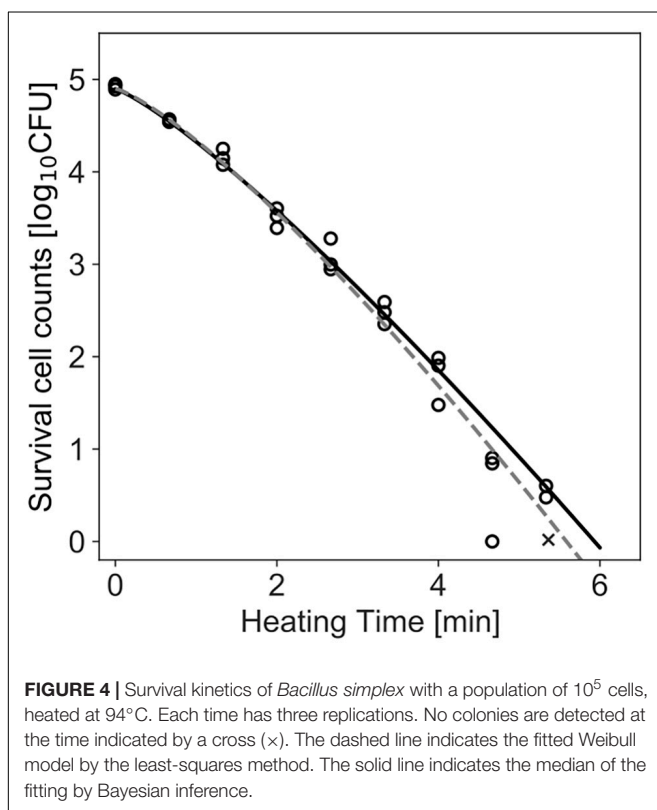
Computation

In this study, the parameters were estimated using Bayesian inference. In Bayesian inference, the obtained data were considered to have been generated from a probability distribution, and all the parameters were estimated as a probability distribution. Bayesian inference can combine priors, even if no prior information is available. In this study, we used a uniform distribution as a non-informative prior distribution because there was no prior information. For each model, inferences were made on 10^4 iterations with four independent chains. The first 5,000 iterations of 10^4 iterations were removed as a warm up period and the rest 5,000 iterations were used as posterior parameters estimation. Convergence was verified by both visually checking the Markov Chain Monte Carlo chain traces and examining the Gelman and Rubin diagnostic called R-hat. The R-hat value should be close to 1.0. Computations were performed using PyStan and Python (version 3.7.7).

Simulation

Inactivation Dataset

Two parameters (δ and p) were obtained in pairs, and 2×10^4 sets (5,000 iterations \times 4 chains) were obtained by conducting Bayesian inference. We simulated the inactivation behavior with average of initial cell numbers of 10^n ($n = 1-3$) using 2×10^4 sets of parameters. We assumed that the initial cell number and the survival cell number at each time followed a Poisson distribution. The time t (min) was set from 0 to 6 at 0.05 (min) intervals. The time and parameter values were substituted into Equation (2), and the survival cell numbers (N_t) were calculated corresponding



to each time and each parameter. We generated random numbers as many as 2×10^4 sets from the Poisson distribution with mean $N_0 \times 10^{-(t/\delta)^p}$. The number obtained here was defined as the number of surviving cells at that time. The 2×10^4 predicted results were arranged in an ascending order, and the points corresponding to the top 2.5% and the bottom 2.5% were plotted. These lines were set as the 95% predicted interval, and the predicted results were compared with the observed values for 10^n ($n = 1-3$) cell inactivation. The procedure for evaluating the predicted results was mostly based on a previous study (Hiura et al., 2020). The 2×10^4 prediction results at each time points were arranged in ascending order. If observed colony count was greater than the prediction corresponding to the lower 2.5% and less than the prediction corresponding to the upper 2.5%, the observed colony count was considered to be within the prediction range. The ratio of the number in the 95% prediction interval among the 60 observed values was calculated as the accuracy.

Growth Dataset

The parameters obtained in section “Computation” were used to predict the growth behavior with average of initial cell numbers of 10^n ($n = -1-1$). We predicted the growth behavior at 10^n ($n = -1-1$) initial cells using 2×10^4 sets of parameters.

(1) Simulation of initial bacterial number

The initial number of bacteria was assumed to follow a Poisson distribution. We generated random numbers as many as 2×10^4 sets from the Poisson distribution with mean N_0 . The number obtained here was set as the number of initial cells.

