Fitting a distribution to microbial counts: making sense of zeros

Ribeiro Duarte, Ana Sofia; Stockmarr, Anders; Nauta, Maarten

Publication date: 2012

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Non-detects or left-censored results are inherent to the traditional methods of microbial enumeration in foods. Typically, a low concentration of microorganisms in a food unit goes undetected in plate counts or most probable number (MPN) counts, and produces “artificial zeros”. However, these “artificial zeros” are only a share of the total number of zero counts resulting from a sample, as their number adds up to the number of “true zeros” resulting from uncontaminated units. In the process of fitting a probability distribution to microbial counts, “artificial” and “true” zeros are usually undifferentiated. This practice may lead to errors in the estimation of the parameters for the distribution of microbial concentrations, most specifically to the underestimation of the mean and overestimation of the variance. Distributions of microbial counts are often used as input in quantitative microbial risk assessment; therefore it is possible that errors related to these distributions have an impact in terms of food safety, if an influence on the estimated risk is observed.

In this study, we developed a method to estimate both the parameters of a lognormal distribution of microbial concentrations (mean and standard deviation) and the prevalence of contaminated food units (one minus the proportion of “true zeros”) from a set of microbial counts. By running the model with in silico generated concentration and count data, we could evaluate the performance of this method in terms of estimation of the three different parameters. In principle, the higher the proportion of zeros in a dataset, the higher the error in the estimation will be, and a lower prevalence contributes to a higher proportion of “true zeros” in microbial counts. Therefore, we also investigated the effect of the prevalence on the estimation of the distribution parameters mean and standard deviation by running the same model for different prevalence scenarios.