



Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicroAnalyzing the power and error of *Listeria monocytogenes* growth challenge studies[☆]Mark R. Powell^{*}

United States Department of Agriculture, Office of Risk Assessment and Cost Benefit Analysis, 1400 Independence Ave., SW, Rm. 4032 SAG, Washington, DC 20250, USA

ARTICLE INFO

Article history:

Received 25 March 2009

Received in revised form 4 August 2009

Accepted 18 September 2009

Available online xxxx

Keywords:

Listeria monocytogenes

Challenge studies

RTE food

ABSTRACT

Domestic and international food safety policy developments have spurred interest in the design and interpretation of experimental growth challenge studies to determine whether ready-to-eat (RTE) foods are able to support growth of *Listeria monocytogenes*. Existing challenge study protocols and those under development differ markedly in terms of experimental design and the acceptance criteria under which a RTE food is determined not to support *L. monocytogenes* growth. Consequently, the protocols differ substantially with respect to the probability of incorrectly determining that growth occurs and the statistical power to detect growth if it does occur. Applying a fixed acceptance criteria exceedance value (e.g., less than a 0.5 log₁₀ or 1 log₁₀ increase) to distinguish real growth from quantitative measurement uncertainty over different experimental designs and/or measurement uncertainty values implies highly inconsistent type I error (α) probabilities. None of the *L. monocytogenes* growth challenge study designs currently being considered are likely to provide an *F*-test with $\alpha=0.05$ and power ≥ 0.8 to detect a 1 log₁₀ increase in mean concentration over the entire range of measurement uncertainty values for enumeration of *L. monocytogenes* reported in food samples in a validation study of ISO Method 11290-2.

Published by Elsevier B.V.

1. Introduction

Domestic and international food safety policy developments have spurred interest in the design and interpretation of experimental growth challenge studies to determine whether ready-to-eat (RTE) foods are able to support growth of *Listeria monocytogenes*. In 2005, the European Commission (EC) defined a food safety criteria limit of 100 colony forming units (CFU)/g for RTE foods “unable to support the growth of *L. monocytogenes*” (European Commission, 2005). The EC regulation also states that as necessary, food business operators shall conduct studies to evaluate the growth of *L. monocytogenes* that may be present in the product during the shelf-life under reasonably foreseeable conditions of storage, distribution, and use. In January 2008, the EU Community Reference Laboratory for *L. monocytogenes* issued a draft guidance document to operationally define acceptance criteria under which a RTE food is determined unable to support *L. monocytogenes* growth on the basis of shelf-life study results and to describe procedures for conducting shelf-life studies to determine compliance with the EC regulatory criteria (EUCRL, 2008a). In November 2008, the laboratory issued a revised working document providing technical guidance on shelf-life studies for *L. monocytogenes*

in RTE foods. Under EUCRL (2008b), a RTE product is determined unable to support growth of *L. monocytogenes* if the difference between the initial and final sample median concentrations is less than 0.5 log₁₀ CFU/g for all batches tested.

In February 2008, the U.S. Food and Drug Administration (FDA) issued a draft compliance policy guide stating that “FDA may regard a RTE food that does not support the growth of *L. monocytogenes* to be adulterated ... when *L. monocytogenes* is present at or above 100 [CFU/g] of food” (Food and Drug Administration, 2008a). Food and Drug Administration (2008a) states that a “listeristic control measure is generally considered to be effective if growth studies show less than a one log increase in the number of *L. monocytogenes* during replicate trials with the food of interest.” Food and Drug Administration (2008b) cites Scott et al. (2005) as an example of guidance for conducting *L. monocytogenes* growth challenge studies. In 2007, the National Advisory Committee on Microbiological Criteria for Foods (NACMCF) was charged with providing guidance to U.S. food safety agencies on inoculated pack and challenge study protocols (NACMCF, 2008). The scope of this charge includes, but is not limited to, *L. monocytogenes* growth challenge studies.

The Codex Committee on Food Hygiene (CCFH) proposed draft microbiological criteria that would establish a limit of 100 CFU/g for RTE foods “in which growth of *L. monocytogenes* will not occur” (CCFH, 2009). The proposed draft states that “a food in which growth of *L. monocytogenes* will not occur will not have an observable increase in *L. monocytogenes* levels greater than (on average) 0.5 log CFU/g for at least the expected shelf life.” In July 2009, the Codex Alimentarius Commission adopted the proposed draft microbiological criteria for *L. monocytogenes* in RTE foods (CAC, 2009).

[☆] The opinions expressed herein are the views of the author and do not necessarily reflect the official policy or position of the U.S. Department of Agriculture. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government.

^{*} Tel.: +1 202 720 9876; fax: +1 202 720 4240.

E-mail address: mpowell@oce.usda.gov.

This paper considers several inter-related issues regarding the design of *L. monocytogenes* growth challenge studies: the acceptance criteria for distinguishing real growth from quantitative measurement uncertainty and the false positive error probability and the statistical power of a study in the context of the acceptance criteria and uncertainty.

2. Materials and methods

2.1. Definitions and assumptions

A type I (false positive) error is the rejection of a true null hypothesis (H_0). The probability of a type I error (α) is often called the test level of significance. A type II (false negative) error occurs if a false H_0 is not rejected. The probability of a type II error is denoted $p(\text{type II error}) = \beta$. The power of a hypothesis test ($1 - \beta$) is the probability that H_0 is rejected when a specific alternative hypothesis (H_a) is true. The power of a test depends on the choice of α , the sample size (n), the magnitude of effect or difference (δ) under H_a , and variance in the population (σ^2). Holding other factors constant, there is a tradeoff between α and β (Rice, 1988). Underscoring this tradeoff, α is often called the producer's risk, and β is often called the consumer's risk (Montgomery, 2005). By convention, α is set at 0.05, and statistical power is considered adequate if $(1 - \beta) \geq 0.8$, although this customary experimental design practice does not consider the severity of type I and type II errors in the context of specific decisions (Di Stefano, 2003).

Let $y(t) = \text{CFU/g}$ at time t and $x = \log_{10}(y)$. Assume that x is normally distributed with mean μ and variance σ^2 , denoted $x \sim \text{Normal}(\mu, \sigma^2)$. Unless stated otherwise, the analysis is simplified by ignoring variability in growth response and assuming that x is subject only to quantitative measurement uncertainty, which includes measurement error as well as the inherent variability (e.g., in subsamples or dilutions) of a strictly unrepeatable measurement process (Corry et al., 2007; Lombard, 2006). Let $t_0 =$ initial sampling time; $t_i =$ final sampling time; $i = 1, \dots, k$, where $k =$ number of sampling times (including t_0); $j = 1, \dots, n$, where $n =$ sample size per sampling time; $l = 1, \dots, b$, where $b =$ number of RTE food batches; $n_{(il)} =$ sample size per sampling time–batch; and $c =$ number of comparisons in a test. Multiple comparisons may be independent or dependent. Orthogonal contrasts are independent. A contrast (C) is a linear combination of two or more treatment totals (T) with coefficients that sum to zero (e.g., $C_1 = -T_1 + T_2$). Multiple contrasts are orthogonal if the products of corresponding coefficients sum to zero (Hicks, 1982). (For example, if $C_2 = -T_3 + T_4$ and $C_3 = -T_5 + T_6$, then C_1 , C_2 , and C_3 are orthogonal.)

