International Journal of Food Microbiology 196 (2015) 40–50

Contents lists available at ScienceDirect

International Journal of Food Microbiology

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

Fitting a distribution to microbial counts: Making sense of zeroes

A.S.R. Duarte [a,*], A. Stockmarr [b], M.J. Nauta [a]

[a] Technical University of Denmark, National Food Institute, Mørkhøj Bygade, 19, Building H, DK-2860 Søborg, Denmark
[b] Technical University of Denmark, Informatics and Mathematical Modelling, Matematiktorvet, Building 303B, DK-2800 Kgs. Lyngby, Denmark

ARTICLE INFO

Article history:
Received 15 May 2013
Received in revised form 25 July 2014
Accepted 22 November 2014
Available online 3 December 2014

Keywords:
Limit of quantification
Zero counts
Poisson-lognormal
Prevalence
Concentration
Raw plate count data

ABSTRACT

The accurate estimation of true prevalence and concentration of microorganisms in foods is an important element of quantitative microbiological risk assessment (QMRA). This estimation is often based on microbial detection and enumeration data. Among such data are artificial zero counts, that originated by chance from contaminated food products. When these products are not differentiated from uncontaminated products that originate true zero counts, the estimates of true prevalence and concentration may be inaccurate. This inaccuracy is especially relevant in situations where highly pathogenic bacteria are involved and where growth can occur along the food pathway. Our aim was to develop a method that provides accurate estimates of concentration parameters and differentiates between artificial and true zeroes, thus also accurately estimating true prevalence. We first show the disadvantages of using a limit of quantification (LOQ) threshold for the analysis of microbial enumeration data. We show that, depending on the original distribution of concentrations and the LOQ value, it may be incorrect to treat artificial zeroes as censored below a quantification threshold.

Next, a method is developed that estimates the true prevalence of contamination within a food lot and the parameters characterizing the within-lot distribution of concentrations, without assuming a LOQ, and using raw plate count data as an input. Counts resulting both from contaminated and uncontaminated sample units are analysed together. This procedure allows the estimation of the proportion of artificial zeroes among the total of zero counts, and therefore the estimation of true prevalence from enumeration results.

We observe that this method yields best estimates of mean, standard deviation and prevalence at low true prevalence levels and low expected standard deviation. Furthermore, we conclude that the estimation of prevalence and the estimation of the distribution of concentrations are interrelated and therefore should be estimated simultaneously. We also conclude that one of the keys to an accurate characterization of the overall microbial contamination is the correct identification and separation of true and artificial zeroes.

Our method for the analysis of quantitative microbial data shows a good performance in the estimation of true prevalence and the parameters of the distribution of concentrations, which indicates that it is a useful data analysis tool in the field of QMRA.

1. Introduction

1.1. Microbial data in the context of QMRA

In food microbiology, the occurrence of a microorganism in a food product is characterized both in terms of (true) prevalence of contamination (the proportion of contaminated food units within a food lot) and microbial concentrations of the contaminated food units (Lorimer and Kiermeier, 2007). These two variables represent together important inputs for quantitative microbiological risk assessment (QMRA) (Commeau et al., 2012; Nauta et al., 2009a,b; Straver et al., 2007). Prevalence is usually determined by qualitative detection methods, whereas concentrations can be determined by semi-quantitative or quantitative enumeration. Hence, microbial analysis of food can consist of a detection test applied to a complete sample set of food products, followed by an enumeration method applied to the positive samples (Pouillot et al., 2013).

In food microbiology, contamination can be interpreted at different levels — food lots, individual food products from one lot (here considered as food units) and test portions taken for microbial analysis from one food unit.

A contaminated food unit is here assumed as a product containing one or more colony forming units (CFU). However, if one food unit is contaminated, it may be that not all units within the food lot are. Also, when a test portion (*e.g.* 10 g) is taken from a contaminated food unit for microbial enumeration, there is a possibility that the resulting count is zero — an artificial zero because it does not depict the true (contaminated) status of the food unit that the portion represents. Hence, in this study, artificial zero relates to the food unit level, as it represents a misleading picture of its true status.