(2) Simulation of the growth number of cells

The estimated parameters (μ and λ), the initial number of bacteria, and the time (t_i) to predict the number of bacteria were substituted into Equation (5), and the random number following a negative binomial distribution was generated. At the desired time (t_i), if $t_i \leq \lambda$, the bacterial population could not grow. When $t_i > \lambda$, we generated a random number following a negative binomial distribution. The obtained value was the division number at each time. The number of bacteria at each time point was predicted by adding this division number and the initial cell number.

The 2×10^4 simulated results obtained by the above procedure were arranged in an ascending order, and the points corresponding to the top 2.5% and the bottom 2.5% were plotted at each time. These lines were set at 95% predicted intervals, and the simulated results were compared with the observed values for 10^n ($n = -1-1$) cell growth behavior. The procedure for evaluating the predicted results was as with section “Inactivation Dataset.” The ratio of the number in the 95% prediction interval among the 50 observed values was calculated as the accuracy.

RESULTS

Bayesian Inference and Prediction of Bacterial Behavior in the Inactivation Process

Figure 4 shows the dataset of the inactivation of *B. simplex* with 10^5 cells, and the fitted results by the least-squares method and Bayesian inference. Both kinetic and Bayesian fitting yielded similar results. As a regression to the Weibull model using the least-squares method in frequentist statistics, δ and p were 1.58 (standard deviation was 0.12) and 1.26 (standard deviation was 0.09), respectively. The root-mean-square error as a goodness-of-fit index was 0.18, which indicates a good fit. Figure 5 shows the posterior distributions of the parameters δ and p with Bayesian inference. Bayesian inference was conducted using data with a survival cell count of 0 colony forming unit (CFU). The R-hat value was 1.0 for each parameter, which indicates a good convergence. The mean values of δ and p were 1.59 (standard deviation was 0.11) and 1.21 (standard deviation was 0.15), respectively. The correlation coefficient between parameters δ and p was 0.18, indicating a poor positive correlation. The average values of the parameters estimated by Bayesian inference were comparable to the results estimated by the least-squares method used in frequentist statistics. However, the parameters were narrowed down to one point in the least-squares method, whereas the parameters were estimated as probability distributions in Bayesian inference.

Figure 6 shows a comparison between the observed data and the simulated results by the model. The rates of validity of the