2.2. Type I error probability and power for a fixed exceedance value

Under current protocols for *L. monocytogenes* growth challenge studies, the objective of distinguishing real growth from quantitative measurement uncertainty is satisfied by setting a fixed exceedance value, or upper limit, that is intended to account for measurement uncertainty. In risk analysis, a probability of exceedance value (M) is defined for a random variable (y) by: $p(y \geq M) = \alpha$ (National Research Council, 2000). In contrast, a fixed exceedance value is defined without specifying an allowable α . Based on current *L. monocytogenes* protocols, two forms of a fixed exceedance value are considered. An exceedance value for a difference in two sample medians is denoted by $(m_{x1} - m_{x2}) < M_m$, where m_{xi} is the 50th percentile of the x_{ij} sample values. An exceedance value for a difference in two sample means is denoted by $(\bar{x}_1 - \bar{x}_2) < M_{\bar{x}}$, where $\bar{x}_i = \sum_{j=1}^n x_{ij} / n$.

The type I error probability for a fixed exceedance limit depends not only on its statistical form (median or mean) and value ($0.5 \log_{10}$ or $1.0 \log_{10}$) but also on the challenge study design specifications (e.g., the enumeration method and sampling plan) and the acceptance

criteria for determining whether a RTE product may support growth of *L. monocytogenes* (e.g., number of sample values allowed to exceed the specified upper limit). Under EUCRL (2008b), a RTE product is determined unable to support growth of *L. monocytogenes* if the difference between the initial and final sample median concentrations $(m_x(t_f) - m_x(t_0)) < 0.5 \log_{10} \text{CFU/g}$ for all batches tested. That is, $M_m = 0.5 \log_{10}$, with zero allowable exceedances of M_m . The protocol calls for testing $b \geq 3$ different batches to account for variability of the RTE food product. (Note that batches may represent a random effect under the protocol.) The protocol results in a test with $c \geq 3$ independent, pair-wise comparisons: $m_{xi}(t_f) - m_{xi}(t_0)$ for $l = 1, \dots, b \geq 3$. (Note that the comparisons are orthogonal.) The median concentration in the l th batch at each of $k = 2$ sampling times is based on a sample of $n_{(il)} = 3$. The sample median is insensitive to outliers and can be calculated if one of the three results is below the limit of enumeration; however, it is a less efficient estimator of μ than the sample mean. (See discussion regarding Eq. (7) below.)

As noted in Sec. 1, the acceptance criteria under Food and Drug Administration (2008a) include a “less than a one \log_{10} increase in the number of *L. monocytogenes* during replicate trials.” Similarly, Scott et al. (2005) concludes that “a $< 1 \log_{10}$ increase above the initial inoculum level throughout the shelf-life of the product and across replicate trials would be an appropriate acceptance criterion” due to the “inherent variation that exists with enumeration of microorganisms.” Food and Drug Administration (2008a) and Scott et al. (2005) do not specify whether the nominal exceedance value of $1 \log_{10}$ refers to a difference in means of the log-transformed enumeration data, as recommended by NSF International (2000). For the purposes of this analysis, assume the criteria are operationally defined as a difference in means, with $(\bar{x}_l(t_i) - \bar{x}_l(t_0)) < M_{\bar{x}} = 1 \log_{10} \text{CFU/g}$ for $i > 1$ over all l . (Note that $t = t_0$ for $i = 1$.)

Scott et al. (2005) recommends a minimum of $k = 5-7$ sampling times and $n = 2-3$ samples per sampling time. Food and Drug Administration (2008a) and Scott et al. (2005) do not set a minimum number of replicate trials to account for variability of the RTE food product. (For example, conditions for *L. monocytogenes* growth may be more favorable in a single, specifically formulated batch than in a small random sample of batches.) Because any one \log_{10} increase above the initial level throughout the study violates the acceptance criteria, analyzing one growth trial involves a test with $c = k - 1$ dependent, pair-wise comparisons and zero allowable exceedances. (Note that the comparisons represent many-to-one, non-orthogonal contrasts. For example, Dunnett's t -test is used for multiple comparisons of treatments with a control rather than the standard t -test to account for the dependency among comparisons (Dunnett, 1964).) As discussed below, this dependency complicates evaluation of α for a fixed exceedance value.

As noted in Sec. 1, CCFH (2009) proposed that “a food in which growth of *L. monocytogenes* will not occur will not have an observable increase in *L. monocytogenes* levels greater than (on average) $0.5 \log_{10} \text{CFU/g}$ for at least the expected shelf life.” That is, $M_{\bar{x}} = 0.5 \log_{10}$ with zero allowable exceedances of $M_{\bar{x}}$. CCFH (2009) specifies no experimental design parameters for *L. monocytogenes* growth challenge studies.

The stated basis for the proposed exceedance value is that “ $0.5 \log$ is two times of the estimated standard deviation (i.e., $0.25 \log$) associated [with] the experimental enumeration [of] viable counting/plate counts” (CCFH, 2009). However, this calculation refers to an approximate upper limit for a two-tailed, 95% confidence interval for the random variable x ($\log_{10} \text{CFU/g}$) subject to known measurement uncertainty (σ_x) (Montgomery, 2005):

$$\bar{x} \pm z_{(1-\alpha/2)} \sigma_x \quad (1)$$

where $\alpha = 0.05$, $z_{(0.975)} = 1.96$, and $\sigma_x = 0.25 \log_{10} \text{CFU/g}$. This interval is equivalent to the “expanded uncertainty” about a measurement

result ($x \pm U$, where $U = 2\sigma_x$), within which the true value of the measurand can be asserted to lie with approximately 95% confidence (Corry et al., 2007; Lombard, 2006).

In general, however, the variance of the difference between two independent random variables (A and B) is (Rice, 1988):

$$\sigma_{A-B}^2 = \sigma_A^2 + \sigma_B^2. \quad (2)$$

Consequently, because both the initial and subsequent pathogen concentrations are subject to measurement uncertainty, the variance of the difference of the means (assuming independence) is calculated as:

$$\sigma_{\bar{x}(t) - \bar{x}(t_0)}^2 = \sigma_{\bar{x}(t)}^2 + \sigma_{\bar{x}(t_0)}^2 \quad (3)$$

where $\sigma_{\bar{x}}^2 = \sigma_x^2/n$ (Rice, 1988).

Assuming equal sample sizes and variance:

$$\sigma_{\bar{x}(t) - \bar{x}(t_0)} = \sqrt{\frac{2\sigma_x^2}{n}}. \quad (4)$$

Therefore, $\sigma_{\bar{x}(t) - \bar{x}(t_0)} = \sigma_x$ only if $n = 2$.

