The interpretation of quantitative microbial data

meeting the demands of quantitative microbiological risk assessment

Ribeiro Duarte, Ana Sofia; Vigre, Håkan; Aabo, Søren; Nauta, Maarten

Publication date:
2013

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

The interpretation of quantitative microbial data:

meeting the demands of quantitative microbiological risk assessment

Ana Sofia Ribeiro Duarte

PhD Thesis

May 2013

Technical University of Denmark

National Food Institute

Division of Epidemiology and Microbial Genomics

Group of Epidemiology and Risk Modelling

Mørkhøj Bygade 28, DK-2860 Søborg

Denmark

Preface

The main goal of this thesis is to investigate the impact on quantitative microbiological risk assessment (QMRA) of methods used for the analysis and description of microbial data. It focus specifically on the probability distributions used to describe concentrations of microorganisms in food and on the predictive models used to describe the behaviour of those microorganisms along the food chain.

The work presented in the thesis was developed between March 2010 and May 2013 mainly at the Group of Epidemiology and Risk Modelling of the National Food Institute, Technical University of Denmark. A part of it was developed during a three month stay at the Tasmanian Institute of Agricultural Research, University of Tasmania, between March and May 2012. Norma and Frode S. Jacobsens Fond financially supported that stay.

I acknowledge all my supervisors for their guidance and critical review of my work, and all the co-authors for their contributions.

I expect this thesis to contribute for the future elaboration of more transparent and accurate QMRA studies.

Søborg, May 2013

Sofia Duarte

Acknowledgments

There is a generalized idea that a PhD project is hard, isolated, individual work. This has some truth in it. Although, I have learned during the past three years that to happily and sanely undergo the process of completing a PhD, much more is needed than a methodic, detail-oriented nature and a taste for working alone most of the time on a project that, you come to believe at a certain point, only sounds exciting to yourself.

It takes regular progress meetings with an inspiring, talented, ever-supportive, challenging and patient supervisor. A supervisor who recognizes individual circadian rhythms, who bears with reviewing "Portuguese-inspired" abstracts, who bravely jumps into two-hour meetings where the word "distribution" is said at a rate of 1/min. Someone who leads you through the first footsteps of a research career, helping you to build your self-confidence and independence, step by step. Finally, someone who understands and respects the fact that your private life cannot be put on hold for three years. I thank Maarten for being that supervisor. This work was only possible with his support.

However, it required a lot more.

A PhD also takes co-supervisors who are always available for meetings where their inspirational thoughts are called in, so that detailed results are kept under real world's perspective. My thanks here go to Håkan and Søren.

I believe enthusiasm about your work is highly dependent on the work environment that surrounds. When the atmosphere is friendly and people are welcoming, helpful and respectful, going to work is an easy thing. I thank everyone in Mørkhøj who have made it easy for me. Special thanks go to everyone who have helped me during the elaboration of this thesis by contributing with proof-reading.

Because a PhD often involves a stay abroad, it helps to be fortunate with the research destination and the scientists hosting you. Ideally, you wish to go to a beautiful far-away land, preferably an English-speaking nation, where the weather is pleasant. Above all, you wish to stay at a research-leading institution and be hosted by internationally recognized scientists, who are among the founders of your area of study. The hosts should help you to feel accommodated and of course involve you in the discussion of forward, innovative, interesting research questions. First of all, I thank Norma and Frode S. Jacobsens Fond for financially supporting my stay at the Tasmanian Institute of Agricultural Research. I am most indebt to Tom Ross for making that stay possible, for being the perfect host and for always finding the time for fruitful discussions. I also thank Lyndal Mellefont for supporting me in the task of data collection. Last, but not least, I thank Tom McMeekin and June Olley for most inspiring scientific discussions

Inspiration to work many times comes unexpectedly during leisure hours, and there is no better time than the one you spend among friends and family. I thank my friends in Portugal, who have supported me in the decision of moving abroad and who have not forgotten me despite the distance. Special thanks go to Solange, for our friendship remains unchanged by all changes, and to Álvaro, who has bear with me through great changes and has remained an even greater friend. I also acknowledge all the friends I have found in Copenhagen. They all contributed to turning this city into my second home. I specially thank Rahul for all the Friday "Pizza&Beer" evenings, Thilo for turning his home into the coziest place in town and Soila, who has been my first and greatest friend abroad. I give special thanks to Cátia, for kindly designing the cover of this thesis.

I will always be indebt to my family for their unconditional support and for adapting to a routine without me. I thank my parents for their encouragement and for all the "farewell" and "welcome back" warm and revitalizing hugs, my three younger siblings for forgiving me for missing all the changes in their lives, and my grandparents for accepting the fact of seeing me so few times a year.

Finally, I thank Norman for bringing such joy into my life that all stress seems manageable and all problems seem solvable, for not being afraid of taking important decisions and for making things happen in a very busy and stressful moment of my life as a PhD.

Summary

Foodborne diseases carry important social, health, political and economic consequences. Quantitative microbiological risk assessment (QMRA) is a science based tool used to estimate the risk that foodborne pathogens pose to human health, i.e. it estimates the number of cases of human foodborne infection or disease due to ingestion of a specific pathogenic microorganism conveyed by specific food products; it is also used to assess the effect of different control measures. In their role of risk managers, public authorities base their policies on the outcome of risk assessment studies. Therefore, they need to be transparent and affected by minimum imprecision.

The potential exposure to and infection by foodborne microorganisms depend, among other factors, on the microbial concentrations in food and on the microbial behaviour (growth, survival and transfer) along the food chain. Both factors are therefore important inputs in QMRA.

Since microbial concentrations vary among different samples of a food lot, probability distributions are used to describe these concentrations in QMRA. As microbial behaviour varies with food storage conditions (because it depends on intrinsic properties of food and extrinsic environmental variables), predictive models of bacterial growth and survival that account for those factors are used in QMRA, to describe expected changes in bacterial concentrations.

Both probability distributions and predictive models may contribute to the imprecision of QMRA: on one hand, there are several distribution alternatives available to describe concentrations and several methods to fit distributions to bacterial data; on the other hand predictive models are built based on controlled laboratory experiments of microbial behaviour, and may not be appropriate to apply in the context of real food. Hence, these models need to be validated with independent data for conditions of real food before use in QMRA.

The overall goal of the work presented in this thesis is to study different factors related to quantitative microbial data that may have an impact on the outcome of QMRA, in order to find appropriate solutions that limit the imprecision of risk estimates. A new method of fitting a distribution to microbial data is developed that estimates both prevalence and distribution of concentrations (manuscript I). Different probability distributions are used to describe concentrations in a simple QMRA model and the risk estimates obtained are compared (manuscript II). The predictive accuracy of a microbial growth model against different literature datasets are compared in order to identify different factors related to experimental data collection with a relevant impact on the model evaluation process (manuscript III).

In manuscript I (“Fitting a distribution to microbial counts: making sense of zeroes”) it is hypothesised that when “artificial” zero microbial counts, which originate by chance from contaminated food products, are not separated from “true” zeroes originating from uncontaminated products, the estimates of prevalence and concentration may be inaccurate. Such inaccuracy may have an especially relevant impact in QMRA in situations where highly pathogenic microorganisms are involved and where growth can occur along the food pathway. Hence, a method is developed that provides accurate estimates of concentration parameters and differentiates between artificial and true zeroes, thus also accurately estimating prevalence. It is demonstrated that depending on the original distribution of concentrations and the limit of quantification (LOQ) of microbial enumeration, it may be incorrect to treat artificial zeroes as censored below a quantification threshold. The method that is presented estimates the prevalence of contamination within a food lot and the parameters (mean and standard deviation) characterizing the within-lot distribution of concentrations, without assuming a LOQ, and using raw plate count data as input. Counts resulting both from contaminated and uncontaminated sample units are analysed together, which allows estimating the proportion of artificial zeroes among the total of zero counts.

The method yields good estimates of mean, standard deviation and prevalence, especially at low prevalence levels and low expected standard deviation. This study shows 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, and that estimation of prevalence and estimation of the distribution of concentrations are interrelated and therefore should be done simultaneously.

In manuscript II ("Impact of microbial count distributions on human health risk estimates") the impact of fitting microbial distributions on risk estimates is investigated at two different concentration scenarios and at a range of prevalence levels. Four different parametric distributions are used to investigate the importance of accounting for the randomness in counts, the difference between treating true zeroes as such or as censored below a LOQ and the importance of making the correct assumption about the underlying distribution of concentrations. By running a simulation experiment it is possible to assess the difference between expected risk and the risk estimated with using a lognormal, a zero-inflated lognormal, a Poisson-gamma and a zero-inflated Poisson-lognormal distribution. The method developed in manuscript I is used in this study to fit the latter.

The results show that the impact of the choice of different probability distributions to describe concentrations at retail on risk estimates depends both on the concentration and prevalence levels, but that in general it is larger at high levels of microbial contamination (high prevalence and high concentration). Also, a zero-inflation tends to improve the accuracy of the risk estimates.

In manuscript III ("Variability and uncertainty in the evaluation of predictive models with literature data – consequences to quantitative microbiological risk assessment") it is assessed how different growth settings inherent to literature datasets affect the performance of a growth model compared to its performance with the data used to generate it. The effect of the number of observations, the ranges of temperature, water activity and pH under which observations were made, the presence or absence of lactic acid in the growth environment, the use of a pathogenic or non-pathogenic strain and the type of growth environment on model performance are analysed. Model performance is measured in terms of Dif_{Af} - the difference between the accuracy factor (A_f) of the model with the data used to generate it and the A_f with an independent dataset. The study is performed using a square root-type model for the growth rate of *Escherichia coli* in response to four environmental factors and literature data that have been previously used to evaluate this model. It is hypothesised that the A_f of the model with the data used to generate it reflects the model's best possible performance, and hence Dif_{Af} is smaller and less variant when the conditions of an independent dataset are closer to the data that originated the model. The distributions of Dif_{Af} values obtained with different datasets are compared graphically and statistically.

The results suggest that if predictive models developed under controlled experimental conditions are validated against independent datasets collected from published literature, these datasets must contain a high number of observations and be based on a similar experimental growth media in order to reduce the variation of model performance. By reducing this variation, the contribution of the predictive model with uncertainty and variability to QMRA also decreases, which affects positively the precision of the risk estimates.

To conclude, this thesis contributes to the clarification of the impact that the analysis of microbial data may have in QMRA, provides a new accurate method of fitting a distribution to microbial data, and suggests guidelines for the selection of appropriate published datasets for the validation of predictive models of microbial behaviour, before their use in QMRA.

Perspectives of future work include the validation of the method developed in manuscript I with real data, and its presentation as a tool made available to the scientific community by developing, for example, a working package for the statistical software R. Also, the author expects that a standardized way of reporting microbial counts that clearly specifies the steps taken during data collection to be adopted in the future. Extending the work presented on manuscript II will allow obtaining more sound conclusions about the general impact of different frequency distributions on risk estimates. Following manuscript III, a simulation study could help investigate to what level QMRA-targeted development and validation of predictive models

are necessary for the accurate estimation of risk. Future needs in food microbiology and QMRA include the development of appropriate statistical methods to summarize novel data obtained from different “omics” technologies, adaptation of the current structure of QMRA studies to allow them to make use of such data, and the assessment of the variability and uncertainty attending those data.

Sammendrag

Fødevarebårne sygdomme har betydelige helbredsmæssige, sociale, økonomiske og politiske konsekvenser. Kvantitativ mikrobiologisk risikovurdering (QMRA) er et videnskabeligt baseret værktøj, der anvendes til at estimere antallet af sygdomstilfælde hos mennesker efter indtag af en given fødevare kontamineret med en specifik sygdomsfremkaldende mikroorganisme. Værktøjet kan ligeledes anvendes til at vurdere effekten af forskellige kontrolforanstaltninger i produktionen af den givne fødevare. Risikovurderinger benyttes af fødevaremyndigheder til udarbejdelse af regler og vejledninger, der kan mindske risikoen for fødevarebårne sygdomme. Det er derfor nødvendigt, at vurderingerne er så nøjagtige og gennemskuelige som muligt.

Forbrugereksponeringen og den deraf følgende mulige infektion med en fødevarebåren bakterie afhænger af flere faktorer, herunder antallet af bakterier, der er tilstede i fødevaren samt den mikrobielle økologi (vækst, overlevelse og krydssmitte), som finder sted i alle trin i fødevarekæden. Begge forhold er vigtige input til en QRMA.

Antallet af bakterier vil variere naturligt mellem de prøver, der udtages af et parti fødevarer. I en QMRA tages der højde for denne variation ved at beskrive bakterieantallet ved anvendelse af sandsynlighedsfordelinger. Udviklingen i antallet af bakterier i en fødevare gennem produktionsprocessen afhænger både af fødevarens art, håndtering og opbevaringsforhold, og beregnes ved hjælp af prædiktive mikrobiologiske modeller, der bl.a. kan forudsige ændringer bakteriekoncentrationer under specifikke fysiske og kemiske forhold.

Både de valgte sandsynlighedsfordelinger og prædiktive mikrobiologiske modeller bidrager til usikkerheden af en QMRA. Dels er det muligt at vælge mellem flere forskellige alternative fordelinger for bakteriekoncentrationer samt måder at tilpasse fordelinger til aktuelle data; og dels er prædiktive mikrobiologiske modeller oftest baseret på kontrollerede laboratorieforsøg, der måske ikke i tilstrækkelig grad afspejler forholdene i de fødevarer, som forbrugerne indtager. Resultaterne af disse modeller bør derfor valideres med uafhængige data indsamlet fra "rigtige" fødevarer inden de indgår i en QMRA.

Det overordnede mål med denne afhandling er at undersøge forskellige faktorer relateret til kvantitative mikrobiologiske data, som kan påvirke resultaterne af en QMRA med henblik på at finde løsninger, der kan minimere usikkerheden på risikoestimaterne. Til dette formål er der udviklet en metode, der kan tilpasse en fordeling til mikrobiologiske data og som angiver både et estimat for prævalens og en fordeling for antallet af bakterier (manuskript I). Forskellige sandsynlighedsfordelinger er derefter blevet anvendt til at beskrive bakterietantallet i en simpel QMRA model og de forskellige risikoestimater er blevet sammenlignet (manuskript II). Endelig er nøjagtigheden af resultaterne af de prædiktive mikrobiologiske modeller blevet undersøgt på basis af litteraturdata og sammenlignet med henblik på at identificere faktorer relateret til eksperimentelle data, der kan have afgørende indflydelse på evalueringen af en model (manuskript III).

I manuscript I ("Fitting a distribution to microbial counts: making sense of zeroes") er hypotesen, at en manglende adskillelse af "falsk negative" mikrobiologiske tællinger, som opstår ved tilfældighed selvom fødevaren reelt er forurenet, fra "sandt negative" tællinger medfører, at estimater for prævalens og bakterieantal bliver unøjagtige. Sådanne unøjagtigheder kan især have betydning for en QMRA, når det drejer sig om særligt virulente bakterier, der kan opformeres i fødevarekæden. Der er derfor udviklet en metode, der kan tilvejebringe nøjagtige estimater for koncentrationen af bakterier og som kan skelne mellem falske og sande negative bakterietællinger og dermed også give mere nøjagtige prævalensestimater. Metoden demonstrerer, at det, afhængigt af den oprindelige fordeling af bakteriekoncentrationen og den aktuelle detektionsgrænse, kan lede til fejlbehæftede resultater, hvis falske 0-prøver ukritisk tolkes som negative. Den udviklede metode estimerer prævalensen af en forurening i et fødevareparti samt parametrene (middelværdi og standardafvigelse) for fordelingen af bakterieantallet på baggrund af direkte bakterietællinger på agarplader og uden antagelse af en detektionsgrænse. Ved at analysere bakterietællinger fra forurenede og ikke forurenede prøver samlet, kan proportionen af falsk negative tællinger ud af det totale antal negative tællinger estimeres.

Metoden frembringer gode estimater over middelværdier, standardafvigelser og prævalenser, i særdeleshed ved lave prævalensniveauer og forventeligt lave standardafvigelser. Undersøgelsen viser, at en af de vigtigste faktorer til en nøjagtig karakterisering af den samlede mikrobiologiske forurening er en korrekt identifikation og adskillelse af sande og falske negative prøver, og at estimater over prævalens og bakteriekoncentrationer er afhængige og at disse derfor skal estimeres samtidigt.

I manuskript II ("Impact of microbial count distributions on human health risk estimates") undersøges det, hvilken indflydelse den tilpassede fordelingen for bakteriekoncentrationen har på det endelig risikoestimat. Dette er gjort ved to forskellige scenarier for bakteriekoncentrationer og en række forskellige prævalensniveauer. Fire forskellige parametriske fordelinger er blevet anvendt til at undersøge betydningen af at inddrage tilfældige variationer knyttet til bakterietællinger, påvise forskellen mellem at behandle sandt negative som sådan eller som under en given detektionsgrænse, samt vise vigtigheden af at anvende korrekte antagelser om de underliggende fordelinger for bakteriekoncentrationer. Ved at gennemføre et simuleringseksperiment er det muligt at angive forskellen mellem den forventede risiko og det risikoestimat, der opnås ved at anvende en lognormal, en zero-inflated lognormal, en Poisson-gamma og en zero-inflated Poisson-lognormal fordeling. Metoden, beskrevet i manuskript I, er anvendt til at tilpasse den sidstnævnte fordeling.

Resultatet viser at valget af sandsynlighedsfordeling til at beskrive bakteriekoncentrationen i fødevaren i detailleddet har betydning for risikoestimatet og afhænger både af bakteriekoncentration og prævalens, men at valget generelt betyder mere jo højere prævalensniveauet og koncentrationen er. Anvendelse af zero-inflation har også en tendens til at forbedre nøjagtigheden af risikoskøn.

I manuscript III ("Variability and uncertainty in the evaluation of predictive models with literature data – consequences to quantitative microbiological risk assessment") vurderes det, hvordan forskellige vækstvilkår, som anvendt i publicerede datasæt, påvirker resultaterne af en vækstmodel sammenlignet med de resultater, der opnås med de data der blev anvendt til at udvikle selve modellen. Betydningen af antal observationer, temperaturforhold, vandaktivitet og pH, tilstedeværelse eller fravær af mælkesyre i vækstmiljøet, anvendelse af en patogen stamme eller ej, samt typen af vækstmiljø på modellens resultater blev analyseret. Modellens præstationsevne blev målt som Dif_{Af}, forskellen mellem modellens nøjagtighedsfaktor udregnet med de data der blev anvendt til at lave modellen ($A_{f\ original}$) og en nøjagtighedsfaktor, bestemt på basis af et uafhængigt datasæt ($A_{f\ evaluation}$). Undersøgelen er lavet med en "square root-type model" for vækstraten af *Escherichia coli* på baggrund af fire miljøfaktorer og de samme litteraturdata som tidligere blev anvendt til at evaluere modellen. Det er hypotesen, at $A_{f\ original}$, vil afspejle den optimale præstation af modellen, og at Dif_{Af} reduceres og bliver mindre variabel jo mere betingelserne bag et uafhængigt datasæt nærmer sig det datasæt, der blev anvendt til at udvikle modellen. Fordelingen af Dif_{Af} værdier, opnået på baggrund af forskellige datasæt sammenlignes grafisk og statistisk.

Resultaterne indikerer at når anvendelse af prædiktive modeller, der er udviklet under kontrollerede eksperimentelle vilkår, bliver valideret med uafhængige datasæt fra litteraturen, så er det en forudsætning for at minimere variation i model outputtet, at datasættene indeholder et stort antal observationer og at de er baseret på tilsvarende vækstvilkår som den prædiktive model er udviklet under. Ved at mindske denne variation, mindskes også usikkerhed og variation fra de prædiktive modeller i den samlede QMRA analysen, hvilket øger præcisionen af risikoestimatet.

Det konkluderes at denne afhandling: bidrager til at afdække hvilken betydning analyse af de mikrobiologiske data kan have på en QMRA, fremlægger en ny og nøjagtig metode til at tilpasse fordelinger til mikrobiologiske data, og foreslår retningslinjer for, hvordan man kan vælge egnede publicerede datasæt til validering af prædiktive modeller for mikrobiel vækst og overlevelse, før de anvendes i en QMRA.

Perspektiver for det fremtidige arbejde inkludere validering af metoden udviklet i Manuskript I med data indsamlet fra 'den virkelige verden', og at præsentere metoden som et værktøj til andre forskere fx som en arbejdspakke i statistikprogrammet R. Ligeledes bør man blive enige om en standardiseret metode til rapportering af kvantitative mikrobiologiske data, det tydeligt beskriver dataindsamlingsprocessen. En

videreudvikling af arbejdet i Manuskript II vil gøre det muligt, at underbygge konklusionerne om hvilken indflydelse forskellige fordelinger har på det endelige risikoestimat. Som en opfølgning på Manuskript III kan der gennemføres et simuleringsstudie med henblik på undersøge i hvilken grad målrettet udvikling af QMRA metoder og validering af prædiktive modeller er nødvendige for et retvisende risikoestimat. Fremtidige behov i fødevaremikrobiologi og QMRA omfatter udviklingen af egnede statistiske metoder til at analysere data fra de forskellige "omics" teknologier, tilpasning af den nuværende struktur i QMRA modeller, så disse kan håndtere sådanne data, samt vurdering af variation og usikkerhed på disse data.

Resumo

Doenças de origem alimentar acarretam consequências de saúde pública, sociais, políticas e económicas importantes. A avaliação de risco microbiológico quantitativa (quantitative microbiological risk assessment - QMRA) é uma ferramenta científica utilizada para estimar o risco que microorganismos patogénicos de origem alimentar representam para a saúde humana. QMRAs estimam o número de casos de infecção humana de origem alimentar ou o número de casos de doença devido à ingestão de um microorganismo patogénico específico, transmitido por meio de um alimento específico; são também utilizados para avaliar o efeito de diferentes medidas de controlo na cadeia alimentar. As autoridades públicas baseiam as suas políticas de gestão de risco nos resultados dos estudos de avaliação de risco. Consequentemente, estes precisam ser transparentes e afetados por um nível mínimo de imprecisão.

A possível exposição e infecção por microorganismos de origem alimentar dependem, entre outros fatores, das concentrações microbianas nos produtos alimentares e do comportamento microbiano (crescimento, sobrevivência e transferência) ao longo da cadeia alimentar. Ambos os fatores são, portanto, dados importantes para a QMRA. Como concentrações microbianas variam entre diferentes amostras do mesmo lote de alimentos, distribuições de probabilidade são utilizadas para descrever essas concentrações na QMRA. E como o comportamento microbiano depende de propriedades intrínsecas dos alimentos e de variáveis ambientais extrínsecas e consequentemente varia de acordo com as condições de armazenamento dos alimentos, modelos preditivos de crescimento e sobrevivência bacteriana que têm esses fatores em consideração são utilizados em QMRA para descrever mudanças esperadas nas concentrações bacterianas.

Tanto as distribuições de probabilidade como os modelos preditivos podem contribuir para a imprecisão da QMRA: por um lado, existem várias alternativas de distribuição disponíveis para descrever concentrações e vários métodos para aplicar as distribuições a dados microbiológicos; por outro lado, os modelos preditivos são construídos com base em experiências que investigam o comportamento microbiano sob condições de laboratório controladas, e podem não ser adequados à aplicação no contexto de produtos alimentares reais. Por isso, estes modelos têm que ser validados com dados independentes de condições reais antes do seu uso em QMRA.

O objetivo geral do trabalho apresentado nesta tese é estudar diferentes fatores relacionados com os dados microbiológicos quantitativos que podem ter um impacto no resultado da QMRA, a fim de encontrar soluções adequadas que limitem a imprecisão das estimativas de risco. Um novo método de aplicacao de uma distribuição a dados microbiológicos, que estima tanto a prevalência como a distribuição de concentrações é desenvolvido (manuscrito I). Distribuições de probabilidade diferentes são utilizadas para descrever as concentrações num modelo simples de QMRA, e as estimativas de risco obtidas são comparadas (manuscrito II). A precisão da previsão de um modelo de crescimento microbiano com diferentes *datasets* é comparada, a fim de identificar os diferentes fatores relacionados com a colheita experimental de dados que têm um impacto relevante sobre o processo de avaliação do modelo (manuscrito III).

No manuscrito I ("Fitting a distribution to microbial counts: making sense of zeroes "), a seguinte hipótese foi definida e testada: quando contagens de zero microorganismos com origem ao acaso em alimentos contaminados (zeros “artificiais”) não são separadas dos zeros "verdadeiros" provenientes de produtos não contaminados, as estimativas de prevalência e concentração podem ser imprecisas. Esta imprecisão pode ter um impacto relevante em QMRA, especialmente em situações em que microrganismos altamente patogénicos estão envolvidos e em que o crescimento microbiano pode ocorrer ao longo da cadeia alimentar. Por isso, um método que fornece estimativas precisas dos parâmetros de concentração e diferencia entre zeros artificiais e verdadeiros é desenvolvido, estimando também de forma precisa a prevalência.

É demonstrado que, dependendo da distribuição original de concentrações e do limite de quantificação (limit of quantification - LOQ) de enumeração microbiana, pode ser incorrecto tratar os zeros artificiais como “censurados” abaixo de um limite de quantificação. O método que é apresentado estima a prevalência de contaminação dentro de um lote de alimentos e os parâmetros (média e desvio padrão) que caracterizam a distribuição das concentracões dentro de um lote, sem assumir um LOQ, e usando os dados brutos de contagem directa como input. Contagens resultantes tanto de unidades amostrais contaminadas como de não contaminadas são analisadas em conjunto, o que permite estimar a proporção de zeros artificiais no total de zeros.

O método produz boas estimativas de média, desvio padrão e prevalência, principalmente a baixos níveis de prevalência real e baixo desvio padrão esperado. Este estudo mostra que uma das chaves para uma caracterização exacta da contaminação microbiana em geral é a identificação correcta e a separação dos zeros verdadeiros e artificiais, e que a estimativa da prevalência e a estimativa da distribuição de concentrações estão interligadas e, por conseguinte, devem ser efectuadas em conjunto.

No manuscrito II ("Impact of microbial count distributions on human health risk estimates "), o impacto da aplicação de distribuições a contagens microbianas nas estimativas de risco é investigado em dois cenários de concentração diferentes e numa variedade de níveis de prevalência. Quatro distribuições paramétricas diferentes são usadas para investigar a importância de contabilizar a aleatoriedade das contagens, a diferença entre tratar os zeros verdadeiros como tal ou como observações censuradas abaixo de um LOQ, e a importância de pressupor corretamente a forma da distribuição subjacente às concentrações. Ao efectuar um exercício de simulação, é possível avaliar a diferença entre o risco esperado e o risco estimado com o uso de uma distribuição lognormal, lognormal zero-inflacionada, Poisson-gama e Poisson-lognormal zero-inflacionada. O método desenvolvido no manuscrito I é utilizado neste estudo para aplicar a última destas distibuições.

Os resultados mostram que o efeito nas estimativas de risco da escolha de diferentes distribuições de probabilidade para descrever concentrações em unidades de retalho depende tanto da concentração como dos níveis de prevalência, mas que, em geral, é maior sob elevados níveis de contaminação microbiana (elevada prevalência e elevada concentração) . Adicionalmente, adoptar a zero-inflação tende a melhorar a precisão da estimativa do risco.

No manuscrito III ("Variability and uncertainty in the evaluation of predictive models with literature data – consequences to quantitative microbiological risk assessment ") é avaliada a forma como diferentes factores associados ao crescimento microbiano inerentes a *datasets* de literatura afetam o desempenho de um modelo de crescimento em relação ao seu desempenho com os dados utilizados para gerá-lo. Os efeitos do número de observações, das gamas de temperatura, actividade da água e do pH sob as quais foram feitas as observações, da presença ou ausência de ácido láctico no meio de crescimento, da utilização de uma estirpe patogénica ou não patogénica e do tipo de meio de crescimento utilizado sobre o desempenho do modelo são analisados. O desempenho do modelo é medido em termos de Dif_{Af} - a diferença entre o factor de precisão do modelo com os dados utilizados para gerá-lo ($A_{f\ original}$) e o factor de precisão com um conjunto de dados independentes ($A_{f\ evaluation}$). O estudo é realizado utilizando um modelo do tipo raiz quadrada para a taxa de crescimento de *Escherichia coli* em resposta a quatro factores ambientais, e os dados de literatura que foram anteriormente utilizados para avaliar o mesmo modelo. A hipótese inicial é a de que o $A_{f\ original}$ reflete o melhor desempenho possível do modelo e, portanto, Dif_{Af} é mais reduzida e apresenta menor variância quando as condições de um conjunto de dados independentes estão mais próximas dos dados que deram origem ao modelo. As distribuições de valores Dif_{Af} obtidos com diferentes conjuntos de dados são comparadas gráfica e estatisticamente.

Os resultados sugerem que, na eventualidade de modelos preditivos desenvolvidos sob condições experimentais controladas serem validados com dados independentes obtidos a partir da literatura, estes conjuntos de dados têm de conter um número elevado de observações e basear-se num meio de

crescimento semelhante, com o fim de reduzir a variação de desempenho do modelo. Ao reduzir essa variação, a contribuição do modelo preditivo com incerteza e variabilidade para a QMRA também diminui, o que afeta positivamente a precisão das estimativas de risco.

Em conclusão, esta tese contribui para o esclarecimento do impacto que a análise dos dados microbianos pode ter em QMRA, fornece um novo método preciso de aplicar uma distribuição a dados microbianos, e sugere diretrizes para a seleção de conjuntos de dados de literatura adequados à validação de modelos preditivos de comportamento microbiano, antes da sua utilização em QMRA.

Perspectivas de trabalho futuro incluem a validação do método desenvolvido no manuscrito I com dados reais, e a sua apresentação como uma ferramenta ao dispor da comunidade científica, por exemplo, por meio do desenvolvimento de uma *package* para o software estatístico R. Além disso, a autora espera ver adotada no futuro uma forma padronizada de relatar contagens microbianas, que especifique claramente as medidas tomadas durante a colheita dos dados. A continuação do trabalho apresentado no manuscrito II permitirá a obtenção de conclusões mais sólidas sobre o impacto geral de diferentes distribuições de frequência nas estimativas de risco. Após a apresentação do manuscrito III, um estudo de simulação poderá ser útil para investigar até que nível são necessários para a estimativa precisa do risco o desenvolvimento e a validação de modelos preditivos direcionados para QMRA. Necessidades futuras em microbiologia alimentar e QMRA incluem o desenvolvimento de métodos estatísticos apropriados para resumir novos tipos de dados obtidos com o uso de diferentes tecnologias "omics", a adaptação da estrutura atual dos estudos de QMRA que lhes permitirá fazer uso desses dados e a avaliação da variabilidade e incerteza inerente aos mesmos.

List of abbreviations

A_f – *accuracy factor*

B_f – *bias factor*

CFU – colony forming units

CPM – consumer-phase model

CTM – cardinal temperature model

CTPM – cardinal temperature pH model

DR – dose-response

ELISA - enzyme-linked immunosorbent assay

GT – generation time

LAB – lactic acid bacteria

LN - lognormal

LOD – limit of detection

LOQ – limit of quantification

MC – Monte-Carlo

MLE- maximum likelihood estimation

MPD – maximum population density

MPN – most probable number

MPRM - modular process risk model

MRA – microbiological risk assessment

PCR - polymerase chain reaction

PGM – Poisson-gamma

PLN – Poisson-lognormal

QMRA – quantitative microbiological risk assessment

sQMRA - swift quantitative microbiological risk assessment

SIEFE - stepwise and interactive evaluation of food safety by an expert system

SSP – seafood spoilage predictor

ziLN – zero-inflated lognormal

ziPLN – zero-inflated Poisson-lognormal

TABLE OF CONTENTS

OUTLINE AND OBJECTIVES

This thesis consists of an introduction section that gives an overview of the subjects under study, and three different parts, each corresponding to an individual manuscript. Each manuscript is followed by a closing section of general conclusions and the thesis ends with a short discussion on future perspectives. The introduction section introduces the subject of quantitative microbiological risk assessment (QMRA), explains the nature of microbial data and its use in QMRA and gives a summarized overview of the past, present and future of the field of predictive microbiology and its integration in QMRA. For the sake of readability, in the sections of discussion, conclusions and future perspectives referencing has been kept to a minimum. The information presented in these sections has been adequately referenced in other sections of the thesis (introduction and manuscripts); no additional references have been used.

Part I presents the study "*Fitting a distribution to microbial counts: making sense of zeros"*, which consisted on the development of a new method of analysis of raw microbial counts that estimates both the prevalence and the distribution of microbial concentrations. Part II presents the study "*Impact of microbial count distributions on human health risk estimates"*, where the impact on risk estimates of the frequency distribution used to describe microbial concentrations in food was assessed. Part III presents the study "V*ariability and uncertainty in the evaluation of predictive models with literature data – consequences to quantitative microbiological risk assessment",* which investigated how factors inherent to experimental data collection can affect the measure that is commonly used to assess the accuracy of predictive microbiology models.

INTRODUCTION

i. Quantitative microbiological risk assessment

Zoonotic diseases are often transmitted to humans through production, handling and consumption of contaminated foods. Examples of pathogenic microorganisms that cause important foodborne disease in humans are *Campylobacter jejuni* and *coli*, *Salmonella* spp., *Yersinia enterocolitica*, Verocytotoxinogenic *Escherichia coli* (VTEC) and *Listeria monocytogenes*. The presence of these microorganisms in foods needs to be controlled through a complete farm-to-fork approach (Nørrung and Buncic, 2007). In most food processes, there are critical points where food can become a microbiological hazard. Once identified, these points can be specifically controlled and monitored (Havelaar et al., 2008). For their identification, the changes that occur in the microbial composition of food during processing must be understood (Kilsby and Pugh, 1981). Quantitative microbiological risk assessment (QMRA) provides a framework to model microbial changes in food along the food chain, and consequently to estimate the risk of foodborne infection associated to consumption of contaminated food (Havelaar et al., 2008). Such information is essential to optimize risk management activities (Nørrung and Buncic, 2007).

Risk assessment in general is integrating part of the process of risk analysis, which also includes risk management and risk communication. All three components are interdependent in risk analysis.

Risk assessment combines the knowledge on the nature of the hazard with the likelihood of exposure to that hazard. It includes the steps of *hazard identification*, *hazard characterization*, *exposure assessment* and *risk characterization* (Bassett et al., 2012; CAC, 1999). Additionally, the “Codex Alimentarius” (CAC, 1999) has introduced the steps of statement of purpose, risk profile, documentation and re-assessment as part of risk assessment. The outcome of risk assessment varies and can include risk estimates, ranking of risks, identification of key controlling or risk-generating factors and highlighting of data gaps (Bassett et al., 2012).

Governmental agreements (such as the Agreement on the Application of Sanitary and Phytosanitary Measures, SPS) and the new challenges that food safety managers had to face in recent years, such as an increasing food demand, a demand for fresher, healthier and less processed foods, new developments in food processing and packaging and the international sourcing of products, led them to move towards a risk-based approach (Andersen et al., 2007, Christensen et al., 2013; EFSA, 2010; Havelaar et al., 2008; Nauta, 2001; Nauta et al., 2003, 2005; Ross et al., 2009; Sanaa et al., 2004). For example, in Denmark, after implementation of a risk-based approach for the control of *Campylobacter* in broiler meat (Rosenquist et al., 2003) it was possible to identify important data gaps and the most appropriate risk management actions, which included initiatives to reduce the number of *Campylobacter* infected broiler flocks, directing negative flocks for the production of chilled meat, disseminating consumer information on handling and cooking practices and initiating projects for reducing the number of *Campylobacter* in meat. Furthermore, after adopting a risk-based approach, the allocation of resources for microbiological analysis has been optimized and projects have been developed specifically aimed to monitor the effect of the established management actions (Andersen et al., 2007).

Besides assisting risk managers in their decisions, risk assessment can also be used to assist industry in the design of new and safe foods (Bassett et al., 2012).

Microbiological risk assessment (MRA) can be qualitative or quantitative. Qualitative MRA works with descriptions of the factors affecting risk, and is particularly useful during the step of risk profile, to give an idea of the potential magnitude of the risk and to indicate if a more detailed, quantitative analysis is needed. QMRA works with numerical data and can be classified as deterministic or stochastic. While deterministic studies do not include any element of randomness in their characterization of the process, stochastic studies include randomness in the form of probability distributions, therefore giving a better representation of natural processes, given the randomness that is inherent to real life (Bassett et al., 2012). Figure 1 provides an overview of the types of MRA, from the less informative (top) to the more complex, data-demanding and informative (bottom).

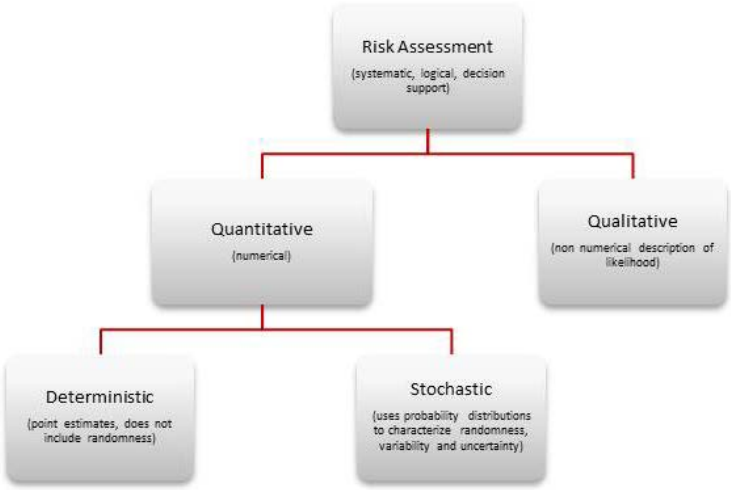

Figure 1: Overview of microbiological risk assessment types. (Adapted from Bassett et al., 2012)

1. QMRA elements

1.1. Hazard identification

Hazard identification is the first step of QMRA and consists on a qualitative description of important information about the hazard, i.e. the agent that potentially causes an adverse health effect. This information includes the pathogen, food product and host interface (types of disease caused, susceptible populations, mode of pathogen-host interaction) (Bassett et al., 2012).

1.2. Hazard characterization

In this step, a description of the adverse effects that may result from the ingestion of the pathogen or its toxin in food is provided. This is usually done using a microbiological dose-response model, which describes the probability of a specified response (infection or illness) from exposure to an ingested dose of a specific

pathogen (or its toxins) in a specified population. In dose-response modeling a single-hit non-threshold approach is often preferred as a more cautious alternative to a threshold based approach. It assumes that a single viable infectious microorganism is able to induce infection. The single-hit approach also assumes the absence of interaction between microorganisms, and the independence of the probability to provoke illness from the size of the dose, the state of the microorganisms, the host and its previous exposure to the microorganism. The dose-response models most frequently used in QMRA are the exponential and the Beta-Poisson (Bassett et al., 2012).

1.3. Exposure Assessment

At this step, the extent of human exposure to the specified microbiological hazard is assessed, by estimating the likelihood and the level of occurrence of the hazard in foods at the time of consumption (Bassett et al., 2012). To fulfill this goal, a food chain risk model is often part of an exposure assessment. This is a common practice both because data are usually available higher up in the chain but not at the point of exposure and because often the objective is to assess the effect of interventions in the food chain. The outcome of exposure assessment is the expected (distribution of the) dose of microorganism that is ingested by the consumer.

1.4. Risk characterization

Risk characterization integrates the information of the three previous steps to provide a sound estimate of risk for a specified population, usually in the form of the expected number of illness cases (Bassett et al., 2012). This estimate is obtained by combining the outcome of the exposure assessment (dose ingested) with the dose-response model of risk characterization (probability of illness dependent on dose), and consumption data (number of servings consumed in the population).