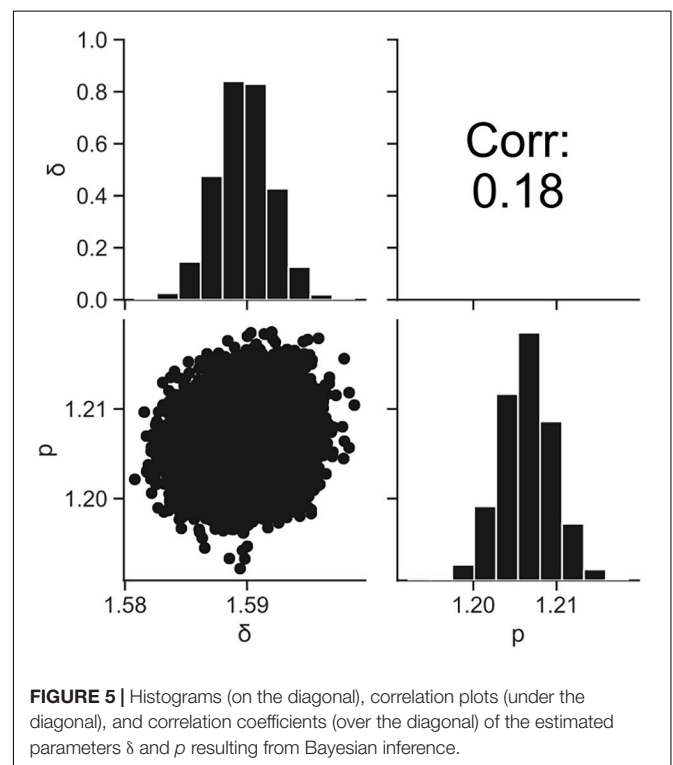
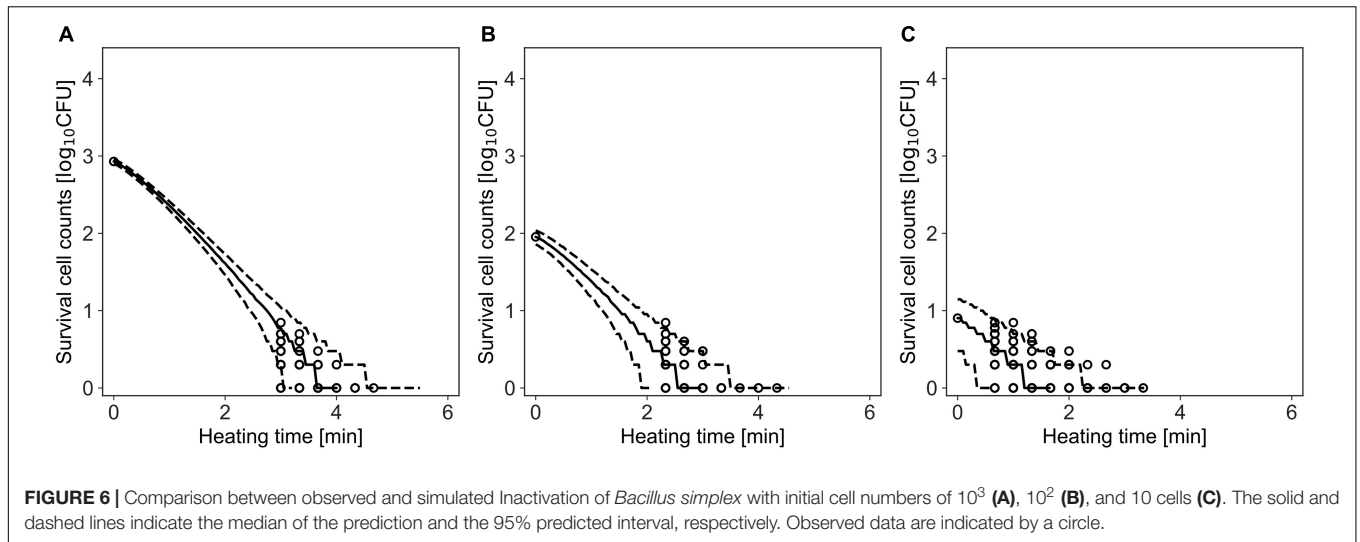


FIGURE 5 | Histograms (on the diagonal), correlation plots (under the diagonal), and correlation coefficients (over the diagonal) of the estimated parameters δ and p resulting from Bayesian inference.



number of cells within the 95% prediction band were 96, 99, and 96% for initial cell numbers of 850, 90, and 8, respectively. This result indicates that the simulation by this model covers almost the entire variation in the inactivation behavior.

Bayesian Inference and Simulation of Bacterial Behavior in the Growth Process

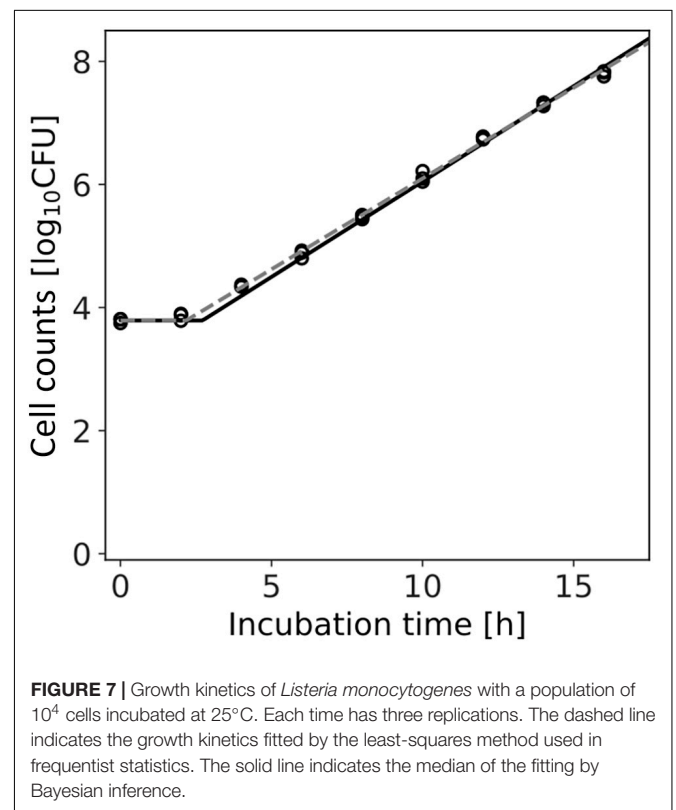
Figure 7 shows the dataset of the growth behavior of *L. monocytogenes* with a population of 10^4 cells incubated at 25°C, and the fitted results by the least-squares method and Bayesian inference. Both kinetic and Bayesian fitting yielded similar results. As a regression to the kinetic model using the least-squares method, μ and λ were 0.68 (standard deviation was 0.01) and 2.18 (standard deviation was 0.15), respectively. The RMSE was 0.05. Figure 8 shows the posterior distributions of the parameters μ and λ with Bayesian inference. The R-hat value was 1.0 for each parameter, which indicates a good convergence. The mean values of μ and λ were 0.71 (standard deviation was 0.14) and 2.72 (standard deviation was 0.01), respectively. The correlation coefficient between the parameters μ and λ was 0.69, indicating a positive correlation. The average values of the parameters estimated by Bayesian inference were comparable to the results estimated by the least-squares method used in frequentist statistics.