Consequently, α for a test based on a fixed $M_{\bar{x}}$ is calculated as follows. Assuming equal variances and sample sizes, if $H_0: \mu(t) - \mu(t_0) \leq 0$ is true, then for a single pair-wise comparison:

$$p(\bar{x}(t) - \bar{x}(t_0)) > M_{\bar{x}} \leq 1 - \Phi\left(M_{\bar{x}} | \mu = 0, \sigma = \sqrt{\frac{2\sigma_x^2}{n}}\right) = \alpha' \quad (5)$$

where Φ = normal cumulative distribution function. (Note that because $\mu < 0$ also satisfies H_0 , this calculation provides an upper bound on α' .) For a test with c independent comparisons, the experiment-wise type I error probability (α) is (Sidak, 1967):

$$\alpha \leq 1 - (1 - \alpha')^c. \quad (6)$$

(Note that for $c = 1$, $\alpha = \alpha'$.)

Assuming $x \sim \text{Normal}(\mu, \sigma^2)$, both \bar{x} and m_x are consistent estimators of μ , although the sampling distributions of the mean and median differ. Asymptotically, the mean is more efficient by a factor of (Greene, 1997):

$$r = \frac{\sigma_m}{\sigma_{\bar{x}}} \rightarrow \sqrt{\frac{\pi}{2}} = 1.25. \quad (7)$$

However, the value of r depends on n . For $n = 2$ and 3, $r = 1.00$ and 1.16, respectively. (The ratios were calculated using the SAS© STMED(n) function.)

Therefore, for a single pair-wise comparison based on a fixed M_m , α is calculated as above (Eqs. (5) and (6)), substituting:

$$p(m_x(t) - m_x(t_0)) > M_m \leq 1 - \Phi\left(M_m | \mu = 0, \sigma = \sqrt{\frac{2r^2\sigma_x^2}{n}}\right) = \alpha'. \quad (8)$$

The straightforward (Sidak, 1967) adjustment provides control of experiment-wise α only when multiple comparisons are independent. For multiple dependent comparisons $\bar{x}(t) - \bar{x}(t_0)$, α is estimated by Monte Carlo simulation. Assuming that $x \sim \text{Normal}(\mu, \sigma_x^2)$, the true mean concentration is held constant over time ($\mu(t) = \mu \log_{10}$ CFU/g), and simulated sample mean values $\bar{x}(t)$ vary randomly only due to measurement uncertainty (σ_x) in the observations (x_{ij}). The proportion of simulated growth study replications resulting in an exceedance of $M_{\bar{x}}$ (and therefore a decision to falsely reject $H_0: \mu(t) - \mu(t_0) = 0$) provides an upper bound estimate of α (for $H_0: \mu(t) - \mu(t_0) \leq 0$).

Monte Carlo simulations were performed with Latin hypercube sampling using Palisades© @Risk™, Ver. 4.5.2, an add-on to Microsoft© Excel™ 2000 (Palisade Corporation, Ithaca, NY, USA). For each value of σ_x , a simulation with 10,000 iterations was performed.

European Commission (2005), Food and Drug Administration (2008a), and CCFH (2009) identify ISO Method 11290-2 (ISO, 2004) as the reference method to enumerate *L. monocytogenes* in RTE foods. In a multi-laboratory validation study of ISO Method 11290-2, Scotter et al. (2001) reported standard deviation of reproducibility (s_R , an intra-laboratory measure of quantitative measurement uncertainty (Lombard, 2006)) values in food samples with a range from 0.17 log₁₀ CFU/g to 0.45 log₁₀ CFU/g. Therefore, consider σ_x values in the range from 0.15 log₁₀ CFU/g to 0.50 log₁₀ CFU/g by increments of 0.05. (Note that under EUCRL (2008b), if at t_0 the sample standard deviation within a batch $\geq 0.3 \log_{10}$ CFU/g, then the trial for the batch is unacceptable and must be repeated to achieve acceptable contamination homogeneity and enumeration precision.) In calculating α for a fixed exceedance value, consider challenge studies with $c = 1-6$ comparisons and $n = 2-3$ samples per sampling time. Consistent with EUCRL (2008b), consider only independent comparisons for the case of $M_m = 0.5 \log_{10}$ CFU/g.

In general, the power for a test depends on the magnitude and pattern of growth assumed under H_a . To illustrate the power of a test based on a fixed exceedance value, consider the simplest case involving a single comparison. For $c = 1$, the power of a test with a fixed $M_{\bar{x}}$ is calculated as follows. Assuming equal variances and sample sizes, if $H_a: \mu(t) - \mu(t_0) = \delta$ is true, then:

$$p(\bar{x}(t) - \bar{x}(t_0)) > M_{\bar{x}} = 1 - \Phi\left(M_{\bar{x}} | \mu = \delta, \sigma = \sqrt{\frac{2\sigma_x^2}{n}}\right) = (1 - \beta). \quad (9)$$

Similarly, the power of a single comparison test based on a fixed exceedance value M_m for a difference in medians is calculated as:

$$p(m_x(t) - m_x(t_0)) > M_m = 1 - \Phi\left(M_m | \mu = \delta, \sigma = \sqrt{\frac{2r^2\sigma_x^2}{n}}\right) = (1 - \beta). \quad (10)$$

For the power of a test based on fixed exceedance values, consider: $M_{\bar{x}} = 0.5 \log_{10}$ CFU/g and $1.0 \log_{10}$ CFU/g and $M_m = 0.5 \log_{10}$ CFU/g; $\sigma_x = 0.15 \log_{10}$ CFU/g to $0.50 \log_{10}$ CFU/g (by increments of 0.05); $n = 2-3$; and $\delta = 0.5 \log_{10}$ and $1.0 \log_{10}$ CFU/g.

2.3. Power of F-test in one-way ANOVA

Analysis of variance (ANOVA) is commonly used in the peer-reviewed literature for statistical analysis of *L. monocytogenes* growth challenge studies (Pal et al., 2008; Thompson et al., 2008). One-way ANOVA applies to an experiment with a single treatment factor and a continuous response variable. The model for a simple growth trial is:

$$x_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (11)$$

where τ_i is a fixed treatment effect for time with k discrete levels and $\varepsilon \sim \text{Normal}(0, \sigma_\varepsilon^2)$. In contrast to regression methods, ANOVA requires no assumptions about growth model form and may be applied to data that exhibit a lag phase or no lag phase, linear or non-linear growth, and monotonic or non-monotonic growth patterns (e.g., adapted cells may resume growth after an initial, partial die-off).