Figure 2 shows a summary of the basic steps of QMRA, with a list of key questions that indicate the scope of each step.

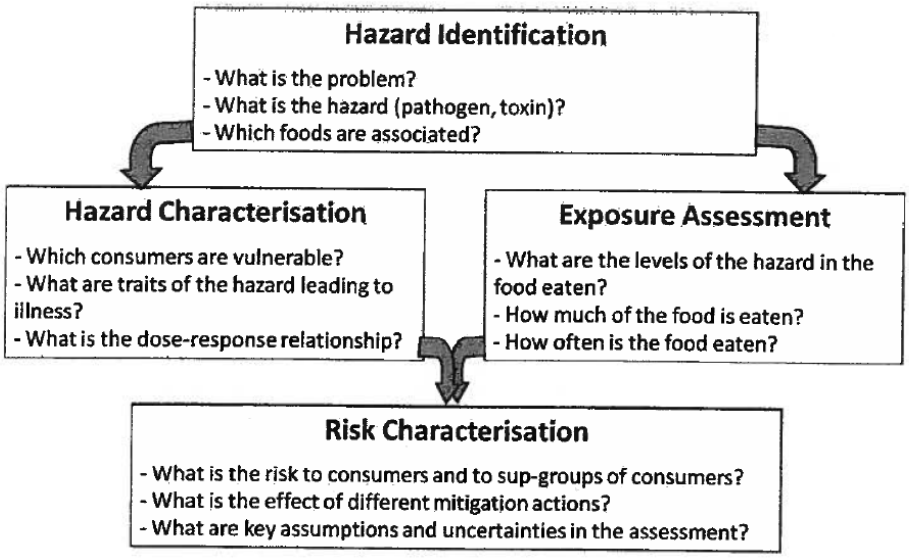

Figure 2: Basic steps of QMRA. (Adapted from Brul et al., 2012)

2. Microbial data in QMRA

For a robust estimation of risk, it is important that the data used as input in QMRA, particularly in the data-demanding exposure assessment step, is abundant, accurate and representative of the reality (Bassett et al., 2012; Busschaert et al., 2011; Christensen et al., 2013). These data include microbial, food, process and consumer data. The variability and uncertainty in all data needs to be clearly reported (Nauta, 2007). To ensure the transparency of the study, the data sources should be well described and the assumptions made should be specified and their potential impact on the outcome needs to be made explicit. Data gaps should be evaluated preferably at the end of the risk assessment in order to investigate the gaps that are most relevant in terms of their contribution to the overall uncertainty of risk (Bassett et al., 2012).

For the exposure assessment, the most data-demanding step of QMRA, key data are 1) the frequency of contamination of foods (prevalence) and the microbial levels in those foods (distribution of microbial concentrations), 2) the survival and growth of microorganisms against processing factors and factors intrinsic to the foods (predictive microbiology models), 3) the transfer of microorganisms between food units, 4) patterns of consumption (serving size and frequency of consumption) and 5) consumer behaviour (consumer-phase models).

This thesis focuses on the effect of prevalence, distribution of microbial concentrations and predictive microbiology models on risk. However, consumer-phase models (CPM) may also have an impact on risk estimation, since they describe the food chain between purchase of the food product at retail and exposure (Christensen et al., 2005; Fisher et al., 2005; Nauta et al., 2008; Nauta and Christensen, 2011). As these models include information on consumer hygiene and handling practices, which is difficult to obtain, they are subject to a high degree of uncertainty that can have a significant impact on the estimate of risk.

3. Variability and uncertainty in QMRA

Data used in QMRA is often derived from laboratory measurements and from samples representing the population of interest. Under ideal circumstances, if a sample is perfectly representative of a population and a measurement method is perfectly precise and unbiased, the variation of the observed data is exclusively affected by variability (Reinders et al., 2003). However, in reality, samples are never fully representative and experiments are imperfect. Also, there is uncertainty associated with the modelling process (although this is difficult to quantify and hence is usually not considered). As a result, QMRA inputs are affected by uncertainty. Variability and uncertainty should hence be integrated in QMRA to quantify imprecision associated to estimates. Those factors should be treated separately (Nauta, 2000, 2007, 2008; Pouillot and Delignette-Muller, 2010); besides the computational advantage that it represents, this also simplifies interpretation of results by risk managers. If only variability is integrated in the calculation of the mean risk, and the uncertainty of the risk estimate is expressed as a confidence interval for that mean (Clough et al., 2005), it will be more straightforward to decide on the need for measures that either 1) decrease the mean risk, by impacting variability, or 2) decrease the uncertainty of risk, by for example collecting additional information where data gaps exist. For example, if uncertainty shows to have a minor impact on risk estimation, no further expensive and time consuming laboratory measurements need to be carried out (Busschaert et al., 2011).

Variability and uncertainty are typically represented with probability distributions in stochastic QMRA (Nauta, 2000, 2007, 2008). Monte-Carlo (MC) simulation is used when distributions are used as input parameters. A MC simulation consists on the repetition of the same calculation multiple times (iterations). In each iteration, a value is selected from the probability distribution of each input, and the resulting outcome is calculated. In the process of sampling values from the input distributions, the values that are most likely to occur according to the distributions are selected more frequently. When sufficient iterations are performed, a probability distribution of the outcome is obtained (Nauta, 2008; Straver et al., 2007).

In order to account for both variability and uncertainty in a QMRA outcome, a two-dimensional (or second-order) MC simulation needs to be adopted. In this type of simulation, the distributions reflecting variability and uncertainty are sampled separately for calculating the risk. Hence, it is essential to specify for each input in the model if the associated randomness represents uncertainty or variability (Nauta 2000, 2007, 2008; Pouillot and Delignette-Muller, 2010).

3.1. Uncertainty

Uncertainty in QMRA consists on the lack of perfect knowledge about the adequacy of the exposure model (process and consumer phase models) and the dose-response model to reflect real-life conditions and about the parameters used as inputs (Marks et al., 1998; Nauta, 2007; Pouillot and Delignette-Muller, 2010). Whereas it is difficult to know, and therefore to express, the uncertainty about the adequacy of the models to mimic reality, it is usually more manageable to assess the uncertainty associated with input parameters. For example, a probability distribution representing the variability (i.e. the natural variation) of microbial concentrations in food products is inferred from a dataset of experimental observations, therefore the estimates of the distribution parameters are uncertain. This parameter uncertainty causes a deviation between the observed and the true distribution of variability and it is jointly caused by measurement and sampling errors. Measurement error is a consequence of the fact that results are obtained under imperfect experimental conditions and using an arbitrary sample unit size for analysis. Sampling error is a consequence of observing a limited subset of the true population under study (Gonzales-Barron and Butler, 2011).

Usually, the uncertainty associated with absolute risk estimates is high. In order to decrease the impact of uncertainty on risk, the relative risk has been adopted as an alternative outcome of QMRA (Evers and Chardon, 2010; Christensen et al., 2013; Nauta and Christensen, 2011). The relative risk has a smaller level of uncertainty than absolute risk estimates because similar uncertainties are cancelled out if two absolute risk estimates are divided (Nauta and Christensen, 2011).

For a quantitative expression of uncertainty, several methods can be used: Bootstrap resampling method, Bayesian derivation, or expert elicitation (Kennedy and Hart, 2009; Pouillot and Delignette-Muller, 2010; Vicari et al., 2007).

Bootstrap resampling (or bootstrapping) can be non-parametric or parametric. It consists on the random sampling of a number of new samples either from the original dataset (non-parametric) or from a distribution that has been fitted to those data (parametric). The statistic (e.g. mean of a fitted distribution) estimated from each bootstrap sample will differ between samples due to the effect of random sampling. The total number of estimates of that statistic forms a probability distribution that reflects the uncertainty of the statistic (Zhao and Frey, 2004).

Examples of tools that can be used to perform bootstrapping are the packages “mc2d” (Pouillot et al., 2011) and “fitdistrplus” (Delignette-Muller et al., 2010), which have been developed for the statistical software R (R Development Core Team). The first can perform non-parametric bootstrapping from empirical distributions and the latter can derive parametric or non-parametric bootstrap samples from distributions that have been fitted to quantitative or semi-quantitative data (Pouillot and Delignette-Muller, 2010). If uncertainty is defined from expert knowledge, it can also be expressed by probability distributions. Triangular and PERT distributions are commonly used for this purpose, as they are both defined by the parameters minimum, most likely and maximum values (van Hauwermeiren and Vose, 2009). The Bayesian approach to describe uncertainty in QMRA can be used at two different levels of complexity. It can be limited to describe the uncertainty of parameters defining distributions of variability. For example, if a lognormal distribution describes the variability of concentrations, its parameters mean and standard deviation can be defined as hyperparameters, i.e. be described by probability distributions (Nauta, 2008). At a more complex level,

Bayesian inference can be used for propagating both uncertainty and variability throughout the full QMRA model (Greiner et al., 2013; Rigaux et al., 2013).

3.2. Variability

Variability in general can be defined as the true, irreducible variation or heterogeneity of the population of subjects considered (Nauta, 2000, 2007). In QMRA, a typical source of variability is the natural variation of the microbial concentrations among food units. When concentration is first used as an input in a QMRA model, its variability has been assigned to various factors: true heterogeneity between food units, stochasticity or randomness between food units and clustering at the cellular level within each food (Jongenburger et al., 2012a; Nauta, 2007, Reinders et al., 2003). However, as the QMRA develops further "down" along the food process, the variability of microbial concentrations suffers from additional influences, such as cross-contamination (or microbial transfer between units), microbial growth and death and the effect of food processing practices (Gonzales-Barron and Butler, 2011).

Therefore, variability in QMRA needs to be described at different levels. For example, in the beginning of the process, variability of microbial concentrations among food units is usually described by a frequency distribution. The choice of this distribution must account for the existence of contamination clustering, in case it is present (Gonzales-Barron and Butler, 2011), for the existence of a high number of uncontaminated food units, when the contamination level is low, and for the possibility of the existence of rare but highly contaminated food units, as these have a large influence on risk (Nauta and Christensen, 2011). Then, at stages of the process where microbial growth or death can occur, it is important to identify the variability associated with different aspects of those microbial behaviours (such as initial microbial population size or specific growth rate) (Shorten et al., 2006). Finally, when the product reaches the consumer's home it is still important to account for the variability in the effects of meal preparation, as these can still largely impact the risk estimation (Nauta and Christensen, 2011).

4. Structured approaches to QMRA

Since the development of QMRA models can easily become cumbersome, structured and simplified approaches to modelling have been developed that aim to facilitate the task of risk assessors.

The stepwise and interactive evaluation of food safety by an expert system (SIEFE) (van Gerwen and Zwietering, 1998; van Gerwen et al., 2000a; van Gerwen et al., 2000b) is a stepwise approach that begins with rough risk estimations using simple models in order to identify the most important phenomena. Those are subsequently described by more sophisticated models and data to improve the predictions. In a last stage, stochastic aspects can be included (Zwietering and Nauta, 2007).

As the transmission of microorganisms through the food chain is often complex, in part due to the existence of a large variety of processes that need to be modelled independently, Nauta (2001, 2002, 2008) introduced the modular process risk model (MPRM) as a structured approach to exposure assessment, extending the earlier introduced concept of Process Risk Model (Cassin et al., 1998). The MPRM concept consists on splitting the food chain into modules corresponding to the independent processes that need to be modelled, which facilitates the data collection and modelling for the construction of the final QMRA. This concept assumes that there are six fundamental events that may cause the change of the microbial numbers during a food process or module. These include two microbial events - growth and inactivation - and four food handling events – mixing, partitioning, removal and cross-contamination. For each MPRM module, the basic processes that are in place must be defined and quantitatively modelled (Nauta, 2008). Table 1 presents an example of a MPRM model structure defined as a function of the occurrence of the basic processes.

Processing step (module)	Basic process	Basic process implemented in MPRM model
Adding ingredients	Mixing/ Inactivation	Start
Vaccum packing	Partitioning	Partitioning
Pasteurization	Inactivation	Inactivation
Cooling	Growth / Inactivation	Growth / Inactivation
Storage	Growth	Growth
Transport	Growth	Growth
Retail	Growth	Growth
Transport	Growth	Growth
Home fridge	Growth	Growth

Table 1: MPRM model structure of the food pathway of an exposure assessment of *Bacilus cereus* in a vegetable puree (starting after the cooking and blending process steps), as a series of basic processes (Adapted from Nauta, 2008).

Although the fact that MPRM approach simplifies the food chain as a series of linked modules can be seen as a disadvantage, this is also what makes it a tool for a simpler and more manageable way of conducting QMRA studies. Additionally, it has the great advantage that at the end of each module outputs of microbial number and prevalence can be obtained, which are dependent on the basic processes occurring at the module (see table 2).

Basic process	Effect on prevalence	Effect on total number of microorganisms
Growth	=	+
Inactivation	-	-
Mixing	+	=
Partitioning	-	=
Removal	-	-
Cross-contamination	+	=

Table 2: Qualitative effect on prevalence and total number of microorganisms in the system of the basic processes of MPRM (= no effect; + increase; - decrease) (Adapted from Nauta, 2008)

As the elaboration of a QMRA study is a time-consuming and expensive process, it is important to prioritize the food hazards that demand such kind of approach. To facilitate this selection, tools for risk ranking can be used. A list of risk ranking tools is presented in Brul et al. (2012). An example of such a tool is the swift quantitative microbiological risk assessment (sQMRA) (Evers and Chardon, 2010). It is a “simplified QMRA model especially aimed at comparing the risk of pathogen-food product combinations”. It is a deterministic model, from retail to consumer, which provides relative risk estimates (compared with results from a full-scale QMRA).

Ultimately, it must be questioned to what extent QMRA models need to describe exhaustively the process that leads to exposure and disease. The answer will be most likely: *it depends.* While to assess effects of different inputs in terms of relative risk a simple model may be enough, if the aim is to study the effect on risk of specific control measures applied at strategic stages of the process, a detailed model with a thorough description of that process may be necessary (Nauta and Christensen, 2011; Zwietering, 2009).

5. Towards Bayesian modelling in QMRA

Although MC simulation has been widely used to implement QMRA models, it is not free from disadvantages. It requires precise probability distributions for the input parameters and a large number of iterations in complex models, it is not appropriate to model nonlinear dependencies between variables, it may be inadequate when large events have a major effect on the estimated risk (Nauta, 2008) and it is unidirectional (the dependency relationship between the variables cannot be inverted) (Rigaux et al., 2013).

The Bayesian method has therefore been increasingly indicated as a good alternative to MC simulation (Donald et al., 2009; Greiner et al., 2013; Parsons et al., 2005; Rigaux et al., 2013). Whereas in MC, many distinct probabilistic models are combined to obtain the outcome, with Bayesian inference the knowledge from different parts of the QMRA model is inter-related, as the outcome consists on a joint posterior distribution that is formed across all parameters, processes and data that are part of the QMRA model (Greiner et al., 2013). Another important difference between MC and Bayesian implementation of QMRA models is the way how variability and uncertainty are described. With MC, two-dimensional models are developed where variability and uncertainty are described in distinct loops, while with Bayesian inference the structure of the model allows to define uncertainty and variability at the same level (Greiner et al., 2013). Finally, a major advantage of Bayesian models is the fact that they allow the update of prior beliefs, i.e. the inference about information entered at the top of the model from results obtained further down in the model structure. This may contribute to a considerable reduction in the impact of uncertainty in QMRA (Rigaux et al., 2013). Although implementation challenges may delay the adoption of Bayesian methods, a gradual transition from MC- to Bayesian-based approaches is expected to happen in the future of QMRA.

ii. Microbial data

There are several reasons that justify microbial examination of food. This includes determining its microbiological quality, verifying if it meets established microbiological criteria and investigating the cause of its spoilage or the presence of pathogenic microorganisms (Adams and Moss, 1995).

The definition of the microbiological composition of food includes the types of microorganism present, their levels, their spatial distribution in food units and their distribution throughout a food lot (Kilsby and Pugh, 1981). When microbiological examination is carried out, detection methods are used to investigate the presence or absence of a microorganism, and enumeration methods are applied to measure its quantity.

1. Detection methods

Detection methods are usually qualitative, as their result is expressed as presence or absence of a specific microorganism in a test portion of a certain size (in g or ml) (ISO 10272-1, 2006; NMKL119, 2007). Typically, in a detection test a sample unit goes subsequently through the experimental steps of pre-enrichment, selective enrichment (where the selective media contains compounds that specifically select for the multiplication of the desired microorganism or represses the competing microflora), isolation of the colonies of interest (to obtain a pure-culture of the microorganism), and biochemical confirmation of the isolated colonies.

2. Enumeration methods

The result of an enumeration test is expressed as the number of microorganisms per ml or g of product (ISO 10272-2, 2006; NMKL119, 2007). This result can be quantitative or semi-quantitative (interval data). Two main enumeration methods are used in food microbiology to obtain counts of bacterial colony forming units (CFU) that represent, respectively, examples of a quantitative and a semi-quantitative enumeration test: plate counts and most probable number (MPN) counts. MPN counts (Cochran, 1950) are mostly used to determine low bacterial numbers (FDA, 2006) below the quantification level of direct plating, whereas direct plate counts are used for higher contamination levels.

Usually, the enumeration by MPN principle includes culturing encouraging the bacteria to multiply during growth in selective or non-selective media and either in broth (liquid) medium or on agar plates. Whether isolation and characterization is a part of the procedures depends on the organisms and the media used.

2.1. Plate counts

In direct plate counting, inocula from several dilutions of a homogenised sample are directly plated on agar plates, which medium is selective when appropriate. Incubation allows the bacteria to grow, with one CFU building a colony on the agar plate that is detectable by the human eye. Incubation can be simultaneously seen as necessary and limiting, since one of the factors that limit the precision of a colony count is the number of colonies present on a plate. Hence, it is generally accepted that reasonably accurate results are obtained when plates contain between a minimum and a maximum number of colonies (e.g. 30-300 colonies) (Adams and Moss, 1995). The upper limit is due to the expected problems in reading caused by plate crowding, while the lower limit is related to an arbitrarily chosen limit of quantification (LOQ) (Niemelä, 2002). To obtain plates within these numbers of colonies, it is often necessary to dilute the sample units and to select only the plates containing an acceptable number of colonies after incubation. The most commonly used dilution technique is the ten-fold dilution series (Adams and Moss, 1995). Since preparation of decimal dilutions causes an increased variation of the observed counts, multiple platings per dilution are usually performed to decrease that error (Reinders et al., 2002). Furthermore, not all colonies present are of the

microorganism of interest. Therefore, a subsample of the presumptive colonies counted is often subjected to a confirmatory test.

The outcome of plate counts is commonly a point estimate of the original bacterial concentration, back-calculated as a weighted average of the observed number of CFU on the plates of the serial dilutions, where the dilution factors are accounted for in the calculation.

Extrapolation from colony counts to number of bacteria depends on the assumption that a colony is derived from a single CFU. Additionally, for the estimation of CFU concentration from CFU counts, it is necessary to assume that the food is homogeneous so that the test portion actually studied is representative of the whole product (Adams and Moss, 1995).

2.2. Most probable number

The MPN method implies the dilution of a sample unit to a degree where inocula in the tubes will only sometimes contain viable microorganisms. The outcome of the method is the number of tubes with growth at each dilution. The interpretation of this number leads to a statistical-based estimate of the original concentration of microorganisms in the sample unit – the MPN is the concentration value that makes the observed outcome (number of tubes with growth per dilution) most probable (FDA, 2006). This number is usually associated with a 95% confidence interval, so that there is at least 95% probability that the confidence interval associated with the MPN will enclose the actual concentration (Garthrigth and Blodgett, 2003). Hence, the MPN method is usually classified as semi-quantitative, rather than quantitative.

Although often used to enumerate microorganisms in food, this method is associated to several assumptions: that the microorganisms are homogeneously distributed (not clustered) within the sample unit, that microorganisms do not repel each other, that in every inoculum containing at least one viable microorganism growth will be observed and that individual tubes of the same sample unit are independent (FDA, 2006). Also, it has been associated with an overestimation of the original concentration when the estimated concentrations are not expressed as logarithms (FDA, 2006; Krämer et al., 2011). Two possible explanations for this overestimation may be that 1) the liquid media used in MPN provides better growth conditions than the agar plates in plate counts and 2) the MPN is interpreted as expressing CFU concentration when in fact it expresses cell concentration, because one CFU can consist of more than one cell (Krämer et al., 2010). Finally, the estimates can be influenced by improbable outcomes (e.g. multiple positive tubes at high dilutions and very few or no positives at lower dilutions) and inconclusive tubes (FDA, 2006).

2.3. Alternative methods

There are alternative methods to detect and count specific microorganisms in food, such as the immunological method enzyme-linked immunosorbent assay (ELISA) and the DNA based method polymerase chain reaction (PCR). These methods have some drawbacks: ELISA commercial kits require the presence of a very high number of microorganisms and cross reactivity with bacteria other than that of concern cannot be discarded (Adams and Moss, 1995); PCR also requires an enrichment step for the target microorganism to produce sufficient nucleic acid to reach a certain threshold of copies of the target sequence (Adams and Moss, 1995). However, they have been used as a complement to plate counts or MPN; for example, PCR has been typically used as a confirmatory test. Recently, PCR has also been indicated as an alternative to the classical MPN culture procedure to enumerate small numbers of CFU, in combination with a fast enrichment step (Krämer et al., 2010). For a sensitive quantification of bacteria with a PCR-based method, an enrichment step prior do DNA extraction which inhibits the growth of background flora and simultaneously recovers and multiplies the cells of interest in a standardized manner is necessary.

Compared to MPN, the new enrichment-PCR method has the advantage of having a higher throughput, less laborious work, faster results and a lower cost (since selective enrichment, selective plating and confirmation are not necessary) (Krämer et al., 2010).

3. Microbial data in the context of food microbiology

In food microbiology, microbial data are collected with two main purposes: to investigate the presence or absence of microorganisms and to determine their concentration. The first is the goal of prevalence studies, and the latter the goal of quantification studies (Lorimer and Kiermeier, 2007). To fully characterize the microbial contamination of food, results of both studies are needed, as microbial contamination is often reported as the prevalence and the mean concentration of positive samples (Lorimer and Kiermeier, 2007). In order to obtain these results, in particular if the microorganisms occur at a low prevalence, microbial analysis can be performed as a detection test applied to a complete sample set of food products, followed by an enumeration method applied to the positive sample units (Pouillot et al., 2013).

3.1. Limits of detection and quantification

Both detection and quantification tests are affected by limits, below which it is not possible to detect or quantify the microorganism of interest (Lorimer and Kiermeier, 2007). Those limits are defined as the limit of detection (LOD) - the minimum concentration required in a food product for a detection test to result as "presence" (Busschaert et al., 2010; Evers et al., 2010) – and the limit of quantification (LOQ) - the minimum concentration required to obtain a count higher than zero in an enumeration test (Busschaert et al., 2010). LOD and LOQ can be established either experimentally (by determining the highest concentration in which the microorganism is detected/quantified) or theoretically (as the presence of one microorganism in the analyzed test portion – e.g. 1 CFU in 10g = 0.1 CFU/g LOD) (Evers et al., 2010).

Such system of defining thresholds most likely leads to an arbitrary interpretation of results, since detection and quantification are both probabilistic processes affected by randomness. This means that irrespective of the limit value defined, "a negative test result does not imply with great certainty that the concentration of the microorganism is lower than the limit" (Evers et al., 2010). Hence, "absence" in a detection test and "zero" in an enumeration test may consist on artificial below-the-limit results. It is not clear in literature what the most appropriate designation for each type of artificial result is. In this thesis, for the sake of clearness, we decided to differentiate them as *non-detect* when referring to an artificial absence and *artificial zero* when referring to an artificial count of zero. Both are *left-censored* results – observations that are not quantified, but are assumed to be less than a threshold value (Helsel, 2006).

The amount of non-detects and artificial zeroes obtained in microbiological analysis depends mostly on the sensitivity and specificity of the method of analysis used (Currie, 1968; Nauta et l., 2009a), on sample size and test portion size (Straver et al., 2007) and on randomness. Some studies have demonstrated these relationships. For example, Nauta et al. (2009b) observed a difference of 20% in the estimated prevalence of *Campylobacter* on broiler meat between detection protocols with and without an enrichment step and Evers et al. (2010) demonstrated that detection is a probabilistic process and hence is influenced by randomness.

It would therefore be beneficial for the accuracy of the interpretation of microbial data if instead of using arbitrary LOD and LOQ values, factors like sensitivity and specificity, test portion size and randomness would be accounted for in data analysis.

3.2. From plate counts to concentrations

The measurement of interest in food microbiology is not the absolute number of CFU (microbial count) but the bacterial concentration, or the number of CFU per area (cm^2), volume (ml) or weight (g) of food, depending on the food product of interest.

There are several ways of estimating bacterial concentrations given microbial count data. The most common method is to convert CFU counts into concentration values for each sample unit analysed and then find the average and standard deviation of the concentrations of all units. The concentration estimate from an individual sample unit is derived from the CFU counts observed for that unit at the different dilution steps. A weighted average of the counts across the different dilutions that considers the different test portions, is the method used to obtain a concentration estimate. Often there are assumptions behind this estimation procedure, such as excluding plates with counts lower than 10 or higher than 100 CFU from the calculations.

This method has two main caveats: the arbitrary selection of dilution steps to include in the calculation and the use of summarized data (the weighted average of counts) with the consequent loss of information that it represents.

In an effort to overcome these drawbacks, the scientific society has recently started to develop methods of estimating concentrations that make use of raw (non-summarized) count data (Gonzales-Barron et al., 2010; Gonzales-Barron and Butler, 2011) and in some cases, extra laboratory information. Clough et al. (2005) developed a method that incorporates both plate counts and information from PCR confirmatory tests, and several recent studies have used raw plate counts together with detection results (Commeau et al.,2012; Pouillot et al., 2013; Williams and Ebel, 2012a).

4. Variability and uncertainty of microbial counts

Quantitative determinations derived by cultivation of microorganisms, such as microbial counts, are affected by both variability and uncertainty. Variations caused by methodology represent uncertainty, whereas variations due to true differences between sample units represent variability.

It is important to quantify the contribution of different sources of uncertainty and variability to the overall variation of microbial counts, in order to prioritize the factors that must necessarily be assessed to obtain fairly accurate count estimates. For example, different homogenization methods (stomacher and blender) contribute with different systematic errors or bias (uncertainty) for microbial enumeration. However, those errors might be negligible when compared to the natural variation between sample units (variability) (Reinders et al., 2002). On the other hand, the natural variation between sample units assumes different importance depending on the food and microorganism under analysis. When the target microorganism occurs in clusters, there is often a significant error associated to sampling due to a high variation in counts between sample units (Reinders et al., 2002). Furthermore, different aspects that can contribute to a specific kind of uncertainty or variability must be accounted for, and their relative importance assessed. For example, when counting colonies on a plate, individual experimenters have their own sensitivity and specificity. However, whereas reliable estimates require high experimental sensitivity, the assumed specificity, on the other hand, appears to have relatively little effect on the uncertainty of the results (Clough et al., 2005).

4.1. Uncertainty

Because a count usually results from a succession of laboratory steps, and each of these steps contributes individually to the overall uncertainty of the observation, the total uncertainty of that count results from the combination of the uncertainties inherent to each step.

In microbiology, several factors can impact the total (or combined) uncertainty: the sampling procedure, the inoculum volume, the dilution factor, particle statistics (or the assumption of a perfectly mixed suspension where the number of CFU observed can be described by a Poisson distribution), reading and confirmation of counts (Niemelä, 2002, 2003; Reinders et al., 2002). Additionally to these sources of individual uncertainty, there are also uncertainties related to systematic errors. Systematic errors that may cause a bias between the observed number of colonies and the true number of viable bacteria on a plate are the occasional inability of a viable cell to express itself as a recognisable colony, the overlap of neighbouring colonies on a plate, systematically deviant style of counting, the decreased yield of a certain selective medium used for enrichment or the method used for sample homogenization (Niemelä, 2002, 2003; Reinders et al., 2002).

4.2. Variability

Variability, or the natural variation of concentrations, can be measured at different levels. One type of variability is the so-called stochastic variability, which causes different sample units to have different numbers of microorganisms simply due to randomness. Then, when a specific food lot is selected for analysis, there is usually a distribution of concentrations inherent to this lot. Different food units from the same lot will not have exactly the same concentration. This variation consists on the within-lot variability. And last, different food lots may present different levels of contamination due to different food origins or different histories of processing and storage. Therefore, there is also between-lot variability.

The spatial distribution of microorganisms in food contributes to a share of the observed variation in microbial counts, which may be assigned to the stochastic variability. Different spatial distributions (clustered, random or regular) affect the dispersion of the count data and therefore correspond to different frequency distributions (Bassett et al., 2010; Jongenburger et al., 2012a),. Hence, when analysing specific food products it is important to work with its natural contamination instead of an artificial contamination, so that the investigated population has natural characteristics.

5. Future microbial data

Tools currently used to predict the fate of pathogenic microorganisms in the food chain, such as detection and enumeration methods, do not provide insights into the underlying molecular mechanisms that define microbial behaviour. However, recent technological advances in microbiology have enlarged the understanding of those mechanisms. These advances include whole genome sequencing analysis, transcription and protein analysis and assessment of metabolic profiles – tools jointly termed as “omics” technologies.

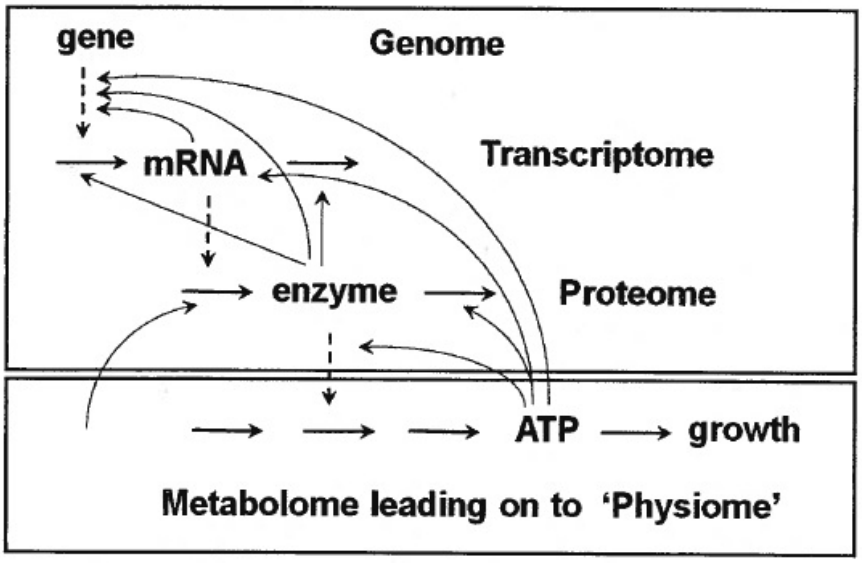

Figure 3: Levels of cellular organisation reflecting sites of action of the "omics" technologies (Genome, Transcriptome, Proteome, Metabolome and Physiome). The "Physiome" reflects the incorporation of all levels of analysis in a description of cellular physiology. (Adapted from Brul et al., 2012)

In order to structure the "omics" data and appropriately link the different kinds of information, bioinformatics is a crucial discipline, which is increasingly being introduced in food microbiology, as "omics" data are expected to play an important role both in microbial detection and enumeration in the near future (Brul et al., 2012).

iii. Microbial data in quantitative microbiological risk assessment

In quantitative microbiological risk assessment (QMRA), microbial data is used as input in exposure assessment in two different ways: as prevalence and as concentration (Fischer et al., 2005; Havelaar et al., 2008; Rieu et al., 2007). These inputs are integrated with other components of an exposure model, leading to a final output in the form of a dose per food serving, to which consumers are exposed. When biased 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). Therefore, in a stochastic approach to risk assessment where uncertainty and variability are taken into consideration, point estimates of either prevalence or concentration are undesired. For one lot of food products, prevalence needs to be reported together with a confidence interval that expresses uncertainty, and concentrations are typically input as a probability distribution that pictures the within-lot variability of concentrations. Despite the need for both inputs, it is not clear how prevalence and concentration should be modelled while analysing microbial data, and there are several methods available. Frequently, prevalence and concentration studies are conducted separately. Prevalence is estimated with detection tests and therefore it is not taken into account while fitting a frequency distribution to concentrations obtained with enumeration methods. These are characterized by fitting a parametric distribution to concentration estimates back-calculated from colony forming units (CFU) counts, obtained from a sample of food products. This sample, in turn, represents the population of products, among which the concentrations are distributed with a certain (unknown) frequency. The fitted distribution is the nearest guess of that original distribution.

1. Fitting distributions to microbial data

Counts obtained by enumeration are commonly converted to base 10 logarithms before any mathematical calculations (Niemelä, 2002) and microbial concentrations in food have been often considered to be lognormally distributed, i.e., their base 10 logarithms follow a normal distribution (Busschaert et al., 2010; Crépet et al., 2007; Gilchris et al., 1973; Kilsby and Pugh, 1981; Reinders et al., 2002, 2003). However, a lognormal (LN) distribution has been considered empirical rather than mechanistic, i.e. it does not provide insight into the causes of variation of the data beyond randomness. This type of information is of particular relevance for data to be used in QMRA (Reinders et al., 2002). It is important to know which mechanisms contribute, and in which magnitude, to the final distribution of microorganisms in food, since the adoption of more strict control measures of these mechanisms can eventually contribute to a decrease in risk. Also, although the LN seems to be an acceptable assumption for highly contaminated populations where there is a negligible probability of obtaining a zero CFU count from a contaminated unit (Bassett et al., 2010), populations of foodborne pathogens pose a challenge to this approach. Their frequency distribution in foods is characterized by a high probability of low concentrations, which likely originate zero counts. These counts result in zero back-calculated concentration estimates. This characteristic complicates the fulfilment of lognormality, as the LN distribution does not allow zero as an outcome and assigns probability to fractional numbers, which is not realistic at the low concentration level, where the difference between successive integers is high (Bassett et al., 2010, Jongenburger et al., 2012b).

Recently, many alternatives to the LN distribution have been discussed and proposed to represent more appropriately microbial contamination data with low concentrations (Bassett et al., 2010; Gonzales-Barron et al., 2010). Discrete (or count) distributions, particularly generalizations of the Poisson, like the Poisson-gamma (PGM) (also called negative binomial) and the Poisson-lognormal (PLN), have been indicated as better alternatives to model concentrations with a considerable amount of low numbers, due to their ability to model count data with over-dispersion (large variance). Important advantages of these distributions are that they are fit directly to CFU counts and not to back-calculated concentration estimates and that their mechanistic nature meets the demands of QMRA models. Although generalized Poisson distributions may account for a large variance of the count data, they do not necessarily account for an excess of zeroes (Gonzales-Barron et al., 2010). Therefore, zero-inflated distributions have been considered appropriate to

model data with a substantial amount of zeroes, as they are more adapted to extra zero counts than simple count distributions.

Bassett et al. (2010) presented a list of criteria that probability distributions must meet to be suitable to describe microbial data: 1) the outcome should not be negative, 2) it should allow zero as an outcome, 3) the outcome should be discrete numbers only, 4) the distribution should generally approximate the Poisson and 5) the distribution should approximate the LN at high numbers of microorganisms. In the same report, the authors compared several frequency distributions and concluded that the PLN was the most suitable with regards to the five proposed criteria, followed by the PGM. The continuous LN and gamma, and the discrete Poisson distributions failed the suitability criteria.

1.1. Lognormal distribution

Although it does not provide an insight in the underlying mechanisms that cause variation of microbial concentrations, the lognormal distribution has proved in many occasions to be a robust choice for the description of that variation (Reinders et al., 2002). This may be explained by the fact that this distribution is not entirely empirical and also provides a mechanistic approach to model bacterial concentrations. As bacteria usually follow exponential growth, when a sample is taken from a growing culture with a normal distribution of growth rates, the microbial counts are lognormally distributed (Nauta, 2001). Additionally, this distribution has the practicality of reducing the scale of magnitude of counts by taking their logarithm, and the normality assumption allows the use of a wide range of statistical techniques and inferential tests. (Gonzales-Barron and Butler, 2011).

The lognormal distribution form most often used in food microbiology corresponds to the normal distribution of the continuous base 10 logarithms of microbial concentrations (x). It is defined by two parameters – mean (μ) and standard deviation (σ) (van Hauwermeiren and Vose, 2009). The probability density function of a normally distributed x is given by

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where the expected value (or mean) of X is

$$E(X) = \mu \quad (2)$$

and its variance is

$$Var(X) = \sigma^2 \quad (3)$$

Kilsby and Pugh (1981) described the implications that a LN distribution of microbial concentrations may have in food microbiology. They pointed out that, due to the relationship between the geometric mean (mean of the LN distribution μ) and the log of the arithmetic mean of counts (α)

$$log_{10}(\alpha) = \mu + \frac{\sigma^2 \times ln(10)}{2} \quad (4)$$

(where σ^2 is the variance of the LN distribution), the geometric mean will always be smaller than the arithmetic mean, unless the variance is zero. Consequently, the mean logarithm of counts obtained with a LN distribution is likely to be below the logarithm of the true count average. The degree of this underestimation depends upon the variance of the distribution. This implies that changes in the level of microorganisms throughout a process can only be compared in geometric mean of counts if the variance at

different steps of the process stays constant, which is rarely observed at different stages of a food process (Kilsby and Pugh, 1981).

Despite the recent progress in food microbiology towards more mechanistic approaches to model microbial counts, with the consequent questioning and withdraw of the LN distribution for that purpose, the LN has shown in different studies to provide fits comparable to more descriptive distributions (Gonzales-Barron and Butler, 2011; Reinders et al., 2003). However, is it all about goodness of fit? Reinders et al. (2003) concluded that although a LN and a PGM distribution fitted equally well the data, only the latter identified whether the microorganisms were clustered or randomly distributed. Also, Kilsby and Pugh (1981) stressed that investigating a food process without accounting for the changes that occur to the distribution of microorganisms during processing, may fail to identify critical control points in the process. The relevance of mechanistic distributions in food microbiology hence depends on the use that is given to the distribution.

1.2. Generalized Poisson distributions

If experimental uncertainty is negligible, microbial counts show a variation which results from the sum of the actual variation in the quantity of microorganisms in food (variability) and the variations due to the method of enumeration (randomness) (Reinders et al., 2002, 2003).

The assumption that the total variation due to the method of enumeration is Poisson distributed is acceptable for plate counts, since at most of the steps of the method of plate counting, the variation that occurs can be described by an individual Poisson distribution. This is mostly due to the homogenization step that is performed at the beginning of the enumeration process, which contributes to the random nature of counts (Kilsby and Pugh, 1981).

A Poisson distribution models the number of occurrences of a discrete event (e.g. y = CFU count) in a given interval of area or volume (e.g. M=10 g) with a known average rate of λ events (e.g. λ = yM in CFU/g). It is defined by the parameter rate (λ) (van Hauwermeiren and Vose, 2009). With parameter $\lambda > 0$, and for $k = 0, 1, 2, \ldots$ the probability density function of y is given by

$$f(y) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (5)$$

The rate λ is equal to the expected value (or mean) of Y and also to its variance

$$\lambda = E(Y) = Var(Y) \quad (6)$$

It is important to note that describing the whole enumeration process with a single Poisson distribution might not be descriptive enough. Reinders et al. (2002) showed that when decimal dilutions are made, each dilution step is “a random event that influences the number of CFU in subsequent dilutions”. Therefore, a repeated Poisson distribution for each dilution step should be used instead of one simplified Poisson curve. This means that for each dilution step, the rate of λ events in CFU/g is multiplied by the corresponding dilution factor.