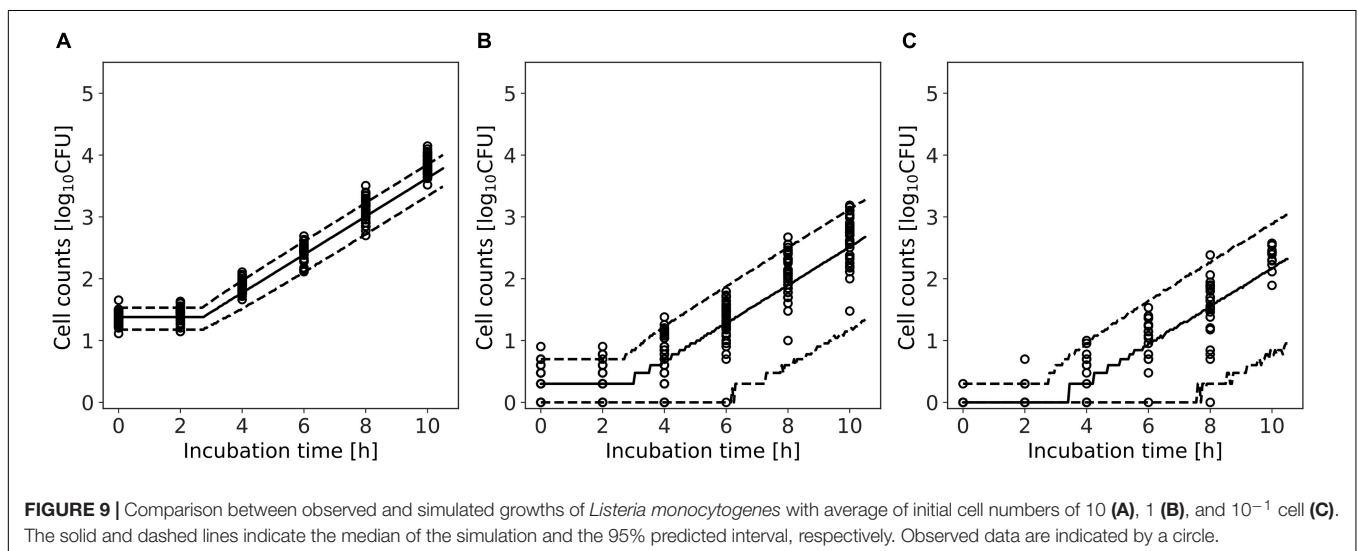
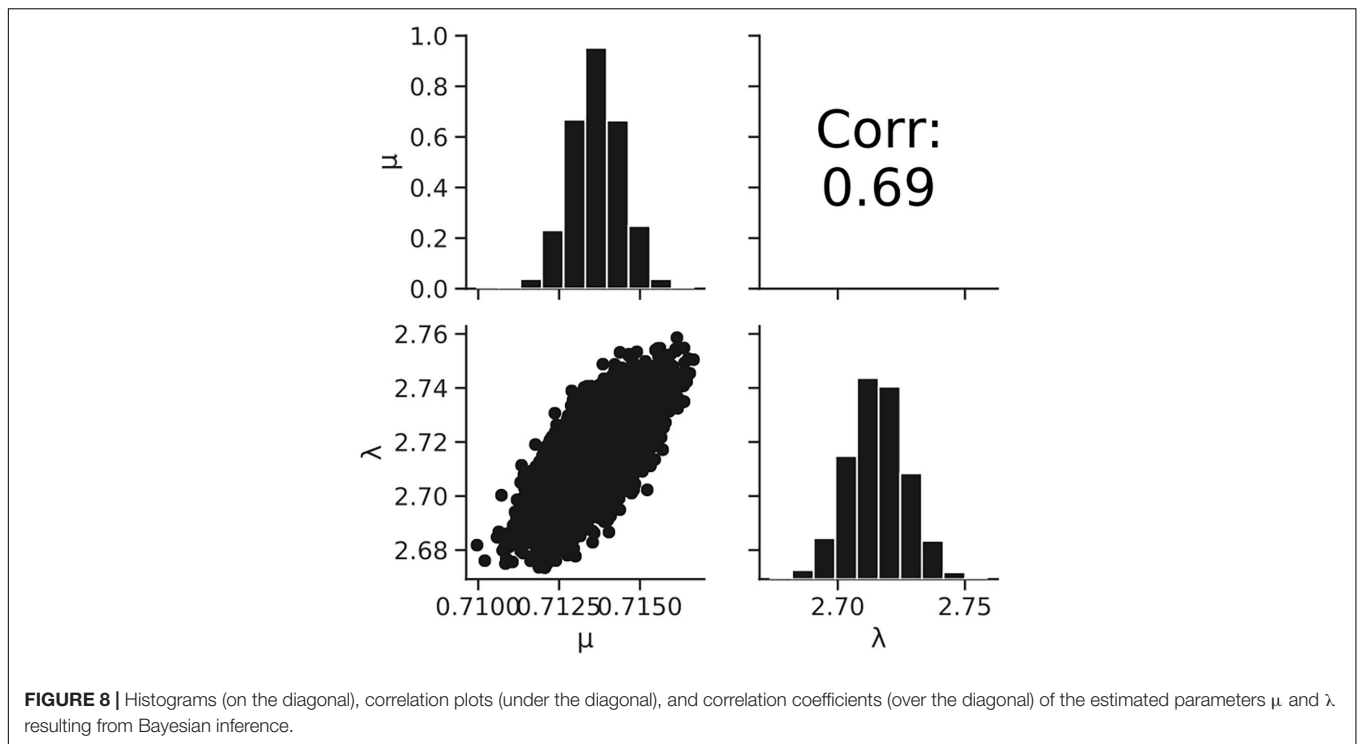
Figure 9 shows a comparison between the observed data and the results simulated by the model. The rates of validity of the number of cells within the 95% predicted interval were 80, 94, and 96% for average of initial cell numbers of 24, 2, and 0.3 cell, respectively. For 0.3 and 2 cell, the accuracy was calculated for observed values greater than 0 CFU. This result indicates that the model simulation covers almost the entire variation in the growth behavior of the small initial cell number.