The F -test in one-way ANOVA assumes that the observations are samples from k normally distributed populations with the same variance (σ_ε^2) but possibly different means and tests $H_0: \mu_i = \mu$ for all i . The alternative hypothesis is that at least one μ_i is different, i.e. H_a : at

least one $\tau_i \neq 0$. (Note that the F -test is inherently two-sided.) For a balanced one-way ANOVA, the F -test statistic is (Hicks, 1982):

$$F = \frac{MS_{\tau}}{MS_{\varepsilon}} \quad (12)$$

where MS_{τ} = mean square for sampling times (between groups), MS_{ε} = mean square residual (within groups). The F -test is a generalization of the two-sample t -test for simultaneously comparing more than two sample means. In contrast to the z -test where the variance is considered known (and assumed here to be limited to measurement uncertainty (σ_x)), the F -test applies where the unknown variance (σ_{ε}^2) is estimated from the sample data.

The power of the F -test is given by (Taylor and Muller, 1995):

$$(1-\beta) = 1 - F_{\omega}[F_{\text{crit}}(1-\alpha)|df_1, df_2, \omega] \quad (13)$$

where: F_{ω} = non-central F distribution, $F_{\text{crit}}(1-\alpha)$ = critical F value for rejecting $H_0 = (1-\alpha)$ percentile of the F distribution with df_1 (numerator degrees of freedom) = $k-1$ and df_2 (denominator degrees of freedom) = $k(n-1)$ (assuming equal sample size), and ω = non-centrality parameter. For balanced one-way ANOVA:

$$\omega = \frac{n \sum_{i=1}^k (\mu_i - \mu)^2}{\sigma_{\varepsilon}^2} \quad (14)$$

where μ_i = i th group mean and μ = overall mean specified under H_a , and σ_{ε}^2 = within-group variance. The power of the F -test increases with ω . For any scenario in which two elements on the right hand side of Eq. (14) are fixed, we can solve for the third to obtain the desired power. Note that for a given level of measurement uncertainty, assuming that the variance within groups reflects no heterogeneity in growth response at time t provides an upper bound on the power of the F -test.

In analyzing the power of the overall F -test for growth challenge studies, denote the ordered group means by $\mu_{(1)} \leq \dots \leq \mu_{(k)}$, and let $\delta = (\mu_{(k)} - \mu_{(1)})$. For $k > 2$ sampling times, the power of the F -test depends on the specific configuration of the k group means (i.e., the pattern of growth) postulated under H_a (George, 2001). For a given δ , power is minimized (maximized) by minimizing (maximizing) the sum of squares between groups (the second term in the numerator of Eq. (14)). The configuration that minimizes power is $\mu_{(i)} = (\mu_{(1)} + \mu_{(k)})/2$ for $i \neq 1, k$ (e.g., growth under H_a has a plateau at μ between t_0 and t_f). The configuration that maximizes power is two polarized clusters of maximally equal size with one cluster mean = $\mu_{(1)}$ and the second cluster mean = $\mu_{(k)}$ (e.g., growth under H_a is a step function with lag time $\approx t_f/2$). These two cases represent atypical growth patterns but provide upper and lower bounds on the power of the F -test for a given δ .

A series of power curves is constructed to plot the probability of rejecting $H_0: \mu_i = \mu$ at $\alpha = 0.05$ for $\delta = 1 \log_{10}$ CFU/g against σ_{ε} values by combinations of k and n . Consider $\sigma_{\varepsilon} = 0.15 \log_{10}$ CFU/g to $0.50 \log_{10}$ CFU/g; $k = 2-7$ sampling times; and $n = 2-3$ samples per sampling time. Calculations were performed with SAS© PROC POWER.

2.4. Power of a planned comparison

If the F -test in ANOVA rejects the null hypothesis, the result does not indicate which group means are significantly different or the direction of the difference (increase or decrease). A variety of *post hoc* tests are available for conducting multiple comparisons while maintaining the experiment-wise α level. For example, Dunnett's test evaluates pair-wise comparisons of $k-1$ treatments with a control, Tukey's test evaluates all $k(k-1)/2$ pair-wise comparisons, and Scheffé's test allows comparison of all possible (not only pair-

wise) combinations of k treatments (Shavelson, 1981). After accounting for a large number of comparisons to maintain the experiment-wise α level, however, the power of such tests may be insufficient.

Alternatively, a planned (*a priori*) comparison may provide more power than the overall F -test and can be applied irrespective of the omnibus test result without impacting α because the hypotheses tested are not equivalent (Rutherford, 2001). A logical planned comparison for growth challenge studies is to test whether the final concentration is significantly greater than the initial concentration. For comparing two sample means, the ANOVA F -test is equivalent to the two-sided t -test, with $F_{(1-\alpha)} = t_{(1-\alpha/2)}^2$ (Evans et al., 1993). For foodborne pathogens, however, the concern is with processes leading to unacceptably high values, and this implies a one-sided limit (AOAC International, 2006). An advantage of a one-sided limit or hypothesis test is that for a given sample size, it has more power than its two-sided counterpart (Rosner, 2005).

Therefore, consider a one-sided t -test of $H_0: \mu(t_f) - \mu(t_0) \leq 0$. The power of the test is calculated for $\alpha = 0.05$; $\delta = 0.5 \log_{10}$ CFU/g and $1 \log_{10}$ CFU/g; $\sigma_{\varepsilon} = 0.15 \log_{10}$ CFU/g to $0.50 \log_{10}$ CFU/g (by increments of 0.05); and $n = 2-3$ samples per sampling time. In addition, the required sample size (n^*) for $(1-\beta) \geq 0.8$ is obtained. (Note that $n \geq 2$ is required to calculate the sample mean.) Calculations were performed with SAS© PROC POWER.

2.5. Power of F -test in two-way mixed model

Growth challenge studies may be designed to include factors such as batches or pathogen strains to account for potential differences in *L. monocytogenes* growth response across such factors. The ANOVA F -test and its power depend on whether such factors are considered to have fixed or random effects. The duration of a growth challenge study is not randomly determined but related to the intended shelf-life of the RTE product. Time represents a fixed effect. If the levels of other factors (e.g., batches) are selected at random and inferences from the challenge study are to be generalized to the underlying population of a RTE food product, then these factors are random effects. Models that contain fixed and random effects are mixed models (Hicks, 1982).

Consider the following mixed model:

$$x_{ij} = \mu + \tau_i + B_j + \tau B_{ij} + \varepsilon_{j(i)} \quad (15)$$

where τ_i is a fixed effect for time with k levels, B_j is a random effect for batch with b levels, and $j = 1, \dots, n_{(ij)}$. Based on the expected mean square (EMS), the F -test statistic for time (the fixed effect) in the mixed model is (Hicks, 1982):

$$F_{df_1, df_2} = \frac{MS_{\tau}}{MS_{\tau B}} \quad (16)$$

where MS_{τ} = mean square between sampling times with $df_1 = k-1$ and $MS_{\tau B}$ = mean square of the τB interaction with $df_2 = (k-1)(l-1)$. By comparison, the denominator for a fixed two-way ANOVA model would be MS_{ε} (mean square residual) with $df_2 = kl(n-1)$ (Hicks, 1982). As the interaction mean square is generally larger than MS_{ε} and has fewer degrees of freedom, the F -test for the mixed model is generally less powerful than the F -test for the fixed model. This reduction in power illustrates the inherently greater challenge of expanding inferences to the entire population rather than just the specific treatments contained in a given study (Chapman and Seidel, 2008). (Batches may be considered to have fixed effects if the challenge study is specifically designed to address key formulation factors known to affect *L. monocytogenes* growth, as recommended by Scott et al. (2005).)