Although the Poisson distribution has been appointed as an acceptable choice to describe the variation of CFU counts between different sample units, in reality the counts exhibit more variation than that expected under a Poisson distribution (variance = mean) (Gonzales-Barron et al., 2010). This condition is called *over-dispersion* (variance > mean) and to model it, more flexible models are needed. Over-dispersion in counts exists because the concentrations between sample units taken from one food lot are also subject to variation - the within-lot variability (Gonzales-Barron and Butler, 2011; Reinders et al., 2002). As concentrations (in CFU per g or ml or cm^2) consist of a continuous variable, continuous distributions are indicated to describe their variability.

Generalized or heterogeneous Poisson distributions represent a good choice to model the total variation of observed counts with over-dispersion. In a generalized Poisson model, the Poisson distribution describes the variation of plate counts and a continuous distribution is used to loosen the Poisson restriction in terms of variance, i.e. it allows the expected number of counts (Poisson rate λ) to follow a continuous distribution (Gonzales-Barron et al., 2010) that provides insight into the variability of concentrations of the sample units from one lot (Reinders et al., 2002). Typically, in food microbiology, a gamma or a lognormal continuous distribution is used to represent the variability of microbial concentrations between different food products; this results respectively in a discrete Poisson-gamma or Poisson-lognormal distribution of observed counts (Gonzales-Barron and Butler, 2011; Jongenburger et al., 2012b). Hence, generalized Poisson distributions are useful to model variations of bacterial counts beyond randomness, by harmonising the randomness of enumeration for plate count methods (a Poisson process), the variability of concentrations (assumed as gamma or lognormal) and the observed distribution of counts (PGM or PLN) (Gonzales-Barron and Butler, 2011; Jongenburger et al., 2012b; Reinders et al., 2002).

In some situations though, a simple Poisson distribution may be more appropriate. The choice between a simple Poisson and a generalized Poisson distribution to describe microbial counts is dependent on the type of food product and processing steps under study. For example, Reinders et al. (2002) observed that in whole beef sample units, a PGM distribution approximated significantly better the observed counts than a Poisson distribution with a fixed mean, which indicated that the bacteria were not homogeneously distributed. However, after two grinding steps, the distribution was more or less random and the simple Poisson became adequate to describe observed counts (figure 4).

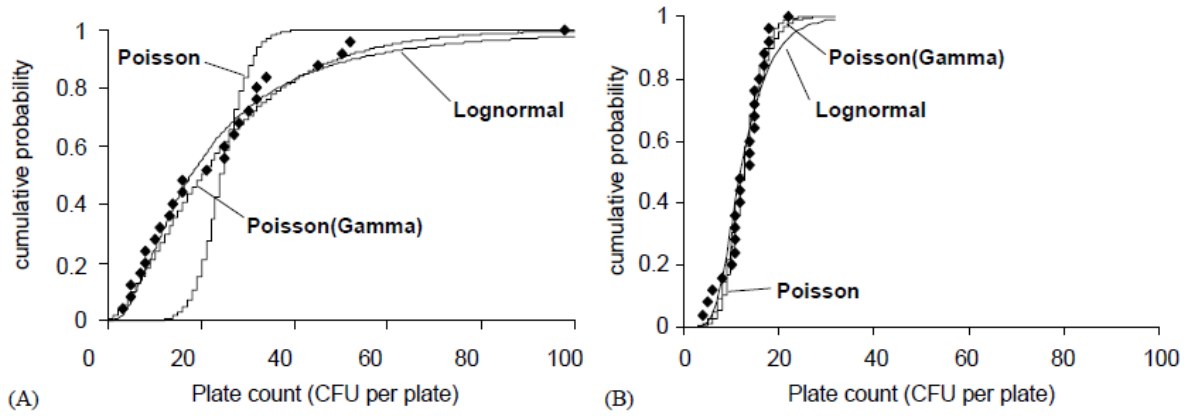

Figure 4: Comparison of the empirical distribution function (black points) and the fitted Poisson, Poisson-gamma and lognormal distributions in minced beef, after grinding meat with a clustered contamination once (A) or twice (B) (Adapted from Reinders et al., 2003).

The choice between a Poisson-gamma and a Poisson-lognormal is also not irrelevant. It has been shown that a PLN fits better to high counts and the PGM fits better to low counts and a higher number of zeroes (Gonzales-Barron and Butler, 2011). In general, the probability of zero in a generalized Poisson distribution is larger than in a simple Poisson (Ridout et al., 1998); however, it might not be large enough when there is an excess of zeroes in the data. In such a situation, a zero-modified distribution may be necessary.

1.3. Zero-modified distributions

With low contamination levels, a high number of zero counts is obtained with enumeration. This contributes highly for the over-dispersion of the data, which may not be accounted for by the variance of the generalized Poisson distributions. Therefore, to model the excess of zeros, zero-modified distributions have been suggested. These represent a mixture of two distributions, i.e., individual counts can originate from two stochastic processes (Gonzales-Barron et al., 2010).

There are two types of zero-modified distributions: zero-inflated and hurdle models. With zero-inflated models one stochastic process will always give a count of zero, and the other, which follows a count distribution, is responsible for positive and additional zero counts. With hurdle models a binomial distribution determines if a count outcome is zero or positive, and a truncated-at-zero count distribution models the positive counts (Gonzales-Barron et al., 2010). A disadvantage of hurdle-models is that the mean of a zero-truncated distribution depends on its parametric form (e.g. a Poisson and PGM distribution with the same mean have different means when zero-truncated) (Ridout et al., 1998).

Note that, in food microbiology, zero counts can arise from two situations: absence of contamination and low concentration. This implies that some zeroes are “true”, hence defining the prevalence of contamination, and others “artificial” and related to low concentration levels. Zero-inflated models may therefore be considered more appropriate than hurdle models to describe this reality. Zero-inflated distributions can be used to model prevalence and concentration together in a single model. They allow specifying in the same model the probability of obtaining an uncontaminated unit and the concentration distribution for the contaminated ones. However, there is little evidence available in the literature about the advantage of zero-inflated distributions over their non-zero-inflated counterparts. This advantage is probably dependent on the proportion of zeros in the dataset. For example, Gonzales-Barron et al. (2010) observed that for a dataset with 42% zeroes, the zero-inflated PGM distribution was comparable to the simple PGM.

Despite the lack of evidence, this descriptive way of modelling the overall microbial contamination, i.e. including prevalence and concentration, may represent an advantage for the use of microbial data in QMRA. Prevalence and concentration are important inputs in exposure assessment and are often interrelated (e.g. the lower the prevalence, the lower the concentration), therefore their description with one single model, while depicting this interrelation between both variables, may contribute to the parsimony of exposure models.

2. Dealing with non-detects and artificial zeroes

Food units with a very low pathogen concentration, although contaminated, can be interpreted as pathogen-free with detection or enumeration methods (Busschaert et al., 2010; Lorimer and Kiermeier, 2007), which leads to left-censored data. To find an appropriate way of dealing with these data has been a concern in food microbiology, especially when microbial data is meant for use in QMRA. Although products with low concentration are generally considered negligible contributors to the risk of foodborne illness, in situations where microbial growth along the risk pathway is a possibility, or in case of highly infective pathogens, the concentration in those products may eventually rise to levels of concern before they reach the consumers’ tables (Pérez-Rodriguéz et al., 2007; Straver et al., 2007).

2.1. Imputation

An approach to deal with artificial zeroes frequently used in microbiological studies, which has been successively abandoned, consists on the substitution of left-censored results by arbitrary values related to the limit of quantification (LOQ) - a practice known as *imputation*. It has been shown that this approach leads to a biased estimation of the distribution parameters (Helsel, 2006; Lorimer and Kiermeier, 2007; Shorten et

al., 2006), with the bias magnitude depending on the arbitrary value adopted for substitution. Substitution by the LOQ value most commonly leads to an overestimation of the mean concentration (Busschaert et al., 2010; Gonzales-Barron et al., 2010). However, the adoption of different fractions of the LOQ can result in either underestimation or overestimation of parameters (figure 5).

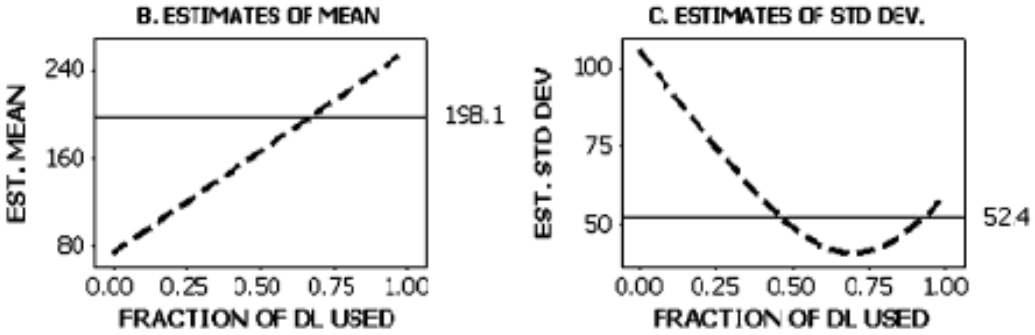

Figure 5: Estimated values for statistics mean and standard deviation of censored data (y axis) as a function of the fraction (0 to 1) of the LOQ value (x axis) used to substitute artificial zeroes. Horizontal lines are at expected values of mean and standard deviation (adapted from Helsel, 2006).

Figure 5 shows that depending on the value used to substitute artificial zeroes, the parameters mean and standard deviation were either underestimated or overestimated. The different values used for substitution corresponded to different fractions of the LOQ. Although for a specific dataset substituting data using a certain fraction of the LOQ might appear to mimic the expected mean or standard deviation, the fraction that is “best” to estimate each parameter is not the same in all situations, and for another dataset with different characteristics, other fractions might seem more appropriate (Helsel, 2006). Obviously, the bias effect shown in figure 5 might not be as pronounced as the size of the dataset increases or if the percentage of left-censored observations is lower. However, it has been shown that using imputation in the analysis of microbial data is not a good approach (Helsel, 2006). The adoption of alternative approaches has been therefore a necessity.

2.2. Maximum likelihood estimation

Recently, a maximum likelihood estimation (MLE) that deals with censored results when fitting a parametric distribution has been introduced in food microbiology to analyse data with censored observations (Busschaert et al., 2010; Busschaert et al, 2011; Delignette-Muller et al., 2010; Lorimer and Kiermeier, 2007; Shorten et al., 2006), and has progressively substituted the imputation method. It analyses not only left-censored, but also semi-quantitative (or interval-censored) results. This fitting method has been mostly used with the assumption that the underlying frequency distribution of $\log_{10}$ concentrations approximates a LN, but it has been recently adapted to fit a PLN distribution to CFU counts as well (Williams and Ebel, 2012b).

In practice, the MLE method assumes that the data follows a certain distribution *a priori*, and then estimates the combination of values for the distribution parameters that results in the maximum log-likelihood of occurrence of the observed data.

Every interpretation method applied to censored data uses the relative ordering of data, either in the form of their cumulative probability (percentiles) or in the proportions of data falling below each censoring threshold (such as the LOQ or the two limits of an interval) (Helsel, 2005). For example, when dealing with interval-censored data, the probability that a data point is between the lower and the upper limit of an interval is equal to the cumulative probability of the upper limit minus the cumulative probability of the lower limit (Zhao and Frey, 2004). Figure 6 shows an example of a MLE fit to two different datasets containing: quantitative data + left-censored data (*a*) and interval-censored + left-censored data (*b*), using the package called "fitdistrplus" for the statistical software R (R Development Core Team).

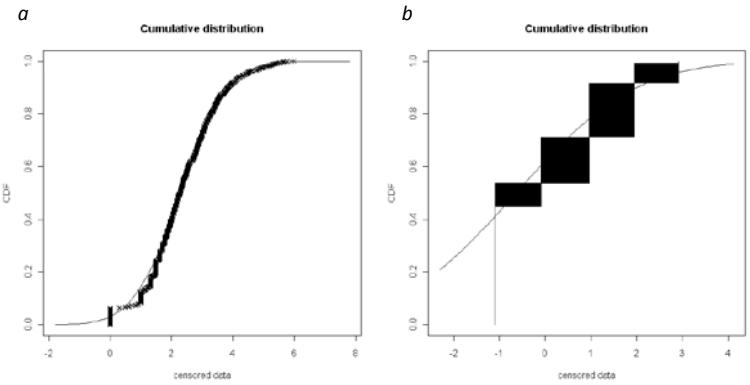

Figure 6: Cumulative distribution function of the MLE fit to two different datasets of microbial concentrations (log_{10} CFU/g). Data in *a* contains quantitative observations and data in *b* contains interval-censored observations. Both datasets contain left-censored observations. The curve lines represent the fitted distributions and the crosses (*a*) and dark areas (*b*) represent the observed data.

The "fitdistrplus" package was developed to fit parametric distributions both to non-censored and censored data (Delignette-Muller et al., 2010). In a mixed dataset, quantitative observations contribute to the likelihood function of MLE as probability density functions, i.e. the probability of single values, whereas the censored data contribute to the likelihood function with the cumulative distribution function, i.e. the probability of being below a certain value. In order to use "fitdistrplus" fitting tools, the data must hence be stored with the method of interval endpoints. With this method, all observations are represented by two variables, the high and low endpoints of quantification. The comparison of values in the first variable to those in the second variable determines whether the observation is left-censored, right-censored, interval-censored or uncensored. For quantitative (uncensored) observations, the values of the two variables are identical, whereas for censored observations the values represent the lower and upper limits of quantification. For left- and right-censored observations, one of the limits correponds to infinity. When censored data is present in a dataset, the function "fitdistcens" included in the package "fitdistrplus" estimates the parameters of a distribution, using the MLE approach for censored data. The uncertainty of the estimated parameters can be estimated by non-parametric bootstrap, using the function "bootdistcens" of the same package. With the non-parametric bootstrap the original dataset is resampled with replacement and a distribution is fitted to each new sample. The output is a 95% confidence interval for the parameter estimates. This fitting tool treats all left-censored observations as originating from contaminated units with a low concentration, i.e., prevalence is assumed as 100%.

Alternatively, a MLE approach can be used that also estimates prevalence, additionaly to the distribution parameters. Using the Solver add-in for Excel 2010, a MLE approach has been implemented to estimate the parameters of a normal distribution of log_{10} concentration values and the most likely prevalence, from interval-censored concentration data, including left-censored observations (BIOHAZ, 2011; Boysen et al., 2013).

As an illustration, the impact of prevalence (*p*) estimation on the obtained parameters of a lognormal distribution (mean (μ) and standard deviation (σ)), is assessed here. Results obtained with the MLE method for censored data implemented in Excel were compared in three different scenarios: 1) assuming all zeroes as uncontaminated units (*p* fixed at the observed prevalence); 2) assuming all zeroes as contaminated units with low concentration (*p* fixed at 100%); 3) estimating *p* as one of the parameters of the MLE. For this exercise, semi-quantitative enumeration data of *Yersinia enterocolitica* from both faecal and skin swab samples collected at pig slaughterhouses were used. The interval censored data including left- and right-censored observations are presented in table 3 and the results are presented in table 4.

Faeces

	Concentration (CFU/g)						
	<0.067	**[0.067-0.67[**	**[0.67-7.46[**	**[7.46-74.63[**	**[74.63-671.14]**	**>671.14**	**Total**
Observations	338	37	52	26	11	13	477
% of total	0.71	0.08	0.11	0.05	0.02	0.03	1

Skin

	Concentration (CFU/ml)						
	<0.08	**[0.08-0.81[**	**[0.81-8.97[**	**[8.97-89.7[**	**[89.7-807.3]**	**>807.3**	**Total**
Observations	322	64	126	148	55	5	720
% of total	0.45	0.09	0.17	0.20	0.08	0.01	1

Table 3: Semi-quantitative estimates of the concentration of Yersinia enterocolitica in faecal and skin swab samples collected at 4 pig slaughterhouses in Denmark.

Faeces

Scenario	***p***	**μ**	**σ**
p=observed	29.14%	0.3950	1.4876
p=100%	100.00%	-2.7715	2.9539
p=estimated	43.72%	-0.3391	1.9396

Skin

Scenario	***p***	**μ**	**σ**
p=observed	55.28%	0.9088	0.9932
p=100%	100.00%	-0.6370	2.0528
p=estimated	56.55%	0.9038	0.9987

Table 4: Estimates of the parameters of a lognormal distribution fitted to the data presented in table 3 with MLE for censored data, with three different approaches (scenario) for prevalence estimation.

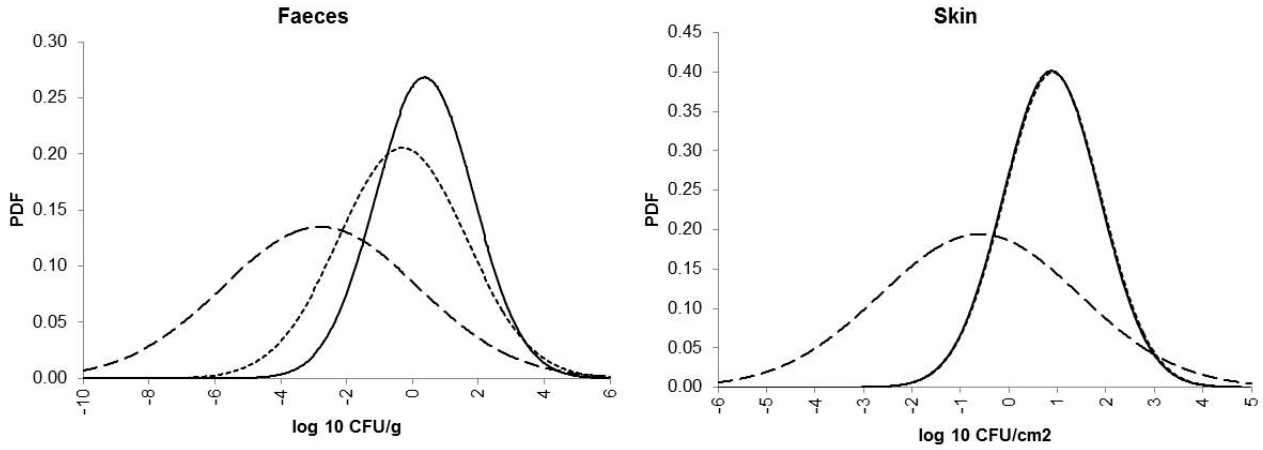

Figure 7: Probability density function of the lognormal distributions presented in table 4, for concentrations of *Yersinia enterocolitica* in faeces and on skin. The three prevalence scenarios are represented: (—) observed *p*; (---) *p*=100%, (····) estimated *p*.

Based on these results, there is a marked difference between a MLE method that assumes a prevalence of 100% (treats all left-censored observations as positive results) and a MLE that allows for the estimation of prevalence, and therefore treats some of the left-censored results as negative. The distance between observed and estimated prevalence appears to depend on the proportion of left-censored observations among the data, but also on the proportion of observations with high concentration values. There is a higher deviation with the dataset of concentrations in faeces (71% of left-censored and 3% of right-censored observations) than with the data of concentrations on skin (45% of left-censored and 1% of right-censored observations) As explained by Helsel (2006), MLE might be problematic with datasets in which one or two outliers throw off the estimation of parameters.

Additionally to the results obtained with the MLE adapted for prevalence estimation, the same data was fit to a lognormal distribution with the package "fitdistrplus", and the confidence intervals of mean and standard deviation were estimated with non-parametric bootstrap (table 5).

	μ	**μ 95% C.I.**	**σ**	**σ 95% C.I.**
Faeces	-2.7700	[-3.3169;-2.3148]	2.9500	[2.5400;3.4492]
Skin	-0.638	[-0.8249;-0.4551]	2.0530	[1.9308;2.1841]

Table 5: Estimates and 95% confidence intervals of the parameters of a lognormal distribution fitted to the data presented in table 3 with MLE for censored data, assuming prevalence=100%.

The results in table 5 show that the parameter estimates of mean and standard deviation are almost identical to the estimates obtained with the MLE implemented in Excel, when the prevalence is fixed at 100%. The confidence intervals for the parameters do not comprise the values that were obtained with the latter method, when the prevalence was also estimated.

Despite of the improvement that MLE for censored data represents in relation to the imputation method, this exercise showed the impact that assumptions in matters of prevalence may have in the estimated

parameters of the distribution of concentrations. There is therefore a need to develop a method that does not assume the totality of left-censored results as originating from contaminated units, preferably by providing a prevalence estimate.

Also, it is important to note that MLE has its limitations. It generally does not work well for small data sets (fewer than 30 to 50 detected values), in the presence of influential outliers and where there is insufficient evidence to know whether the assumed distribution is appropriate to characterize the data (Helsel, 2006).

iv. Predictive microbiology

Mathematical modelling is one of the possible approaches to predict the behaviour of microorganisms, additionally to challenge trials and expert judgment (Adams and Moss, 1995). Predictive microbiology can be defined as the development and use of mathematical models which can predict the rate of growth or decline of microorganisms under a given set of environmental conditions (McMeekin et al., 1993). In food microbiology, these conditions include intrinsic properties of the food product and extrinsic properties of the storage environment. McDonald and Sun (1999) present a list of intrinsic and extrinsic factors that affect microbial growth (table 6).

Intrinsic factors	Extrinsic factors
pH, acidity, acidulant identity	Temperature
% buffering power	Relative humidity
Water activity and content	Light intensity and wavelength
Humectant identity	Atmospheric gas composition and ratio
Redox potential	Packaging characteristics and interactions
Presence of antimicrobials	Processing characteristics and interactions
Identity and distribution of natural microbial flora	Storage, distribution and display considerations
Presence of physical structures	
Presence of biological structures	
Availability of nutrients	
Colloidal form	
Substrate surface to volume ratio	

Table 6: Some intrinsic and extrinsic factors affecting microbial growth. (Adapted from McDonald and Sun, 1999)

1. Model building

A reliable predictive model is usually developed in four steps: 1) definition of a specific question, appropriate for the problem at hand; 2) collection of experimental data; 3) testing of different mathematical models to fit the data and choosing the most appropriate option; 4) evaluate the selected model with different data collected under similar experimental conditions (McMeekin et al., 1993; Adams and Moss, 1995).

1.1. Planning

To clearly define the question that is being asked, this question should define a) the microorganism of concern, b) the controlling factors to use as inputs and their levels and c) the limit of the acceptable output (e.g. 1 $\log_{10}$ increase/decrease in the cell count). An example of an appropriate growth question would be the minimum time needed at a certain temperature for a specified log increase in the population of a specific microorganism in a specific product with determined water activity and pH (Legan , 2007).

1.2. Data collection

Predictive models often aim at studying the impact of multiple factors on a specific response. For this purpose, three approaches are possible: 1) *one factor at a time* – one factor is varied and all others are kept constant; 2) *full factorial design* - all factors are studied in combination; 3) *fractional factorial design* - only a fraction of the factor combinations is tested. The first approach is useful to estimate the curvature of the effect of various factors but disregards the interaction between factors. The second is appropriate to investigate interaction terms but the number of experiments needed goes quickly unmanageable. Therefore,

the third approach is often preferred as a manageable alternative to investigate interaction terms (van Boekel and Zwietering, 2007).

1.3. Model selection

Once a model is fitted to the collected data, goodness of fit must be assessed based on the analysis of residuals (residual distribution, test for runs, serial correlation), the coefficient of determination (or R^2) or the residual mean square. Once a model is identified as a good fit to the data, its performance must be compared against other models. Model discrimination can be based on R^2, *F*-test, *t*-test, Akaike Information Criterion or Bayesian Information Criterion (López et al., 2005; van Boekel and Zwietering, 2007). If model discrimination indicates that two or more models can satisfactorily describe the data, the model with less parameters is recommended over less parsimonious models. Barany and Roberts (1995) question the benefit of upgrading models by including more parameters in order to obtain better fits, and note that "generally, the more parameters a model contains the better the fit it can produce". However, the better performance is often related to a more appropriate description of random errors that occur in individual datasets, and, as the authors conclude, "our aim should be to eliminate these random errors rather than to fit them". Other authors consider that to accurately predict microbial behaviour in food it is essential to include all important parameters in the model (Mejlholm and Dalgaard, 2009). Ultimately, the criteria for choosing a model are not only statistical goodness-of-fit and parsimony, but also convenience (López et al., 2004). Therefore, more parsimonious models might often be preferred as they are simpler and therefore easier to use (Zwietering et al.,1990).

Furthermore, mathematical models can be either mechanistic or empirical. Empirical models represent mathematical functions that fit the data without describing the mechanism behind the response. Mechanistic models represent the processes controlling the response. Mechanistic models should be preferred as they are more informative regarding the conditions governing growth and hence less dataset-specific.

1.4. Model evaluation

As noted by McMeekin et al. (1993), "the true value of a model ultimately relies on how well it can predict microbial responses under novel conditions". Particularly, when the novel conditions correspond to a real food environment, model evaluation (or *validation*) should demonstrate that microbial behaviour in laboratory media is similar to behaviour in a real food system (McDonald and Sun, 1999). This evaluation can be done using data collected in challenge trials or data collected from literature.

A standard method of model assessment was lacking in predictive microbiology in the early years of this discipline. Delignette-Muller et al. (1995) studied the distribution of prediction errors of 14 different growth models with 14 published datasets and identified problems of robustness of models when tested under different conditions. The authors recognized back then that model developers generally did not provide any information regarding the average error on predicted variables. Shortly after, Ross (1996) introduced the concepts of accuracy (A_f) and bias factors (B_f) to assess model performance under new conditions, and these parameters have been adopted as a standard method for model evaluation in the field ever since. These are measures based on the average deviation between predicted and observed microbial responses.As the absolute difference between predicted and observed responses for large values has a greater impact in the averaging process, the ratio between the responses was adopted instead, as a relative standardized measure of deviation. Additionally, since over-prediction and under-prediction should have equal weight in determining the average deviation, the logarithm of that ratio was chosen (Ross, 1996). Hence, the bias factor is given by

$$B_f = 10^{\left(\sum log\left(GT_{predicted}/GT_{observed}\right)/n\right)} \quad (7)$$

where $GT_{predicted}$ and $GT_{observed}$ are the predicted and observed generation times, respectively, and n is the number of observations used in the calculation. In equation 7 the under-predictions and over-predictions tend to cancel out due to opposite signs of the logarithms. Hence, the accuracy factor is calculated by using the absolute value of the logarithm of the ratio

$$A_f = 10^{(\Sigma|log(GT_{predicted}/GT_{observed})|/n)} \quad (8)$$

In a situation of perfect agreement between observations and predictions, both A_f and B_f equal 1.
The B_f answers by how much, on average, the predicted values over- or under-predict the observed ones, i.e., it tells whether the model is "fail-safe" (B_f <1) or "fail-dangerous" (B_f >1).
The A_f averages the distance between each predicted and observed value, hence it tells how close, on average, predictions are to observations. It is an indication of the general predictive accuracy of the model. The greater the deviation from 1 (A_f >1), the poorer the average predictive ability of the model.
As noted by McDonald and Sun (1999), "in practice, the issue is not necessarily how well a model fits data, but the accuracy with which it mimics the microbial response", particularly so if a model is used in quantitative microbiological risk assessment (QMRA).

2. Predictive modelling in the context of food microbiology

The need for predictive models in food microbiology was first recognized as a tool to predict time to spoilage (Olley et al., 1988). Before the advent of predictive modelling, those predictions were made through expensive and slow challenge tests, which needed to be repeated for different combinations of factors (Baranyi and Roberts, 1995; McDonald and Sun, 1999). This approach did not provide an understanding of the microbial responses to key controlling individual factors or factor interactions. Predictive models revolutionized food microbiology by allowing the prediction of microbial behaviour based on combinations of several environmental factors and by their ability to predict for combinations of factors where no experimental data exists (McDonald and Sun, 1999). Furthermore, mathematical modelling has in occasions also contributed to the identification of major growth or inactivation controlling factors or key microorganisms responsible for food spoilage (Dalgaard, 2004).
Predictive models are categorized under several schemes at different hierarchical levels. According to the described microbial behaviour, they are categorized as growth models and inactivation (or survival) models. Here the focus is on growth models, which describe the increase in microbial population over time. Microbial growth can be described has having three phases: lag, exponential growth and stationary phase. Plate counts that are used to measure bacterial growth often require a log transformation because of their heterocedasticity (López et al., 2004). When the growth curve is defined as the logarithm of the number of organisms or colony forming units (CFU) plotted against time, it results in a sigmoidal curve representing those three phases - lag followed by exponential growth and then a stationary phase (Zwietering et al., 1990). Microbial death starts after the stationary phase (figure 8).

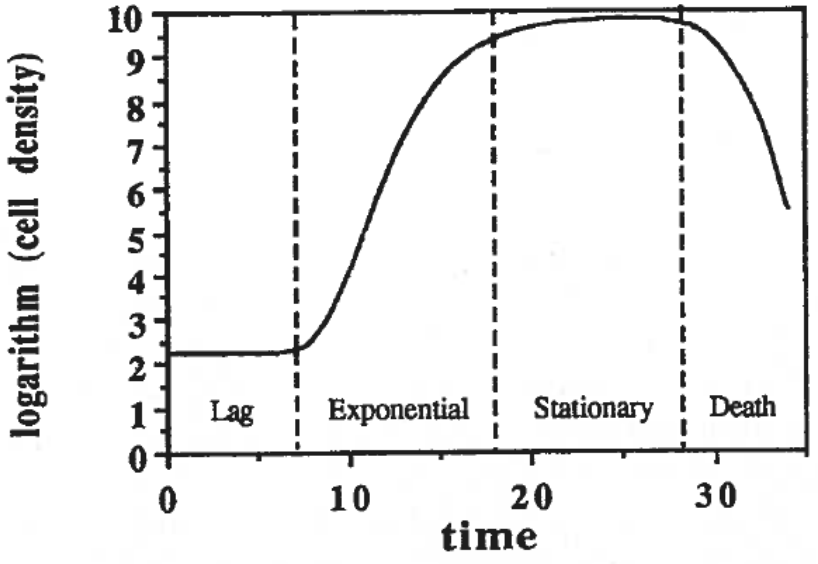

Figure 8: Different phases of microbial growth. The vertical lines mark the end of the lag, the exponential and the stationary phases (Adapted from McMeekin et al., 1993).

The lag phase is a period of metabolic adjustment of the cell to the new environment during which only the intracellular conditions change. After the lag, the cell does not necessarily divide yet, but the exponential period of balanced growth begins (Baranyi, 1998). In the exponential growth phase, the logarithm of the cell concentration increases linearly with time (Baranyi, 1998). In the stationary phase the population is under a metabolic adjustment to enhance survival; it has reached its maximum density, as the growth rate is in balance with the death rate (cryptic growth).

The lag phase duration is difficult to estimate due to the lack of physiological understanding of the lag, its varying definitions, the lack of knowledge on the relationship between the cell lag and the population lag (because of the inter-individual cell lag variability) (Baty and Delignette-Muller, 2004) and its dependence on previous cell history, which is hard to characterize. Hence, the initially proposed method to do growth estimations was to describe a dataset with a primary growth model and then characterize the three growth phases by inferring from the model the lag time (λ), the maximum specific growth rate (μ_{max}) and the maximum population density (MPD). The μ_{max} is given by the slope of the line when the organisms grow exponentially, the λ is the x-axis intercept of the tangent in the inflection point of the curve and the asymptote represents the MPD (Zwietering et al., 1990). Additionally to the MPD, the lag time and the maximum specific growth rate, the initial bacterial density can also be considered as a parameter of growth kinetics (Baty and Delignette-Muller, 2004), although it is practically impossible to determine when working with real food. McMeekin et al. (1993) illustrate how to determine the three parameters graphically (figure 9) but note that it represents a very subjective estimation method.

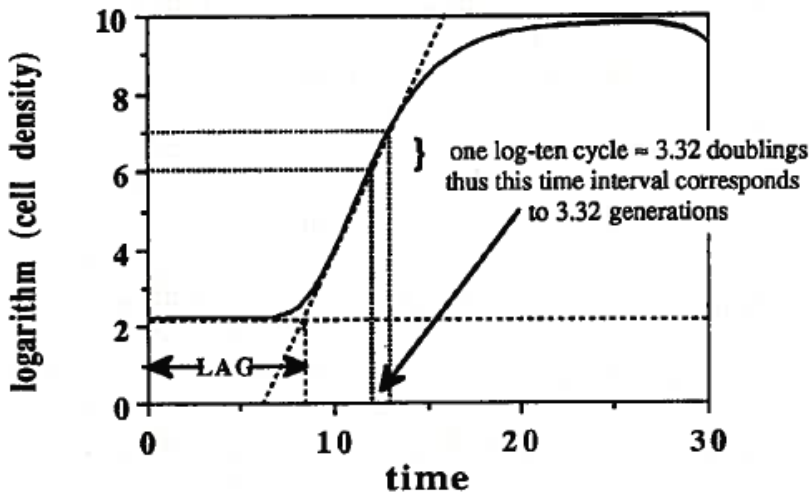

Figure 9: Graphical method for the estimation of generation and lag time from a growth curve. The slope of the tangent to the steepest part of the curve estimates the exponential growth rate (or maximum specific growth rate μ_{max}). The generation time can be calculated from this tangent as the time for a 0.301 (i.e. a doubling of the population or $\log_{10}(2)$) unit increase in the $\log_{10}$(cell density). The intercept of the tangent with the lower asymptote (i.e. $\log_{10}$(cell density) at time=0) is taken as the lag time. (Adapted from McMeekin et al., 1993)

The growth parameter generation time (GT) represents the time taken for the doubling of the population and is related to the maximum growth rate (McMeekin et al., 1993) according to the equation (for microbial counts expressed as $\log_{10}$ CFU)

$$GT = log_{10}(2)/\mu_{max} \quad (9)$$

According to the units of the response variable, there are kinetic models and probability models. Kinetic models can be primary or secondary, according to the type of predicted response and explanatory variables. Primary and secondary kinetic models can be categorized according to the type of mathematical model used.

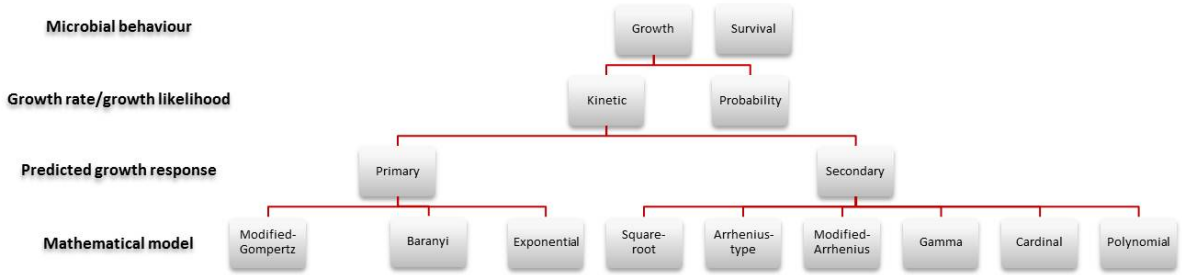

Figure 10: Classification of predictive microbiology models

Kinetic models are used to predict the rate of growth or the time to a critical amount of growth whereas probability models predict the likelihood of growth within a given time period. Ross and McMeekin (1994) highlighted that the delineation between these two types of model is artificial. On one hand, probability models, whilst indicating the absolute likelihood of growth to occur, need to include information about the variability of growth rates. On the other hand, kinetic models must consider the probability of growth occurring at a predicted growth rate (McDonald and Sun, 1999).

In fact, predictive microbiology has evolved towards the unification of probability and kinetic models when growth-boundary models were created. These are kinetic models built from qualitative or quantitative measurements of growth and no-growth over time, that predict the limits of conditions permitting growth (Legan, 2007; McDonald and Sun, 1999; Ratkowsky and Ross, 1995), i.e., the region of no growth encloses the combinations of growth factors that result in a specific growth rate of zero. Besides this probabilistic approach derived from kinetic models, Le Marc et al. (2002; 2005) introduced a deterministic approach to model the growth/no growth interface. This approach is based on the assumption that the experimental range can be divided in three areas where 1) conditions act independently on bacterial growth, 2) conditions interact and the growth rate reduction is greater than in the “independent effects” area, and 3) the interaction between conditions prevents growth (no growth area).

2.1. Probability models

To develop a model that predicts the probability of growth, replicate samples of known inocula are incubated under defined environmental conditions for a fixed period of time and the proportion of replicates where growth occurs is assessed. The measured growth is dependent upon the time for it to be detectable, which is a function of the lag time, the rate of growth of the microorganism (parameters that are usually predicted by kinetic models) and the number of cells initially present (Genigeorgis et al., 1971; Lindroth and Genigeorgis, 1986).

Probability models have proven useful predicting the probability of toxin production, where any toxin presence is unacceptable, however they have been criticized for the difficulty of translating probabilities into practical values that can be used to set safe shelf lives when a certain amount of microbial growth may be tolerated (Gibson et al., 1988; McDonald and Sun, 1999; Ross and McMeekin, 1994).

2.2. Kinetic models

Nutrients are usually not considered a limiting factor for bacterial proliferation in food products (Zwietering et al., 1990); therefore the rate and extent of microbial growth are dependent on a combination of product intrinsic factors (such as pH, water activity and preservatives) and environmental extrinsic factors (such as temperature and gaseous atmosphere). Kinetic models can be used to predict the changes in microbial numbers over time for a specific set of conditions and therefore to define growth parameters - *primary models*, or to describe the effect of intrinsic/extrinsic factors on specific growth parameters - *secondary models*. In a third category, *tertiary models*, primary and secondary models are combined with user-friendly application software in order to allow the prediction of microbial behaviour under (untested) specified conditions (Whiting, 1995). Dalgaard (2004) and McMeekin et al. (2013) present a list of available sources of tertiary models. An example of a tertiary model is the Seafood Spoilage Predictor (SSP), developed to predict the effect of temperature on the growth of specific spoilage microorganisms in seafood (Dalgaard et al., 2002).

2.2.1. Primary models

There are two distinct approaches in the building of primary models: the first is to model the growth rate from observed data and use an exponential function with the fitted rate to make predictions, where the number of bacteria at time *t* is a function of the growth rate and the initial inoculum. When the interest is only to predict growth during the exponential growth phase, the exponential model has proven to be sufficient (Zwietering et al., 1996; van Gerwen and Zwietering, 1998).

The second approach is to fit a sigmoid curve to the observed population growth and derive lag time and specific growth rate from the fitted model. Sigmoid curves have the advantage of not assuming a constant growth rate and providing a more objective characterization of the lag and generation times that are part of microbial growth. However, they have the disadvantage that a relatively large amount of data points (at least ten according to Bratchell et al. (1989)), are required for their fit (Gibson et al.,1988; Whiting,1995). The modified-Gompertz model (Gibson et al.,1988; Ross and McMeekin, 1994) and the Baranyi model (Baranyi et al., 1993a, 1993b) are the two most famous sigmoidal primary growth models. The modified-Gompertz model is a modification of the original Gompertz model that describes mortality as a function of age (Gompertz, 1825). Several criticisms have been addressed to the modified-Gompertz model. First, it describes the exponential growth as a continuous curve with an inflexion point at the maximum growth rate, although the exponential growth is in fact represented by a straight line. This characteristic of the model suggests a maximum specific growth rate that is higher than might be expected. Second, the fitting of the function frequently estimates a negative lag time due to the geometrical definition of the lag phase. And third, the use of the logarithm of the cell concentration as the time-dependent variable makes it a strictly empirical model (Baranyi et al., 1993a; Whiting, 1995). As an alternative to the empirical Gompertz model, Baranyi et al. (1993a, 1993b) developed a more mechanistic logistic model. The essential differences between this model and the latter are that in the Baranyi model 1) the lag period is defined as the time needed for a critical product to reach a constant value (concept that is represented in the model by an adjustment function), 2) at time *t*=0 the slope is zero because the value of the adjustment function at the origin is zero, 3) the exponential phase of growth is represented by a straight line and not by a curvature and 4) the curvature before the stationary phase is more pronounced (Baranyi et al., 1993b). Baranyi et al. (1993b) compared their model to the Gompertz model and observed that for lower growth rates the Gompertz function was more advantageous, whereas for higher growth rates the Baranyi model provided better fits.