DISCUSSION

In the present study, we introduced Bayesian GLM to incorporate the variability and uncertainty into a predictive

model. The estimated results of the parameter uncertainty by both the inactivation and growth models were represented in the form of a probability distribution (Figures 5, 8), which has not been considered in the conventional least-squares method. The estimated parameters enabled to predict the inactivation behavior at various initial cell numbers, such as 10^3 , 10^2 , and 10 cells, with an accuracy of over 90% (Figure 6). The estimated parameters of the growth model enabled to predict the growth behavior at various





initial cell numbers, such as 24, 2, and 0.3 cell (Figure 9). In particular, for less than 2 cells, the growth behavior was predicted with a high accuracy of over 90%. The present model expresses the variation in the bacterial behavior at low cell concentrations, which is remarkable given the individual cell heterogeneity (Koutsoumanis and Lianou, 2013; Aspidou and Koutsoumanis, 2015). The Bayesian GLM was able to fit the inactivation and growth of the bacterial population and predict the bacterial behavior considering variability and uncertainty. This modeling procedure allows to consistently consider the variations in the actual bacterial behavior, from fitting to prediction.

As a means of expressing the variation in bacterial behavior, a stochastic model has been developed, that expresses the variability in bacterial behavior with a probability distribution (FAO/WHO, 2008). Several models that can represent variability in bacterial behavior have been developed. Previous studies have clarified that variability due to individual cell heterogeneity can be expressed using a probability distribution and Monte Carlo simulation (Poschet, 2003; Aspidou and Koutsoumanis, 2015). Others have suggested combining kinetic models with computer simulations to demonstrate variability in bacterial behavior (Abe et al., 2019; Hiura et al., 2020). Even if the variability is expressed by a Monte Carlo simulation after fitting the model, such

a method is mathematically inappropriate because there are discrepancies in the model residuals during and after fitting. Therefore, in the present study, a consistent procedure, from fitting to prediction, was implemented by introducing the GLM and model fitting to the data considering the variability due to individual cell heterogeneity. The model fitting to the data and the bacterial behavior simulation can be conducted under the assumption of the same probability distribution in the process of fitting the model to the data and the simulation (Equations 3 and 5). It is reasonable to consider variability and uncertainty while fitting to the data instead of doing so after the fitting.

In the least-squares method, the logarithmic number of bacterial populations is treated as a continuous number because a normal distribution is used as the error distribution. The cell count is logarithmically analyzed in data analyses in the field of microbiology. Because the logarithm cannot be taken for 0 CFU, the count of 0 CFU is omitted from the dataset. O'Hara and Kotze (2010) suggested that distributions designed to deal with counts, such as the Poisson distribution or negative binomial distribution, should be used to fit count data instead of using a continuous distribution such as a normal distribution. O'Hara and Kotze (2010) also insisted that a log-transformation of count data should be considered when dealing with zero observations. With a discrete distribution, it is possible to make fitting and predictions, including for data with a cell number of 0 CFU. Therefore, in the present study, the number of bacteria was treated as a discrete number when fitting using the GLM (Equations 3 and 5). We expressed the variability in the bacterial cell number by introducing a discrete distribution, i.e., a Poisson distribution (Equations 3 and 5), and a negative binomial distribution (Equation 5). We were able to use data with a survival cell count of 0 CFU for parameter estimation (Figure 4) and prediction, since the Poisson distribution was used as the error distribution of the cell count. With the use of discrete distributions, the data of bacterial counts, including 0 CFU, can be used for analyses. Thus, data loss during analysis can be prevented. A discrete probability distribution is useful for expressing the number of bacteria counts, particularly in the case of a low dose, including zero.

Variability and uncertainty simultaneously appear experimentally (Poschet, 2003; Park and Lee, 2008). It is necessary to consider both these factors when conducting exposure assessments (FAO/WHO, 2008). Bacterial behavior is characterized by variability and uncertainty, and the need to consider both has been pointed out (Nauta, 2000). It is relatively easier to define variability using an equation than uncertainty, since variability is derived from some exact factors such as individual cell inactivation time as individual cell heterogeneity (Aspidou and Koutsoumanis, 2015). Therefore, the definition of variability such as individual cell inactivation time and between-strains is an important first step, since understanding variability can help determine the degree of uncertainty. The better we know the variability, the clearer the uncertainty. We defined the variability in the number of bacteria during the inactivation and growth processes using the Poisson distribution and negative binomial distribution (Equations 3 and 5), which can be a fundamental assumption for the further analysis of the variability and uncertainty in bacterial behavior.

Some other distributions such as Poisson-lognormal and Poisson-gamma distributions were used to describe number of cells in food production, where many factors affect the heterogeneity of microbial numbers among food units (Gonzales-Barron and Butler, 2011). Poisson-lognormal and Poisson-gamma distributions are used to describe over dispersion of count data (Congdon, 2006). These probability distributions may be possible choice to predict bacterial population behavior in food.

Only colony count data have been used for constructing kinetic models in Bayesian GLM that considers variability and uncertainty in bacterial behavior. Colony count data can be found not only in literature but also in databases such as ComBase¹. So far, risk related to food borne pathogens has been assessed using these accumulated data. The model proposed in the present study can help represent the variability and uncertainty in bacterial behavior using existing published data, providing a more realistic quantitative exposure assessment compared to using the conventional least-squares method. The proposed modeling procedure can help account for the variability and uncertainty in risk-based modeling.

CONCLUSION

The present study illustrated the construction of a Bayesian GLM considering the variability and uncertainty in bacterial inactivation and growth behavior. This modeling procedure allowed to consistently assume a probability distribution representing the variation in bacterial behavior throughout the fitting process for simulating bacterial behavior. The developed models enable a more explicit illustration of the variation in bacterial behavior via probability distributions, because the models are based on probabilistic theory. For example, the variation in bacterial numbers following a Poisson distribution was derived from experimentally prepared bacterial cells via a dilution process. In addition, the probability distributions of the growth or inactivation processes were assumed to be independent of other biological events. Thus, the models developed in the present study provide a reliable foundation for representing the variability and uncertainty. The Bayesian GLM can separately describe the variability and uncertainty, which cannot be done using the conventional least-squares methods used in frequentist statistics.