SAS© PROC POWER does not contain a feature for power analysis of mixed models; however, SAS© PROC MIXED performs ANOVA for mixed models. Therefore the power of the F -test in the two-way

mixed model was estimated by simulating data from multiple replications of the experiment with known inputs and calculating the proportion of *F*-tests rejected at α using SAS© PROC MIXED. The simulations were performed by parametric bootstrapping (Vose, 2000) using a SAS© DATA step. Based on the EMS for the fixed effect *F*-test, the denominator degrees of freedom were calculated using the containment method in SAS© PROC MIXED.

Consistent with EUCRL (2008b), consider a two-way mixed model with $k=2$ sampling times (t_0 and t_f), $b=3$ batches, and $n_{(ij)}=3$. Bootstrap samples were generated assuming for t_0 : $x_{ij} \sim \text{Normal}(\mu_0, \sigma_\varepsilon^2)$ and for t_f : $x_{ij} \sim \text{Normal}(\mu_0 + \delta_i, \sigma_\varepsilon^2)$. Under H_a , the difference in means varies among batches, with $\delta_i \sim \text{Normal}(\mu_\delta, \sigma_\delta^2)$. In all simulations, the average difference is fixed at $\mu_\delta = 1.0 \log_{10}$ CFU/g. To assess the sensitivity of power to the degree of variability in growth among batches, the standard deviation of δ_i (σ_δ) was increased across simulations from $0.1 \log_{10}$ CFU/g to $0.5 \log_{10}$ CFU/g by increments of 0.1. (For example, specifying $\delta_1 = 0.5 \log_{10}$ CFU/g, $\delta_2 = 1.0 \log_{10}$ CFU/g, and $\delta_3 = 1.5 \log_{10}$ CFU/g obtains $\mu_\delta = 1.0 \log_{10}$ CFU/g and $\sigma_\delta = 0.5 \log_{10}$ CFU/g. Regarding the lower bound for σ_δ , the difference in power between $\sigma_\delta = 0.01$ (results not shown) and $\sigma_\delta = 0.1$ is negligible. Thus, the coefficient of variation for batch differences ($cv = \sigma_\delta / \mu_\delta$) in the sensitivity analysis ranged from 10% to 50% across simulations. (Regarding the upper bound $\sigma_\delta = 0.5$ in the sensitivity analysis, by comparison, Cornu et al. (2006) reported *L. monocytogenes* growth rates among batches of cold-smoked salmon with $cv \approx 60\%$.) Consider $\sigma_\varepsilon = 0.15 \log_{10}$ CFU/g to $0.50 \log_{10}$ CFU/g (by increments of 0.05). For each of the 40 combinations of σ_δ and σ_ε , a simulation with 10,000 bootstrap iterations was performed. For each simulation, the power of the fixed effect *F*-test at $\alpha = 0.05$ under the alternative growth hypothesis was estimated as the proportion of the 10,000 iterations with $p(\text{greater } F) \leq 0.05$.

3. Results

3.1. Type I error probability and power for a fixed exceedance value

Tables 1–3 present the α levels for fixed exceedance values of $M_m = 0.5 \log_{10}$, $M_{\bar{x}} = 0.5 \log_{10}$ CFU/g, and $M_{\bar{x}} = 1.0 \log_{10}$ CFU/g, respectively. The results demonstrate that applying any fixed exceedance value to distinguish real growth from quantitative measurement uncertainty over different experimental designs and/or measurement uncertainty values implies highly inconsistent allowable type I error probabilities. In general, α increases with measurement uncertainty (σ_x , std. dev.) and number of comparisons (c) and decreases with increasing sample size (n) and fixed exceedance value (M). The α level is lower for dependent relative to independent comparisons and for $M_{\bar{x}}$ relative to M_m .

Table 4 presents the power of a single comparison test for fixed exceedance values of $M_m = 0.5 \log_{10}$, $M_{\bar{x}} = 0.5 \log_{10}$ CFU/g, and

Table 1
Type I error probability for difference in medians fixed exceedance value (M_m) = $0.5 \log_{10}$ CFU/g.

Std. Dev. (log ₁₀ CFU/g)	Sample size (n) = 2						Sample size (n) = 3					
	Independent comparisons (c)						Independent comparisons (c)					
	1	2	3	4	5	6	1	2	3	4	5	6
	p(type I error) ≤ α											
0.15	**	**	**	**	**	**	**	**	**	**	**	**
0.20	0.01	0.01	0.02	0.02	0.03	0.04	**	0.01	0.01	0.02	0.02	0.02
0.25	0.02	0.04	0.07	0.09	0.11	0.13	0.02	0.03	0.05	0.07	0.08	0.10
0.30	0.05	0.09	0.14	0.18	0.22	0.25	0.04	0.08	0.11	0.15	0.18	0.21
0.35	0.08	0.15	0.21	0.27	0.33	0.38	0.07	0.13	0.18	0.24	0.29	0.34
0.40	0.11	0.20	0.28	0.36	0.43	0.49	0.09	0.18	0.25	0.32	0.39	0.44
0.45	0.13	0.25	0.35	0.44	0.51	0.58	0.12	0.23	0.32	0.40	0.47	0.54
0.50	0.16	0.29	0.40	0.50	0.58	0.65	0.15	0.27	0.38	0.47	0.54	0.61

**p < 0.01.

Table 2
Type I error probability for difference in means fixed exceedance value ($M_{\bar{x}}$) = $0.5 \log_{10}$ CFU/g.

Std. Dev. (log ₁₀ CFU/g)	Sample size (n) = 2						Sample size (n) = 3					
	Independent comparisons (c)						Independent comparisons (c)					
	1	2	3	4	5	6	1	2	3	4	5	6
	p(type I error) ≤ α											
0.15	**	**	**	**	**	**	**	**	**	**	**	**
0.20	0.01	0.01	0.02	0.02	0.03	0.04	**	**	**	**	0.01	0.01
0.25	0.02	0.04	0.07	0.09	0.11	0.13	0.01	0.01	0.02	0.03	0.04	0.04
0.30	0.05	0.09	0.14	0.18	0.22	0.25	0.02	0.04	0.06	0.08	0.10	0.12
0.35	0.08	0.15	0.21	0.27	0.33	0.38	0.04	0.08	0.12	0.15	0.19	0.22
0.40	0.11	0.20	0.28	0.36	0.43	0.49	0.06	0.12	0.18	0.23	0.28	0.32
0.45	0.13	0.25	0.35	0.44	0.51	0.58	0.09	0.17	0.24	0.30	0.36	0.42
0.50	0.16	0.29	0.40	0.50	0.58	0.65	0.11	0.21	0.30	0.37	0.44	0.50

**p < 0.01.

$M_{\bar{x}} = 1.0 \log_{10}$ CFU/g for a difference in the mean (δ) = $0.5 \log_{10}$ and $1.0 \log_{10}$ CFU/g. Comparing Tables 1–4 illustrates the tradeoff between the α and β . Note that while the fixed exceedance value $M_{\bar{x}} = 1.0 \log_{10}$ CFU/g has the lowest producer's risk (Table 3) over the range of measurement uncertainty in food samples for ISO Method 11290-2 reported by Scotter et al. (2001), it also has the highest consumer's risk (lowest power) when growth occurs (Table 4).