Note that the primary growth models can be defined both in function of the base 10 logarithm or the natural logarithm of the microbial numbers, and the value of μ_{max} used must be chosen accordingly.

Other approaches where microbial growth, which has been measured as biomass, turbidity or conductance, is modelled against time can also be used instead of models that, like the Gompertz or the Baranyi, use cell concentration as the time-dependent variable. Note that the two types of growth curves cannot be directly compared, before finding the quantitative connection between the measured quantity and the cell concentration (Baranyi et al., 1993a). Similarly, when using a predictive model, it is important to know the type of data that was used to develop the model, as microbial growth measured with indirect methods often differs to that determined with plate counts (Neumeyer et al., 1997); hence appropriate calibration factors must be applied in order to use that model with data collected using different methods.

2.2.2. Secondary models

There are several model types recognized in kinetic modelling: Bělehrádek or square-root model, Arrhenius-type model, modified Arrhenius or Davey model, gamma model, cardinal model and polynomial or response-surface model. Independently of the model type, the response variable in secondary kinetic models is always a growth parameter (typically the growth rate) modelled against factors that influence growth.

The Arrhenius-type model used in predictive microbiology is derived from the classical Arrhenius equation used in physical chemistry, which describes the relationship of the logarithm of the rate vs. the reciprocal of

the temperature. When applied to microbial growth, the original form of the model was found inadequate to describe the growth response to temperature (Ross and McMeekin, 1994). Therefore it has been developed towards more sophisticated models, among which, the nonlinear model of Schoolfield et al. (1981) is the most popular. The original form of this model has been extended to include the effects of water activity (Broughall et al., 1983) and pH (Broughall and Brown, 1984). It has also been re-parameterized by Adair et al. (1989) to include the lag or generation time.

Since the success of nonlinear models is dependent on the ability to obtain good initial parameter estimates for nonlinear regression fitting, a modified, linear type of Arrhenius model was introduced by Davey (1989) to allow for explicit solution of the best parameter estimates. Later, Davey (1991) described a similar function to model the lag phase duration.

A model was introduced by Zwietering et al. (1993) that considers each variable determining growth as a separate hurdle, hence modelling the relative effect of each of these variables on the growth rate independently. This approach is known as the *gamma concept*. It assumes that there are no interactive effects of the various growth variables, and that their combined effect is multiplicative. For example, an effect of 10% in the growth rate due to temperature, together with an effect of 10% due to pH results in a combined effect of 1% (10% x 10%) (Zwietering et al., 1996). This model estimates however the effect of growth variables on a growth factor (γ), also called the relative growth rate due to its calculation: μ_{max}/μ_{opt}, where μ_{opt} is the optimum growth rate (occurring when growth variables are optimum for growth).

The growth factor is equal to 1 at optimum growth conditions and between 0 and 1 for all other conditions. It is assumed that each variable that is not at its optimum value can reduce growth rate. Hence, γ is calculated by multiplying individual γ values defined for each growth variable separately, independent of the value of the other variables (Zwietering et al., 1993).

The cardinal model is relatively similar to the gamma model, as it assumes independence between the effects of different growth variables on the relative growth rate, which are characterized by independent growth factors, hence using the multiplicative approach. However, the authors who developed this model (Rosso et al., 1993, 1995) aimed to eliminate structural correlation between parameters and to reduce the parameters strictly to those with a biological meaning – the *cardinal* growth parameters. A cardinal temperature model (CTM) was initially developed (Rosso et al., 1993) and later expanded to a cardinal temperature pH model (CTPM) (Rosso et al., 1995). Le Marc et al. (2002) later expanded the CTPM including a term to describe the interactive effect of growth conditions near growth limits. In this model, it is hypothesised that the contribution of each variable to the interaction term can be derived from its separate effect on the growth rate.

Finally, polynomial models represent an empirical approach to summarize growth responses, where multiple linear regression is used to determine the best fit values for parameters describing the effect of each explanatory variable on the response variable. They have been used, for example, to estimate the parameters of the primary modified-Gompertz model (Gibson et al., 1988). Polynomial models have been criticized for having no theoretical basis and for being entirely descriptive of the particular dataset used to calculate an equation (Whiting, 1995), therefore not contributing to any knowledge to mechanisms underlying a process and not allowing for extrapolation of predictions (McDonald and Sun, 1999).In this thesis the focus is on the square-root model, as it is the kinetic model that was selected for the study described in manuscript III.

Square-root model

Ratkowsky et al. (1982) introduced a square-root model to describe the linear relationship between the square-root of the maximum specific growth rate and temperature. This initial model was posteriorly extended to include temperatures superoptimal for growth or T_{max} (Ratkowsky et al., 1983), water activity a_w (McMeekin et al., 1987), and pH (Adams et al., 1991). McMeekin et al. (1992) summarized the combined effect of temperature, a_w and pH on microbial growth in a simple four parameter model. Recently, Ross et al. (2003) extended the model to include the effect of lactic acid concentration

$$\sqrt{\mu_{max}} = c(T - T_{min}) \times \left(1 - exp\left(d(T - T_{max})\right)\right) \times \sqrt{(a_w - a_{wmin})} \times \sqrt{\left(1 - exp\left(g(a_w - a_{wmax})\right)\right)} \times$$
$$\sqrt{(1 - 10^{(pH_{min} - pH)})} \times \sqrt{(1 - 10^{(pH - pH_{max})})} \times \sqrt{\left(1 - [LAC]/\left(U_{min}(1 + 10^{(pH - pK_a)})\right)\right)} \times$$
$$\sqrt{\left(1 - [LAC]/\left(D_{min}(1 + 10^{(pK_a - pH)})\right)\right)} \quad (10)$$

where μ_{max} is the rate of growth, *c*, *d* and *g* are coefficients to be estimated, T is the temperature, T_{min} the theoretical minimum temperature below which growth is not possible, T_{max} the theoretical maximum temperature beyond which growth is not possible, a_w the water activity, a_{wmin} the theoretical minimum water activity below which growth is not possible, a_{wmax} the theoretical maximum water activity above which growth is not possible, pH has its usual meaning, pH_{min} is the theoretical minimum pH below which growth is not possible, pH_{max} the theoretical maximum pH beyond which growth is not possible, [LAC] the lactic acid concentration (in mM), U_{min} the minimum concentration (mM) of undissociated lactic acid which prevents growth when all other factors are optimal, D_{min} the minimum concentration (mM) of dissociated lactic acid which prevents growth when all other factors are optimal, and pKa the pH for which concentrations of undissociated and dissociated lactic acid are equal.

There have been many discussions about the most appropriate transformation of the growth rate to use in this type of model. Zwietering et al. (1990) concluded that the proper response variable should be μ_{max} for the model to have homogenous variances, whereas Alber and Schaffner (1992) proposed the natural logarithm of μ_{max} for the same reason (Whiting, 1995).
Despite of the criticism, several authors compared the performance of square-root and Arrhenius-type models and favoured the square-root model (Whiting, 1995). Adair et al. (1989) concluded the opposite and their conclusions were disputed by authors (Davey, 1989; Ross, 1993) who observed that the square-root model "fitted the data well, was close to linear, had good estimates of the parameters, was appropriate to the stochastic properties of growth rates and was easy to use" (Whiting, 1995).

3. Variability and uncertainty in predictive microbiology

In the history of predictive modelling of microbial growth there has been scepticism that models derived in an experimental system can reliably predict growth in food. This mistrust is likely related to the awareness of the complexity and heterogeneity of the bacterial/food system, the disbelief that models derived from experiments at static conditions can be applied to fluctuating conditions and to the publication of studies where there is disagreement between predictions and observed growth in food. Specific factors that may affect model performance have been identified (Ross and McMeekin, 1994). Each factor usually represents a source of either variability or uncertainty (or both) in model predictions.
While variability represents a true heterogeneity, consequently to the physical system and irreducible by additional measurements, uncertainty represents a lack of perfect knowledge and may be reduced by further gathering of information (Kelly and Campbell, 2000; Nauta, 2007).

The lack of precision of model predictions is sometimes presented with a confidence interval of the mean response. However, model imprecision is a consequence of variability and uncertainty together, and a confidence interval does not provide any insight into what causes variation in model predictions. Ideally, sources of variability and uncertainty would be identified, their impact on predictions quantified, and Monte-Carlo simulations would allow the incorporation of variability and uncertainty of different origins in model parameters (Havelaar and Nauta, 2007; Kelly and Campbell, 2000; Marks et al., 1998; Nauta, 2007; Vicari et al., 2007). In practice, both are very difficult to quantify; hence assumptions have to be made. As explained by Nauta (2007), if all parameter imprecision is assumed as true uncertainty, only stochastic variability is maintained, whereas if all is interpreted as variability, there is no uncertainty. Both these assumptions are extreme and unrealistic, and should therefore be avoided. As indicated by Kelly and Campbell (2000), "in

theory, every component (…) could be represented as having both variability and uncertainty", but in practice "we choose which components should be modeled with both uncertainty and variability", based on the purposes of the study and the importance of the component. Indeed, efforts have been made to characterize sources of variability and uncertainty in predictive microbiology. Some of the identified sources of variability, such as the initial number of cells (Coleman et al., 2003), previous cell history and consequential lag time duration (Baranyi et al., 1993a), between- and within-strain variability (Coleman et al., 2003; Nauta, 2007), are directly related to the growth response of microorganisms. Other sources of variability are associated to growth conditions, and include variability in atmospheric conditions (Ratkowsky et al., 1991), variability in process conditions (Ritz et al., 2007), variability of food properties (Dalgaard et al., 2002; Dens and Van Impe, 2001; Nauta, 2007; Wilson et al., 2000) and variability of microbial interactions in food due to variability of competitive flora (Pin and Baranyi, 1998; Van Impe et al., 2005; Vold et al., 2000;). As to the most common sources of uncertainty in predictive microbiology, there are experimental uncertainty (Ritz et al., 2007), the number of observations used to build a model (Ratkowsky et al., 1991) and the model used to describe model behaviour (Ratkowsky et al., 1991).

The total observed variation in microbial growth, which includes both variability and uncertainty, can be described by growth variance. Ratkowsky et al. (1991) studied the total variance of microbial growth and observed that it was higher at slower growth rates. They characterized the relationship between the variance of a (inverse Gaussian distributed) response and its magnitude as

$$variance = 4 \times (response\ time)^3 \times variance\left(\sqrt{k}\right) \quad (11)$$

where *k* is the reciprocal of the response time. Posteriorly, Ratkowsky (1992) presented the relationship between the variance in growth response time and its mean for a range of possible distribution types as $V = c\mu^n$, where μ is the mean of the probability distribution, *V* is the variance of the probability distribution, *c* is a constant and *n* is an integer exponent having values 0,1,2 or 3, corresponding to the normal, Poisson, gamma or inverse Gaussian distributions, respectively (Ross and McMeekin, 1994).

Nauta (2000; 2007) introduced the concept of a parameter *α* – the *uncertain attributable fraction* – to translate the total variance into variance due to uncertainty and variance due to variability. If a mean estimate *m* is obtained for the parameter *x*, with standard deviation *s*, then $s\sqrt{\alpha}$ represents the standard deviation of the uncertainty distribution and $s\sqrt{(1-\alpha)}$ the standard deviation of the variability distribution.

While the above mentioned studies focused on the total variance of growth responses, other authors have focused on the study of a specific source of variability or uncertainty. Baty and Delignette-Muller (2004) showed that the estimations of lag time vary more between different models when the data are sparse and that the exponential and the Baranyi models seem less influenced by data quality compared to the Gompertz model. Nauta and Dufrenne (1999) focused on between-strain variability of growth. The authors recognized that the variability in growth may be large, even within the same species. They experimentally determined growth characteristics of 75 strains of *E. coli* O157:H7 and characterized the variability of each growth characteristic by a probability distribution. Furthermore, they accounted for variability in a Gamma model that estimates the actual specific growth rate. Their results showed no significant difference between the optimum growth rate of different strains but considerable differences in the actual growth rate (Nauta and Dufrenne, 1999). The authors stressed however that without a separation between within-strain variability and between-strain variability and between biological variability and experimental uncertainty, all responsible for the total variation in growth, it is difficult to indicate that their predictions of between-strain variability are realistic. Coleman et al. (2003) studied the effect of the temperature, agitation, initial microbial density and strain on three growth parameters and observed that, in an environment similar to ground beef, strain variability appeared to have a relatively minor effect on growth, compared to agitation and initial density. The authors identified the lower probability of cell clustering and the greater availability of dissolved oxygen as two factors that may contribute to shorter lag and higher growth rate and maximum population density in shaken media. On the other hand, they indicated that a higher inoculum is theoretically associated with a

higher likelihood of including at least one cell in the proper physiological state for immediate growth, and therefore longer lag times for lower densities were observed as expected.

4. Future of predictive microbiology

4.1. Mechanistic vs. parsimonious models

Since the advent of predictive microbiology, a mechanistic approach to model microbial behaviour has been favoured to an empirical one. Baranyi and Roberts (1995) go even further and consider that the term *model* can only refer to a mechanistic one, as it refers to "a set of basic hypotheses on the studied processes, some of which are possibly expressed by means of functions". Hence, those authors consider that empirical models are merely *functions* with the aim of providing a "smooth representation of experimental results". Similarly, McDonald and Sun (1999) note that only mechanistic models allow generating predictions based on hypotheses. Additionally, models with a mechanistic basis are easier to develop further as the information from the system increases. For example, although the Baranyi model represents a mechanistic description of the growth process, by defining an adjustment function that describes the lag duration dependent upon previous cell history, Baranyi et al. (1993b) observed that it was not the most appropriate model choice in some situations when compared to an empirical model. However, later it has been advanced that this is due to an incomplete description of the growth process. As observed by Van Impe et al. (2005) and Standaert et al. (2004), the inappropriateness of the Baranyi model in some circumstances could be due to the existence of empirical factors within the model structure. Van Impe et al. (2005) extended therefore the concept introduced by Baranyi et al. (1993a) and developed an even more mechanistic model that describes the transition from the exponential growth phase to the stationary phase, where both substrate exhaustion and the accumulation of toxic products can have an effect on the maximum population density.

Indeed, the forecast of an evolution towards mechanistic approaches has been present in the field of predictive microbiology. In their review on predictive microbiology, Ross and McMeekin (1994) stressed that the full potential of predictive modelling had not yet been realized and that ultimately, the benefits of predictive modelling all derive from a better understanding of the microbial ecology of foods. Similarly, Cole (1991) had proclaimed that predictive microbiology research was evolving towards more mechanistic approaches. Recently, McMeekin et al. (2013) state that "we envision the future of predictive microbiology in which models morph from empirical to mechanistic underpinned by microbial physiology and bioinformatics to grow into Systems Biology". Those authors cite early works of the Copenhagen school (Kjeldgaard et al.,1958; Schæchter et al., 1958) which showed that the physiological state of cells is determined primarily by the growth rate, and hence they conclude that the focus of predictive models must change from modelling rates to modelling cell physiology. In a previous study, Ritz et al. (2007) also concluded that the data used for predictive models to be used in QMRA need to take into account the physiological state of cells. With the rise of "omics" technologies, systems biology and bioinformatics, the challenge will be data handling and developing models that can comprehend different types of information to accurately describe microbial responses dependent on cell physiology. One of the keys will be to switch from population-based to individual-based models. An individual modelling approach is a good methodology to investigate and test biological theories and assumptions, and therefore to move towards more mechanistic models (Dens et al., 2005). Standaert et al. (2004) enumerate the main advantages of adopting an individual-based modelling approach in predictive microbiology: biological variability can be more easily incorporated into the models, the incorporation of mechanistic knowledge is significantly more straightforward and models can be extended to account for spatial factors affecting microbial responses. However those authors also point out the challenges associated to this type of models: they can rapidly evolve to a high level of complexity, their validation through experiments is difficult to perform and they require considerable computational resources. Despite these challenges, evolution towards mechanistic approaches with incorporation of cell physiology knowledge continues. Recently, Ratkowsky et al. (2005) included thermodynamic terms for protein denaturation in a kinetic model and identified the similarity between the temperature dependence of bacterial

growth and of globular proteins stability. Afterwards, Corckrey et al. (2012) advanced the thermodynamic approach of predictive microbiology and developed a Bayesian model in which growth rates of all unicellular organisms respond to temperature according to the same master reaction hypothesis. This study showed that, as advanced by McMeekin et al (2013), "mechanistic models and systems biology sit naturally together" and "predictive systems biology models have the potential to reveal unifying themes in biological sciences".
The adoption of compartmental models may also ease the integration of new types of data in predictive microbiology. Baty and Delignette-Muller (2004) present examples of published compartmental models: a two-compartment model that separates the evolution of all chromosomal material and the evolution of all non chromossomal material against time (Hills and Wright, 1994); a model that assumes that within a bacterial population some cells will grow exponentially without any delay and some will never grow (MacKellar, 1997) and a two-compartment model where the bacterial population could be divided into cells which are still in the lag phase and cells which are in the exponential phase (Baranyi, 1998).
Along with the development of increasingly mechanistic models, attention should be paid to avoid over-parameterization and creation of unnecessarily complex models (McDonald and Sun, 1999), as simpler approaches can be sufficient in some situations, such as the use of models in QMRA. As a way of considering only the strictly necessary variables impacting significantly microbial behaviour, more product specific models instead of generalist mechanistic models might be required in the future (McDonald and Sun, 1999), which somehow appears to point into a different direction in terms of evolution of predictive microbiology.

4.2. Microbial interactions and "omics" technologies

Microbial interactions are a reality in food environments and an important factor controlling microbial growth (Duffy et al., 1994). The way interactions influence growth is described by the so-called Jameson effect (Jameson, 1962). The Jameson effect describes the effect of the fastest growth microorganism on the remaining ones. In a mixed population, the first group reaching their maximum density will cause competitors to move from the exponential to the stationary phase (McMeekin et al., 2013). Hence, a mixed culture of the most common bacteria found in food is often used by modellers, so that the growth predicted by the model corresponds to the fastest microorganism present (McDonald and Sun, 1999). For example, in cold-smoked salmon the growth of *Listeria monocytogenes* ceases when lactic acid bacteria reach their maximum density and this interaction has been successfully integrated in a growth model (Dalgaard, 1998, 2004).
A biological (mechanistic) explanation for the Jameson effect is the quorum-sensing phenomena. Surette et al. (1999) described quorum-sensing as "the regulation of gene expression in response to changes in cell density". In practice, quorum-sensing bacteria produce and release signalling molecules called autoinducers that accumulate in the environment as the cell density increases and that ultimately, once a threshold concentration is achieved, activate a change in behaviour of the microorganisms responding to it. Furthermore, Surette et al. (1999) identified a gene common to three different bacterial species responsible for the production of a specific autoinducer, which supports the theory of Jameson effect occurring in a natural food environment, where several bacteria species coexist. At the light of this knowledge, and as noted by Coleman et al. (2003), it is obviously important to do an appropriate selection of the component species of the indigenous microflora once simulating a natural food environment in a laboratory experiment. Also, care must be taken once selecting the strain of the pathogen of interest used in such experiments, since domesticated laboratory strains may not produce the same signal molecule as the wild-type strains (Surette et al., 1999). As noted by Nauta (2007) DNA sequencing techniques and studies on gene expression are both promising tools to study strain variability, and are therefore expected to be used in order to aid the integration of such phenomena in future mechanistic predictive models.

Although Jameson effect and quorum-sensing phenomena are often referred to for explaining the transition of a bacterial population from exponential growth to the stationary phase, opinions about the transition between the lag phase and exponential growth seem to be more divergent. Baranyi (1998) states that the

decrease of the population lag with a higher initial population density is caused by the randomness of the process, and not because the cells have information on the size of the population they live in.

v. Predictive microbiology in quantitative microbiological risk assessment

Predictive models are indispensable for the quantification of risk in quantitative microbiological risk assessment (QMRA), since it is impossible to measure microbial contamination at all important parts of the production chain. Each stage of the chain is characterized by specific conditions of the food product and the environment. Those conditions, together with the residence time of the product in the stage, are used as inputs in predictive models to determine the specific rate of microbial behaviour. As a result, the change in log numbers between stages is estimated, which can be used to compare the relative effect of conditions applied in the various stages (Zwietering and Nauta, 2007) and consequently the effect of changes in time and growth conditions on risk.

1. Variability and stochasticity

When implementing predictive models in risk assessment it is crucial to account for the inherent variability in growth and survival, as a high level of exposure can be the consequence of both high infrequent doses and low more frequent doses (Nauta and Dufrenne, 1999). The way to describe variability in risk assessment is typically by adopting a stochastic modelling approach. Consequently, experts of both QMRA and predictive microbiology have identified the inadequacy of deterministic models and the consequent need for stochastic models to use in risk assessment. While deterministic models give merely point estimates of the microbial concentration as a function of time, stochastic models provide a description of the variability in the concentration after a fixed amount of time, or the probability of exceeding a threshold value of the concentration, both more adequate for QMRA purposes (Zwietering and Nauta, 2007). Baranyi (2002) identified the variability of lag time of individual cells as an influencing factor in QMRA due to the existence of *rare* cells that can unexpectedly shorten the population lag time and be paramount for the probability of population survival and growth under changing environmental conditions.

Several studies have focused on the development of stochastic approaches in predictive microbiology, focusing on different growth parameters. Baranyi (1998) concluded that when studying growth by both deterministic and stochastic models, contradictory results can be obtained for lag time. The stochastic concept of bacterial lag should be preferred to the deterministic, especially when studying small populations. Marks and Coleman (2005) developed a simple stochastic growth model, accounting only for inherent cell growth variability (or within-strain variability), i.e. describing the distribution of the microbial density over replicate trials of the same scenario of environmental conditions.

Despite the relevance of lag time studies for unravelling cell behaviour , for the purpose of QMRA variability in lag might not be the most crucial to characterize. In fact, the lag time is often assumed to be zero, either because the microorganism is assumed to be already adapted to the product (especially for studies starting at the retail level), or because the origin of the contamination is unknown and therefore the physiological state of the cells is difficult to define (Zwietering and Nauta, 2007).

However, variability of the maximum specific growth rate, which may be due to different factors such as bacterial strain and composition of the growth environment, is potentially more relevant to consider in QMRA. Its accurate characterization and appropriate integration into exposure models is therefore important.

2. Model validity

Predictive models are typically developed based on data produced under well controlled experimental conditions, with a rich nutrient broth, a high initial inoculum, lack of competing microflora and a high level of mixing. However, in QMRA the same models are used to predict microbial behaviour in non-standardized conditions and for wild-type microorganisms – the real food environment. Before using a predictive model in risk assessment, it is therefore crucial to assess how representative is that model for the conditions under

study. Ross (2000) reported 300 to 400% overestimation of growth for extrapolation from broth culture models to other foods. Hence, there is a need for bridging studies between experimental conditions and food matrices to determine the adjustments necessary to characterize growth kinetics in food, particularly for calculation of variability and uncertainty in exposure models (Coleman et al., 2003). As noted by Delignette-Muller et al. (1995), “the safe use of predictive models in microbiology requires a thorough good knowledge of their accuracy and limitations”. To extend the use of predictive models from research to industrial tools, which might include their use in QMRA, it is important to focus on the practical use of models during their validation (McDonald and Sun, 1999).

3. Simple vs. complex models

Despite all the research efforts to develop more mechanistic models and to characterize variability associated to different parameters, simple predictive models are often preferred in QMRA studies, as a way of guarantying their efficiency (McDonald and Sun, 1999). In any case, the selection of the most appropriate model to use in QMRA is a difficult decision. Therefore, if time allows, it can be useful to compare results of several predictive models in estimating risk. This practice helps to understand the relative influence on the risk estimate of the predictive model used and of other variations in the process of risk assessment. Indeed, it has been shown that the effect of using different predictive models in QMRA may be negligible (Nauta, 2001). If other variations affect the risk more than the choice of the predictive model, it is justifiable to use the simplest model available (van Gerwen and Zwietering, 1998). Ultimately, the criteria for selection of a predictive model to use in QMRA include: model suitability (fit for purpose), model simplicity and number of parameters, ability to look up parameters in literature and databases, practical applicability of the model, biological meaning of the parameters and limits of growth covered by the model (van Gerwen and Zwietering, 1998).

PART I

Manuscript I

Fitting a distribution to microbial counts: making sense of zeros

Duarte, A.S.R. ; Stockmarr, A. ; Nauta, M.J.

Submitted

Title

Fitting a distribution to microbial counts: making sense of zeroes

Authors

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

asrd@food.dtu.dk

anst@dtu.dk

maana@food.dtu.dk

Corresponding author

Sofia Duarte

asrd@food.dtu.dk

+45 23673256

Abstract

The accurate estimation of prevalence and concentrations of microorganisms in foods is an important element of quantitative microbiological risk assessment (QMRA). This estimation is often based on microbial enumeration data. Among such data are artificial zero counts, that originated by chance from contaminated food products. When these products are not separated from uncontaminated products that originate true zeroes, the estimates of 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 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 prevalence of contamination within a food lot and the parameters (mean and standard deviation) characterizing the within-lot distribution of concentrations, without assuming a LOQ, and using raw plate count data as 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 prevalence from enumeration results.

We observe that this method yields better estimates of mean, standard deviation and prevalence at low 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 has been implemented in software and shows a good performance in the estimation of prevalence and the parameters of the distribution of concentrations, which indicates that it is a useful data analysis tool in the field of QMRA.

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

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 prevalence of contamination and microbial concentrations (Lorimer and Kiermeier, 2007). These two variables represent together important inputs for quantitative microbial risk assessment (QMRA) (Commeau et al., 2012; Nauta et al., 2009a; Nauta et al., 2009b; 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 traditionally consists on a detection test applied to a complete sample set of food products, followed by an enumeration method applied to the positive sample units (Pouillot et al., 2013). .

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, the minimum concentration required for a count higher than zero is referred to as limit of quantification (LOQ)(Busschaert et al.,2011). 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 sample portion used for measurement (Busschaert et al., 2011), therefore varying among different experimental protocols, which complicates the comparison of studies performed with different microbial methods.

Although thresholds commonly adopted in microbiological analysis represent artificial concepts as shown by Evers et al. (2010), detection and quantification results are indeed subject to limitations: of test sensitivity and specificity (Currie, 1968; Nauta et al., 2009a) sample size and portion size(Straver et al., 2007) and randomness(Williams and Ebel, 2012). Therefore, "absence" in a detection test and "zero" in an enumeration test may consist on 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. 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); 2) there is no enrichment step performed during detection (Nauta et al., 2009a); 3) the test portion is small

(Straver et al., 2007). The amount of non-detects is hence also dependent on the method of microbiological analysis used (Gonzales-Barron et al., 2010).

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 sample under analysis. When biased 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. 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 for each sample unit, which results in an increased uncertainty of the overall characterization of the unit's contamination. For instance, a sample unit that tests positive in detection may present a true negative result in enumeration (Pouillot et al., 2013). Thus, it is erroneous to interpret all zero counts from enumeration as artificial zeroes.

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.

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 et al., 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., Gilchrist et al., 1973; 2010; 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 on 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 biased 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 datasets with zero counts to a lognormal distribution is the use of alternative parametric forms that allow for the occurrence of zeroes (Bassett et al., 2010; Gonzales-Barron et al., 2010; Gonzales-Barron and Butler, 2011). Gonzales-Barron et al. (2010) studied the use of heterogeneous Poisson distributions to model plate counts with a high occurrence of zeroes under a Bayesian modelling approach. Those authors observed that the

Poisson-gamma (PGM) and the Poisson-lognormal (PLN) distributions were good alternatives to the simple lognormal. However, the two distributions performed differently at distinct contamination levels – the PGM provided a better description of low concentrations and the PLN a better description of high concentrations (Gonzales-Barron and Butler, 2011).

A second challenge is the fact that the data used to fit a lognormal distribution consists on concentration estimates back-calculated from microbial counts. The use of derived estimates results in a loss of information that is initially entailed in the raw count data, and which use can contribute to a more accurate description of microbial contamination (Commeau et al., 2012; Nauta et al., 2009a). Recent advances in microbial analysis have been moving towards the use of raw data - initially with the use of counts instead of back-calculated concentrations (Gonzales-Barron et al., 2010; Gonzales-Barron and Butler, 2011), and more recently with the use of raw enumeration data together with presence/absence results (Commeau et al.,2012; Pouillot et al., 2013; Williams and Ebel, 2012). Commeau et al. (2012) studied in which conditions the use of raw data provided more accurate estimates compared to the use of concentration-like data and concluded that for observed prevalence between 25% and 85% a model using raw data had a better performance. In our study, we used raw count data as input in a model that estimates prevalence and concentration and assumed that enumeration data resulting from plate counting follows a PLN distribution.

1.3.Aim of our study:

With our study, we aim to 1) demonstrate why the LOQ should be excluded from the analysis of microbial data, and 2) develop a model that estimates the parameters of a distribution of microbial concentrations, using quantitative microbial counts generated by plate counting from a complete sample set, without assuming any LOQ. The model should also estimate the proportion of artificial zeroes among the total number of zero counts, thus providing an estimate of the true prevalence without the need for detection results.

2. Materials and Methods

In this study we consider a lot of food units from which a proportion p is contaminated, so p is the true prevalence in the lot. The $\log_{10}$ of the concentration of microorganisms (x=$\log_{10}(C)$, with concentration C in CFU/g) in contaminated food units follows a normal distribution, with mean μ and standard deviation σ, so x has a probability density function (pdf) $f(x)$:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

Next, we consider a plate counting procedure, where one test portion i is taken from a food unit j. If j is contaminated, it has concentration C_j, sampled from eq. (1). The test portion is enumerated in four 10-fold sequential dilution steps with three replicates each, starting with dilution 0.1. It is assumed that y, the total number of colony forming units (CFU) observed on all plates of different dilutions resulting from one test portion i of size m (in grams), taken from a sampled food unit with concentration C_j, can be described by a Poisson distribution with mean $\lambda_{m,j}$, where $\lambda_{m,j} = C_j \times m \times d$ with d the dilution factor resulting from 3x4=12 counted plates from the dilution series ($d=0.3333$). Hence, the probability of observing a total number of y CFU in the test portion is

$$\mathrm{P}(y|C_j) = \frac{\lambda^y e^{-\lambda}}{y!} = \frac{(C_j \times m \times d)^y e^{-(C_j \times m \times d)}}{y!} \quad (2)$$

where we here and in the following have left out indexes on y and λ for notational reasons. Note that here it is also assumed that the microorganisms are homogeneously distributed in each sample unit (and therefore in each test portion), that there are no measurement errors during the enumeration procedure and that every bacterial cell present in a test portion originates a countable colony.

2.1. LOQ is an artificial concept:

We investigate the suitability of the LOQ for the analysis of quantitative microbial enumeration data.

Using the Bayes' theorem, the pdf for a log concentration x=$\log_{10}(C)$ given a count of zero is calculated as

$$P(x|y=0) = \frac{P(y=0|x)P(x)}{P(y=0)} \quad (3)$$

where, applying eq. (2)

$$P(y=0|x) = e^{-(10^x \times m \times d)} \quad (4)$$

For a contaminated food unit P(x) = f(x) is given by eq. (1), and P(y=0) is a normalizing factor.

Hence, the pdf of concentrations that give zero counts, *g(x)*, is given by

$$g(x) = P(x|y = 0) \sim e^{-(10^x \times m \times d)} f(x) \quad (5)$$

Consequently, the pdf of concentrations that give positive counts *h(x)* is

$$h(x) = P(x|y > 0) \sim \left(1 - e^{-(10^x \times m \times d)}\right) f(x) \quad (6)$$

Here we study original concentration distributions representing three different levels of contamination as input, with means μ=1, μ=-2 and μ=2, all with standard deviation σ=1 . For comparison, we used two different LOQ values, 100 CFU/g and 10 CFU/g.

2.2.Data simulation:

In this study, the data were generated *in silico* by applying eqs. (1) and (2) for different values of μ and σ. The reference scenario has μ=1 and σ=1, alternative scenarios are μ=-2, μ=0 and μ=2 with σ=1, and μ=1 with σ=0.1 and σ=2.

Five hundred food units were sampled from each original input distribution of *x* and the corresponding plate counts *y* for one test portion from each food unit were simulated.

Five different scenarios of true prevalence *p* (*p*= 0.1, 0.3, 0.5, 0.7 and 0.9) were then simulated by inflating the simulated sample of 500 counts *y* with zeroes, where the added zeroes represented counts resulting from uncontaminated units (true zeroes).

The frequency of counts (y>0) and the frequency of zeroes (y=0) were extracted from the final set of microbial counts *y* and used as data input in the statistical model.

An alternative way of generating different true prevalence scenarios was analyzed, to assess the sensitivity of the model to sample size. In this case, the size of the final sample of simulated counts *y* was fixed at 500 and randomly selected counts were replaced by zeroes, to obtain the prevalence according to each scenario of *p*.

2.3.Model to estimate concentration and prevalence:

In a general situation, if the concentrations C in the food units follow a probability density function $\phi(C)$, and plate counting is performed as outlined above, the probability of a positive count is

$$P(y>0) = \int P(y>0|C)\varphi(C)dC = \int\left(1-e^{-(C\times m\times d)}\right)\varphi(C)dC \qquad (7)$$

Next, the probability of obtaining a specific plate count $i>0$, given a positive count, is:

$$P(y=i|y>0) = \frac{\int\frac{(C\times m\times d)^i e^{-(C\times m\times d)}}{i!}\varphi(C)dC}{\int\left(1-e^{-(C\times m\times d)}\right)\varphi(C)dC} \qquad (8)$$

In this study, the microbial data was assumed to follow a zero-inflated lognormal distribution, where C=0 with probability 1-p, introducing p as a parameter, and otherwise (with probability p) C is sampled from $\phi_s(C)$, the probability density function of the lognormal distribution for 10-based logarithms:

$$\varphi_s(C) = \frac{1}{C\sigma\sqrt{2\pi\times}ln(10)}e^{-\frac{(log_{10}(C)-\mu)^2}{2\sigma^2}} \qquad (9)$$

Hence, here the probability of obtaining a specific plate count $i>0$ is

$$P(y=i) = p\times\int P(y=i|Conc=C)\times\varphi_s(C)dC = \ p\times\int\frac{(C\times m\times d)^i e^{-(C\times m\times d)}}{i!}\varphi_s(C)\,dC \qquad (10)$$

Finally, the expected frequency n_i of a specific positive plate count $i>0$ was estimated as a product of the probability of obtaining a positive plate count y of size i and N, the size of the input dataset with plate counts.

$$E(n_i) = \ P(y=i)\times N \qquad (11)$$

Similarly, the expected frequency of zero counts $y=0$ was estimated as

$$E(n_0) = \left((1-p)+p\times\ P(y=0)\right)\times N \qquad (12)$$

Using these concepts, a computer program was developed in R software (R Development Core Team) to estimate the mean μ and standard deviation σ, and the true prevalence of contamination p, from a set of plate count data. The program aims to find the set of parameter values that provides the minimum adjusted sum of squared differences (SSD) between the observed (here *in silico* generated) counts ($n_{obs\,i}$) and their expectations ($E(n_i)$),

$$SSD_{adjusted} = \ \sum_{i=0}^{\infty}\left(\frac{\left(n_{obsi}-E(n_i)\right)^2}{E(n_i)}\right) \qquad (13)$$

where the squared difference between the observed and expected number of counts $y=i$ is divided by the expected number of $y=i$ counts.

For each scenario, the analysed set of parameter values included the input values for the parameters μ, σ and p, and values around it, so that in total 6x5x5 iterations were run for varying values of μ, σ and p respectively. For each concentration/prevalence combination scenario, a total of 1000 simulated data sets were analysed. The final output of the model consisted on the averages of best estimates of μ, σ and p, resulting from 1000 simulations. The method was implemented in R (R Development Core Team).

3. Results

3.1.LOQ is an artificial concept:

The pdf curves of concentrations originating artificial zeroes, *g*, and positive counts, *h*, show a consistent overlap of the left-hand tail of *h* with the right-hand tail of *g*. This illustrates how some concentration values have a distinct probability of generating either an artificial zero or a positive count. The results also show that the lower the mean of the within-lot distribution of $\log_{10}$ concentrations, *f*, the larger the intersection between *g* and *h*. Additionally, it is observed that for a high mean scenario (μ=2) there is an almost complete overlap between *f* and *h*, whereas for a low mean scenario (μ=-2) the overlap is higher between *f* and *g* (see figure 1). This indicates that at low means almost all concentrations generate an artificial zero and very few can generate a positive count, whereas at higher means an increasing number will generate a positive count and fewer can generate an artificial zero. Even in extreme situations (very high and very low means), there is always a certain probability of occurrence of the alternative outcome.

The cumulative density functions (cdf) *G* and *H*, together with *F*, are shown for the three different means in figure 1. For μ=-2, G shows that all the concentrations that can generate a zero count are below the LOQ of 10 CFU/g, whereas for μ=2, 1.05% is above that limit. However, for a LOQ of 100 CFU/g, the cumulative probability in G is 100% for all scenarios, so all concentrations generating artificial zeroes are below the LOQ. Still, the cumulative probability of both LOQ values is never 0% in *H*, so below the LOQ there are always concentrations generating positive counts. These results show that, there is no LOQ threshold above which all concentrations generate a positive count and below which all concentrations generate an artificial zero.

3.2.Estimates of concentration and prevalence:

The comparison between the final outputs of the prevalence estimate and the true prevalence was done at six different concentration scenarios (see figure 2). For one scenario (μ=-2, σ=1), the prevalence is always underestimated by approximately

10%. For the remaining, the method provides accurate prevalence estimates at the two lowest true prevalence levels analysed (0.1 and 0.3) and, at the highest levels (0.5, 0.7 and 0.9), it shows a tendency to underestimate p. However, this underestimation never exceeds 5% of the true value. Furthermore, for two of the concentration scenarios ((μ=2, σ=1) and (μ=1, σ=0.1)), the method performs equally well for all true values of prevalence.

For each combination of concentration distribution/ true prevalence, 1000 simulations were performed. In those simulations, the estimates of μ, σ and p varied among a group of six, five and five possible values, respectively, including the expected value. The empirical distributions of the 1000 estimates of each parameter were analysed for the five different prevalence subscenarios and compared with the expected values and the final estimates (average of 1000).

We observe that for the same scenario at which the prevalence is consistently underestimated by 10% (μ=-2, σ=1), the performance of the method is also poor in terms of prediction of mean (overestimated by 0.3 $\log_{10}$) and standard deviation (underestimated by 0.2 $\log_{10}$).

Figure 2 shows the boxplots, for the distribution of estimates of p. We observe that the method is very accurate at the two lowest true prevalence levels (0.1 and 0.3). The shape of these distributions varies between different concentration scenarios. For the distributions of estimates of p corresponding to (μ=1, σ=1), (μ=0, σ=1) and (μ=1, σ=2), it presents an approximately normal shape with a tendency to underestimation in a varying proportion of the simulations, which results in a general underestimation of the final estimate. However the underestimation of the final estimate does not exceed 5% of the true value. For the two remaining concentration scenarios ((μ=2, σ=1) and (μ=1, σ=0.1)), the observed shape has a higher kurtosis, generally with a very high proportion of the estimates corresponding to the expected value.

In general, these results show that the distribution tends to be wider when the prevalence increases and that the model provides better predictions of p when the distribution is less spread.