* Corresponding author. Tel.: +45 23673256.
E-mail addresses: asrd@food.dtu.dk (A.S.R. Duarte), anst@dtu.dk (A. Stockmarr), maana@food.dtu.dk (M.J. Nauta).

http://dx.doi.org/10.1016/j.ijfoodmicro.2014.11.023
0168-1605/

Although they are determined separately, prevalence and concentration are known to be closely related, with higher concentrations most likely occurring at higher prevalence levels (Evers et al., 2010). This relationship between prevalence and concentration is the basis for the established concept of limit of detection (LOD) — the minimum concentration required in a food product for a detection test to result as "presence" (Busschaert et al., 2011; Commeau et al., 2012; Evers et al., 2010). Similarly, in an enumeration test, there is reference to a limit of quantification (LOQ). This LOQ is defined in slightly different ways (*e.g.* Busschaert et al., 2011; ISO 7218:2007). In this paper, the LOQ is interpreted as the minimum concentration required to obtain at least one colony in a plate count with a given dilution, assuming that the precision of the enumeration test is 100%.

LOD and LOQ can be either established experimentally or theoretically (Evers et al., 2010). Independent of the method used to determine them, their values are dependent on the size of the portion used for measurement (Busschaert et al., 2011; Hardin, 2011), therefore varying among different experimental protocols, which complicates the comparison of studies performed with different microbial methods. Furthermore, it is not 100% certain that an observation of zero bacteria, either in detection or enumeration, results from a sample unit with concentration below the LOD or the LOQ (Williams and Ebel, 2012; Pouillot et al., 2013).

Although thresholds commonly adopted in microbiological analysis represent artificial concepts (Evers et al., 2010), detection and quantification results are indeed subject to real limitations: of test sensitivity and specificity (Currie, 1968; Nauta et al., 2009a; Hardin, 2011; Hoelzer and Pouillot, 2013), *i.e.* the ability of a test to correctly classify contaminated and uncontaminated units, respectively, sample size and portion size (Straver et al., 2007; Hardin, 2011) and randomness (Williams and Ebel, 2012). Hardin (2011) provides a detailed description of the numerous factors that may have an influence on the limit of detection and the sensitivity of microbial analysis.

As a result, "absence" in a detection test and "zero" in an enumeration test may consist of artificial negative results (Pouillot et al., 2013). Here we decided to differentiate between the two types of artificial results. Hence, we used the term non-detect when referring to an artificial absence and the expression artificial zero when referring to a count of zero arising from a contaminated unit. The fact that non-detects are not forwarded to enumeration, leads to a situation where a number of contaminated samples are considered as non-contaminated due to their low concentration level (Pouillot et al., 2013). This practice results in an underestimation of the prevalence, particularly if the microbial concentrations are low and 1) the sample size is small (Straver et al., 2007; Hoelzer and Pouillot, 2013); 2) there is no enrichment step performed during detection (Nauta et al., 2009a); 3) the test portion is small (Straver et al., 2007); 4) the sensitivity of the detection test is low (Gardner, 2004; Hoelzer and Pouillot, 2013). The amount of non-detects is hence dependent on the method of microbiological analysis used (Gonzales-Barron et al., 2010; Hardin, 2011).

Although one may argue that products with low concentration might be considered negligible contributors to the estimated risk of certain types of foodborne illness, such as campylobacteriosis (Nauta et al., 2009a), in situations where microbial growth along the risk pathway is a possibility, or in case of more infective pathogens the concentration in those products may eventually rise to levels of concern before they reach the consumers' tables (Perez-Rodriguez et al., 2007; Straver et al., 2007). In such cases, the importance of an accurate prevalence estimate to apply in QMRA increases. Similarly, the estimated distribution of microbial concentrations must be as close as possible to the representation of the true variability within the population under analysis. When biassed estimates of prevalence and concentration are used in QMRA, the correct management of public health by the authorities may be compromised (Pouillot et al., 2013).