3.2. Power of *F*-test in one-way ANOVA

Figs. 1 and 2 present series of upper and lower bound power curves plotting the power of the *F*-test in one-way ANOVA for a difference (δ) = $1 \log_{10}$ CFU/g and $\alpha = 0.05$ against σ_ε (std. dev.) values by k sampling times for sample sizes of $n = 2$ and $n = 3$, respectively. To provide a frame of reference within the bounds of these power curves, it can be shown that the power curve for the case

Table 3
Type I error probability for difference in means fixed exceedance value ($M_{\bar{x}}$) = $1 \log_{10}$ CFU/g.

Std. Dev. (log ₁₀ CFU/g)	Sample size (n) = 2						Sample size (n) = 3					
	Independent comparisons (c)						Independent comparisons (c)					
	1	2	3	4	5	6	1	2	3	4	5	6
	p(type I error) ≤ α											
0.15	**	**	**	**	**	**	**	**	**	**	**	**
0.20	**	**	**	**	**	**	**	**	**	**	**	**
0.25	**	**	**	**	**	**	**	**	**	**	**	**
0.30	**	**	**	**	**	**	**	**	**	**	**	**
0.35	**	**	0.01	0.01	0.01	0.01	**	**	**	**	**	**
0.40	0.01	0.01	0.02	0.02	0.03	0.04	**	**	**	**	0.01	0.01
0.45	0.01	0.03	0.04	0.05	0.06	0.08	**	0.01	0.01	0.01	0.02	0.02
0.50	0.02	0.04	0.07	0.09	0.11	0.13	0.01	0.01	0.02	0.03	0.04	0.04

**p < 0.01.

Table 4
Power of a single comparison test for fixed exceedance values.

Std. Dev. (log ₁₀ CFU/g)	Sample size (n) = 2		Sample size (n) = 3	
	True difference (δ)		True difference (δ)	
	0.5 log ₁₀ CFU/g	1.0 log ₁₀ CFU/g	0.5 log ₁₀ CFU/g	1.0 log ₁₀ CFU/g
Power for difference in medians fixed exceedance value (M _m) = 0.5 log ₁₀ CFU/g				
0.15	0.50	> 0.99	0.50	> 0.99
0.20	0.50	0.99	0.50	> 0.99
0.25	0.50	0.98	0.50	0.98
0.30	0.50	0.95	0.50	0.96
0.35	0.50	0.92	0.50	0.93
0.40	0.50	0.89	0.50	0.91
0.45	0.50	0.87	0.50	0.88
0.50	0.50	0.84	0.50	0.85
Power for difference in means fixed exceedance value (M _x) = 0.5 log ₁₀ CFU/g				
0.15	0.50	> 0.99	0.50	> 0.99
0.20	0.50	0.99	0.50	> 0.99
0.25	0.50	0.98	0.50	0.99
0.30	0.50	0.95	0.50	0.98
0.35	0.50	0.92	0.50	0.96
0.40	0.50	0.89	0.50	0.94
0.45	0.50	0.87	0.50	0.91
0.50	0.50	0.84	0.50	0.89
Power for difference in means fixed exceedance value (M _x) = 1.0 log ₁₀ CFU/g				
0.15	< 0.01	0.50	< 0.01	0.50
0.20	0.01	0.50	< 0.01	0.50
0.25	0.02	0.50	0.01	0.50
0.30	0.05	0.50	0.02	0.50
0.35	0.08	0.50	0.04	0.50
0.40	0.11	0.50	0.06	0.50
0.45	0.13	0.50	0.09	0.50
0.50	0.16	0.50	0.11	0.50

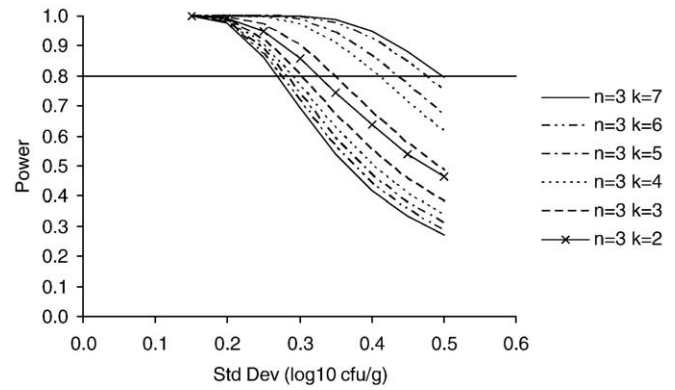


Fig. 2. Power curves for one-way ANOVA F-test ($\alpha=0.05$, $\delta=1 \log_{10}$ CFU/g) with sample size $n=3$ and sampling times $k=2-7$.

Therefore, as k increases, the likelihood of realizing power ≥ 0.8 becomes more contingent on the specific pattern of growth.

3.3. Power of a planned comparison

Table 5 summarizes the power analysis of a planned, one-sided t -test for $\alpha=0.05$ by σ_e (std. dev.) and δ . For difference (δ) = 0.5 log₁₀ CFU/g and 1 log₁₀ CFU/g, the required sample size (n^*) for power ≥ 0.8 increases from 3–11 to 2–4, respectively, over the entire range of measurement uncertainty values reported in food samples by Scotter et al. (2001). The challenge study designs considered, with $n=2-3$, do not have a sample size sufficient to provide power ≥ 0.8 over the reported range of measurement uncertainty.

3.4. Power of F-test in two-way mixed model

Table 6 summarizes the power analysis for the fixed effect F -test in a two-way mixed model study design with two sampling times (fixed), three batches (random), and three samples per time-batch for $\mu_0 = 1.0 \log_{10}$ CFU/g and $\alpha=0.05$. The results demonstrate that for a given value of σ_e (std. dev.), power decreases with increasing variation in growth among batches (σ_b). In general, the two-way mixed model challenge study design does not provide an F -test with power ≥ 0.8 to detect a mean 1 log difference over the entire range of measurement uncertainty values reported in food samples by Scotter et al. (2001). By linear interpolation, for $\sigma_e > 0.38 \log_{10}$ CFU/g, the experimental design cannot obtain power ≥ 0.8 . Similarly, for $\sigma_e > 0.27 \log_{10}$ CFU/g, to obtain power ≥ 0.8 would require that the variation between batches be less than the residual error ($\sigma_b < \sigma_e$).

where growth is log-linear with no lag time falls closer to the minimum power curve. Depending on the specific alternative, the power curve for the case where there is a substantial lag before log-linear growth can fall closer to the maximum power curve.