Figure 3 shows the boxplots for the distribution of estimates of μ. The results vary between concentration scenarios. In comparison to the reference (μ=1, σ=1), the method performs best with a distribution with lower σ (μ=1, σ=0.1), and worse with a distribution with higher σ (μ=1, σ=2). Furthermore, in the first case, the performance decreases with increasing true prevalence, whereas in the latter the final estimates are accurate at the three higher prevalence levels.

The performance of the method in terms of the final estimates of μ does not show a visible association with the spread of the distribution of the estimates from different

simulations – it provides good mean estimates for distributions with high (μ=1, σ=0.1) and low ((μ=1, σ=2) at $p \geq 0.5$) kurtosis.

At the concentration scenarios (μ=1, σ=1) and (μ=2, σ=1), the method consistently overestimates and underestimates the expected μ, respectively, at all prevalence levels. The distributions of the different simulation estimates in these two scenarios have normal-like shapes. At scenario (μ=0, σ=1), the method also consistently overestimates the expected μ, however by varying magnitudes depending on the prevalence level (on average, across all levels, by 0.05 $\log_{10}$). The shape of the distributions in this case deviates from the normal shape. The comparison between these three concentration scenarios, specifically between each alternative scenario and the reference, show that the increase or decrease by 1 $\log_{10}$ of the expected mean, for a constant standard deviation, does not have the same impact in terms of model performance. The increase by 1 $\log_{10}$ causes a shift from a 0.05 $\log_{10}$ overestimation to a 0.03 $\log_{10}$ underestimation, whereas the decrease by 1 $\log_{10}$ has a less marked impact that varies between different prevalence levels.

In summary, regarding the estimation of μ, the model performs clearly better in a concentration scenario with a low standard deviation, but also in the case of a high standard deviation and high prevalence ($p \geq 0.5$).However, we do not observe a clear relationship between different levels of prevalence or the shape and kurtosis of the between-simulations distribution and the performance of the method.

Figure 4 shows the boxplots for the distribution of estimates of σ. In all concentration scenarios with the exception of (μ=1, σ=0.1) the final estimate is an underestimation of the expected σ. This underestimation is mostly of equal magnitude across prevalence levels, except in the case of the distribution (μ=1, σ=2), where it is more marked for $p \geq 0.5$. The minimum underestimation magnitude is 0.03 $\log_{10}$ (for (μ=2, σ=1)) and the maximum is 0.18 $\log_{10}$ (for (μ=1, σ=2) at p=0.9). For the concentration scenario with a standard deviation lower than the reference (σ=0.1), the method overestimates the final estimate, especially at higher prevalence levels ($p \geq 0.7$), by 0.06 $\log_{10}$.

Regarding the estimation of σ, both scenarios with the lowest and the highest expected values of standard deviation ((μ=1, σ=0.1) and (μ=1, σ=2)) show a better performance at lower prevalence levels. However, the shape and kurtosis of the between-simulations distribution and the performance of the method do not seem to have an association.

In the results there is a clear association between the performance of the model in terms of estimation of p and estimation of σ. Figure 5 shows that, for all concentration scenarios analyzed, the trend in the difference between estimated and expected values is the same for both parameters for different prevalence levels.

Furthermore, with the exception of one scenario (μ=1, σ=0.1), these trends mirror negatively what is observed in relation to the estimation of μ.

As a way of assessing the performance of the method and the relevance of the underestimation and overestimation of parameters, the final outputs of the estimates of p, μ and σ were used as inputs in eqs (1), (2) and (10) to generate microbial counts y. The obtained frequencies of specific positive counts ($y=i|y>0$) and zeroes ($y=0$) were compared to the frequencies obtained with the expected values of p, μ and σ. For this comparison we selected three examples of combinations of concentration scenario/prevalence subscenario for which the under/overestimation of the parameters μ and σ was more pronounced.

The frequency distributions of both *in silico* generated sets of positive counts ($y>0$) are very similar, with some minor variation that can be allocated to two different causes - randomness inherent to the simulation and the combined effect of the errors in the estimation of the three parameters μ, σ and p. In general, the overestimation of the frequency of high microbial counts is predominantly associated to the overestimation of μ, even when the σ is underestimated, whereas its underestimation is related either to a combination of underestimation of σ and μ or σ and p.

The difference between the frequencies of zero counts (y=0) is low and less marked at p=0.1 (average difference of 0.025% of the expected number of zeroes) compared to p=0.9 (average difference of 7.5% of the expected number of zeroes). This result can be explained by the accuracy in the estimation of p - lower at high true prevalence and more accurate at low true prevalence.

The sensitivity of the model to sample size was analysed by fixing the size of the final sample of simulated counts y at 500. Results were obtained for the reference concentration scenario and compared to the previous results obtained for the same combinations of expected parameters.

In general, (see figure 6) we observe a wider bias in the performance of the method at p=0.3 (for the estimation of μ) or at the three lowest prevalence levels (for the estimation of σ). Otherwise, the impact of the decrease in the sample size is either negligible (for the estimation of p at low prevalence) or an improvement in the performance of the model (at high prevalence levels, for the estimation of all parameters). For a decrease in sample size of 900%, the maximum additional bias observed was of 0.08 $\log_{10}$ in the estimation of σ.

4. Discussion and Conclusion

Improvements in the analysis of microbial data emerge at a strong pace. The latest advances regard the use of raw data in the estimation of microbial concentrations, and the combined analysis of detection and enumeration results. The value of the MLE method for censored data, which not long ago was considered as an important step forward in the field, is now starting to be questioned (Pouillot et al., 2013; Williams and Ebel, 2012). Moreover, the established concept of LOD was shown to be a purely theoretical concept that does not depict the true nature of microbial detection, *i.e.* that it is a probabilistic event (Evers et al., 2010).

In this study on the analysis of microbial data, we aimed (1) to demonstrate the benefit of excluding the LOQ from the analysis and (2) to develop an alternative method of statistical analysis of enumeration data that does not employ a LOQ value for the interpretation of zeroes, differentiates between artificial and true zeroes and estimates the prevalence directly from enumeration results.
The first part of the study showed that it might be erroneous to assume that above a certain LOQ all concentrations generate a positive count, and below the same threshold, all generate an artificial zero. It was demonstrated that the validity of this assumption is dependent on the original distribution of concentrations and on the established LOQ value. The quantification of microorganisms by an enumeration method like plate counting is a probabilistic process. If the sample is contaminated and hence the concentration C is positive, the count can, in principle, always be both zero and positive. The probability that a count y is zero decreases with increasing concentration C in the sample, however it is never truly zero. We have introduced two probability density functions, g and h. In a food lot, the $\log_{10}$ concentration X (X = $\log_{10}$ C) may be assumed to follow a probability distribution with pdf f. We showed that the probability of a specific X in a unit drawn from a food lot with distribution f, given that a zero count is observed, follows probability distribution with pdf g, and is dependent on f (eq. 5). The same principle applies for the probability distribution of X given that a count is positive - h. The pdf's g and h have intersecting tails (see figure 1), and consequently there is a range of concentrations that can occur both if a zero or a positive count is observed. This range, or the overlap between g and h, is dependent on the distribution f (see figure 2).

We also illustrated why artificial zeroes should not be interpreted as censored below a certain LOQ. Depending on the cumulative density function of the distribution f, F, the cumulative density function of distribution g, G, will vary. Therefore, for the same LOQ value, the probability that a zero count originated from a concentration equal to or below the LOQ will also vary. This means that there might be food units with concentrations above the LOQ that can originate artificial zeroes.
As a consequence of these results, it can be said that the concept of a theoretically established LOQ is an artificial one, and may lead to a false interpretation of

microbial data. These findings corroborate the results of Evers et al. (2010), in relation to the LOD.

The second part of the study consisted of the development and analysis of a method that estimates the prevalence p of microbial contamination and the parameters defining the distribution of microbial concentrations in a food lot, μ and σ. The method does not assume a LOQ threshold and uses raw counts as input. In this study, hypothetical enumeration data were assumed to follow a zero-inflated Poisson lognormal distribution. As prevalence and concentration are often closely related, the method was tested for several concentration scenarios, with different combinations of μ and σ, and for five different p levels. Combinations included the cases of high prevalence-high concentrations and low prevalence-low concentrations as often observed for bacteria in food.

Among the three parameters, the method provided the most accurate estimates for p, independent of the expected concentration and prevalence. The largest inaccuracy found was an underestimation of p by 0.1 in relation to the true value (0.3, 0.5, 0.7 and 0.9), in a concentration scenario with a low mean for which the general performance of the method was poor. Otherwise, the estimate of p always approximated the true value, although it tended towards underestimation with increasing true prevalence. This can be explained by two factors that occurred at lower prevalence levels - the existence of a lower proportion of artificial zeroes among the total number of zero counts in the input dataset, and the larger sample size used as input. While a lower proportion of artificial zeroes represents a lower impact of the misclassification of a zero count on the p estimate, the use of a larger sample of microbial counts as an input also improves the accuracy of the method.

The method was less accurate for the estimation of μ and σ. In most analyses, it showed a tendency to either overestimate or underestimate the parameters by approximately the same magnitude across all simulated p levels. However, the performance varied among different concentrations.

The method overestimated μ at the reference and low mean concentrations, underestimated it at the highest expected mean, and performed better at lowest expected standard deviation. As the estimates of the three parameters are closely related (see eq. (10)), these results need to be interpreted together with the estimates of p, which are dependent on the interpretation of zero counts. At low mean concentration, many artificial zeroes are obtained in microbial enumeration from units with very low concentrations. If a share of these artificial zeroes is interpreted as true zeroes, the prevalence is underestimated and some of the units with very low concentration are not recognized as such. As a consequence, the estimated μ gets too high. In contrast, if the mean concentration is high, there are less artificial zeroes resulting from low concentrations, but due to the probabilistic nature of enumeration,

artificial zeroes will occur. If these zeroes are interpreted as resulting from low-contaminated units, the estimated μ might be lower than the expected.

Furthermore, the method underestimated σ at the reference and highest expected standard deviation and overestimated it at the lowest expected value. Again, estimates of p need to be taken into account, as a high standard deviation allows for the occurrence of very low concentrations that originate artificial zeroes. And oppositely, if the standard deviation is low, less artificial zeroes are produced and the method assumes some of the true zeroes as falsely originating from low concentrations. This causes an overestimation of σ.

In this study, we presented two examples to illustrate the importance of the classification of artificial and true zeroes and the estimation of p.

First, we analysed two different concentration scenarios with 1) a high proportion (scenario with μ=-2) and 2) a low proportion (scenario with μ=2) of artificial zeroes. These two scenarios represented opposite challenges to the performance of the method. We observed that it was particularly challenged when the proportion of artificial zeroes was very high (97%). Here, the method underestimated p and σ and overestimated μ. In a scenario with μ=2, (average of 15% of zeroes at 90% prevalence), the p estimates were accurate and μ and σ were negligibly underestimated.

Secondly, two concentration scenarios represented distributions with low (σ=0.1) and high (σ=2) standard deviation. It was observed that in the first scenario, with the lowest proportion of artificial zeroes, the method negligibly overestimated the three parameters μ, σ and p, whereas in the scenario with the highest proportion of artificial zeroes, all parameters were underestimated.

We conclude that the overall performance of the method is satisfactory for the interpretation of microbial counts under different conditions of true prevalence and concentration, with the best estimates obtained for prevalence and negligible bias observed in the estimation of mean and standard deviation.

The performance of the method is dependent on the fact that it was developed and evaluated with *in silico* generated data that followed the underlying assumptions. With real count data this performance will be dependent on the fulfilment of the assumptions that plate counting follows a Poisson distribution, that the microorganisms are homogeneously distributed within each sample unit, that there are no measurement errors during enumeration and that the choice of the parametric distribution to describe the within-lot occurrence of concentrations is appropriate.

Another limitation of the method is that the inference of each parameter is dependent on the initial guess provided. Here we adopted the expected value as the initial guess around which the iterations varied. Once the method is applied to real count data, the

expected values of μ and σ or the true p are unknown, and therefore the initial guesses must be derived from the input data. Further studies are needed to assess the performance of the method in such a situation.

A first investigation about the impact of the decrease in sample size showed that with small samples the method might be inaccurate especially at low prevalence levels (< 0.3) and mostly for the estimation of μ and σ. However, the impact we observed on the estimation of p was negligible. Further studies are needed to explore the decrease in performance of the method with smaller sample sizes.

Recent studies have presented new statistical methods for the analysis of microbial data. The method presented here has added value to these studies by offering an approach that does not assume a LOQ value and does not combine presence/absence results with enumeration data. It makes use of raw plate count data, to which it fits a discrete zero inflated PLN distribution, and estimates the prevalence of contamination.

When using real data, a PGM distribution might be a better choice to describe the counts than the PLN applied here (Gonzales-Barron and Butler, 2011). Therefore, the method was built in a way that allows the assumption of alternative parametric distributions to describe the distribution of concentrations. It can also be adapted to alternative experimental settings, by adjusting the number of dilution steps and the number of replicates per dilution (d) or the size of the test portion (m).

This method has the advantage of using raw count data from all performed dilutions, i.e. it does not imply the arbitrary selection of specific dilution steps to estimate concentrations, thus decreasing the impact of experimental uncertainty. Also, it does not need to assume a “higher-than” LOQ threshold, since high counts are grouped together in one interval during the fitting procedure.

It is of particular use in situations where information regarding both prevalence and concentration is necessary, such as in QMRA studies. A good differentiation between true and ”artificial” zeroes can for example be essential in QMRA studies of pathogens with a risk pathway that includes one or more growth steps. Here, it is important to reduce the number of artificial negative results that may arise from samples with low microbial concentration, since even low contaminated units may represent a risk.

Although the parameters obtained did not match exactly the correspondent original input distribution and true prevalence, we showed that, provided the assumptions behind microbial counting apply, two samples of size N simulated with both expected and estimated parameters would produce very approximate frequencies of microbial counts. In general, one can claim that the expected conditions and the estimated ones would represent similar inputs in an exposure assessment.

This method represents thus a new analytical tool to be used for the analysis of microbial enumeration data and eventually a possible tool for the future interpretation of quantitative results of alternative enumeration techniques that have started to be used in food microbiology, such as enrichment real-time PCR (Krämer et al., 2011).

Acknowledgements

The authors thank Dr. Sara Monteiro Pires for critically reading the manuscript.

References

Bassett, J., Jackson, T., Jewell, K., Jongenburger, I., Zwietering, M. (2010). Impact of microbial distributions on food safety. ILSI Europe Report, Brussels

Busschaert, P., Geeraerd, A. H., Uyttendaele, M., Van Impe, J. F. (2010). Estimating distributions out of qualitative and (semi)quantitative microbiological contamination data for use in risk assessment. International Journal of Food Microbiology, 138(3), 260-269.

Busschaert, P., Geeraerd, A. H., Uyttendaele, M., Van Impe, J. F. (2011). Hierarchical bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. Food Microbiology, 28(4), 712-719.

Cochran, W. G. (1950). Estimation of bacterial densities by means of the most probable number. Biometrics, 6(2), 105-116.

Commeau, N., Parent, E., Delignette-Muller, M., Cornu, M. (2012). Fitting a lognormal distribution to enumeration and absence/presence data. International Journal of Food Microbiology, 155(3), 146-152.

Currie, L. A. (1968). Limits for qualitative detection and quantitative determination - application to radiochemistry. Analytical Chemistry, 40(3), 586-593.

Evers, E. G., Post, J., Putirulan, F. F., Wal, F. J. (2010). Detection probability of campylobacter. Food Control, 21(3), 247-252.

FDA. (2006). Bacteriological Analytical Manual. Appendix 2: Most Probable Number from serial dilutions.

Gilchrist, J.E., Campbell, J.E., Donnelly, C.B., Peeler, J.T., Delaney, J.M. (1973). Spiral plate method for bacterial determination. Applied and Environmental Microbiology, 25(2), 244-252.

Gonzales-Barron, U., Butler, F. (2011). A comparison between the discrete poisson-gamma and poisson-lognormal distributions to characterise microbial counts in foods. Food Control, 22(8), 1279-1286.

Gonzales-Barron, U., Kerr, M., Sheridan, J. J., Butler, F. (2010). Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. International Journal of Food Microbiology, 136(3), 268-277.

Helsel, D. R. (2006). Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. Chemosphere, 65(11), 2434-2439.

Kilsby, D. C., Pugh, M. E. (1981). The relevance of the distribution of micro-organisms within batches of food to the control of microbiological hazards from foods. Journal of Applied Bacteriology, 51(2), 345-354.

Krämer, N., Löfström, C., Vigre, H., Hoorfar, J., Bunge, C., Malorny, B. (2011). A novel strategy to obtain quantitative data for modelling: Combined enrichment and real-time PCR for enumeration of salmonellae from pig carcasses. International Journal of Food Microbiology, 145(1)

Lorimer, M. F., Kiermeier, A. (2007). Analysing microbiological data: Tobit or not tobit? International Journal of Food Microbiology, 116(3), 313-318.

Marks, H., Coleman, M. (1998). Estimating distributions of numbers of organisms in food products. Journal of Food Protection, 61(11), 1535-1540.

Müller, A., Hildebrandt, G. (1990). Sampling errors and systematic errors in bacterial count determination. The accuracy of colony counting. Fleischwirtschaft, 70(6), 680-684.

Nauta, M., Hill, A., Rosenquist, H., Brynestad, S., Fetsch, A., van der Logt, P., Fazil, A., Christensen, B., Borck, B., Havelaar, A. (2009a). A comparison of risk assessments on campylobacter in broiler meat. International Journal of Food Microbiology, 129(2), 107-123.

Nauta, M. J., van, d. W., Putirulan, F. F., Post, J., van, d. K., Bolder, N. M. (2009b). Evaluation of the testing and scheduling strategy for control of campylobacter in broiler meat in the Netherlands. International Journal of Food Microbiology, 134(3), 216-222.

Nauta, M. J. (2002). Modelling bacterial growth in quantitative microbiological risk assessment: Is it possible? International Journal of Food Microbiology, 73(2-3), 297-304.

Perez-Rodriguez, F., Asselt, E. D., Garcia-Gimeno, R., Zurera, G., Zwietering, M. H. (2007). Extracting additional risk mangers information from a risk assessment of listeria monocytogenes in deli meats. Journal of Food Protection, 70(5), 1137-1152.

Pouillot, R., Hoelzer, K., Chen, Y., Dennis, S. (2013). Estimating probability distributions of bacterial concentrations in food based on data generated using the most probable number (MPN) method for use in risk assessment. Food Control, 29(2), 350-357.

Shorten, P. R., Pleasants, A. B., Soboleva, T. K. (2006). Estimation of microbial growth using population measurements subject to a detection limit. International Journal of Food Microbiology, 108(3), 369-375.

Straver, J. M., Janssen, A. F. W., Linnemann, A. R., Boekel, M. A. J. S., Beumer, R. R., Zwietering, M. H. (2007). Number of salmonella on chicken breast fillet at retail level and its implications for public health risk. Journal of Food Protection, 70(9), 2045-2055.

Williams, M. S., Ebel, E. D. (2012). Methods for fitting a parametric probability distribution to most probable number data. *International Journal of Food Microbiology, 157*(2), 251-258.

Figures

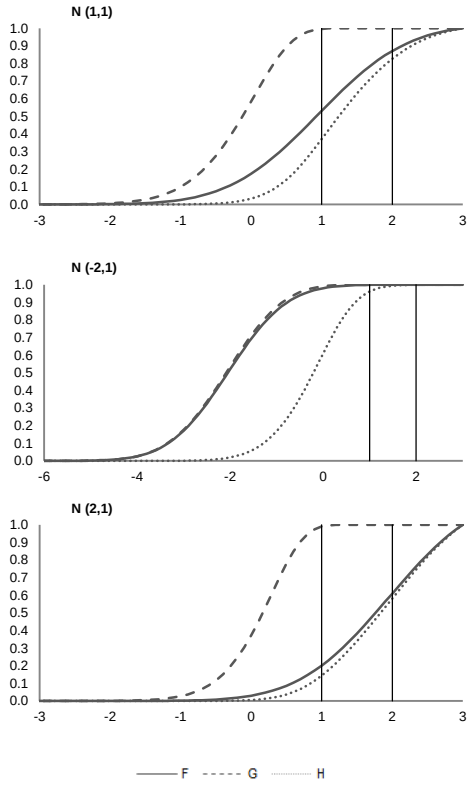

Figure 1: Cumulative density functions of microbial concentrations in a food lot (*F*), concentrations originating artificial zeroes (*G*), and concentrations originating positive counts (*H*), for scenarios with different concentration means (μ=1, μ=-2 and μ=2). The horizontal axis represents microbial concentration ($\log_{10}$ CFU/g) and the vertical axis represents cumulative probability density. The vertical lines represent a low (10 CFU/g) and a high (100 CFU/g) LOQ.

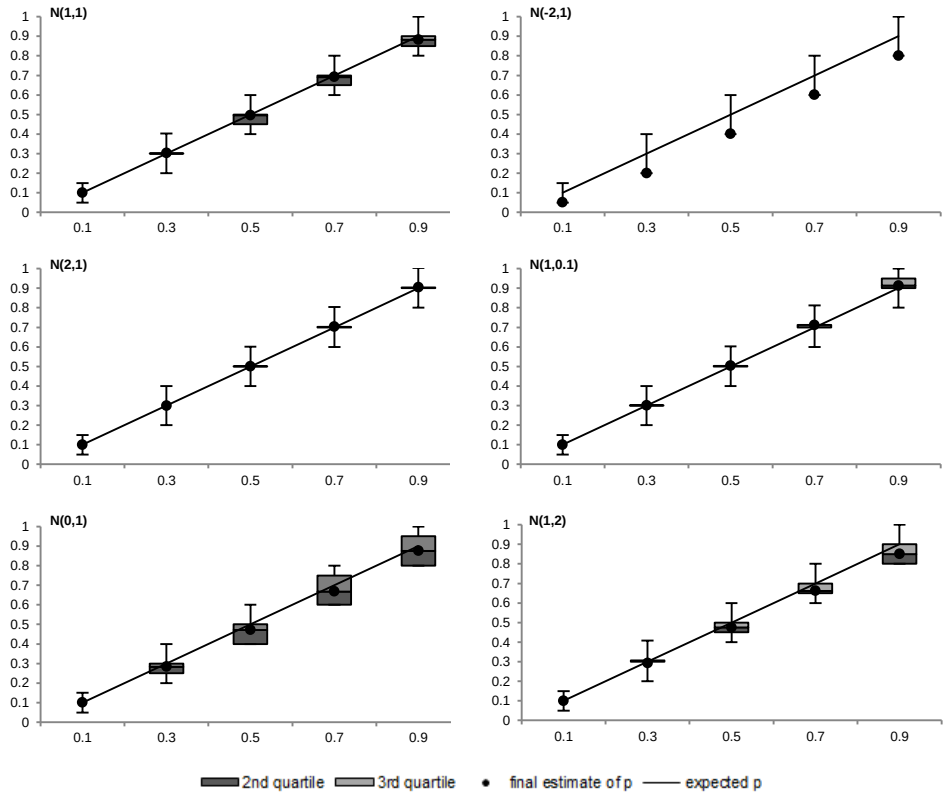

Figure 2: Distribution of the prevalence estimates of 1000 simulations (vertical axis), for five different prevalence scenarios and six different distributions of the parametric form Normal (μ,σ) of x=log10(C). Values on the horizontal axis represent the true prevalence of the scenario. The boxplots show the mean or final estimate of p(dot) , the expected p(line), quartiles and the minimum and maximum values.

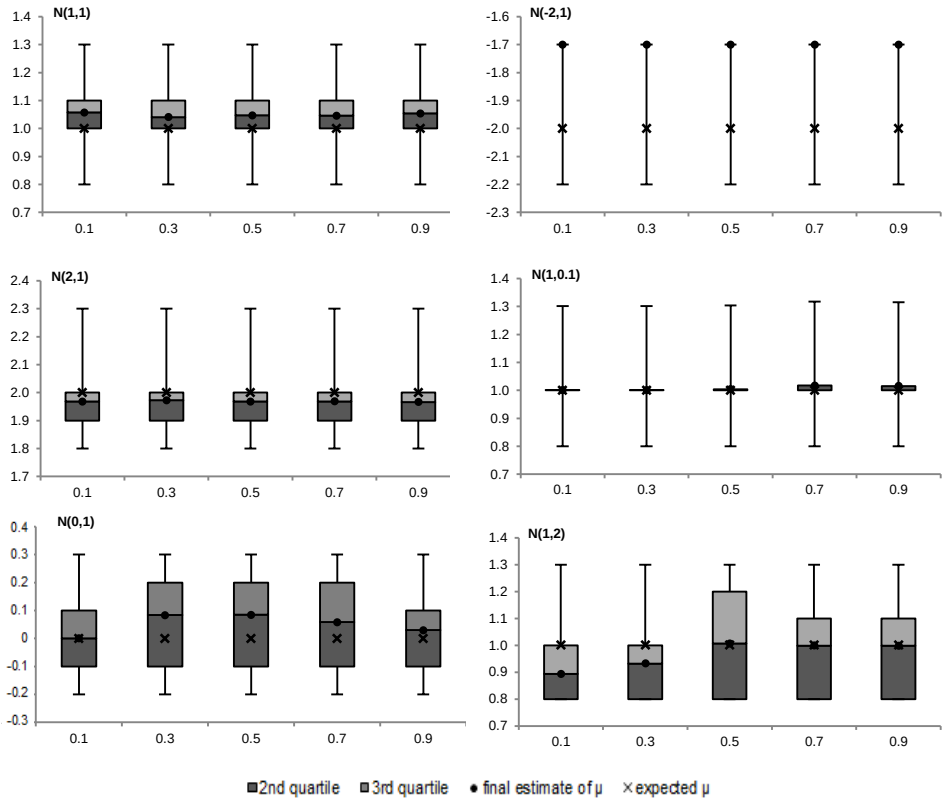

Figure 3: Distribution of the estimates of μ of 1000 simulations (vertical axis), for five different prevalence scenarios and six different distributions of the parametric form Normal (μ,σ) of x=log10(C). Values on the horizontal axis represent the true prevalence of the scenario. The boxplots show the mean or final estimate of μ(dot) , the expected μ(cross), quartiles and the minimum and maximum values.

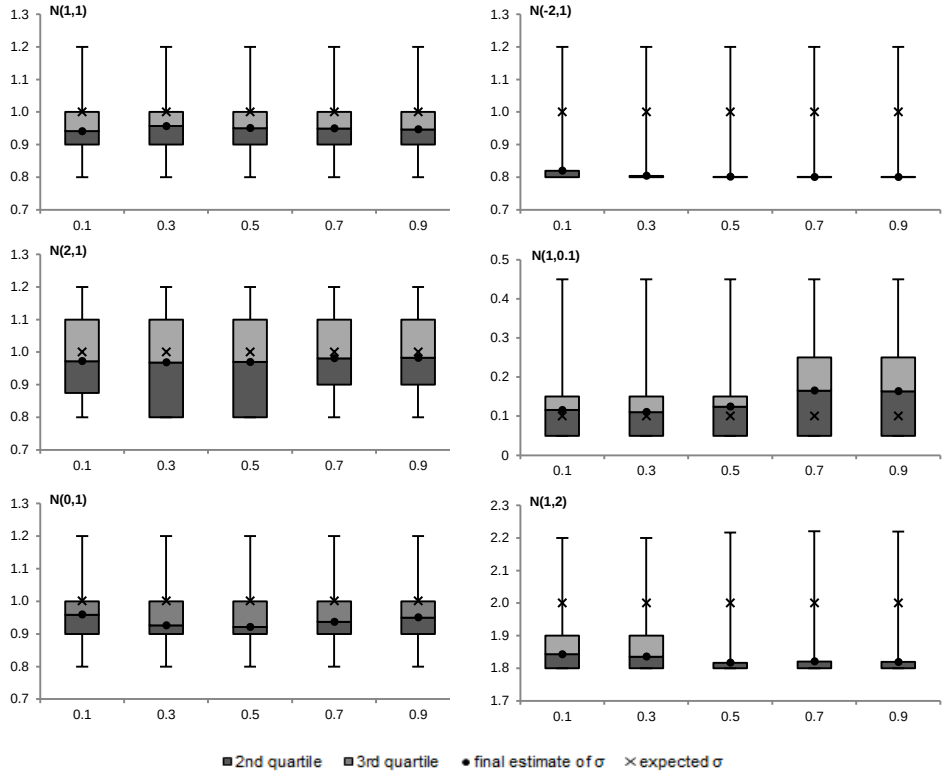

Figure 4: Distribution of the estimates of σ of 1000 simulations (vertical axis, for five different prevalence scenarios and six different distributions of the parametric form Normal (μ,σ) of x=log10(C). Values on the horizontal axis represent the true prevalence of the scenario. The boxplots show the mean or final estimate of σ(dot) , the expected σ(cross), quartiles and the minimum and maximum values.

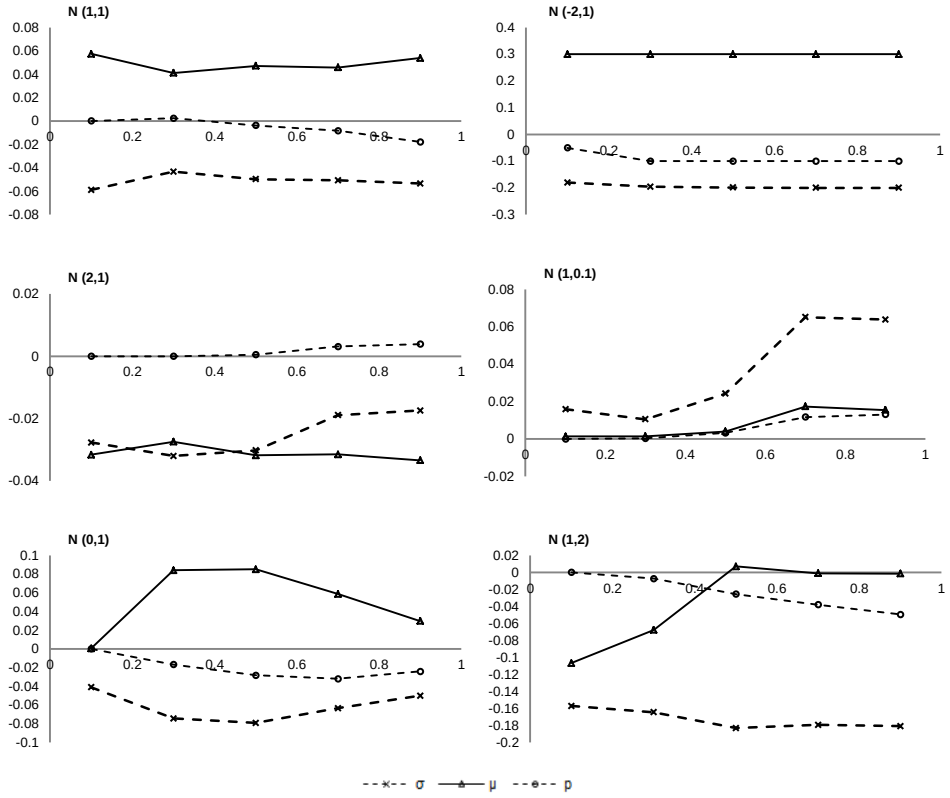

Figure 5: Difference between the final estimates and the expected values of the three parameters p, μ and σ (vertical axis) for each prevalence subscenario and for each input distribution of concentrations. Values on the horizontal axis represent the true prevalence of the scenario. Values on the vertical axis represent the difference in log10 CFU for μ and σ and in percentage for p.

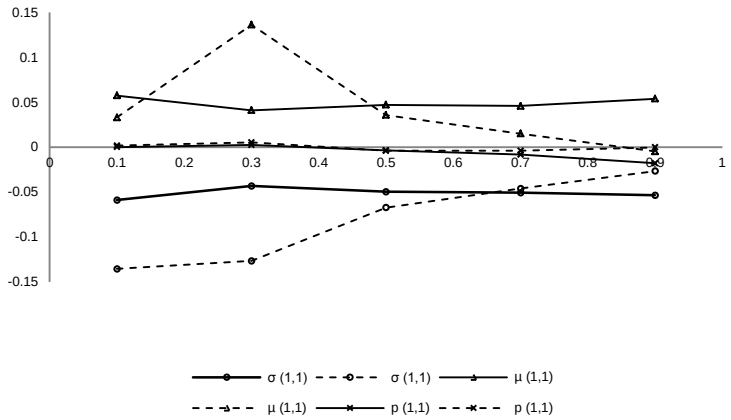

Figure 6: Difference between the final estimates and the expected values of the three parameters p, μ and σ (vertical axis) for each prevalence subscenario, for the reference input distribution of concentrations N(1,1), with two different sampling procedures. Values on the horizontal axis represent the true prevalence of the scenario. Values on the y-axis represent the difference between in log10 CFU for μ and σ and in percentage for p. Each symbol represents a different parameter. Continuous lines represent the results obtained with larger samples and dashed lines the results obtained with smaller samples.

Discussion

Manuscript I presents a new method of analysis of raw microbial counts that estimates both the prevalence and the distribution of microbial concentrations.

Microbial contamination in foods, especially when used as input in QMRA, is described both in terms of prevalence and distribution of concentrations. These are usually detected separately, with detection and enumeration methods, respectively. Detection and enumeration methods are both subject to limitations that cause the observed number of positive sample units or the observed number of CFU to deviate from what exists in reality. The main limitations are the natural randomness in testing, the sensitivity and specificity of the methods used and the size of the test portion used for analysis. Also, the fact that microbial detection and enumeration can be preceded by an enrichment step consists on a source of uncertainty in the analysis of the resulting data; it has been shown that, for example, different prevalence estimates were obtained by using a protocol with or without an enrichment step (Nauta et al., 2009b). The enrichment media used, the duration of the enrichment step and the natural microflora present in the sample (and the adaptation of the different microorganisms to the enrichment media) are factors that can cause uncertainty in observations performed after enrichment.

The above mentioned limitations of microbial testing are the factors that should be considered during the analysis of microbial data, instead of artificially defined threshold values for detection or quantification of microorganisms, such as the LOD and LOQ. The statistics derived from microbial data when such limits are applied in the analysis can be more or less biased depending on the data itself and on the threshold value adopted. In this study, a method for the analysis of microbial counts was developed where no LOQ value is applied. Although the method does not describe the enrichment conditions and assumes a100% recovery on plate counts, it does include as parameters the size of the test portion used for analysis and the homogeneous nature of plate counting.

The method was developed and tested with simulated data, in order to evaluate its performance. With simulated counts, there are no such limitations as those implied in plate reading. In fact, it is desirable that such limitations are also not applicable in reality. Therefore, the method was built in a way that allows including the reading of all plates produced during enumeration. All counts are treated together: zeroes are differentiated into true or artificial; too-high, uncountable numbers of CFU are pooled together during the estimation of the distribution of concentrations, as right-censored observations.

The limitations of the MPN method of enumeration have been presented in section ii of introduction. The new statistical method is therefore aimed to deal with quantitative plate counts instead. A challenge for its application in real life might be the fact that MPN is often preferred for microbial analysis of foods because it is difficult to enumerate microorganisms by direct plate counts when the concentrations are low.

On the other hand, microbial analysis is commonly performed separately in terms of prevalence and concentration estimation. It was shown in a previous study that contaminated food units can give a negative result in a detection test (Evers et al., 2010), whereas the present study showed that contaminated units with the same concentration can give either zero or positive microbial counts, with different probabilities depending on the original distribution of concentrations. For both reasons, microbial analysis will be less sensitive to uncertainty if positive and negative sample units are not separated before microbial enumeration. The joint analysis of both types of units carries important information that helps the accurate estimation of prevalence and distribution of concentrations. This information is otherwise ignored once only the assumed positive units are forwarded to enumeration after detection. If prevalence can be estimated directly from enumeration results, detection tests may even become unnecessary to perform. However, such suggestion contradicts the often economic reasoning of microbial analysis. Microbial enumeration is expensive and time-consuming, hence detection tests are usually performed in order to reduce the number of units enumerated to the strictly necessary – the positive ones – especially at low prevalence levels. A cost-utility study can help

to clarify the actual feasibility of adopting a routine of microbial analysis where only enumeration is performed.

The challenge will lie on finding an appropriate number of sample units that once enumerated will allow for an accurate estimation of both prevalence and distribution of concentrations with the new method. This number must be both acceptable in terms of the accuracy of results but also in terms of the economic and time-resource burden that it might represent in terms of laboratory work.

Once this challenge has been overcome, the new method will have the advantage of using raw data, i.e. data that has not been summarized in any way, such as back-calculated concentrations are. It does not account however so far with measurement uncertainty, and the sampling uncertainty was negligible with the sample size used in this study. Before application in reality, it is advisable that the impact of uncertainty on the method's performance is assessed.

The estimation of prevalence and concentration using the same data source and the same statistical method is one of the main advantages of this study. It has been demonstrated in section iii of introduction that considering all zeroes as belonging to contaminated units results in a different description of the distribution of concentrations compared to considering some zeroes as true, by estimating the prevalence. Care should be taken though on the application of statistical methods that, like the one here presented, provide estimates of prevalence. Ultimately, the choice of the method for analysis of microbial data must depend on how such data were collected. If one deals with enumeration data that was obtained strictly from units that tested positive to detection, it might not make sense to estimate prevalence from those data.

However, it is known that prevalence and concentration are closely interrelated. It thus makes sense that they are estimated together. A possible flaw of this new method of analysis is that it does not account for this correlation between prevalence and concentration. To do so might however impact negatively the outcome of the method, since the degree of such correlation is difficult to determine and therefore its modelling would be most likely based on disputable assumptions.

Here the data was generated and fitted to a zero-inflated Poisson-lognormal (ziPLN) distribution. This has obviously an impact on the performance of the model. Poorer performances would be expected in case the concentrations did not follow a lognormal distribution, the CFU counts did not follow a Poisson process and there would be no extra variation caused by an extraordinary number of zero counts. The method can easily be adapted to alternative distribution realities. However, the challenge is to guess correctly what the original distribution behind the observed data might be. The PLN has been so far the distribution indicated as the most appropriate to describe microbial count data (Bassett et al., 2010). It has the advantage of describing CFU counts instead of back-calculated concentrations, which allows the use of raw count data instead of summarized data.

PART II

Manuscript II

Impact of microbial count distributions on human health risk estimates

Duarte, A.S.R. ; Nauta, M.J.

Submitted

Title

Impact of microbial count distributions on human health risk estimates

Authors

Duarte, A.S.R.[a]; Nauta, M.J.[a]

[a] Technical University of Denmark – National Food institute

Mørkhøj Bygade, 19, Building H

DK-2860 Søborg

Denmark

asrd@food.dtu.dk

maana@food.dtu.dk

Corresponding author

Sofia Duarte

asrd@food.dtu.dk

+45 23673256

Abstract

Quantitative microbiological risk assessment (QMRA) is influenced by the choice of the probability distribution used to describe pathogen concentrations, as this may eventually have a large effect on the distribution of doses at exposure. When fitting a probability distribution to microbial enumeration data, several factors may have an impact on the accuracy of that fit. Analysis of the best statistical fits of different distributions does not provide a clear indication of the impact in terms of risk estimates.

Thus, in this study we focus on the impact of fitting microbial distributions on risk estimates, at two different concentration scenarios and at a range of prevalence levels. By using four different parametric distributions, we investigate whether different characteristics of a good fit are crucial for an accurate risk estimate. Among the factors studied are the importance of accounting for the randomness in counts, the difference between treating "true" zeroes as such or as censored below a threshold value and the importance of making the correct assumption about the underlying distribution of concentrations.

By running a simulation experiment with zero-inflated Poisson-lognormal distributed data and an existing QMRA model from retail to consumer level, it was possible to assess the difference between expected risk and the risk estimated with using a lognormal, a zero-inflated lognormal, a Poisson-gamma and a zero-inflated Poisson-lognormal distribution.