DATA AVAILABILITY STATEMENT

The datasets and codes in this article can be found online at: <https://github.com/Satoko-Hiura/Bayesian-generalized-linear-model-for-simulating-bacterial-inactivation-growth.git>.

AUTHOR CONTRIBUTIONS

SH, HA, KK, and SK designed the study and wrote the manuscript. SH performed the study and analyzed the data.

¹www.combase.cc

All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a grant from the Food Safety Commission, Cabinet Office, Government of Japan (Research Program for Risk Assessment Study on Food Safety, number

2004). This work was also supported by JSPS KAKENHI (Grant No. 19K23655).

ACKNOWLEDGMENTS

We would like to thank Editage for the English language editing (<https://www.editage.jp>).

REFERENCES

- Abe, H., Koyama, K., Kawamura, S., and Koseki, S. (2019). Stochastic modeling of variability in survival behavior of *Bacillus simplex* spore population during isothermal inactivation at the single cell level using a Monte Carlo simulation. *Food Microbiol.* 82, 436–444. doi: 10.1016/j.fm.2019.03.005
- Aguirre, J. S., Pin, C., Rodríguez, M. R., and de Fernando, G. D. G. (2009). Analysis of the variability in the number of viable bacteria after mild heat treatment of food. *Appl. Environ. Microbiol.* 75, 6992–6997. doi: 10.1128/AEM.00452-09
- Aspidou, Z., and Koutsoumanis, K. P. (2015). Individual cell heterogeneity as variability source in population dynamics of microbial inactivation. *Food Microbiol.* 45, 216–221. doi: 10.1016/j.fm.2014.04.008
- Buchanan, R. L., Whiting, R. C., and Damert, W. C. (1997). When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiol.* 14, 313–326. doi: 10.1006/fmic.1997.0125
- Coleman, M. E., and Marks, H. M. (1999). Qualitative and quantitative risk assessment. *Food Control* 10, 289–297. doi: 10.1016/S0956-7135(99)00052-3
- Congdon, P. (2006). *Bayesian Statistical Modelling*. Chichester: John Wiley & Sons Ltd, doi: 10.1002/9780470035948
- Crépet, A., Stahl, V., and Carlin, F. (2009). Development of a hierarchical Bayesian model to estimate the growth parameters of *Listeria monocytogenes* in minimally processed fresh leafy salads. *Int. J. Food Microbiol.* 131, 112–119. doi: 10.1016/j.ijfoodmicro.2009.01.028
- Dey, D. K., Ghosh, S. K., and Mallick, B. K. (2000). *Generalized Linear Models: A Bayesian Perspective*. Boca Raton, FL: CRC Press.
- Dickman, P. W., and Coviello, E. (2015). Estimating and modeling relative survival. *Stata J.* 15, 186–215. doi: 10.1177/1536867X1501500112
- Dickman, P. W., Sloggett, A., Hills, M., and Hakulinen, T. (2004). Regression models for relative survival. *Stat. Med.* 23, 51–64. doi: 10.1002/sim.1597
- Dolan, K. D., and Mishra, D. K. (2013). Parameter estimation in food science. *Annu. Rev. Food Sci. Technol.* 4, 401–422. doi: 10.1146/annurev-food-022811-101247
- Faille, C., Tauveron, G., Le Gentil-Lelièvre, C., and Slomianny, C. (2007). Occurrence of *Bacillus cereus* spores with a damaged exosporium: consequences on the spore adhesion on surfaces of food processing lines. *J. Food Prot.* 70, 2346–2353. doi: 10.4315/0362-028x-70.10.2346
- FAO/WHO. (2008). *Exposure Assessment of Microbiological Hazards in Food*. Rome: FAO.
- Garre, A., Zwietering, M. H., and Besten den, H. M. W. (2020). Multilevel modelling as a tool to include variability and uncertainty in quantitative microbiology and risk assessment. Thermal inactivation of *Listeria monocytogenes* as proof of concept. *Food Res. Int.* 137, 109374. doi: 10.1016/j.foodres.2020.109374
- Gil, M. M., Miller, F. A., Brandão, T. R. S., and Silva, C. L. M. (2017). Mathematical models for prediction of temperature effects on kinetic parameters of microorganisms' inactivation: tools for model comparison and adequacy in data fitting. *Food Bioprocess. Technol.* 10, 2208–2225. doi: 10.1007/s11947-017-1989-x