The characterization of microbial contamination in two distinct steps — detection followed by enumeration — contributes to the inaccuracy of the estimates of prevalence and concentration, and eventually to the distortion of the assumed relationship between those variables, and may lead to the occurrence of highly improbable outcomes (Pouillot et al., 2013). When a sample unit is split into two test portions, one for detection and the other for enumeration (Pouillot et al., 2013), sampling and measurement errors (Marks and Coleman, 1998; Müller and Hildebrandt, 1990), as well as the effect of randomness (Williams and Ebel, 2012), occur in duplicate, which results in an increased uncertainty of the overall characterization of the unit's contamination.

In this study, we consider that the key to the generation of accurate estimates of prevalence and concentration lies in the separation between artificial negative results (non-detects and artificial zeroes) from true negative results, without the employment of theoretical thresholds, such as the LOQ.

Furthermore, we believe that it is possible to limit the uncertainty in the analysis of microbial data by performing a single-step characterization of microbial contamination. Therefore, we developed a model that estimates both prevalence and concentration from the same set of quantitative enumeration data, hence avoiding the need for collection of detection data and its combined analysis with enumeration data.

1.2. Analysis of microbial data

For QMRA purposes, microbial concentrations should preferably be characterized as a probability distribution describing population variability, instead of as a point estimate (Nauta, 2002). In order to derive such type of distribution from microbial data, a certain parametric form is assumed as adequate *a priori*, to which count data or concentration estimates obtained with enumeration methods are fitted, usually by maximum likelihood estimation (MLE). The lognormal distribution has been often adopted as the parametric choice to describe variability of concentrations (Busschaert et al., 2011; Gilchrist et al., 1973; Kilsby and Pugh, 1981; Shorten et al., 2006), especially at high contamination levels (Bassett et al., 2010). In that approach, the $\log_{10}$ of concentration estimates inferred from semi-quantitative or quantitative microbial counts are fitted to a normal distribution, and estimates of mean $\log_{10}$ and standard deviation $\log_{10}$ are obtained. The challenges of this approach have been long recognized (Kilsby and Pugh, 1981) and many authors have studied alternative ways of analysing microbial data (Bassett et al., 2010; Busschaert et al., 2010; Commeau et al., 2012; Gonzales-Barron et al., 2010; Lorimer and Kiermeier, 2007; Pouillot et al., 2013; Shorten et al., 2006; Williams and Ebel, 2012).

A first challenge consists of the observation of artificial zeroes in enumeration tests, which represents a problem to the fit of a lognormal distribution that does not allow the occurrence of zero values. As a first solution to this problem, artificial zeroes were substituted by LOQ-related values. However, this approach was shown to produce biassed estimates (Lorimer and Kiermeier, 2007; Shorten et al., 2006). Alternatively, artificial zeroes started to be interpreted as censored values. A MLE method to use with censored data, had to be implemented to fit a lognormal distribution to microbial datasets involving "less-than-LOQ" values (Helsel, 2006; Lorimer and Kiermeier, 2007; Pouillot et al., 2013; Shorten et al., 2006). Later on, this method has been extended to deal with even more complex datasets, containing different types of censored information, resulting from a combination of qualitative detection tests and semi-quantitative and quantitative enumerations (Busschaert et al., 2010). This technique represented an important step forward in the interpretation of microbial data, as it allows the use of presence/absence results together with counts, for the fit of a concentration distribution. However, it is still dependent on the assumption of a LOD and a LOQ. These thresholds have an influence on the performance of the statistical method (Busschaert et al., 2010) and have been demonstrated to be artificial theoretical concepts (Evers et al., 2010). Another solution that has been applied to the challenge of fitting