We show that the impact of the choice of different probability distributions to describe concentrations at retail on risk estimates is dependent both on concentration and prevalence levels. In general a zero-inflation improves the risk estimates.

Keywords: Count-distribution; Zero-inflation; Poisson-lognormal; Prevalence; Censored data; Quantitative microbiological risk assessment

1. Introduction

Quantitative microbiological risk assessment (QMRA) depends on consistent descriptions of the distributions of pathogen concentrations in food products (Bassett et al., 2012; Busschaert et al., 2010; Busschaert et al., 2011; Nauta et al., 2007; Straver et al., 2007), since these distributions eventually have a large effect on the distribution of doses at exposure. Furthermore, the concentrations in the high value tail of the distribution often have most impact on the risk (Bassett et al., 2010; Straver et al., 2007). Usually, concentrations are represented by probability distributions fit through enumeration data obtained from a set of samples. One of the most frequently used distributions for this purpose is the lognormal (LN), due to its generally good fit to enumeration data and its attractiveness for statistical testing (Gonzales-Barron & Butler, 2011; Kilsby & Pugh, 1981) .

When fitting a distribution to enumeration data from a sample of food products, several factors have an influence on the accuracy of the fit. First, low contaminated sample units can give zero counts ("artificial" zeroes) that add up to the number of "true" zeroes resulting from non-contaminated units, thereby inflating the total number of zeroes in a sample of microbial counts. To deal with this issue, two approaches can be adopted: to treat the total number of zeroes as left-censored data (results below a limit of quantification (LOQ)) (Busschaert et al., 2010; Busschaert et al., 2011; Delignette-Muller et al., 2010; Helsel, 2005a, 2005b; Lorimer & Kiermeier, 2007; Shorten et al., 2006), or alternatively to use zero-modified distributions, such as zero-inflated models (Bassett et al., 2010; Gonzales-Barron et al., 2010; Ridout et al., 1998) to model prevalence and concentration. Zero-inflated distributions specify the probability of obtaining a non-contaminated unit and the concentration distribution for the contaminated ones, thus allowing for a separation between "true" and "artificial" zeroes.

Second, the test portion taken from each sample unit is homogenized prior to serial dilution and enumeration, and the enumerated colony forming units (cfu) can be assumed to be generated by a Poisson process. The resulting distribution of cfu has been referred to as the "measurement distribution" (Gonzales-Barron & Butler, 2011) and is theoretically close to a Poisson. Continuous distributions, like the LN or the gamma, although considered appropriate to model the distribution of concentrations at retail, do not account for this Poisson process at the measurement level. They exclusively describe the heterogeneity in concentrations between food products. Generalized Poisson distributions, like the Poisson-gamma (PGM) or the Poisson-lognormal (PLN) are considered more mechanistic as they describe both realities (Reinders et al., 2004) – distribution of concentrations and "measurement distribution".

Third, microbial concentrations in food are often considered to be lognormally distributed (Busschaert et al., 2010; Crepet et al., 2007; Kilsby & Pugh, 1981). However, the frequency distribution of pathogens in food is commonly characterized by a high probability of zeroes and low concentrations, i.e. it presents clustering or over-dispersion (Bassett et al., 2010). This complicates the fulfilment of lognormality, as the LN distribution does not allow zero as an outcome. Recently, several alternatives to the LN distribution have been proposed to represent microbial contamination data with low prevalence and low concentrations more appropriately (Bassett et al., 2010; Gonzales-Barron et al., 2010; Gonzales-Barron & Butler, 2011). Among these are the discrete generalizations of the Poisson distribution, PGM and PLN, due to their ability to model count data with over-dispersion and the zero-inflated distributions for their ability to model data with a substantial amount of zeroes.

Fourth, data sets usually consist of concentration estimates that were back-calculated from enumeration data from samples of food products. The uncertainty associated to food product sampling (Zhao & Frey, 2004) and the measurement uncertainty (Marks & Coleman, 1998) influence the difference between the "true" distribution of concentrations in the food products and the fitted one(s). Only the selection of an extremely representative sample and a perfect enumeration procedure (without measurement errors), both unlikely to achieve in reality, could minimize those uncertainties.

Recently, many authors have explored different solutions to these challenges by studying choices of the probability distribution to fit through enumeration data (Bassett et al., 2010; Gonzales-Barron et al., 2010; Gonzales-Barron & Butler, 2011; Reinders et al., 2004). However, most research on fitting distributions to microbial data has focused on the best statistical fits of different models to existing datasets. Bassett et al. (2010) additionally investigated the impact of microbial distributions on public health, but without considering different levels of prevalence of contamination.

In this study, we took this research question one step further and investigated the impact of fitting microbial distributions on QMRA estimates at different concentration and prevalence levels, using a hypothetical example of *Campylobacter* on broiler meat. We assessed the impact of the use of different parametric distributions and of different ways of dealing with zeroes when fitting those distributions, either by separating "true" from "artificial" zeroes (i.e. by estimating prevalence) or by treating all zeroes as left-censored results below a certain LOQ (i.e. by assuming 100% prevalence). By running a simulation experiment at different prevalence scenarios while keeping the same original concentration distribution, we assessed the advantage of using zero-inflated distributions to model prevalence

separately from concentrations, depending on the probability of occurrence of "true" zeroes.
Furthermore, we investigated whether the impact of fitting different distributions on the risk changed depending on the contamination level at retail. For that purpose, risk estimates obtained for two realistic levels of contamination were compared: the lowest and highest contamination levels of broiler meat at Danish retail.

2. Materials and Methods

2.1.Simulating different scenarios of retail concentration

Two scenarios were analysed, representing "true" distributions of *Campylobacter* concentrations (*C*) in broiler meat at retail. For each scenario, we defined a LN distribution:

$$log_{10}(C) \sim Normal(\mu, \sigma) \quad (1)$$

with geometric mean μ and geometric standard deviation σ.
Values for μ and σ were based on the analysis of Danish retail data (data not shown). The highest contamination scenario (μ= 0.75 $\log_{10}$ cfu/g meat) was based on the chilled meat data for high prevalence seasons in years 2004 to 2007; the lowest contamination scenario (μ= 0 $\log_{10}$ cfu/g meat) was based on the frozen meat data for the low prevalence seasons in the same period (Boysen et al., 2011). The adopted standard deviation (σ= 1) was based on the analysis of the same retail data and did not differ substantially between the low and the high season.

From each distribution, we randomly sampled 500 units, representing a set of "true" concentrations in 500 broiler meat retail products.

Next, we simulated the experimental enumeration procedure for that sample. We assumed four serial dilutions with three replicates each, and a standard portion weight of 10 g taken for analysis. In the first dilution step (*j=0*), we simulated the homogenization of the 10 g portion in a 90 ml volume, using a stomacher. Consequently, the number of cfu (N_{ij}) for each sample unit *i*, at each dilution step *j* resulted from a Poisson process with mean λ, where $\lambda = C_{ml\_ij}$ x 3.

$$N_{ij} \sim Poisson(C_{ml\_ij} \times 3) \quad (2)$$

and

$$C_{ml\_ij} = \frac{C_i}{10} \times d_j \quad (3)$$

where C_i represents the "true" concentration in cfu/g of a sample unit *i* (*i =1,2,…,500*), C_{ml_ij} represents the concentration in cfu/ml of that unit at dilution step *j* (*j = 0,1,2,3*) and d_j represents the dilution factor of the concentration in cfu/ml (10^{-j}) at dilution step *j*.
The estimate of the concentration in cfu/ml ($C_{ml\_est\_i}$) for each sample unit *i* was calculated as a weighted average of the results

$$C_{ml\_est\_i} = \frac{\sum_{j=1}^{3} N_{ij}}{\sum_{j=1}^{3}(d_j \times 3)} \quad (4)$$

and the back-calculated estimate of the concentration in cfu/g (C_{est_i}) for the same unit corresponded to $C_{ml\_est\_i}$ x 10.

2.2. Simulating different levels of prevalence

For each concentration scenario, the sample sets were zero-inflated to investigate the effect of prevalence (p). Ten levels of p were constructed by inflating the initially drawn samples with zeroes. For a 100% prevalence level there was no zero-inflation, therefore zero counts originated exclusively from sample units with extremely low concentrations ("artificial" zeroes). As p decreased, the zero-inflated samples were constructed by adding an increasing number of zeroes to the initial sample of 500 concentration values; in these cases, the zero counts resulted both from sample units with very low concentration and from uncontaminated units ("true" zeroes).

2.3.Fitting distributions to the concentration estimates

The twenty datasets of cfu counts (*N)* and back-calculated concentration estimates (C_{est}), obtained for the two scenarios with different prevalences ,were used for distribution fitting with four different probability distributions : a lognormal (LN), a zero-inflated lognormal (ziLN), a zero-inflated Poisson-lognormal (ziPLN) and a Poisson-gamma (PGM).

The continuous distributions LN and ziLN were fit to C_{est} datasets (cfu/g) with a MLE approach for censored data (Busschaert et al., 2010; Busschaert et al., 2011; Delignette-Muller et al, 2010; Helsel, 2005a, 2005b; Lorimer & Kiermeier, 2007; Shorten et al., 2006) using the Solver add-in for Excel 2010 to maximize the log-likelihood. For this purpose, the data was arranged in semi-quantitative intervals of 1 $\log_{10}$ and the total number of zeroes was treated as censored between -∞ and a LOQ threshold. We investigated the use of two different LOQ values: 0 $\log_{10}$ and -1 $\log_{10}$ cfu/g. The main results were produced with the highest LOQ and compared to the results obtained with the lowest LOQ.

In this approach we assumed that C < LOQ occurs with probability:

$$P(C < LOQ) = (1 - p) + \left(p \times \Phi(LOQ|\mu, \sigma)\right) \quad (5)$$

where p is the prevalence, and Φ(LOQ|μ, σ) is the probability that the concentration in a contaminated food product is below the LOQ, calculated from the normal cumulative distribution function with parameters μ and σ. Hence, the first term of eq.(5) (1-p) represents the probability of "true" zeroes and the second term represents the probability of positive concentrations below the LOQ, or "artificial zeroes".

Otherwise, for non-zero concentration estimates, we assumed that, for all concentration intervals {L,H}, with $L \geq$ LOQ, a concentration in this interval is found with probability

$$P(L < C < H) = p \times (\Phi(H|\mu,\sigma) - \Phi(L|\mu,\sigma)) \quad (6)$$

Using eqs (5) and (6) the MLE finds the most likely values of p, μ and σ for the simulated concentration data. For the fit of LN, p was fixed at 1, i.e. all concentrations were assumed to originate from contaminated units.

The discrete ziPLN distribution was fit using a similar method. This method also estimates μ and σ of a normal distribution of log concentrations, and the "true" prevalence of contamination p, but using a set of plate count data N_{ij} directly. It has been implemented in a computer program recently developed in R software (Duarte et al., in prep.).
Finally, the PGM distribution was fit to datasets of simulated counts N (cfu) by using the software @RISK 5.7 (Palisade) to fit a negative binomial distribution.

2.4. *Implementing zero-inflated distributions*

To implement the zero-inflated distributions ziLN and ziPLN, we defined the probability of obtaining a non-zero concentration or count, respectively, as

$$P(+) \sim Binomial\,(1, p_{est}) \quad (7)$$

This binomial process generated an output of one with probability p_{est} and an output of zero (or an extremely low $\log_{10}$ value, for the ziLN) with probability 1-p_{est}. An output of 1 implied a positive result that was sampled from $f(x)$ with parameters μ_{est} and σ_{est} and an output of 0 corresponded to a "true" zero.

2.5. *Estimating risk with different distributions*

To compare the impact of the use of different distributions on risk estimates, each fitted probability distribution was used as an input in an existing QMRA model (Nauta & Christensen, 2011; Nauta et al., 2012). This model combines a consumer-phase model for broiler meat with a dose response model for *Campylobacter* (Teunis & Havelaar, 2000) to obtain estimates for the probability of illness (P_{ill}). The consumer-phase model includes an empirical distribution for the transfer rate of *Campylobacter* from raw meat to salad (Nauta et al., 2008). The initial number of cfu in a portion of raw meat ($N_{portion}$) depends on a Poisson process which mean is the concentration on the retail product (C in cfu/g)) times the serving size (w_C in g). The model used here is equal to that published by Nauta et al. (2012), except that the serving size w_c is now fixed at 100 g, and the transition factor τ=0.7 (Christensen et al., 2013). It provides a relation between the concentration in the samples and the probability of illness, as in a microbiological dose response relationship.

The estimates of P_{ill} for the fitted distributions were compared with the expected P_{ill} for the "true" distribution of C. With zero inflation, the decrease in P_{ill} was proportional to the decrease in prevalence p.
The QMRA model was implemented in @RISK 5.7 (Palisade).

3. Results

The risk estimates, or mean probabilities of illness, were obtained for each fitted distribution and compared with the expected risk, at 10 prevalence levels and two concentration scenarios (see figure 1).

At the low concentration (μ=0; σ=1) the ziLN and ziPLN provide the most accurate estimates of risk, with a maximum deviation from the expected risk of approximately 0.005% (overestimation by the ziPLN at 100% prevalence). The LN and the PGM distributions supply comparable risk estimates. Both overestimate the risk at all p levels, especially at highest prevalence (0.007 to 0.009% overestimation at p >50%). In general, all distributions with the exception of ziPLN, overestimate the risk at all p levels.

At the high concentration (μ=0.75; σ=1), the LN distribution performs best across all prevalence levels. However, despite of the accurate estimates obtained, this distribution has a tendency to marginally overestimate the risk at lowest prevalence (maximum 0.009% overestimation at 30% level) and to underestimate it at highest prevalence (maximum 0.01% underestimation at 100% level). The performance of the distributions ziLN and ziPLN is similar across all levels – comparable risk estimates and a general trend of increasing underestimation with increasing p (maximum 0.03% underestimation by ziPLN at 90% level). In contrast to what is observed at the low concentration, the results obtained with the PGM distribution are not similar to those observed with the LN. The PGM consistently overestimates the risk, at increasing magnitude for increasing p (maximum 0.06% overestimation at 100% level).

In general, we observe that the deviation between estimated and expected risk, at both concentration scenarios, is higher at high prevalence levels independently of the distribution used to describe the microbial concentrations. The individual performance of each fitted distribution varies between the two concentration scenarios. At the low concentration, the risk estimates obtained with different distributions are more comparable than at the highest concentration. In the latter, the performance of the PGM deviates from the performance of the remaining distributions.

The fit of the distributions LN and ziLN is dependent on the assumption of a LOQ threshold. To investigate the consistency of the results obtained with those

distributions, they were fit to the same data, assuming a lower LOQ value (-1 $\log_{10}$ cfu/g) than the one originally used (0 $\log_{10}$ cfu/g). The results (see figure 2) show that the risk estimates obtained with the two ziLN distributions are extremely similar. However, for the LN the decrease of the LOQ value causes a high increase in the deviation between estimated and expected risk.
To understand these results, we analysed the estimates of the distribution parameters μ_{est}, σ_{est} and p_{est} for the fits of ziLN and LN. With the ziLN distribution (figure 3), the performance is best for the highest LOQ. For the lowest LOQ, the overestimation of μ and underestimation of σ increase by the same magnitude at all p levels. The bias trend of the p_{est} in relation to the expected value is the same for both ziLN distributions but the lowest LOQ underestimates the prevalence most – the relative underestimation ((p_{est} - p)/p) was 23% for the highest LOQ and 36% for the lowest, at all p levels. The results for the LN distribution (figure 4) show that the bias trend of μ_{est} and σ_{est} in relation to their expected values is also the same for both fits, however, in this case the deviation between estimates of different fits is higher as the prevalence decreases. At all prevalence levels, and contrarily to what is observed with the ziLN distribution, the assumption of a lower LOQ threshold causes a general underestimation of μ and an overestimation of σ.

4. Discussion

In this study we simulated the generation of a microbial data set, which allowed us to compare the risk estimate of the "true" distribution of concentrations in food products, with the risk estimate of the fitted distribution, for different probability distributions. As "true" distribution we chose a set of zero-inflated Poisson Lognormal distributions. We investigated whether different characteristics of a good fit are crucial for an accurate risk estimate. For that purpose, among the fitted distributions in this study there were 1) a ziPLN, which represented the expected best fit, 2) a LN, to investigate the importance of accounting for the randomness in counts, 3) a ziLN, to investigate the difference between treating "true" zeroes as such or as censored below a LOQ threshold and 4) a PGM, to assess the importance of making the correct assumption about the underlying distribution of concentrations.

The results for the ziPLN showed that, as expected, this distribution provided accurate risk estimates in all situations. However, these results were comparable to those of the ziLN and, for the lowest investigated LOQ value, did not differ markedly from those generated by the LN. The results of the PGM were always the most deviated from the expected. The similar performance of the distributions ziLN and ziPLN suggested that, for obtaining a risk estimate, it may not be necessary to account for the randomness in counts. However, the result for the PGM showed that the choice of the underlying distribution of concentrations may be relevant. The latter result was anticipated, since it has been previously shown that analysing

lognormally-distributed data with a LN and a PGM distribution gives different results (Wiens, 1999). Furthermore, Wiens (1999) showed that large outliers have more impact when using a PGM distribution, whereas the LOQ value assumed for censored analysis of "artificial" zeroes has more impact when using a LN distribution.

It has been previously shown that in the absence of zeroes, fitting a PLN can be closely approximated by the fitting of a LN (Gonzales-Barron & Butler, 2011) . We showed here that even in situations where zeroes are present in a very high proportion (>90%) small differences were observed between the zero-inflated distributions ziPLN and ziLN and the non-zero-inflated LN. In a low concentration scenario, ziPLN and ziLN estimated the risk somewhat more accurately and the LN overestimated it, whereas in a high concentration scenario, the LN distribution appeared to provide slightly better estimates compared to the zero-inflated forms that underestimated the risk. However, when using a lower LOQ threshold to fit the LN distribution to the same data, we observed that the risk estimates became highly biased towards overestimation. This shift was due to an increase in the standard deviation of the fitted LN, which has a high influence on the increase of the arithmetic mean and the degree of clustering, determinant for the increase of the estimated risk (Bassett et al., 2010). Since it is difficult to adopt the most appropriate LOQ value for each particular case, it is preferable to use a distribution which performance is not dependent on an artificial threshold. The fit of the zero-inflated form of the LN, although fitted with a similar MLE censored approach, was not affected by the change of LOQ in a way that impacted the accuracy of risk estimation.

Between the two different concentration scenarios ($\mu = 0$ and $\mu = 0.75$) we observed that the difference between risk estimates of different distributions was lower at low mean concentration. This may be caused by the impact of the misinterpretation of "artificial" zeroes as "true" zeroes. In a low concentration scenario, "artificial" zeroes arise from contaminated samples that inevitably have a low concentration and hence represent low or negligible risk. Therefore, the misinterpretation of "artificial" zeroes as "true" zeroes has little or no impact in the estimated mean risk. However, in a high concentration scenario, "artificial" zeroes may randomly arise from concentrations that have a larger impact on the risk. Therefore, if "artificial" zeroes are misinterpreted as "true" zeroes in a fitting procedure, the risk estimated with the fitted distribution will be an underestimation of the expected risk.

Similarly, between different prevalence levels, we observed an almost negligible difference between distributions at the lowest level (p=10%) and increasing deviations from the expected risk, with all distributions, for increasing prevalence. This result may also be explained by the different impact of the misinterpretation of "artificial" zeroes at low and high prevalence levels. At a low level, the proportion of "true" zeroes among the data is inevitably high; therefore, the impact of the

misinterpretations of a relatively small number of “artificial” zeroes will not have a high impact on the mean risk. Contrarily, at high prevalence there are few “true” zeroes among the data, therefore considering “artificial” zeroes as “true” might have a significant impact towards the underestimation of the mean risk. It might be relevant to note that the results of absolute difference between risk estimates and expected risk contrast with the results in terms of relative difference ((estimated P_{ill}-expected P_{ill})/expected P_{ill}). In this case, the relative difference tended to decrease with increasing *p* for the distributions LN and PGM and remained approximately constant for the distributions ziLN and ziPLN.

The analysis of LN and ziLN distributions fitted using two distinct LOQ values showed that an extremely good fit of a distribution to the data might not be fundamental to obtain an accurate risk estimate. Measurement and sampling uncertainty are two factors that also contribute to the distance between the “true” distribution of counts and the fitted one(s). In this study we simulated a procedure that includes sampling uncertainty but is free from measurement errors. Therefore, a part of the difference between the expected risk and the risk estimates was anticipated to be due to sampling uncertainty. This difference was negligible (0.0003% on average) for the sample size and method used in this study. Evidently, a more pronounced impact is expected as the sample decreases in size and if the sampling cannot be described as a Poisson process

The results presented here might be dependent on the QMRA model, i.e. both the consumer phase model and the dose-response model used. However, as it has been previously shown it is possible that the range of doses that are most responsible for causing disease (higher probability of illness) remains similar with a different dose-response relationship, and therefore the expected results would be comparable (Bassett et al., 2010).

5. Conclusions

We concluded that choosing an appropriate parametric form of continuous distribution to describe the variability of microbial concentrations at retail has a higher impact in the correct estimation of risk than considering the randomness inherent to microbial enumeration by using more mechanistic distributions, such as generalizations of the Poisson. This conclusion was based on the observation that using a PGM distribution to describe PLN distributed counts provided poorer risk estimates than using a LN distribution.

Furthermore, we showed that zero-inflated distribution forms should be preferred to non-zero-inflated forms, particularly at high prevalence and/or high concentration scenarios, where the correct separation between “artificial” and “true” zeroes is particularly crucial for an accurate estimation of risk. Also, the most commonly used

alternative to zero-inflated forms, the simple lognormal distribution, proved to be inadequate for risk estimation, as its performance was highly influenced by the choice of the LOQ value for censored MLE.

In a previous study it was concluded that the choice of the distribution might or might not impact significantly on the magnitude of the risk (Bassett et al., 2010). In our study we concluded that that impact is more pronounced at high prevalence and/or high concentration compared to low prevalence and/or low concentration.
Moreover, it has been argued that the typical simplifying assumption of normality of log counts generally adopted in risk assessment needs to be revised (Gonzales-Barron & Butler, 2011). This study showed that even in a situation where "true" microbial concentrations originate from a lognormal distribution, the assumption of simple lognormality might not be enough to provide accurate risk estimates. Instead, the use of a zero-inflated form of the lognormal distribution was more appropriate for the purpose of QMRA because zero-inflated distributions, contrarily to non-inflated forms, allow estimating the prevalence and therefore the existence of "true" zeroes among the data is accounted for. This correct estimation of prevalence with zero-inflated distributions proved to have a positive influence on risk estimation.

Acknowledgements

The authors thank Dr. Sara Monteiro Pires for critically reading the manuscript.

References

Bassett, J., Jackson, T., Jewell, K., Jongenburger, I., & Zwietering, M. (2010). Impact of microbial distributions on food safety. ILSI Europe Report, Brussels.

Bassett, J., Nauta, M., Lindqvist, R. and Zwietering, M. (2012). Tools for Microbiological Risk Assessment. ILSI Europe Report, Brussels.

Boysen, L., Vigre, H., & Rosenquist, H. (2011). Seasonal influence on the prevalence of thermotolerant campylobacter in retail broiler meat in denmark. *Food Microbiology, 28*(5), 1028-1032.

Busschaert, P., Geeraerd, A. H., Uyttendaele, M., & Van Impe, J. F. (2010). Estimating distributions out of qualitative and (semi)quantitative microbiological contamination data for use in risk assessment. *International Journal of Food Microbiology, 138*(3), 260-269.

Busschaert, P., Geeraerd, A. H., Uyttendaele, M., & Van Impe, J. F. (2011). Hierarchical bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. *Food Microbiology, 28*(4), 712-719.

Crepet, A., Albert, I., Dervin, C., & Carlin, F. (2007). Estimation of microbial contamination of food from prevalence and concentration data: Application to listeria monocytogenes in fresh vegetables. *Applied and Environmental Microbiology, 73*(1), 250-258.

Christensen, B., Nauta, M., Korsgaard, H., Sørensen, A.I., Rosenquist, H., Boysen, L., Perge, A., Nørrung, B. (2013). Case-by-case risk assessment of broiler meat batches: An effective control strategy for *Campylobacter*, *Food Control, 31*, 485-490.

Delignette-Muller, M. L., Pouillot, R., Denis, J., & Dutang, C. (2010). *fitdistrplus: help to fit of a parametric distribution to non-censored or censored data.R package version 0.1-3.* Hentet fra http://riskassessment.r-forge.r-project.org.

Duarte, A.S.R., Stockmarr, A., Nauta, M.J. (in prep.). Fitting a distribution to microbial counts: making sense of zeros.

Gonzales-Barron, U., Kerr, M., Sheridan, J. J., & Butler, F. (2010). Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. *International Journal of Food Microbiology, 136*(3), 268-277.

Gonzales-Barron, U., & Butler, F. (2011). A comparison between the discrete poisson-gamma and poisson-lognormal distributions to characterise microbial counts in foods. *Food Control, 22*(8), 1279-1286.

Helsel, D. R. (2005a). Insider censoring: Distortion of data with nondetects. *Human and Ecological Risk Assessment, 11*(6), 1127-1137.

Helsel, D. (2005b). *Nondetects and data analysis: Statistics for censored environmental data.* USA: Wiley Interscience.

Kilsby, D. C., & Pugh, M. E. (1981). The relevance of the distribution of micro-organisms within batches of food to the control of microbiological hazards from foods. *Journal of Applied Bacteriology, 51*(2), 345-354.

Lorimer, M. F., & Kiermeier, A. (2007). Analysing microbiological data: Tobit or not tobit? *International Journal of Food Microbiology*, *116*(3), 313-318.

Marks, H., & Coleman, M. (1998). Estimating distributions of numbers of organisms in food products. *Journal of Food Protection, 61*(11), 1535-1540.

Nauta, M. J., Jacobs-Reitsma, W., & Havelaar, A. H. (2007). A risk assessment model for campylobacter in broiler meat. *Risk Analysis, 27*(4), 845-861.

Nauta, M. J., Fischer, A. R. H., van Asselt, E. D., de Jong, Aarieke E. I., Frewer, L. J., & de Jonge, R. (2008). Food safety in the domestic environment: The effect of consumer risk information on human disease risks. *Risk Analysis, 28*(1), 179-192.

Nauta, M., & Christensen, B. (2011). The impact of consumer phase models in microbial risk analysis. *Risk Analysis, 31*(2), 255-265.

Nauta, M. J., Sanaa, M., & Havelaar, A. H. (2012). Risk based microbiological criteria for campylobacter in broiler meat in the european union. *International Journal of Food Microbiology, 158*(3), 209-217.

Reinders, R. D., De Jonge, R., & Evers, E. G. (2004). A statistical method to determine whether micro-organisms are randomly distributed in a food matrix, applied to coliforms and escherichia coli O157 in minced beef (vol 20, pg 297, 2003). *Food Microbiology, 21*(6), 819-819.

Ridout, M., Demétrio, C., & Hinde, J. (1998). Models for count data with many zeros. *International Biometric Conference.* Cape Town.

Shorten, P. R., Pleasants, A. B., & Soboleva, T. K. (2006). Estimation of microbial growth using population measurements subject to a detection limit. *International Journal of Food Microbiology, 108*(3), 369-375.

Straver, J. M., Janssen, A. F. W., Linnemann, A. R., Boekel, M. A. J. S., Beumer, R. R., & Zwietering, M. H. (2007). Number of salmonella on chicken breast fillet at retail level and its implications for public health risk. *Journal of Food Protection, 70*(9), 2045-2055.

Teunis, P. F. M., & Havelaar, A. H. (2000). The beta poisson dose-response model is not a single-hit model. *Risk Analysis, 20*(4), 513-520.

Wiens, B. L. (1999). Statistical practice - when log-normal and gamma models give different results: A case study. *American Statistician, 53*(2)

Zhao, Y., & Frey, H. C. (2004). Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. *Risk Analysis, 24*(4), 1019-1034.

Figures

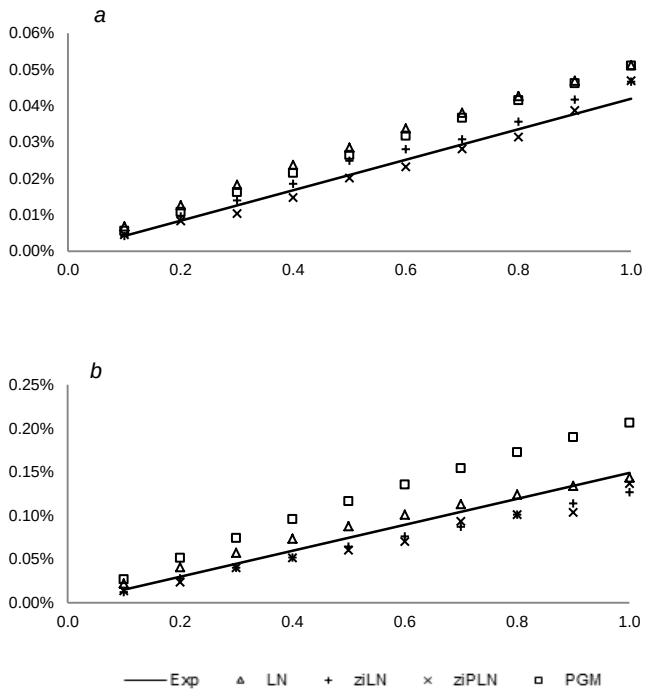

Figure 1: Probability of illness estimated with four distributions of microbial concentrations fitted to low (*a*) and high (*b*) concentration data, at ten different prevalence scenarios. The horizontal axis represents prevalence and the vertical axis represents the mean probability of illness obtained with distinct distributions (*symbols*) in a QMRA model and the expected probability of illness using the same model (*line*).

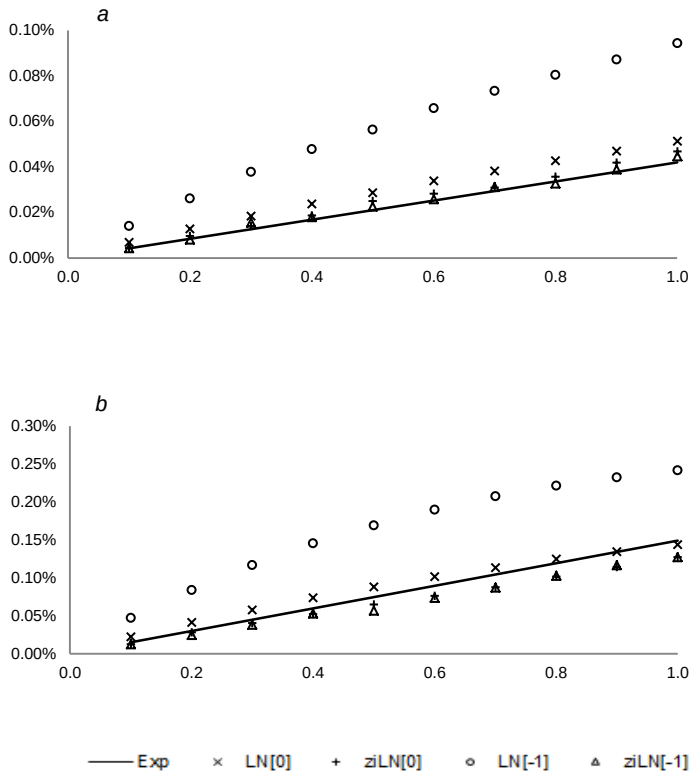

Figure 2: Probability of illness estimated with a LN and a ziLN distributions fitted to low (*a*) and high (*b*) concentration data, at ten different prevalence scenarios. The horizontal axis represents prevalence and the vertical axis represents the mean probability of illness obtained with distinct distributions (*symbols*) in a QMRA model and the expected probability of illness using the same model (*line*). Shape symbols represent distributions fitted assuming a LOQ of -1 $\log_{10}$ cfu/g and cross symbols represent distributions fitted assuming a LOQ of 0 $\log_{10}$ cfu/g.

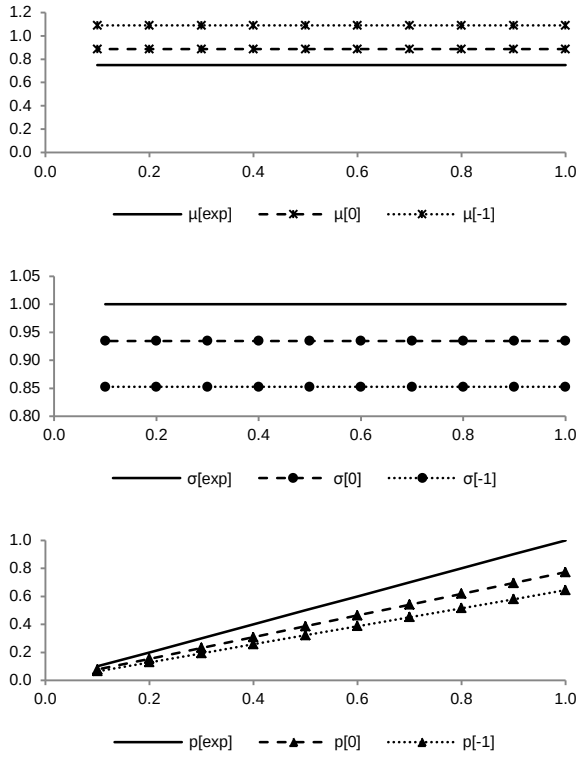

Figure 3: Estimated values of mean (cross), standard deviation (circle) and prevalence (triangle) of a ziLN distribution fitted at a high concentration scenario (μ=0.75, σ=1), assuming two different LOQ values (0 and -1 log10 cfu/g). The continuous lines represent the expected values of the parameters and dashed lines represent estimates obtained with two different fits. The horizontal axis shows the prevalence scenario and the vertical axis shows the parameter values (cfu/g for μ and σ; proportion for prevalence).

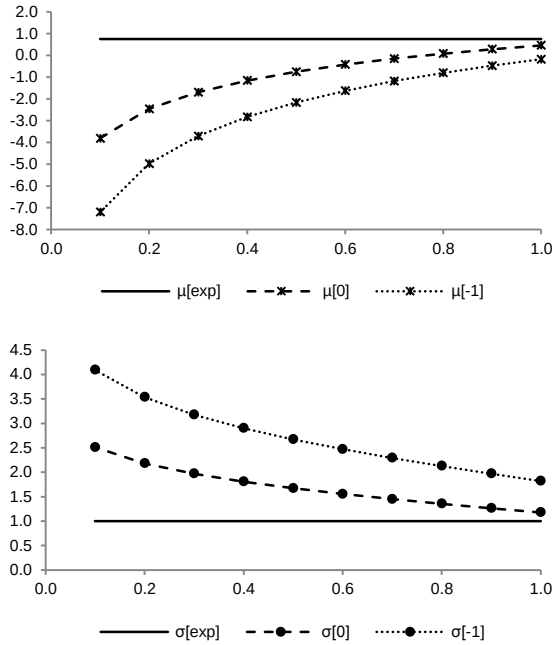

Figure 4: Estimated values of mean (cross) and standard deviation (circle) of a LN distribution fitted at a high concentration scenario (μ=0.75, σ=1), assuming two different LOQ values (0 and -1 log10 cfu/g). The continuous lines represent the expected values of the parameters and dashed lines represent estimates obtained with two different fits. The horizontal axis shows the prevalence scenario and the vertical axis shows the parameter values (cfu/g).

Discussion

In this study, the impact on risk estimates of the frequency distribution used to describe microbial concentrations in food was assessed. It is relevant to investigate this impact because the distributions that are fit to microbial data and used as input in QMRA often deviate from the true distribution of concentrations. Numerous factors may have an impact on the collection of microbial data and on the frequency distribution that is fitted to observations, therefore influencing the distance between distributions. During the analysis of microbial data, one should consider 1) that the enumeration procedure induces homogeneity in the distribution of CFU within a sample unit, 2) the assumptions adopted for the back-calculation of concentrations from CFU counts, 3) the assumption about the parametric form corresponding to the true distribution and 4) the level of contamination (including prevalence and concentration). Furthermore, 5) measurement uncertainty and 6) sampling uncertainty may also have an influence on the inference of distributions from microbial data.

All these factors were taken under consideration in the present study, as follows:

1) The comparison between a zero-inflated lognormal (ziLN) and a zero-inflated Poisson-lognormal (ziPLN) distribution demonstrated the importance of accounting for homogeneity due to enumeration;
2) There were no assumptions made in the back-calculation of concentrations from CFU counts, since in a computer simulation there are no limitations of the human-eye; therefore plates with too-low or too-high counts do not need to be excluded from the calculations;
3) The comparison between a Poisson-gamma (PGM) and a LN distribution served to assess the importance of making the correct assumption about the underlying distribution of concentrations (originally lognormal in this case);
4) Different levels of concentration and prevalence were simulated;
5) There was no measurement uncertainty, as the simulated enumeration experiment was precise and unbiased;
6) The sampling uncertainty was negligible for a sample size of 500 units.

It can be argued that simulated conditions should better approximate real-life conditions. In order to achieve that goal, the same assumptions that are made during enumeration should be adopted in the simulated experiment, for example, by excluding plates with CFU numbers that cannot be precisely counted by a human operator. Also, measurement errors should be integrated in the simulated experiment at all levels where it is expected to occur in a real experiment. Finally, smaller samples should be taken for the analysis, so that sampling error approximates what is encountered in reality. The variation that is caused by these factors was not included in the experiment so the observed changes in risk would be allocated exclusively to the change in the distribution of concentrations. However, it would be interesting to explore the effect of additional factors inherent to microbial enumeration on risk estimation. In that case, uncertainty should be quantified, for example using bootstrapping, and expressed as a confidence interval for the estimated mean risk.

This study had two other major assumptions. First, the data was generated from a zero-inflated PLN distribution, which has likely an impact on the result observed in terms of risk estimation with other distributions, particularly the PGM. Fitting a Poisson-gamma distribution is data-demanding, especially at low contamination levels where the proportion of zero counts is high, and it is more affected by outliers than fitting a LN distribution (Wiens, 1999). A similar study should be performed where the data is generated, for example, with a PGM distribution, in order to investigate how reproducible the conclusions here presented are. Second, the simulated experiment does not consider a detection step prior to enumeration. Therefore, contaminated and uncontaminated units are used together when fitting the frequency distribution to microbial data. In reality, many times only the supposed positive units are consequently enumerated. However, it is not certain that all “negative” units are in fact uncontaminated, which impacts the estimation of true prevalence. Hence, it is preferable to analyse all units together, without separating contaminated from uncontaminated

units. Provided that both types of units have been enumerated, there are available statistical distributions, the zero-inflated distributions, which allow for the estimation of prevalence and the estimation of distribution of concentrations with a single model.

Distribution of microbial concentrations is only one of the factors in QMRA that can contribute to the imprecision of the final risk. The consumer-phase model (CPM) and the dose-response (DR) model are two other factors often considered. It has been shown that the use of different CPMs in the same QMRA study with the same DR model produces small differences in terms of calculated relative risk (Nauta and Christensen, 2011). Here the focus was merely on the effect of the concentration distribution on the risk. Therefore, other sources of variation in the model were reduced to a minimum. For that reason, and because it has been proven sufficient to assess the relative risk (Nauta and Christensen, 2011), the simple data-based "Nauta" CPM was used to compute the risk estimates. Note that in this study the outcome was not in terms of relative risk but absolute risk. Hence the results obtained with different CPMs in terms of risk estimates may differ. Since the difference in the tail of input distributions is an important cause for differences between results obtained with distinct CPMs, and in this study input distributions with different tails were used, it is likely that the results here presented may not be reproducible when using another CPM.