None of the challenge study designs considered provide an F -test with $\alpha=0.05$ and minimum power ≥ 0.8 to detect a 1 log difference in the mean over the entire range of measurement uncertainty values reported in food samples by Scotter et al. (2001). For $n=2$ (Fig. 1), minimum power ≥ 0.8 for all k with $\sigma_e \leq 0.177 \log_{10}$ CFU/g. For $n=3$ (Fig. 2), minimum power ≥ 0.8 for all k with $\sigma_e \leq 0.268 \log_{10}$ CFU/g. For $n=3$ and $k \geq 6$, maximum power ≥ 0.8 over the entire range of measurement uncertainty. Note, however, that the difference between the maximum and minimum power increases with k .

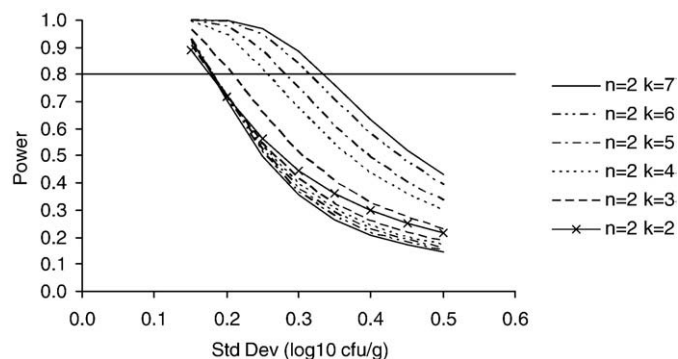


Fig. 1. Power curves for one-way ANOVA F-test ($\alpha=0.05$, $\delta=1 \log_{10}$ CFU/g) with sample size $n=2$ and sampling times $k=2-7$.

Table 5
Power analysis of a one-sided t -test ($\alpha=0.05$).

Std. Dev. (log ₁₀ CFU/g)	True increase (δ) = 0.5 (log ₁₀ CFU/g)			True increase (δ) = 1.0 (log ₁₀ CFU/g)		
	Sample size (n)			Sample size (n)		
	n = 2	n = 3	n*	n = 2	n = 3	n*
0.15	0.69	0.95	3	0.99	> 0.99	2
0.20	0.50	0.81	3	0.92	> 0.99	2
0.25	0.38	0.65	4	0.80	0.99	2
0.30	0.31	0.52	6	0.69	0.95	3
0.35	0.26	0.42	7	0.59	0.89	3
0.40	0.22	0.36	9	0.50	0.81	3
0.45	0.19	0.31	11	0.44	0.72	4
0.50	0.17	0.27	14	0.38	0.65	4

n^* = Required sample size for power ≥ 0.8 .

Table 6
Power analysis of *F*-test in two-way mixed model ($\alpha=0.05$).

Std. Dev. (Root mean square error) (log ₁₀ CFU/g)	True mean increase (μ_0) = 1.0 log ₁₀ CFU/g				
	Std. Dev. growth among batches (σ_0) (log ₁₀ CFU/g)				
	0.1	0.2	0.3	0.4	0.5
	Power				
0.15	> 0.99	> 0.99	0.94	0.57	0.14
0.20	> 0.99	0.98	0.86	0.54	0.21
0.25	0.99	0.94	0.79	0.52	0.25
0.30	0.95	0.87	0.71	0.49	0.27
0.35	0.85	0.78	0.64	0.46	0.28
0.40	0.73	0.66	0.55	0.40	0.28
0.45	0.58	0.53	0.45	0.35	0.25
0.50	0.46	0.42	0.35	0.28	0.21

4. Discussion

To the extent that *L. monocytogenes* growth response is heterogeneous among samples, the observed variance in data from challenge studies performed using ISO Method 11290-2 may exceed the upper end of the range of measurement uncertainty values reported in food samples by [Scotter et al. \(2001\)](#). Holding other factors constant, as variance increases, the statistical power of a hypothesis test decreases. Recognizing this limitation of the analysis reinforces the central conclusion that none of the *L. monocytogenes* growth challenge study designs currently being considered are likely to support statistical tests that provide $\alpha=0.05$ and power ≥ 0.8 to detect a 1 log₁₀ increase in mean pathogen concentration over the range of measurement uncertainty associated with the enumeration method referenced by [European Commission \(2005\)](#), [Food and Drug Administration \(2008a\)](#), and [CCFH \(2009\)](#). Satisfying conventional experimental design criteria would require larger sample sizes, lower measurement uncertainty, or both.

With regard to growth challenge studies designed to account for random differences in *L. monocytogenes* growth response among batches of RTE food, an investigator may select multiple samples per batch due to an interest in estimating the magnitude of the batch effect. However, for a fixed total sample size, power is maximized by selecting a single sample from multiple batches rather than drawing multiple samples from a small number of batches. The stronger the batch effect, the greater the gain in power by sampling more batches ([Holson et al., 2008](#); [National Toxicology Program, 2001](#)).

Alternatively, power may be gained by not seeking to generalize from the challenge study to the underlying population of a RTE food product. Preliminary sampling of RTE food batches can serve to characterize the variability in pH, water activity, and other physicochemical properties known to effect *L. monocytogenes* growth. Batches specifically formulated to provide favorable conditions for *L. monocytogenes* growth (as discussed by [Scott et al. \(2005\)](#)) may be considered to have fixed treatment effects, resulting in more power than a mixed model. This presumes, however, that the conditions favoring *L. monocytogenes* growth, including interactions among multiple physicochemical parameters, are well understood in the RTE product under study and that the variability of the key physicochemical factors in the RTE product is well characterized. When confronted with uncertainty, it may be more practical and expeditious in some circumstances to simply conduct a larger growth challenge study using randomly selected batches.

Finally, sampling plans and methods commonly developed as part of microbiological criteria are intended for lot-by-lot acceptance decisions ([ICMSF, 2002](#)). In contrast, a determination of whether a RTE food product supports growth of *L. monocytogenes* is a non-recurrent, one-off decision that may establish the regulatory standard applied to all lots produced henceforth. Practices associated with

routine acceptance sampling where microbiological testing is ongoing are not necessarily suitable for generating data for standard-setting purposes. Therefore, protocols for conducting and interpreting *L. monocytogenes* growth challenge studies to establish regulatory criteria should strive for consistency in application and give careful consideration to experimental design criteria, including the importance and magnitude of producer's and consumer's risks.