- Gonzales-Barron, U., and Butler, F. (2011). A comparison between the discrete Poisson-gamma and Poisson-lognormal distributions to characterise microbial counts in foods. *Food Control* 22, 1279–1286. doi: 10.1016/j.foodcont.2011.01.029
- Helmond, M., Nierop Groot, M. N., and van Bokhorst-van de Veen, H. (2017). Characterization of four *Paenibacillus* species isolated from pasteurized, chilled ready-to-eat meals. *Int. J. Food Microbiol.* 252, 35–41. doi: 10.1016/j.ijfoodmicro.2017.04.008
- Hiura, S., Abe, H., Koyama, K., and Koseki, S. (2020). Transforming kinetic model into a stochastic inactivation model: statistical evaluation of stochastic inactivation of individual cells in a bacterial population. *Food Microbiol.* 91:103508. doi: 10.1016/j.fm.2020.103508
- Jaloustre, S., Cornu, M., Morelli, E., Noël, V., and Delignette-Muller, M. L. (2011). Bayesian modeling of *Clostridium perfringens* growth in beef-in-sauce products. *Food Microbiol.* 28, 311–320. doi: 10.1016/j.fm.2010.04.002
- Koutsoumanis, K. P., and Lianou, A. (2013). Stochasticity in colonial growth dynamics of individual bacterial cells. *Appl. Environ. Microbiol.* 79, 2294–2301. doi: 10.1128/AEM.03629-12
- Koyama, K., Aspidou, Z., Koseki, S., and Koutsoumanis, K. (2019). Describing uncertainty in *Salmonella* thermal inactivation using Bayesian statistical modeling. *Front. Microbiol.* 10:2239. doi: 10.3389/fmicb.2019.02239
- Koyama, K., Hokunan, H., Hasegawa, M., Kawamura, S., and Koseki, S. (2016). Do bacterial cell numbers follow a theoretical Poisson distribution? Comparison of experimentally obtained numbers of single cells with random number generation via computer simulation. *Food Microbiol.* 60, 49–53. doi: 10.1016/j.fm.2016.05.019
- Lammerding, A. M., and Fazil, A. (2000). Hazard identification and exposure assessment for microbial food safety risk assessment. *Int. J. Food Microbiol.* 58, 147–157. doi: 10.1016/s0168-1605(00)00269-5
- McKellar, R. C., and Lu, X. (2003). *Modeling Microbial Responses in Food*. New York, NY: CRC Press.
- Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *Int. J. Food Microbiol.* 57, 9–18. doi: 10.1016/S0168-1605(00)00225-7
- Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A* 135, 370–384. doi: 10.2307/2344614
- O'Hara, R. B., and Kotze, D. J. (2010). Do not log-transform count data. *Methods Ecol. Evol.* 1, 118–122. doi: 10.1111/j.2041-210X.2010.0021.x
- Park, J.-P., and Lee, D.-S. (2008). Analysis of temperature effects on microbial growth parameters and estimation of food shelf life with confidence band. *J. Food Sci. Nutr.* 13, 104–111. doi: 10.3746/jfn.2008.13.2.104
- Poschet, F. (2003). Monte Carlo analysis as a tool to incorporate variation on experimental data in predictive microbiology. *Food Microbiol.* 20, 285–295. doi: 10.1016/S0740-0020(02)00156-9
- Pouillot, R., Albert, I., Cornu, M., and Denis, J. B. (2003). Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to *Listeria monocytogenes*. *Int. J. Food Microbiol.* 81, 87–104. doi: 10.1016/s0168-1605(02)00192-7

- Ratkowsky, D. A., Ross, T., Macario, N., Dommert, T. W., and Kamperman, L. (1996). Choosing probability distributions for modelling generation time variability. *J. Appl. Bacteriol.* 80, 131–137.
- Renshaw, E. (1993). *Modelling Biological Populations in Space and Time*. New York, NY: Cambridge University Press.
- van Boekel, M. A. J. S. (2020). On the pros and cons of Bayesian kinetic modeling in food science. *Trends Food Sci. Technol.* 99, 181–193.
- Ver Hoef, J. M., and Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 88, 2766–2772. doi: 10.1890/07-0043.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hiura, Abe, Koyama and Koseki. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.