Here a combined CPM-DR model was selected to estimate the risk because it provides a direct relationship between the concentration at retail and the probability of illness (figure 11). Therefore, it allows a straightforward comparison of the impact on risk of different distributions of concentrations at retail. The impact of the dose-response relationship on the results was assumed negligible; however it is possible that different DR models would produce different results.

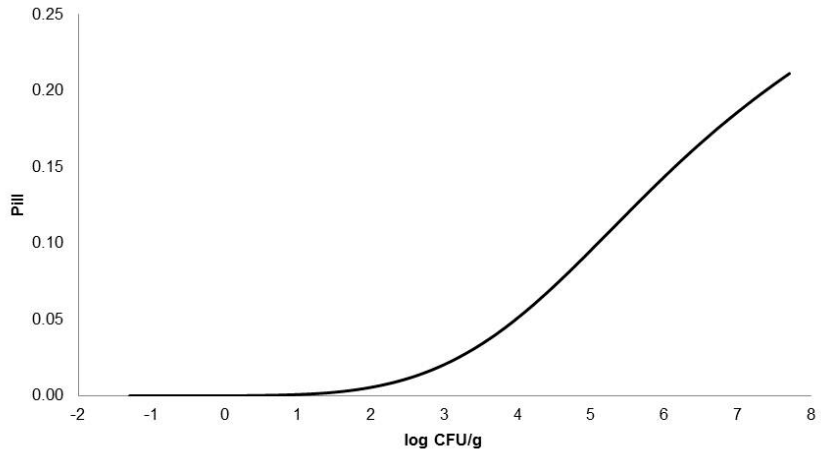

Figure 11: Outcome of the combined CPM-DR model for *Campylobacter* – probability of illness (P_{ill}) against concentration at retail (log CFU/g).

Simulated data is generated for the same reason as chemists prepare standard solutions – so that the original conditions are known (Helsel, 2006). This way, when statistical methods (such as fitting a distribution) are applied to the simulated data, the similarity to the known values can be assessed. However, among the scientific community, the use of real data is many times preferred; under the arguments that they depict real-life conditions and that the generation of simulated data is subjective. The fact that results may

lack a sound confirmation is then ignored. Real-life conditions contain numerous sources of variation, several of which unknown, that make it difficult to draw clear conclusions. When simulated data is used, the effect of such variations is eliminated and therefore results can be attributed exclusively to the factors under study. This approach is particularly useful if different statistical methods are compared, as the differences observed can be attributed to the method used without any influence of unknown, and therefore not accounted for, sources of variation. This study used simulated microbial data, i.e. the original (or "true") distribution of concentrations at retail was known, which allowed estimating the expected risk in the QMRA model for comparison with the mean risk estimated with fitted distributions. Another advantage of using *in silico* data is the fact that numerous scenarios can be tested, for example in terms of prevalence; real data is usually restricted to more homogeneous conditions. A good example is the "high" and "low" concentration scenarios adopted in this study that were based on Danish retail reality, where both scenarios are not significantly different from each other. Hence, the goal of investigating whether the impact of using different distributions on the risk depended on the concentration level at retail might have not been completely achieved. It would probably be more conclusive to repeat this study in a reality where there is a larger difference between high and low concentrations. Another pitfall of the study is that prevalence levels from 10% to 100% were combined with the two concentration distributions representing high and low levels to generate samples. In reality, there is usually a correlation between concentration and prevalence, which was not accounted for in this protocol. For a closer approximation of the simulated data to reality, it could be considered to repeat this study under circumstances where prevalence and concentration are correlated. Also, sampling uncertainty was reduced to a negligible level in this study by using a large sample size (500 units). However, it should be investigated if the results in terms of the risk obtained with different fitted distributions change once sampling error plays a role.

PART III

Manuscript III

Variability and uncertainty in the evaluation of predictive models with literature data – consequences to quantitative microbiological risk assessment

Duarte, A.S.R.; Ross, T.; Mellefont, L.; Nauta, M.J.

To be submitted

Title

Variability and uncertainty in the evaluation of predictive models with literature data – consequences to quantitative microbiological risk assessment

Authors

Duarte, A.S.R. [a]; Ross, T. [b]; Mellefont, L. [b]; Nauta, M.J [a].

[a] Technical University of Denmark – National Food Institute

Mørkhøj Bygade, 19, Building H, DK-2860 Søborg, Denmark

[b] University of Tasmania/School of Agricultural Science – Tasmanian Institute of Agriculture

Hobart, TAS 7001, Australia

asrd@food.dtu.dk

Tom.Ross@utas.edu.au

Lyndal.Mellefont@utas.edu.au

maana@food.dtu.dk

Corresponding author

Sofia Duarte

asrd@food.dtu.dk

+45 23673256

Abstract

Predictive microbiology is an important tool for quantitative microbiological risk assessment (QMRA). However, for reliable predictions, models must be validated for the product of concern and realistic conditions of contamination and handling.

Many studies have shown that a change in experimental settings, such as the bacterial strain or the growth media, leads to different estimates of growth parameters or to different 'performance' of the same model. Consequently, stochastic integration of variability and uncertainty of microbial growth is indispensable for the application of predictive models in QMRA. However, if model evaluation is targeted at the use of a model in a specific QMRA, and part of the impact of uncertainty and variability in growth is controlled for by selecting appropriate evaluation datasets, the number of sources of variability and uncertainty that need to be modeled in QMRA may be reduced. When a model is validated against literature data, it is crucial to determine the effect of data-related factors on model performance, to characterize variability and uncertainty.

Here we assess how different growth settings inherent to literature datasets affect the performance of a growth model compared to its performance with the data used to generate it. For each evaluation dataset, we registered the number of observations, the ranges of temperature, water activity and pH under which they were made, the presence or absence of lactic acid, the use of a pathogenic or non-pathogenic strain and the type of growth environment.

We use the accuracy factor (A_f) as a measure of the performance of a published square root-type model for the growth rate of *Escherichia coli* in response to four environmental factors and literature data that have been previously used to evaluate this model. The A_f of the model with the data used to generate it was assumed as its best possible performance. We compare graphically and statistically the distribution of A_f values obtained with different datasets, by testing the hypothesis that mean and variance is lower in groups for which conditions are closer to those of the original data used to generate the model. We assess the difference in variances using a one-sided F-test, and in the cases where no statistical significance is found, the difference in means with ANOVA.

Results show that selecting large datasets and datasets produced with a growth environment comparable to that used to generate the original data has more impact on the reduction of performance variation than selecting datasets with a similar bacterial strain or the complete overlap between the growth conditions and the interpolation region of the model.

Keywords: *Square-root model; Accuracy factor; Model validation; Uncertainty; Variability; Dataset*

1. Introduction

Predictive models of microbial growth are important tools for quantitative microbial risk assessment (QMRA)(Nauta, 2000; Ross and McMeekin, 2003). They allow the inference of doses to which consumers are exposed from the initial microbial contamination and the exposure pathway. However, these models are a simplification of reality, having therefore a limited predictive ability (Ross et al., 2000) that is affected by sources of uncertainty and variability. Prediction errors may be due to the model's parameterization, the type and range of conditions that the model is based on, the number of data points used to generate it and ultimately due to an inappropriate characterization of the uncertainty and variability in the microbial response (Ross et al., 1999).

Here we focus on the impact of data-related factors on the prediction ability of a model. Those factors considered first, are the growth settings covered by a model which represent the conditions under which the data used to generate it were collected. This includes the type and range of extrinsic and intrinsic growth conditions considered, and the growth media and bacterial strain(s) that were used in the experiment; and second, the number of observations made during growth at a constant set of conditions which also determines the precision of the estimates of growth parameters (Bratchell et al., 1989; Ratkowsky et al., 1991).

The combination of the range of all conditions controlled in the experiment that generates a model constitutes the *interpolation region* of that particular model. Ideally, to maximize predictive accuracy, the interpolation region should 'equal' the entire microbial growth domain (Baranyi et al., 1996). However that demands laborious, expensive and time-consuming experiments. Instead, data are usually collected over a limited range of each growth condition. This approach defines the interpolation region of the model and can affect its reliability and robustness (Bratchell et al., 1989). For the same reason, the selection of growth settings for the experiment should also ideally approximate the real circumstances for which the model will be applied in the future. These settings include the selected strain(s) for which growth is measured, the selected growth media or food product, and the type of extrinsic and intrinsic conditions of food that are monitored. Several authors have studied the effect of using different bacterial strains on the accuracy of predictive models (Lebert et al., 1998; Lianou and Koutsoumanis, 2011). Nauta and Dufrenne (1999) observed that between-strain variability is expressed at different levels for different growth parameters and Lebert et al. (1998) observed the same for different ranges of the same growth condition. Similarly, the development of models using growth media may limit their applicability to food due to the influence of competitive microbiota and food matrix (Tamplin, 2002) or specific compounds or nutrients in the food that are not represented in simple growth media (Dalgaard and Jørgensen, 1998).

Ideally, some of the limitations of predictive microbiology would be obviated with the development of predictive models applicable to all foods, for all bacteria and over the entire microbial growth domain. This demanding task is the ultimate goal of predictive microbiology. However, to date, the common approach of predictive microbiology has been the development of multi-factor product- and species-oriented models (Dalgaard et al., 1997, 2002; Mejlholm and Dalgaard, 2009; Wemmenhove et al., 2011). These models can be used in exposure assessment provided that they have been evaluated for the product of concern and realistic conditions of contamination and handling (Dalgaard and Jørgensen, 1998; Mejlholm et al., 2010; Ross et al., 2000).

This evaluation of a predictive model has been defined as the "comparison of predicted responses to observations in product, independent of those used to generate the model" (Ross, 1996). Typically, a growth model is developed using a set of growth observations and evaluated against sets of independent literature data or data obtained from challenge studies for the same microorganism, under relatively close growth conditions (Walls and Scott, 1997).

Considering the need for objective performance measures for model evaluation, Ross (1996) introduced the concepts of accuracy factor (A_f) and bias factor (B_f), later adapted by Baranyi et al. (1999) to better describe systematic deviations between predicted and observed values. A_f is a measure of how close, on average, predictions are to observations and B_f is a measure of the extent of under- or over-prediction by the model of the observed microbial response. An ideal model would have $A_f = B_f = 1$. However, biological systems exhibit unexplained variability and therefore A_f usually deviates from 1, even when calculated with the observations that originated the model. Considering the performance of a model with independent literature data, Baranyi et al. (1996) stated that for three variable models, "predictive accuracy in the normal physiological range is likely to be limited to no better than approximately 25% of the observed response". Nevertheless, evaluation of models using published data has been a common practice in predictive microbiology (Carlin et al., 2013; Dalgaard and Vigel Jørgensen, 1998; Dominguez and Schaffner, 2007; Mellefont et al., 2003a; Sutherland et al., 1994), despite awareness that data derived under well-defined conditions may be more appropriate to assess model performance (Ross, 1996). The accuracy of a model with literature data may be affected by a number of differences between these data and those used to develop the model: the use of a different strain, a different experimental approach (such as including, or not, competitive microbiota or other growth-inhibiting factors), literature data collected under inadequately controlled or defined conditions, use of media of various compositions, measurements performed under different ranges of growth conditions and different numbers of observations (Lebert et al., 1998; Ross, 1996). All these factors represent sources of variability and/or uncertainty in the evaluation process. Nonetheless, it is important to note that the ultimate arbiter of model performance is

the match between the goodness-of-fit to the data used to generate it and the goodness-of-fit to reliable independent results (Baranyi et al., 1996). Hence, should literature data continue to be used for model evaluation, it is important to investigate the effect of different data-related factors on model performance. This knowledge can benefit the establishment of criteria for the choice of appropriate evaluation datasets, which will contribute to a more transparent and reliable evaluation process.

This study aims to assess how factors inherent to independent literature datasets affect the performance of a growth model compared to its performance with the data used to generate it. For this analysis we use A_f as the measure of performance, and a published square root-type model for the growth rate of *Escherichia coli* as a function of four growth conditions (Ross et al., 2003) that has been previously evaluated with independent literature data (Mellefont et al., 2003a). We compare A_f of the model with data generated under different growth settings to identify conditions inherent to data collection that may affect model performance. We hypothesize that:

i) variation in model performance is lower among datasets collected under conditions closer to those of the original data

ii) the higher the overlap between the interpolation region of the model and the range of growth conditions tested in an evaluation dataset, the better the model performance with that dataset, and

iii) in principle, the larger the dataset, the better is the expected performance of the model.

2. Materials and Methods

2.1.Growth model

In this study, we used the square-root model introduced by Ross et al. (2003) describing the specific growth rate (μ_{max}) of *Escherichia coli* as a function of temperature (T) expressed in Celsius or Kelvin, water activity (a_w), pH and lactic acid concentration ([LAC]) expressed in mM.

$$\sqrt{\mu_{max}} = 0.2345 \times (T - 4.14) \times \left(1 - exp\left(0.2636 \times (T - 49.55)\right)\right) \times \sqrt{(a_w - 0.9508)} \times \sqrt{(1 - 10^{(3.909 - pH)})} \times \sqrt{(1 - 10^{(pH - 8.860)})} \times \sqrt{\left(1 - [LAC]/\left(10.433(1 + 10^{(pH - 3.86)})\right)\right)} \times \sqrt{\left(1 - [LAC]/\left(995.509(1 + 10^{(3.86 - pH)})\right)\right)} \quad (1)$$

This model was generated from 236 generation time ($GT=1/\mu_{max}$) observations, resulting from the pooling of three different datasets (Mellefont, 2000; Presser, 2001; Salter, 1998). For the pooled data, the following interpolation region applies to the model: 7.63–47.43 ℃ for *T*, 0.951–0.999 for a_w, 4.02-8.28 for pH and 0-500mM for [LAC]. The definition of interpolation region adopted here follows the assumptions explained by Baranyi et al. (1996) that a) measured points are by definition in the interpolation region and b) if two points are in the interpolation region, then the interval enclosed by these two points is also comprehended. The model has an A_f of 1.21 when assessed against the data used to generate it, which represents a 21% average deviation between predicted and observed GT.

2.2.Evaluation data

Mellefont et al. (2003a) previously evaluated this model by using independent literature data. Here we used the same published datasets used in that study, the three individual datasets that were used to generate the model, additional literature data (Mellefont et al., 2003b), unpublished data (Mellefont, *personal communication*) and data collected from ComBase (2012) (see appendix). This resulted in 63 datasets including 1340 individual observations. For each evaluation dataset, we recorded the number of observations (*n*) and the ranges of *T*, a_w and pH under which observations were made. Additionally, we also recorded the following factors as growth settings: presence or absence of lactic acid, pathogenic or non-pathogenic *E. coli* strain and the growth environment (liquid media, meat or other foods) (see appendix).

The lower and upper limits of the ranges of *T*, a_w and pH of each dataset were calculated by assuming measurement errors around the measured conditions of ± 0.5 ℃ for *T*, ± 0.003 for a_w and ± 0.2 for pH. This assumption intends to minimize the impact of any possible inadequate control of growth conditions that may be present in independent literature data, and the systematic error that varies between different operators.

2.3.Performance evaluation

The performance of the model was assessed by comparing the model predictions of GT to the observed GT in the evaluation data, through the calculation of the A_f of the model for each individual dataset. The accuracy factor was calculated as:

$$Af = 10^{(\Sigma|log(GT_{predicted}/GT_{observed})|/n)} \quad (2)$$

The difference between the A_f of each evaluation dataset ($A_{f\ evaluation}$) and the original A_f of the model (1.21) was calculated:

$$Dif_{Af} = Af_{evaluation} - 1.21 \quad (3)$$

In theory, the original dataset used to build a predictive model represents the data that it is best able to describe. We assumed therefore that A_f =1.21 corresponds to the measure of best performance of this particular model. Consequently, we expect that $A_{f\ evaluation}$>1.21, and hence that $Dif_{Af} > 0$ with independent literature data. Therefore, we excluded from the analyses datasets with a negative value of Dif_{Af} , as well as an outlier for which Dif_{Af} =3.9.

2.4.Growth settings affecting predictive accuracy

We compared the distribution of Dif_{Af} of datasets with presence and absence of lactic acid, pathogenic and non-pathogenic strains and different growth environments (liquid media, meat and other foods). First we calculated the mean (μ) and variance (σ^2) of Dif_{Af} for different groups and observed graphically its distribution. Based on this analysis, we defined the hypotheses to be tested. Next, we assessed the difference in variance between groups using a one-sided F-test. In the cases where no statistical significance was found, an ANOVA was performed to assess the difference in means. For these analyses, the independent variable Dif_{Af} was square-root transformed to approach normality.

Furthermore, we assessed how the interpolation region of a model can influence its performance when used with independent data. For this purpose, we defined the overlap between interpolation region of the model and the range of growth conditions of the evaluation data as follows: we first identified the range of a growth condition (e.g. T) in an evaluation dataset, then we identified the number of observations in the data used to build the model that were within that range ($n_{within-range}$), and finally we calculated the proportion of the 236 observations in the data used to build the model that $n_{within-range}$ represented (e.g. $T_{represented}$):

$$T_{represented} = n_{within-range}/236 \quad (4)$$

This procedure was performed for each evaluation dataset, and for the separate and combined ranges of three growth conditions (T, a_w and pH). The created variables were plotted against Dif_{Af} to observe the relationship between the representativeness of growth conditions in the original data and the performance of the model with independent data.

Finally, Dif_{Af} was plotted against the number of observations. Due to data gaps in terms of dataset size, we created artificial sets of observations of various sizes, by randomly sampling and pooling individual observations from the total 1340 individual observations from all 63 evaluation datasets. This procedure served exclusively to analyze the impact of the dataset size on model performance.

3. Results

Figure 1 shows the boxplots for the distribution of Dif_{Af} values of groups of observations from literature data generated under different growth settings. Table 1 shows the size and calculated median, mean and variance for each group.

In terms of spread, the results of figure 1 and table 1 (σ^2) show lowest values for observations with a non-pathogenic strain, presence of lactic acid and use of meat as growth environment. With the exception of growth environment categories, these observations also apply to the results of medians and means (table 1). In the growth environment categories, though, observations from meat have a higher median and μ compared with observations from liquid media.

Based on these results, we specified the hypotheses for the one-sided F-test for comparison of σ^2 (see table 2). A significance level of 5% was used. We observe a statistically significant difference ($p < 0.05$) between presence and absence of lactic acid and between meat and liquid media growth environments. The difference between meat and other foods is borderline significant (p=0.0571). This shows that there is less variance in Dif_{Af} for datasets with presence of lactic acid (opposed to absence) and with meat as growth environment (opposed to liquid media and other foods). In contrast, there is no indication of a significant difference in the σ^2 between the use of pathogenic and non-pathogenic strains, and between the use of liquid media and other foods as growth environment. In these two cases, a two-way ANOVA (table 2) was performed to compare means. At a 5% significance level, there is no significant difference between those groups.

In figure 2, the isolated (plots *a*, *b* and *c*) and combined (plot *d*) representativeness of growth conditions *T*, a_w and pH are plotted against Dif_{Af}. There is no clear indication of a trend in terms of μ or σ^2 for any of the isolated growth conditions or for the combination of the three, meaning that model performance does not increase with a higher representativeness level.

The number of observations *n* per (artificial) dataset was also plotted against Dif_{Af} (Figure 3). There is an obvious trend of decrease in spread as *n* increases in number. Mean and standard deviation were calculated for Dif_{Af} of datasets of size varying by 10 observations. There is an abrupt decrease in both mean and standard deviation as the size of the datasets increases from n=1 to $2<n<10$. Later, the changes become less dramatic. It can also be seen that the mean of Dif_{Af} becomes constant (0.08 to 0.09) for n >40, even if calculated with varying numbers of datasets. The standard deviation however, shows a decrease by 0.01 for every increase of approximately 100 observations. Accordingly, the correlation coefficient between standard deviation and *n* for datasets with n>1 is -72%. Furthermore, we observe that negative values of Dif_{Af} tend to occur exclusively with smallest datasets.

4. Discussion and Conclusions

In theory, a predictive model is expected to perform best with the data that was used to create it (original data). Therefore, it is pertinent to investigate which factors inherent to other sources of data may significantly affect model performance. This knowledge is of special importance when literature data is used to evaluate a model, as it increases the transparency of the model evaluation process and may prevent misleading conclusions about predictive ability. Different factors may represent either sources of variability or uncertainty. While the impact of the first can only be reduced by selecting data generated under conditions more appropriate to the model in question, the impact of the latter can be reduced by increasing the number of observations or by increasing the control of experimental settings.

Here we have shown how differences in several experimental conditions influence the accuracy factor parameter, commonly used for assessment of model performance. Using a published square-root growth model for *E. coli* (Ross et al., 2003) and different sources of literature data (see appendix), we assessed the extent to which the growth conditions of the evaluation data were represented in the experiments that generated the original data affect the apparent performance of the model. We investigated the effect of the ranges of T, a_W and pH used, the presence or absence of lactic acid, the growth environment and the chosen strain of *E. coli*.

We expected the model performance, in terms of absolute value of Dif_{Af} and its variation, to be lower in groups in which conditions approximated those of the original data, i.e., liquid media including lactic acid or a meat environment, and a non-pathogenic strain. Additionally, a higher representativeness in the original data of the values of T, a_W and pH tested in an evaluation dataset, was expected to produce better apparent model performance. Finally, the larger the dataset the better we expected the performance of the model to be.

The results show that the inclusion of lactic acid data concentration in an evaluation dataset approximates the performance of this specific model to its theoretical best performance (lower Dif_{Af} mean). Additionally, the model accuracy varies less between datasets (lower Dif_{Af} variance) when lactic acid is present. Lactic acid has been identified as a growth inhibiting factor for *Escherichia coli* (Brashears and Durre, 1999; Bredholt et al., 1999) and was a controlled variable in the original data that generated the model. These results indicate therefore the importance of the correspondence between the growth inhibiting factors used in the original data and the ones used in selected evaluation datasets.

Similarly, Dif_{Af} has less variation with datasets where meat was used as growth environment, as opposed to liquid media and other foods. No significant difference is observed between datasets where liquid media or other foods were used as growth environment. These results might be explained by 1) the presence of lactic acid in a

natural meat environment and 2) the large variability in the composition of liquid media and other foods used in distinct studies. By interpreting together the results relative to the factors presence/absence of lactic acid and growth environment, we conclude that when a model is built with broth-based data but aimed at predicting growth in a specific type of food commodity, evaluation datasets produced both with liquid media and the food product in question can be used for an accurate model assessment. The key to the selection of appropriate datasets lies in the prior knowledge of the type of product that the model is directed at and the representation in those datasets of the growth conditions that have been considered during model generation, either in the form of addition of growth inhibitory substances to liquid media (e.g., lactic acid), or in the form of cultures of bacteria indigenous to the specific food product (e.g., LAB in vacuum packed products) as has been illustrated by Mejlholm and Dalgaard (2009).

Salter et al. (1998) showed that the growth rate of *E. coli* in foods predicted with three different models derived from independent datasets built from liquid media, was always over-predicted. Apart from considering all the growth conditions enclosed in a model, other possible causes for the discrepancy between growth in liquid media and growth in foods can be the difference in variables such as initial microbial density, agitation of the growth environment (Coleman et al., 2003), presence of competing microbiota and food structure (Dourou et al., 2007). Consequently, growth models built from data generated in broth cultures are often conservative (or "fail-safe") and over-predict growth under typical growth conditions of food. In the context of QMRA, this is an important source of inaccuracy of predictive microbiology models that needs to be considered before using a model as part of an exposure assessment.

We showed here that the performance of a predictive model varies between different growth environments. We conclude that, when the goal is to evaluate a model for use in QMRA, it is crucial to select evaluation data exclusively relevant to the food commodity of interest, i.e., growth measured in food or in liquid media mimicking the growth conditions encountered in the food product of interest for QMRA. Model evaluation with data collected using different growth environments may either mask poor performance of a model with the food of interest, or contribute to a poor evaluation of a model that actually performs well with that product, or lead to unrealistically high estimates of variability of model predictions leading to greater uncertainty in the risk estimates. The use of a pathogenic or a non-pathogenic strain of *E. coli* did not seem to have a significant influence in the performance of the model used here, although differences were observed in terms of mean and variance that suggested that the model performed best (lower Dif_{Af} mean and variance) with datasets where, similar to the original data (Mellefont, 2000; Presser, 2001; Salter, 1998), a non-pathogenic strain was used. However, those differences were not statistically significant. This result corroborates findings of previous authors who

analyzed the impact of strain variability on growth. Salter et al. (1998) studied the suitability of a square-root model for the growth of *E. coli* against temperature, built for the non-pathogenic strain M23, to predict growth of pathogenic strains. They concluded that there is a high level of similarity in the growth characteristics of most *E. coli* strains, especially among the fast-growing ones, and that a generic model is applicable to both pathogenic and non-pathogenic strains. Coleman et al., (2003) observed that the behavior of different strains of *E. coli* O157:H7 in shaken and unshaken media differed in the same way, and was very similar between strains in both cases. Furthermore, they observed that strain variability appeared as a minor effect on growth, compared to the significance of media agitation and initial microbial density. However, both studies identified that the variability of growth between strains increased as the temperatures approached the suboptimal temperature region (Coleman et al., 2003; Salter et al., 1998), which indicates that the applicability of a generic model to all strains is possibly restricted to temperatures that are near to optimal for growth. Also, this conclusion cannot be generalized for every microorganism. A recent study has shown that within *Bacillus cereus*, the adaptation to temperature, pH and a_w differs between different phylogenetic groups, and therefore, growth of two strains from different groups cannot be predicted with the same model (Carlin et al., 2013).

We conclude therefore that for the specific growth model used here, which has been built for a non-pathogenic strain of *E. coli*, datasets of pathogenic strains could be used for its evaluation, provided that the observations had not been performed at suboptimal or superoptimal growth conditions. However, one would expect that a food product (e.g. meat) is held (most of the time) at appropriate conservation temperatures, which most often deviate from the optimal temperature for *E. coli* growth (near to 39°C) and may approximate suboptimal temperatures. For example, home refrigeration temperatures may have their 99^{th} percentile at 10°C (Coleman et al., 2003). In practice, this means that if this model is used in QMRA, and handling practices deviate, as expected, from the optimum growth conditions, between-strain variability of growth may still need to be stochastically modeled.

In summary, the evaluation of a model for QMRA purposes needs to account for the realistic conditions of growth encountered along the risk pathway, as the guidelines for the choice of appropriate evaluation data may be more or less strict, depending on those conditions.

We show here also that for the growth model and literature datasets considered, the model performance is not significantly influenced by the degree of representativeness of the original data in terms of the ranges of T, a_W and pH observed in an evaluation dataset. Although all evaluation datasets were within the interpolation region of the model for T, a_W and pH,, some conditions were more extensively represented in the original data than others, due to more observations

situated within the range of that condition represented in the evaluation data. This did not affect the value of Dif_{Af}, neither when considering the range of T, a_W and pH separately, nor the combination of the three ranges. Thus, we suggest that the original data does not necessarily need to have observations within the range of all growth conditions from an evaluation dataset, for the model to have an acceptable performance. This may not be true though with all types of models, for example with some polynomial models strongly adapted to the original data. However, we believe that, despite the overlap between the interpolation region of the model and the growth region of an evaluation dataset may be small, it is still important that the evaluation data is collected within the interpolation region of the model for all growth conditions. It has been previously shown that the extrapolation of a model to a situation where even only one growth condition is outside its interpolation region, may result in underestimation of growth (Cappuyns et al., 2011), which is not acceptable in the context of risk estimation.

We conclude that, when selecting datasets for model evaluation, it is mostly relevant to check the inclusion of the minimum and maximum values of each growth condition in the interpolation region of the model, but it appears irrelevant for model performance to investigate how many observations in the original data represent exactly the conditions present in the evaluation dataset. These findings contribute to simplify the process of selecting literature data for model evaluation and seem to demonstrate the correctness of the assumption that "if two points are in the interpolation region, then the interval enclosed by these two points is also comprehended" (Baranyi et al., 1996).

We show that the number of observations is negatively correlated with the spread of Dif_{Af} – the larger the dataset, the smaller the standard deviation. This indicates that sampling uncertainty needs to be accounted for if a model is evaluated with literature data, and hence large datasets should be preferred to obtain results with less variation in terms of Dif_{Af}. However, in terms of the mean value of Dif_{Af} across several datasets, we show that an average size of 40 to 50 observations per dataset provides the same mean estimate as using larger datasets.

Furthermore, we show that datasets of single observations must be avoided in model evaluation, as there is a large variation and an overestimation of the mean associated with Dif_{Af} calculation. Our study also shows that with randomly assembled datasets, negative values of Dif_{Af} occur predominantly with datasets of very few observations (n< 12 in 75% of the cases). A negative Dif_{Af} implies that the performance of the model with evaluation data exceeded its performance with the original data. Considering the high variation of Dif_{Af} values with small dataset size, one may hypothesize that negative results are due to a random presence of observations with very accurate predictions and probably not due to a true outperformance of the

model compared to the original data. This reinforces the previous observation that small datasets should not be used for model evaluation.

In summary, for reliable evaluation of predictive microbiology models, validation should be based on either a small number of large datasets (>200 observations) or a large number of datasets with an average of 40 to 50 observations. Especially, datasets of single observations must be avoided in model evaluation. When these recommendations are not followed, a model cannot be truly considered as "validated", as the outcome of its assessment may be highly influenced by sampling uncertainty. This implies that a model evaluated in such a way is not suitable for use in QMRA, without stochastically integrating the uncertainty of the model's predictions.

To conclude, we show that if literature data is used to evaluate a predictive model, factors associated with variability and uncertainty in the experimental conditions under which those data were generated need to be accounted for. Specifically, we identify the growth environment as an important source of variability and the number of observations in the dataset as an important source of uncertainty with influence on model performance. According to our results, we conclude that selecting large datasets and datasets produced in a growth environment comparable to that used to generate the original data are important conditions for a transparent and reliable model evaluation process. Furthermore, if the impact of these factors is controlled during model evaluation, the related uncertainty and variability of growth need not be accounted for in QMRA's where the validated model is used.

Acknowledgements

We thank Prof. McMeekin and Prof. Olley for inspiring scientific discussions and the Norma and Frode S. Jacobsens Fond for granting financial support for a research stay at the Tasmanian Institute of Agriculture.

References

Baranyi, J., Ross, T., McMeekin, T. A., & Roberts, T. A. (1996). Effects of parameterization on the performance of empirical models used in `predictive microbiology'. *Food Microbiology, 13*(1), 83-91.

Baranyi, J., Pin, C., & Ross, T. (1999). Validating and comparing predictive models. *International Journal of Food Microbiology, 48*(3), 159-166.

Brashears, M.M., Durre, W.A. (1999). Antagonistic action of *Lactobacillus lactis* toward *Salmonella* spp. and *Escherichia coli* O157:H7 during growth and refrigerated storage. *Journal of Food Protection, 62(11),* 1336-1340.

Bratchell, N., Gibson, A. M., Truman, M., Kelly, T. M., & Roberts, T. A. (1989). Predicting microbial growth - the consequences of quantity of data. *International Journal of Food Microbiology, 8*(1), 47-58.

Bredholt, S., Nesbakken, T., Holck, A. (1999). Protective cultures inhibit growth of *Listeria monocytogenes* and *Escherichia coli* O157:H7 in cooked, sliced, vaccum- and gas- packaged meat. *International Journal of Food Microbiology, 53,* 43-52.

Cappuyns, A.M., Vermeulen, A., Paelinck, H., Devlieghere, F., Van Impe, J.F. (2011). Accurate assessment of microbial safety in food industry: adapting predictive models for specific food products. Conference Proceedings 7th International Conference on Predictive Modelling of Food Quality and Safety, Dublin, Ireland.

Carlin, F., Albagnac, C., Rida, A., Guinebretière, M.H., Couvert, O., Nguyen-the, C. (2013). Variation of cardinal growth parameters and growth limits according to phylogenetic affiliation in the *Bacillus cereus* Group. Consequences for risk assessment. *Food Microbiology, 33,* 69-76.

Coleman, M.E., Tamplin, M.L., Phillips, J.G., Marmer, B.S. (2003). Influence of agitation, inoculum density, pH, and strain on the growth parameters of Escherichia coli O157:H7 - relevance to risk assessment. *International Journal of Food Microbiology, 83,* 147-160.

ComBase - Combined dataBase for predictive microbiology (2012). [online] Available at: http://www.combase.cc [Accessed: 25 April 2012].

Dalgaard, P., Mejlholm, O., & Huss, H. H. (1997). Application of an iterative approach for development of a microbial model predicting the shelf-life of packed fish. *International Journal of Food Microbiology, 38*(2-3), 169-179.

Dalgaard, P., & Vigel Jørgensen, L. (1998). Predicted and observed growth of listeria monocytogenes in seafood challenge tests and in naturally contaminated cold-smoked salmon. *International Journal of Food Microbiology, 40*(1-2), 105-115.

Dominguez, S.A., & Schaffner, D.W. (2007). Development and validation of a mathematical model to describe the growth of *Pseudomonas* spp. in raw poultry stored under aerobic conditions.

Dourou, D., Stamatiou, A., Koutsoumanis, K., Nychas, G.J.E. (2007). Effect of food structure (type of growth), composition and microbial interaction on the growth kinetics of *L. monocytogenes*. Conference Proceedings 5th International Conference on Predictive Modelling in Foods, Athens, Greece.

Lebert, I., Begot, C., & Lebert, A. (1998). Development of two *Listeria monocytogenes* growth models in a meat broth and their application to beef meat. *Food Microbiology, 15*(5)

Lianou, A., & Koutsoumanis, K. P. (2011). A stochastic approach for integrating strain variability in modeling salmonella enterica growth as a function of pH and water activity. *International Journal of Food Microbiology, 149*(3), 254-261.

Mejlholm, O., Dalgaard, P. (2009). Development and Validation of an Extensive Growth and Growth Boundary Model for *Listeria monocytogenes* in Lightly Preserved and Ready-to-Eat Shrimp. *Journal of Food Protection, 72(10),* 2132-2143.

Mejlholm, O., Gunvig, A., Borggaard, J., Blom-Hanssen, J., Mellefont, L.A., Ross, T., Leroi, F., Else, T., Visser, D., Dalgaard, P. (2010). Predicting growth rates and growth boundary of *Listeria monocytogenes* - an international validation study with focus on processed and ready-to-eat meat and seafood. *International Journal of Food Microbiology, 141*,137-150.

Mellefont, L. (2000). Predictive model development and lag phase characterisation for applications in the meat industry. PhD Thesis, University of Tasmania, Australia.

Mellefont, L. A., McMeekin, T. A., & Ross, T. (2003a). Performance evaluation of a model describing the effects of temperature, water activity, pH and lactic acid concentration on the growth of escherichia coli. *International Journal of Food Microbiology, 82*(1), 45-58.

Mellefont, L. A., McMeekin, T. A., & Ross, T. (2003b). The effect of abrupt osmotic shifts on the lag phase duration of foodborne bacteria. *International Journal of Food Microbiology, 83*(1), 281-293.

Nauta, M.J., Dufrenne, J.B. (1999). Variability in growth characteristics of different *E. coli* O157:H7 isolates, and its implications for predictive microbiology. *Quantitative microbiology, 1*, 137-155.

Nauta, M. J. (2000). Modelling bacterial growth in quantitative microbiological risk assessment: Is it possible? *International Journal of Food Microbiology, 73*(2-3), 297-304.

Presser, K.A. (2001). Physiology and modelling of *Escherichia coli* growth inhibition due to pH, organic acids, temperature and water activity. PhD Thesis, University of Tasmania, Australia.

Ratkowsky, D.A., Ross, T., McMeekin, T.A., Olley, J. (1991). Comparison of Arrhenius-type and Bělehrádek-type models for prediction of bacterial growth in foods. *Journal of Applied Bacteriology, 71*, 452-459.

Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. Journal of Applied Microbiology, 81(5)

Ross, T., Baranyi, J., McMeekin, T.A. (1999). Predictive Microbiology and Food Safety. In *Encyclopaedia of Food Microbiology*, Eds. Robinson, R., Batt, C. A., Patel, P. Academic Press, London.

Ross, T., Ratkowsky, D. A., Mellefont, L. A., & McMeekin, T. A. (2003). Modelling the effects of temperature, water activity, pH and lactic acid concentration on the growth rate of escherichia coli. *International Journal of Food Microbiology, 82*(1), 33-43.

Ross, T., Dalgaard, P., & Tienungoon, S. (2000). Predictive modelling of the growth and survival of listeria in fishery products. *International Journal of Food Microbiology, 62*(3), 231-245.

Ross, T., McMeekin, T.A. (2003). Modeling microbial growth within food safety risk assessments. *Risk Analysis, 23*,179 - 197.

Salter, M.A., Ross, T., McMeekin, T.A. (1998). Applicability of a model for non-pathogenic *Escherichia coli* for predicting the growth of pathogenic *Escherichia coli*. *Journal of Applied Microbiology, 85*, 357-364.

Salter, M.A. (1998). Effects of temperature and water activity on Escherichia coli in relation to beef carcasses. PhD Thesis, University of Tasmania, Australia.

Sutherland, J.P., Bayliss, A.J., & Roberts, T.A. (1994). Predictive modelling of growth of Staphylococcus aureus: the effects of temperature, pH and sodium cloride. *International Journal of Food Microbiology, 21,* 217-236.

Tamplin, M. L. (2002). Growth of escherichiacoli O157:H7 in raw ground beef stored at 10Â°C and the influence of competitive bacterial flora, strain variation, and fat level. *Journal of Food Protection, 65*(10), 1535-1540.

Walls, I., & Scott, V. N. (1997). Use of predictive microbiology in microbial food safety risk assessment. *International Journal of Food Microbiology*, *36*(2-3), 97-102.

Wemmenhove, E., Zwietering, M.H., van Hooijdonk, A.C.M., & Wells-Bennink, M.H.J. (2011). Risk assessments of *Listeria monocytogenes* in Dutch-type semihard cheese: incorporating variability in both product parameters and microbial growth parameters. Conference Proceedings 7th International Conference on Predictive Modelling of Food Quality and Safety, Dublin, Ireland.

Figures

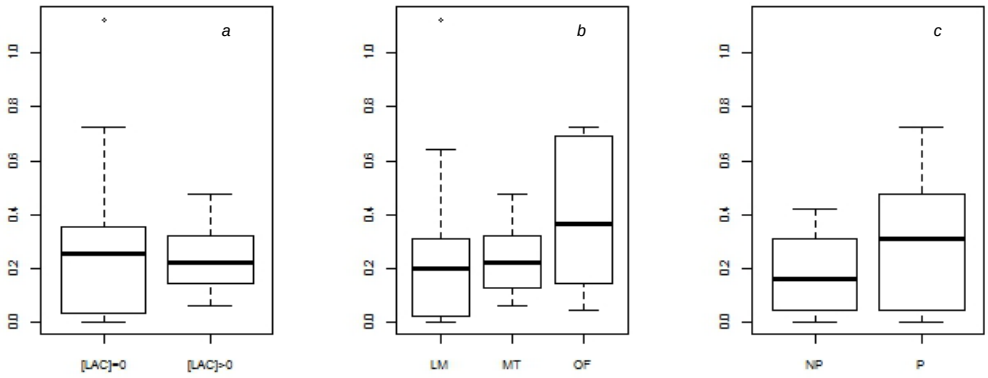

Figure 1: Distribution of *DifAF* (y-axis) for evaluation datasets produced with *a*) absence ([LAC]=0) and presence ([LAC]>0) of lactic acid; *b*) liquid media (LM), meat (MT) and other foods (OF); *c*) pathogenic (P) and non-pathogenic (NP) strains of *Escherichia coli*.