References

- AOAC International, 2006. Final Report and Executive Summaries from the AOAC International Presidential Task Force on Best Practices in Microbiological Methodology. AOAC International, Gaithersburg, MD.
- CAC (Codex Alimentarius Commission), 2009. Report of the Thirty-Second Session, FAO Headquarters, Rome, Italy, 29 June–4 July 2009 (ALINORM 09/32/REP). Food and Agriculture Organization of the United Nations/World Health Organization, Rome.
- CCFH (Codex Committee on Food Hygiene), 2009. Proposed Draft Annex II: Microbiological Criteria for *Listeria monocytogenes* in Ready-to-Eat Foods (At Step 5/8 of the Procedure) (Annex II of the Guidelines on the Application of General Principles of Food Hygiene to the Control of *Listeria monocytogenes* in Ready-to-Eat Foods (CAC/GL61-2007)). Food and Agriculture Organization of the United Nations/World Health Organization, Rome.
- Chapman, P.L., Seidel, G.E., 2008. Experimental design, power and sample size for animal reproduction experiments. *Reproduction, Fertility and Development* 20, 33–44.
- Cornu, M., Beaufort, A., Rudelle, S., Laloux, L., Bergis, H., Miconnet, N., et al., 2006. Effect of temperature, water-phase salt and phenolic contents on *Listeria monocytogenes* growth rates on cold-smoked salmon and evaluation of secondary models. *International Journal of Food Microbiology* 106, 159–168.
- Corry, J.E.L., Jarvis, B., Passmore, S., Hedges, A., 2007. A critical review of measurement uncertainty in the enumeration of food micro-organisms. *Food Microbiology* 24, 230–253.
- Di Stefano, J., 2003. How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Functional Ecology* 17, 707–709.
- Dunnett, C.W., 1964. New tables for multiple comparisons with a control. *Biometrics* 20, 482–491.
- EUCRL (European Union Community Reference Laboratory for *Listeria monocytogenes*), 2008a. Guidance document for conducting shelf-life studies to determine compliance with microbiological criteria for *Listeria monocytogenes* in ready-to-eat foods set out in EC regulation No 2073/2005. AFSSA (French Food Safety Agency), Maisons-Alfort, France.
- EUCRL (European Union Community Reference Laboratory for *Listeria monocytogenes*), 2008b. Technical guidance document on shelf-life studies for *Listeria monocytogenes* in ready-to-eat foods. AFSSA (French Food Safety Agency), Maisons-Alfort, France.
- European Commission, 2005. Commission regulation (EC) No 2073/2005 of 15 November 2005 on microbiological criteria for foodstuffs. *Official Journal of the European Union* 338, 1–26.
- Evans, M., Hastings, N., Peacock, B., 1993. *Statistical Distributions*, 2nd ed. John Wiley & Sons, Inc., NY.
- Food and Drug Administration, 2008a. Draft Compliance Policy Guide Sec. 555.320 – *Listeria monocytogenes*. Federal Register 73, 7293–7298.
- Food and Drug Administration, 2008b. Control of *Listeria monocytogenes* in Refrigerated or Frozen Ready-To-Eat Foods: Draft Guidance for Industry. Center for Food Safety and Applied Nutrition, Washington, DC.
- George, S.L., 2001. Multiple treatment trials. In: *Handbook of Statistics in Oncology*, Marcel Dekker, NY, pp. 149–160.
- Greene, W.H., 1997. *Econometric Analysis*, 3rd ed. Prentice Hall, Upper Saddle River, NJ.
- Hicks, C.R., 1982. *Fundamental Concepts in the Design of Experiments*, 3rd ed. Holt, Rinehart and Winston, NY.
- Holson, R.R., Freshwater, L., Maurissen, J.P.J., Moser, V.C., Phang, W., 2008. Statistical issues and techniques appropriate for developmental neurotoxicity testing: A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. *Neurotoxicology and Teratology* 30, 326–348.
- ICMSF (International Commission on Microbiological Specifications for Foods), 2002. *Micro-organisms in Foods 7: Microbiological Testing in Food Safety Management*. Springer, NY.
- ISO (International Standards Organization), 2004. ISO 11290-2:1998/Amd. 1:2004(E) Microbiology of Food and Animal Feeding Stuffs – Horizontal Method for the Detection and Enumeration of *Listeria monocytogenes* – Part 2: Enumeration Method. ISO, Geneva.
- Lombard, B., 2006. Estimation of measurement uncertainty in food microbiology: the ISO approach. *Accreditation and Quality Assurance* 17, 94–100.
- Montgomery, D.C., 2005. *Introduction to Quality Control*, 5th Ed. John Wiley & Sons, Inc., Hoboken, NJ.
- NACMCF (National Advisory Committee on Microbiological Criteria for Foods), 2008. 2007–2009 Subcommittee: Parameters for Inoculated Pack/Challenge Study Protocols. http://www.fsis.usda.gov/About_Fsis/NACMCF_Subcommittee_Inoculated_Pack/.
- National Research Council, 2000. *Risk Analysis and Uncertainty in Flood Damage Reduction Studies*. National Academy Press, Wash., DC.

- National Toxicology Program, 2001. National Toxicology Program's Report of the Endocrine Disruptors Low-Dose Peer Review. National Toxicology Program, Research Triangle Park, NC.
- NSF International, 2000. Non-potentially Hazardous Foods. NSF International Standard/American National Standard NSF/ANSI 752000. NSF International, Ann Arbor, MI.
- Pal, A., Labuza, T.P., Diez-Gonzales, F., 2008. Shelf life evaluation for ready-to-eat sliced uncured turkey breast and cured ham under probable storage conditions based on *Listeria monocytogenes* and psychrotroph growth. International Journal of Food Microbiology 126, 49–56.
- Rice, J.A., 1988. Mathematical Statistics and Data Analysis. Wadsworth & Brooks/Cole, Pacific Grove, CA.
- Rosner, B.A., 2005. Fundamentals of Biostatistics, 5th ed. Thompson Brooks/Cole, Belmont CA.
- Rutherford, A., 2001. Introducing ANOVA and ANCOVA: A GLM Approach. Sage Publications, London.
- Scott, V., Swanson, K., Freier, T., Pruett, P., Sveum, W., Hall, P., et al., 2005. Guidelines for conducting *Listeria monocytogenes* challenge testing of foods. Food Protection Trends 25, 818–825.
- Scotter, S.L., Langton, S., Lombard, B., Lahellec, C., Schulten, S., Nagelkerke, N., et al., 2001. Validation of ISO method 11290 Part 2. Enumeration of *Listeria monocytogenes* in foods. International Journal of Food Microbiology 70, 121–129.
- Shavelson, R.J., 1981. Statistical Reasoning for the Behavioral Sciences. Allyn and Bacon, Inc., Boston.
- Sidak, Z., 1967. Rectangular confidence regions for the means of multivariate normal distributions. Journal of the American Statistical Association 62, 626–633.
- Taylor, D.J., Muller, K.E., 1995. Computing confidence bounds for power and sample size of the general linear univariate model. The American Statistician 49, 43–47.
- Thompson, R.L., Carpenter, C.E., Martini, S., Broadbent, J.R., 2008. Control of *Listeria monocytogenes* in ready-to-eat meats containing sodium levulinate, sodium lactate, or a combination of sodium lactate and sodium diacetate. Journal of Food Science 73, M239–M244.
- Vose, D., 2000. Risk Analysis: A Quantitative Guide. John Wiley & Sons., Ltd., Chichester, England.