International Journal of Food Microbiology 136 (2009) 10–17

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Analyzing 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

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 $\log_{10}$ or 1 $\log_{10}$ 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 $\log_{10}$ 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 $\log_{10}$ 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 "listeristatic 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.

0168-1605/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.ijfoodmicro.2009.09.019

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; $t_f =$ final sampling time; $i = 1,\ldots, k$, where $k =$ number of sampling times (including t_0); $j = 1,\ldots, n$, where $n =$ sample size per sampling time; $l = 1,\ldots, b$, 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 $(m_x(t_f) - m_x(t_0)) < 0.5 \log_{10}$ 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: $m_{xl}(t_f) - m_{xl}(t_0)$ for $l = 1,\ldots, b \geq 3$. (Note that the comparisons are orthogonal.) The median concentration in the lth 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 $(\bar{x}_l(t_i) - \bar{x}_l(t_0)) < M_{\bar{x}} = 1 \log_{10}$ 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 $\log_{[10]}$ 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 ($\log_{10}$ CFU/g) subject to known measurement uncertainty (σ_x) (Montgomery, 2005):

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

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