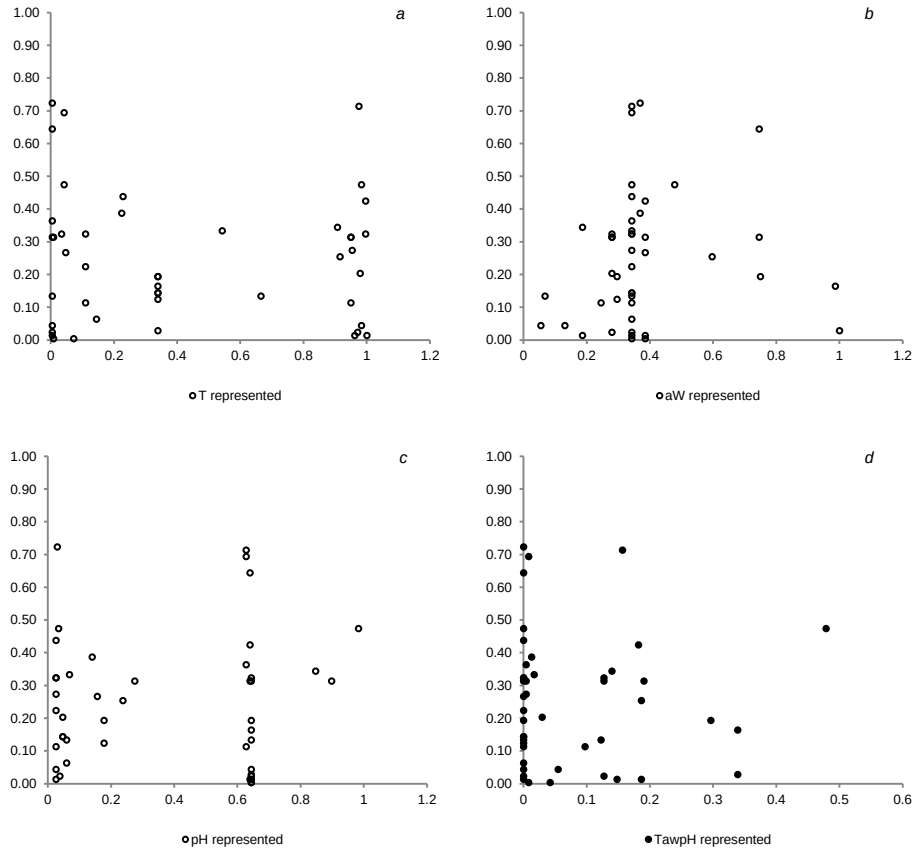

Figure 2: Relationship between representativeness of growth conditions in the original data (x-axis) and model performance (y-axis). Each dot represents one individual evaluation dataset. The x-axis represents the proportion of observations used to build the model that are within the range of T (*a*), a_w (*b*), pH (*c*) and the combination of the three growth conditions (*d*) observed in the evaluation dataset. The y-axis represents model performance measured in *DifAF*.

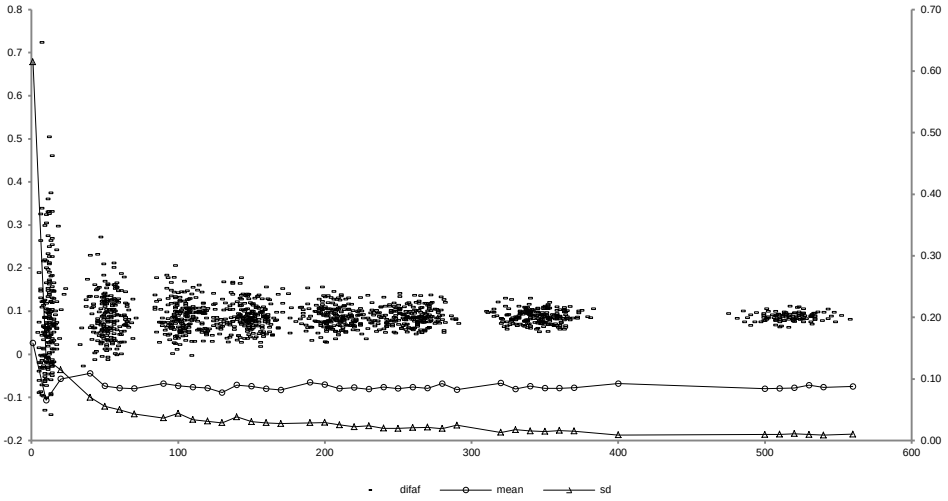

Figure 3: Number of observations in a dataset (x-axis) versus model performance measured in *DifAF* for each individual dataset (left y-axis) and for the mean and standard deviation of datasets of approximate size (right y-axis).

Tables

Growth condition	Group	Datasets	n	μ	σ^2	median
[LAC]	+	15	[1;95]	0.25	0.02	0.22
	-	31	[1;256]	0.28	0.07	0.25
Strain	P	26	[1;256]	0.33	0.08	0.31
	NP	9	[1;218]	0.18	0.02	0.16
Environment	LM	26	[1;256]	0.23	0.06	0.20
	MT	11	[1;67]	0.24	0.02	0.22
	OF	9	[1;230]	0.40	0.07	0.36

Table 1: Number of datasets and dataset size (n), mean (μ), variance (σ^2) and median of *DifAF* calculated for different categories of growth conditions present in evaluation datasets: presence (+) or absence (-) of lactic acid, pathogenicity (P) or non-pathogenicity (NP) of strain, growth environment (liquid media (LM), meat (MT) or other food (OF)). Dataset size represents the interval from the minimum to the maximum number of observations in the datasets included in each group. Datasets with $DifAF<0$ were excluded from the calculations.

F-test				
Growth condition	H_A	*p-value*	σ^2 ratio	95% C.I. σ^2 ratio
[LAC]	$[LAC]^+ < [LAC]^-$	0.002	0.22	[0;0.51]
Strain	NP < P	0.264	0.64	[0;1.99]
Environment	MT < LM	0.0212	0.28	[0;0.78]
	MT < OF	0.0571	0.34	[0;1.05]
	LM < OF	0.5816	1.20	[0;2.80]
ANOVA				
Growth condition	Groups	*p-value*	µ difference	95% C.I. µ difference
Strain	P - NP	0.1326	0.15	[-0.05; 0.35]
Environment	OF - LM	0.0992	0.17	[-0.03; 0.36]

Table 2: Results of the one-sided F-test for comparison of variances and two-way ANOVA for comparison of means. H_A represents the one-sided alternative hypothesis of the F-test and "Groups" represents the different groups compared with ANOVA.

Appendix – Evaluation datasets

Source	For reference details see	*n*	T_{low}-0.5	T_{high}+0.5	a_{w_low}-0.003	a_{w_high}+0.003	pH_{low}-0.2	pH_{high}+0.2	LAC	Strain	Env	Dif_{AF}
Original data												
Mellefont(2000)	Ross et al (2003)	71	24.5	25.9	0.948	1.002	7.2	7.6	-	P	LM	0.029
Presser(2001)	Ross et al (2003)	95	20.38	23.26	0.983	1.001	3.82	8.48	+	NP	LM	-0.019
Salter(1998)	Ross et al (2003)	69	7.13	47.93	0.963	1	7.2	7.6	-	NP	LM	-0.009
Data used in previous model evaluation												
Barber(1908)	Mellefont et al (2003a)	218	9.5	47.3	0.994	1	7.2	7.6	-	NP	LM	-0.036
Barber(1908)	Mellefont et al (2003a)	24	29.6	44.3	0.994	1	7.2	7.6	-	NP	LM	0.004
Barber(1908)	Mellefont et al (2003a)	32	29.5	38	0.994	1	7.2	7.6	-	NP	LM	-0.056
Bernaerts(2000)	Mellefont et al (2003a)	8	14.5	35.5	0.99	0.996	7.2	7.6	-	NP	LM	0.314
Buchanan(1997)	Mellefont et al (2003a)	4	11.5	28.5	0.984	0.99	5.3	7.7	-	P	LM	0.344
Buchanan(1992)	Mellefont et al (2003a)	58	9.5	42.5	0.954	0.99	4.3	8.7	-	P	LM	0.474
Buchanan1993)	Mellefont et al (2003a)	1	11.5	12.5	0.984	0.99	6.3	6.7	-	P	LM	0.014
Buchanan(1993)	Mellefont et al (2003a)	1	41.5	42.5	0.977	0.983	5.7	6.1	-	P	OF	-0.126
Buchanan(1993)	Mellefont et al (2003a)	1	11.5	12.5	0.989	0.995	6.4	6.8	-	P	OF	0.724
Buchanan(1993)	Mellefont et al (2003a)	1	27.5	28.5	0.973	0.979	5.8	6.2	-	P	MT	0.134
Buchanan(1993)	Mellefont et al (2003a)	1	18.5	19.5	0.982	0.988	6.3	6.7	-	P	OF	0.044
Demetz(2000)	Mellefont et al (2003a)	14	13.5	39.5	0.992	0.998	7.1	7.5	-	NP	LM	0.014
Doyle(1984)	Mellefont et al (2003a)	8	24.5	45	0.992	0.998	7.1	7.5	-	P	LM	1.124
Eustace(1998)	Mellefont et al (2003a)	5	7.5	37.5	0.992	0.998	7.1	7.5	-	NP	LM	-0.046
Fratamico(1997)	Mellefont et al (2003a)	6	36.5	37.5	0.994	1	6.8	7.2	-	P	LM	0.024
Gill(1985)	Mellefont et al (2003a)	11	7.2	46.5	0.99	0.996	7.2	7.6	-	NP	LM	0.324
Gill(1985)	Mellefont et al (2003a)	8	9.5	42.8	0.961	0.967	7.2	7.6	-	NP	LM	0.044
Gill(1991)	Mellefont et al (2003a)	6	7.5	30.5	0.994	1	6.3	6.7	+	NP	MT	0.274
Gill(1980)	Mellefont et al (2003a)	2	.	.	0.994	1	5.3	5.7	+	NA	MT	0.114
Gill(1980)	Mellefont et al (2003a)	2	.	.	0.994	1	5.3	5.7	+	NA	MT	0.324
Gill(1980)	Mellefont et al (2003a)	9	.	.	0.994	1	5.3	5.7	+	NA	MT	0.224
Gill(1980)	Mellefont et al (2003a)	4	29.5	30.5	0.994	1	5.3	5.7	+	NA	MT	0.324
Glass(1992)	Mellefont et al (2003a)	5	36.5	37.5	0.951	0.994	7.1	7.5	-	P	LM	0.644
Glass(1992)	Mellefont et al (2003a)	7	36.5	37.5	0.992	0.998	4.3	7.5	-	P	LM	0.314
Grau(1983)	Mellefont et al (2003a)	10	24.5	25.5	0.987	0.993	5.4	7.11	+	NA	MT	0.194
Grau(1983)	Mellefont et al (2003a)	1	24.5	25.5	0.994	1	5.4	5.8	+	NA	OF	0.144
Grau(1983)	Mellefont et al (2003a)	8	24.5	25.5	0.987	0.993	5.3	6.99	+	NA	MT	0.124
Grau(1983)	Mellefont et al (2003a)	1	24.5	25.5	0.994	1	5.4	5.8	+	NA	OF	0.144
Ingraham(1958)	Mellefont et al (2003a)	20	7.5	46.5	0.992	0.998	7.1	7.5	-	NP	LM	0.424
Jason(1983)	Mellefont et al (2003a)	2	36.5	37.5	0.993	0.999	6.3	6.7	-	NP	LM	-0.186
Jennison(1935)	Mellefont et al (2003a)	10	21.5	42.5	0.994	1	7.2	7.6	-	NA	LM	0.134
Kauppi(1996)	Mellefont et al (2003a)	20	8	12.5	0.992	0.998	7.2	7.6	-	P	LM	0.004
Kauppi(1996)	Mellefont et al (2003a)	16	8	12.5	0.99	0.996	7.1	7.5	-	P	LM	0.314
Kauppi(1996)	Mellefont et al (2003a)	14	8	12.5	0.983	0.989	7	7.4	-	P	OF	-0.026
Lowry(1989)	Mellefont et al (2003a)	7	7.69	40.5	0.99	0.996	5.3	5.7	+	NA	LM	0.204
Maxcy(1989)	Mellefont et al (2003a)	5	9.5	30.5	0.995	1.001	7	7.4	-	NA	LM	0.114

Mellefont(2000)	Mellefont et al (2003a)	4	9.5	20.5	0.994	1	5.8	6.2	+	NP	MT	0.064
Mellefont(2000)	Mellefont et al (2003a)	3	9.5	20.5	0.994	1	5.8	6.2	+	NP	MT	-0.056
Mellefont(2000)	Mellefont et al (2003a)	46	9.4	46	0.992	0.998	7.3	7.7	-	P	LM	-0.056
Palumbo(1995)	Mellefont et al (2003a)	3	9.5	37.5	0.99	0.996	7.2	7.6	-	P	LM	0.024
Palumbo(1997)	Mellefont et al (2003a)	13	7.5	37.5	0.994	1	7	7.4	-	P	OF	0.714
Palumbo(1997)	Mellefont et al (2003a)	4	.	.	0.994	1	7	7.4	-	P	OF	0.694
Palumbo(1997)	Mellefont et al (2003a)	3	11.5	12.5	0.994	1	7	7.4	-	P	OF	0.364
Palumbo(1997)	Mellefont et al (2003a)	6	.	.	0.994	1	5.6	6	+	P	MT	0.474
Rajkowsky(1995)	Mellefont et al (2003a)	46	7.5	28.5	0.974	0.994	4.8	7.2	-	P	LM	0.254
Salter(1998)	Mellefont et al (2003a)	256	7.2	47.7	0.994	1	7.2	7.6	-	P	LM	0.014
Smith(1995)	Mellefont et al (2003a)	1	24.5	25.5	0.994	1	7.2	7.6	-	NA	LM	-0.076
Smith(1995)	Mellefont et al (2003a)	1	24.5	25.5	0.994	1	7.2	7.6	-	NA	LM	-0.076
Smith(1985)	Mellefont et al (2003a)	1	7.7	8.7	0.994	1	5.8	6.2	+	NP	MT	-0.056
Smith(p.c.)	Mellefont et al (2003a)	67	9.5	40.5	0.994	1	5.8	6.2	+	NP	MT	-0.046
Sutherland(1995)	Mellefont et al (2003a)	5	9.5	30.5	0.951	0.994	4.29	7.17	-	P	LM	0.314
Walls(1996)	Mellefont et al (2003a)	6	11.5	35.5	0.994	1	5.5	6.6	+	P	MT	-0.096
Wang(1997)	Mellefont et al (2003a)	10	7.5	22.5	0.994	1	6.7	7.3	-	P	OF	0.334
Additional evaluation data												
Mellefont(2003)	Mellefont et al (2003b)	36	24.5	26	0.948	1.002	7.2	7.6	-	NP	LM	0.165
Mellefont(p.c.)	(personal communication)	15	24.2	25.8	0.958	0.996	7.2	7.6	-	P	LM	0.194
Abdul-Raouf(1993)	ComBase (2012)	5	20.5	30.5	0.989	0.995	4.5	5.6	+	P	MT	0.387
Jordan(2001)	ComBase (2012)	9	29.5	30.5	0.961	0.98	4.8	5.2	+	P	LM	3.943
Jordan(2001)	ComBase (2012)	11	29.5	30.5	0.992	0.998	4.8	6.2	+	P	LM	0.267
Kasrazadeh(1995)	ComBase (2012)	3	11.5	30.5	0.994	1	6.3	6.7	+	P	OF	0.438
Miller(1994)	ComBase (2012)	1	9.5	10.5	0.989	0.995	6.8	7.2	+	P	MT	-0.204

n – number of observations; T_{low} – minimum temperature observed; T_{high} – maximum temperature observed; a_{W_low} – minimum water activity observed; a_{w_high} – maximum water activity observed; pH_{low} – minimum pH observed; pH_{high} – maximum pH observed; LAC – presence (+) or absence (-) of lactic acid; strain – pathogenic (P) or non-pathogenic (NP); Env – liquid media (LM), meat (MT) or other foods (OF).

Discussion

The study presented in manuscript III aimed to investigate how factors inherent to experimental data collection can affect the measure that is commonly used to assess the accuracy of predictive microbiology models. It was expected that the model performance would be best with datasets collected under conditions that approximated those of the original data, for example liquid media including lactic acid, a meat environment and a non-pathogenic strain. An example of a square-root growth model for *E. coli* was used, which implies that some of the conclusions of the study cannot be generalized for every predictive model and for all microorganisms. However, this work provided an insight into which factors can affect the performance measure of a predictive model in general, and therefore need to be considered during the selection of published data for model evaluation.

The accuracy factor (A_f) was selected as a measure of model performance because it provides an average measure of the deviation between model predictions and observations. Usually in model evaluation studies, A_f =1 is considered a reference value for perfect model accuracy. However, it is here considered that this represents only a theoretical value. In reality, there is always a share of variation in predictions associated to model uncertainty that cannot be reduced. Therefore, a model will rarely have A_f =1 against the data on which it is based. It is hypothesized that the A_f value obtained against those data describes the best possible performance of a model. Therefore, it is the difference between this value and the value obtained with independent datasets (Dif_{Af}) that should matter for model evaluation, and not the deviation from a perfect agreement. Furthermore, it was shown that A_f values representing a better performance than that obtained against the original data occurred with small datasets, suggesting that a better A_f value was obtained due to chance and not due to a true better performance of the model. Datasets with a negative Dif_{Af} should therefore be avoided in model evaluation, especially if their number of observations is small.

This is only one of the reasons why small datasets cannot be used for model evaluation. The main reason why they should be avoided is the impact of sampling uncertainty in the variation of the performance parameter Dif_{Af}. The values of Dif_{Af} obtained with datasets containing few observations show a high variation and therefore cannot be considered reliable. Although the threshold value of 40 to 50 observations as the minimum sample size required for a reliable validation may be somewhat connected to the particular model used in this study, the general conclusion that small datasets should be avoided is considered valid; especially in cases of datasets with single observations. The advisable minimum number of observations may vary though for alternative models. It is expected that the more a model is linked to the original data (e.g. polynomial models), the more it will be subject to variation during evaluation with alternative datasets, and therefore the larger those datasets should be.

While the bias factor is useful to show whether the model is “fail-safe” or “fail-dangerous”, depending on the direction of the prediction bias, it does not picture the general predictive accuracy of a model since negative and positive biases cancel-out (Ross, 1996). However, it would be interesting to perform a similar study using the bias factor as a measure, in order to investigate the impact of data generation factors on the bias direction of a model. This kind of study would be particularly interesting for the comparison of datasets with different strains and different growth environments. For example, although here no significant difference between pathogenic and non-pathogenic strains was encountered, that might not be true in terms of bias factor.

One of the main conclusions of this study was that it is important to take into account the correspondence between the growth inhibiting factors, such as the presence of lactic acid, used in the original data and the ones used in selected evaluation datasets. The model performed best with both data collected in liquid media with added lactic acid and in meat environments. Lactic acid is present in fresh meat due to its level in the animal at the time of slaughter, residual biochemical activity (glycolysis) after slaughter and, to a lesser extent, due to metabolism of residual glucose by bacteria during storage. These conditions were mimicked in the original data that generated the model by adding lactic acid to the liquid media, in order to mimic a fresh

meat growth environment. This explains the better performance of the model with those two types of environments when evaluated against independent data. Its poor performance with data collected with other foods can be attributed to the fact that those contained various factors that can affect microbial growth and that were not considered in the original data, such as other organic acids or other sources of a_w modification.

The study also showed that the difference between model performance with pathogenic and non-pathogenic strains was negligible. Although this is supported by previous findings with the same model (Salter, 1998), there are also indications that the maximum specific growth rate of *E. coli* may vary significantly between strains at near growth-limiting conditions (Nauta and Dufrenne, 1999), such as the ones often experienced in the food chain. Also, the fact that data was grouped based on strain pathogenicity, and not in terms of individual strains, may have contributed to the result; compared to the strain used to generate the model, faster- and slower-growing individual strains may exist in both groups, which contributes to an overall non-significant difference between them. However, when working with published data, strain information is reported in such various ways that it is a difficult task to correctly group data by strain. Therefore, when a model is evaluated with literature data, between-strain variability should not be discarded.

Other sources of variability in predictive growth models have previously been indicated as the initial cell density, previous cell history, (consequential) lag duration, variability of microflora and microbial interactions and variability of growth conditions (food composition and environmental factors).

Whereas some of these sources are practically impossible to determine when dealing with real food in a food chain, such as initial cell density, previous cell history and (consequently) lag duration, variability of food microflora and of growth conditions can easily be measured. Hence, during the development of a microorganism- and product-specific QMRA, these characteristics can be defined and the predictive models selected for use in the exposure assessment should be tested specifically for those conditions under study. This will improve the precision of the model predictions under the circumstances applicable to that particular QMRA; consequently, the impact of variability on the risk estimation due to the application of predictive models will be reduced, and also easier to quantify, as it will be directly related to the distribution of Dif_{Af}.

Ultimately, variability also results from stochastic randomness. For example, the lag time duration is variable both due to biological reasons (previous cell history) and to randomness (initial inoculum). Both factors do not necessarily need to be exclusive; it is rather that they act in combination in the determination of variability. The individual modelling approach is a tool that can help to clarify the contribution of each condition to the total variability of a microbial response. This will contribute not only to the improvement of the understanding of microbial behaviour but also to a more transparent stochastic representation of variability of growth parameters in predictive microbiology. Nevertheless, one may question what will be the advantage of such development in terms of QMRA. The difference between modelling separately two or more different sources of variability of one growth parameter, or the overall variability of that parameter without specifying its various causes may be lost in the complexity of a QMRA model. Additionally, when the knowledge about a source of variability is in itself uncertain (e.g. previous cell history), the usefulness for QMRA of this mechanistic and extremely detailed approach of modelling the variability of growth parameters should be questioned.

Once the composition of the natural microflora of a food is determined, the variability that it causes in growth parameters can be quantified. Since this variability is mostly due to microbial interactions, which occur through molecular mechanisms such as quorum-sensing, “omics” technologies represent an essential tool to determine and describe this source of variability. Once the molecular microbial interactions that occur in real food are understood, the selection of appropriate microbial “cocktails” to simulate the real food environment in laboratory experiments will be more accurate, and consequently, predictive models developed in laboratory media will be better able to predict the growth occurring in foods. This represents an advantage

for the use of predictive models in QMRA. Additionally, it will be possible to clarify at which stages microbial interactions are in fact contributing to variability in growth. For example, it might be that during the lag phase cells are too far apart to communicate at the molecular level, unlike at the transition between exponential and stationary phase. In any case, there is still a need to assess the relative importance of lag variability in QMRA compared to variability of growth rate.

CONCLUSIONS

This thesis aimed to investigate the impact of different methods of analysis of microbial data on the risk estimates obtained in quantitative microbiological risk assessment (QMRA) studies, where those data are used as input.

The concept of QMRA together with the most common statistical approaches to the analysis of microbial data and the concept of predictive microbiology were introduced in the beginning of the thesis. Next three independent studies were presented in three separate manuscripts.

The first study consisted mainly of the development of a new statistical method for the analysis of raw (non-summarized) plate count data that results in the joint estimation of prevalence of contamination and distribution of microbial concentrations. This method meets one of the latest advances in the analysis of microbial data, i.e. the use of raw count data. Furthermore, it discards the use of the concept of limit of quantification, which has been demonstrated in this same study to be artificial, since enumeration of microorganisms is a probabilistic process that cannot be interpreted by means of threshold values. The validity of the assumption of a particular value as limit of quantification (LOQ) was here demonstrated to be dependent on the original distribution of concentrations and on the LOQ value considered. It is concluded that other factors with a real impact on enumeration should be considered during analysis of microbial counts, instead of the LOQ. These factors include specificity and sensitivity of the method of analysis, test portion size, probabilistic nature of enumeration and characteristics of the enrichment step. The new method developed takes some of those factors into account (probabilistic nature of enumeration and test portion size).

During the interpretation of counts, some plates are often excluded from the calculations due to overcrowding. This represents a loss of information that should be avoided, especially because these plates might represent sample units with extremely high concentrations, which might have an important influence on risk estimation. This method represents an approach where highly populated plates do not need to be ignored and can be treated as right-censored; hence there is no loss of important information for risk estimation.

The method was developed with simulated data and assuming a zero-inflated Poisson-lognormal (ziPLN) distribution as the “true” distribution of concentrations. The Poisson-lognormal has previously been indicated as one of the most appropriate choices to describe microbial count data, as it describes the concentrations from which the counts arise as lognormal and the process of plate counting as Poisson. Furthermore, zero-inflation accounts for over-dispersion of counts caused by an extreme number of zeroes, which is frequently observed when measuring contamination of foods by pathogenic microorganisms, due to low prevalence. The ziPLN was therefore considered an appropriate choice to build the new method. However, the method can easily be adapted to other distributions.

Several scenarios of concentrations and levels of prevalence were simulated in order to assess the performance of the method under different contamination realities. In general, the method provided very accurate estimates of prevalence and estimates of mean and standard deviation of concentrations with negligible biases. The bias found seemed to be associated with the proportion of artificial zero counts among all the observed zeroes and on their correct differentiation from true zeroes by the model. A poorer performance was observed at higher proportions of artificial zeroes among the total number of zeroes. A joint estimation of prevalence and concentrations, with the inherent differentiation between true and artificial zero counts is particularly important when microbial data integrates QMRA studies where growth can occur.

The successful application of this method with real data will obviously depend on the fact that those data fulfil the assumptions behind the modelling process (including those behind data generation), on good initial guesses for the parameters estimated and on an appropriate sample size that does not compromise the accuracy of the method. Furthermore, the method assumes no measurement uncertainty or uncertainty

associated to growth during enrichment, and the uncertainty associated to sampling from the "true" distribution of concentrations was negligible. It is advisable to test the method's sensitivity to uncertainty sources when applying it to real data.

In the end, this method can only be applied when the estimation of prevalence and concentrations is based on plate count data. However, direct plate counts are not frequently used due to low concentration levels in foods; the MPN approach is often preferred, despite its flaws. Also, low prevalence levels often lead to the performance of a detection test before proceeding to enumeration, despite the contribution of this practice to the inaccuracy of microbial analysis. QMRA needs accurate descriptions of both prevalence and distribution of concentrations, hence alternative laboratory practices should be considered to meet the specific needs of QMRA. It is considered that the new method can be adapted in the future for the interpretation of quantitative microbial information of other nature than direct plate counting, such as quantification by enrichment real-time PCR.

The second manuscript presents a study that aimed to investigate the impact on risk estimates of the frequency distribution used to describe microbial concentrations in QMRA.

The study was performed with datasets generated by simulation from a set of zero-inflated Poisson-lognormal (ziPLN) distributions. These datasets were fitted to a ziPLN, a lognormal (LN), a zero-inflated lognormal (ziLN) and a Poisson-gamma (PGM) distribution, and these were used to estimate the risk using a combined consumer-phase - dose-response model (CPM-DR) for *Campylobacter* as an example.

The similarity of the risk given by the ziLN and the ziPLN suggested that it might not be necessary to account for the homogeneity in counts (which is described by a Poisson distribution) in order to obtain accurate risk estimates. On the other hand, the difference observed between the LN and the PGM indicated that it is important to choose a frequency distribution that is close to the parametric form of the true distribution of concentrations. In fact, in terms of risk estimation, it is expected that providing a good description of the distribution of microbial concentrations in a food lot (including the level of contamination clustering) has a greater impact on the outcome than accounting for homogeneity in counts obtained at the laboratory level. This study has confirmed this hypothesis. However, generalizations of the Poisson distribution can be used with a different goal in QMRA. They can be applied for the differentiation of two levels of variability of concentrations, such as the between- and within-lot variability. In such case, it is expected that the use of a generalized Poisson distribution provides more accurate risk estimates compared to the simple continuous distributions (gamma or lognormal).

Zero-inflation was demonstrated to be a good approach to the fit of distributions to microbial data at all levels of prevalence. Furthermore, when compared with the non-zero-inflated LN distribution, the ziLN was not significantly influenced by the adoption of alternative LOQ values during the maximum likelihood estimation (MLE) procedure. Another advantage of zero-inflation is that it provides an estimate of prevalence and hence differentiates between true and artificial zeroes during fitting. This differentiation appears to be more relevant as the true prevalence and concentrations increase, since the proportion of artificial zeroes among the total number of zeroes is larger in those situations. It is therefore recommended to use zero-inflated distributions to describe microbial concentrations for use in QMRA.

The LN distribution was considered inappropriate for risk estimation, even when fitted with MLE for censored data, due to its dependence on the LOQ value used.

As for the first study, the application of these results is dependent on the fulfilment by real data of the assumptions behind modelling (including data generation). It is expected that once data originates from a different true distribution of concentrations the results in terms of risk will differ. Additionally, no measurement errors or assumptions in terms of back-calculation from counts to concentrations were assumed in this study, and the uncertainty due to sampling from the "true" distribution of concentrations was negligible. Also,

different observations might occur when using an alternative CPM in the estimation of risk and for different (higher) true concentration levels. To obtain more generic conclusions, this study should be repeated assuming different conditions, such as accounting for the effect of different sources of uncertainty, applying different CPMs, and assuming higher concentration levels at retail in order to obtain sound generalized conclusions in terms of the impact of different distributions on human health risk estimates.

An important characteristic of this study is the fact that the "true" distribution was known. Hence, the impact of distribution choice could be assessed not only in terms of goodness of fit, but most importantly, in terms of the accurate estimation of risk. Indeed, it was here demonstrated that a perfect fit of a distribution to the data is not a crucial factor for an accurate QMRA.

The last study aimed to investigate how factors with an impact on variability and uncertainty inherent to collection of evaluation data, affect the performance evaluation of predictive microbiology models. For this, a published square-root model for the growth of *E. coli* against four growth variables was used as an example. The investigated factors included the relationship between the growth conditions encountered in evaluation data and those that integrate the model's interpolation region, the presence or absence of the growth-inhibiting lactic acid, the type of growth media used and the pathogenicity of the bacterial strain used. It was hypothesized that the model would perform best with evaluation datasets collected under conditions closer to the ones encountered in the original data (the data used to build the model). To investigate model performance, the measure accuracy factor (A_f) was used as a basis.

It was concluded that, while selecting published data to evaluate a model, with the objective of validating it for use in a specific QMRA, it is important to:

1) Account for the presence of the same growth-inhibiting factors that were used in the original data. This is both true for data from liquid media and foods. It is hence important to know the purpose behind the development of the model, i.e. to which kind of product it was targeted, in order to select data either from the same type of food or from liquid media with conditions that mimic that food. Using evaluation data collected without such consideration may contribute to mask a good performance of the model for the purpose to which it was actually developed.
2) Consider the growth conditions, such as temperature, which are expected to be modelled in the QMRA. The importance of between-strain variability of microbial growth rates is expected to be dependent on the growth conditions encountered along the food chain. Near growth-limiting conditions it might be crucial to account for this variability, whereas it might be irrelevant to do so at optimum growth conditions. Here no significant difference on the model performance between pathogenic and non-pathogenic strains of *E. coli* was observed. This observation is however subjected to a few considerations. Data was not analysed at the individual strain level, so it is possible that strains contributing both to a good and a bad performance of the model existed in both groups compared. Also, the growth conditions under which data from both groups were collected were not taken into account, therefore the effect of a higher variability near growth limiting conditions was also not considered in the analysis. Last, for different microorganisms and for different growth models, the impact of strain variability on model performance may vary. Therefore, it is concluded that before further studies bring to light the actual impact of strain variability in model evaluation, variation of growth parameters due to strain should be stochastically accounted for in QMRA.
3) Account for sampling uncertainty, i.e. larger datasets should be preferred to smaller ones, and datasets of single observations should always be avoided in model evaluation. This practice will greatly reduce the uncertainty in the model evaluation process. It will particularly avoid an underestimation of the average A_f of the model, since it was here demonstrated that lower A_f values were most of the times obtained with small datasets and are therefore believed to be associated to randomness.
4) Consider the type of model under evaluation. Highly parameterized models, such as some polynomial models or highly product-specific models, are usually much linked to the original data. Hence,

considerations in terms of selection of appropriate evaluation data in such cases may need to be stricter, for example, in terms of the relationship between the ranges of growth conditions considered both on the evaluation and the original data. It is also expected under those circumstances that larger datasets are needed for a valuable model validation.

One of the main contributions of this study for the future evaluation of model performance is the fact that an alternative accuracy factor-related parameter is used for model evaluation against literature data. It is hypothesized that the best possible performance obtained with a model is given by its accuracy factor calculated with the data used to generate it ($A_{f\ original}$). Since models are imperfect, this value is rarely 1 corresponding to a perfect agreement between predictions and observations. Hence, for a specific model, $A_{f\ original}$ should be adopted as the reference value for its "perfect" performance. As a result, Dif_{Af} (the difference between $A_{f\ original}$ and the A_f calculated with the evaluation data) is used as a measure of performance, instead of the simple A_f. Dif_{Af} should be considered in future evaluation studies, and datasets for which negative values of Dif_{Af} are obtained should be excluded from the evaluation process, since it is highly likely that a negative result occurs due to chance.

The final conclusion of manuscript III is that, since predictive microbiology models are seldomly developed with the aim of serving QMRA purposes, before their use in a specific QMRA study they should at least go through a targeted evaluation process. This means that the data used for evaluation should be selected according to the conditions modelled in the QMRA study. Obviously, the lack of appropriate published data which meet all the above mentioned conditions will represent a considerable challenge to the adoption of such approach. Hence, whenever it is impossible to describe the variability or reduce the uncertainty due to specific factors during the model evaluation process, it is advisable that the variation caused by those factors is stochastically integrated in QMRA studies and their impact on risk investigated.

To conclude, the three studies presented in this thesis, although focused on three individual specific problems, jointly contributed to a clarification of the impact of different factors of microbial data analysis on the outcome of QMRA. The total variation that exists on the final risk estimate is a result of the uncertainty and variability attending the input data. Hence, across this thesis, there was a special focus on the interpretation of the meaning of results in terms of variability and uncertainty. In summary, this work 1) introduced a new method for the analysis of microbial counts that reduces the uncertainty associated with the assumption of a LOQ and the use of summarized data, and provides a more accurate description of the variability in concentrations of bacteria on food products; 2) clarified the importance of assuming the correct form of variability of concentrations at retail while fitting a distribution to microbial data, and the importance of an accurate estimation of prevalence; 3) identified the growth environment as an important source of variability and the number of observations as an important source of uncertainty which need to be considered during the evaluation of predictive microbiology models with published data.

FUTURE PERSPECTIVES

With the rise of “omics” technologies (genomics, transcriptomics, proteomics, metobolomics…), it is expected that microbial analysis will slowly transit from traditional enumeration of microorganisms by direct plate counts and MPN to quantification in terms of “omics” data (e.g. whole genome sequences, target genes, target proteins…). New methods for the basic analysis of such kind of outputs need to be developed, preferably, through a trans-disciplinary approach, so that the same data can be integrated in different kinds of studies, including quantitative microbiological risk assessment (QMRA).

However, great challenges are foreseen in undertaking this task, due to the high throughput of data and the novelty of “omics” technologies in the field of food microbiology. The joint collaboration between bioinformaticians, microbiologists and risk assessors will be a key element for the success of developing ways to interpret “omics” data in a useful and meaningful way to all parties involved (Brul et al., 2012; McMeekin et al., 2013).

The challenge lies not only on the development of alternative statistical ways to summarize the new types of microbial data, but also on adapting the current structure of QMRA studies to make the use of such data possible. Additionally, the variability and uncertainty that may influence the data, and therefore have an impact on the risk, need to be investigated. Ultimately, integration of “omics” data into QMRA models needs to be done in a conscious way, with the objectives of QMRA in mind. The use of such data in QMRA studies will only represent an advantage where it helps improving the accuracy and transparency of risk estimation, for example, by clarifying the role of microbial interactions in foods or important between-strain variability. Otherwise, detailed descriptions of molecular level reactions or genetic expression should be avoided if they only contribute to an overparameterization of QMRA models, which are usually already far from being parsimonious.

Apart from the advent of “omics” data in QMRA, there are other challenges that need to be addressed. There are indications that the modeling of exposure assessment will eventually evolve from a Monte-Carlo simulation approach to a Bayesian approach (Greiner et al., 2013; Rigaux et al., 2013). Similarly, the development of predictive microbiology models has already entered the “Bayesian era” (Corkrey et al., 2012). It is therefore expected that new tools for the analysis of microbial data that can be integrated within a Bayesian modeling structure will be needed in the near future, such as Bayesian approaches to fit distributions to microbial enumeration data.

In the immediate future though, the traditional way of reporting and analyzing microbial counts will still be in place, and it is not certain to what extent it will be substituted by the upcoming methods. Even when these take over, there will still be a need to fit quantitative data for use in QMRA. Therefore, the performance of the method introduced in manuscript I with real data must be assessed and if necessary the appropriate adjustments need to be made, in order to turn this into an accessible tool for a more accurate analysis of microbial counts. Once validated, it would be desirable to make the method available to the scientific community by developing, for example, a working package for the statistical software R.

It was here concluded that the choice of the method for analysis of microbial counts should depend on the way the data was collected, for example by detection followed by enumeration or only by enumeration. This information is crucial for the adoption of a method of data analysis that either does or does not estimate prevalence. There is therefore a need to adopt a standardized way of reporting microbial counts, which clearly specifies the steps taken during data collection. Another sort of (crucial) information that is sometimes missing is the precise amount of sample from each dilution that is actually spread on the plates. Extensions of manuscript II could help obtaining more general conclusions about the impact of different frequency distributions on risk estimates. Scenarios where the data are simulated from a different “true” distribution of concentrations (e.g. a Poisson-gamma), where a different consumer-phase model is used, and where different kinds of uncertainty are accounted for are some of the obvious possibilities.

In the field of predictive microbiology, models seem to be evolving towards more mechanistic ways of describing microbial behavior, following also the trend of introducing “omics” data into modelling (Corkrey et al., 2012). The understanding of systems biology may benefit predictive microbiology by allowing the development of more generalist predictive models applicable to, for example, different microorganisms. However, there will be a need to assess how applicable those models will be in QMRA, since they will probably be so data-demanding and computationally complex that it might be impractical to rely on their predictions in the context of risk estimation. Hence, simpler microorganism- and product-specific models may still be the choice for use in QMRA. These models need to be validated before their application in QMRA. Manuscript III provided an insight into two important sources of variability and uncertainty that may affect performance evaluation of a model. One important conclusion of this study was that there is a need to perform model evaluation targeted at the conditions under which the model will be used, i.e. the conditions modelled in exposure assessment. An alternative approach to “targeted validation” would be to make the development of predictive models an integrant part of the QMRA structure. Both searching for the most appropriate evaluation data for specific conditions and developing a specific growth model for use in QMRA are resource- and time-demanding tasks; thus their beneficial impact on the accuracy of risk estimation needs to be assessed before such approaches are put into place. A simulation study could help investigate to what level QMRA-targeted development and validation of predictive models are necessary for the accurate estimation of risk.

REFERENCES

Adair, C., Kilsby, D. C., Whittall, P. T. (1989). Comparison of the schoolfield (non-linear arrhenius) model and the square root model for predicting bacterial growth in foods. *Food Microbiology, 6*(1), 7-18.

Adams, M. R., Little, C. L., Easter, M. C. (1991). Modeling the effect of ph, acidulant and temperature on the growth-rate of *Yersinia enterocolitica*. *Journal of Applied Bacteriology, 71*(1), 65-71.

Adams, M.R., Moss, M.O. (1995). *Food microbiology*. Cambridge: The Royal Society of Chemistry.

Alber, S. A., Schaffner, D. W. (1992). Evaluation of data transformations used with the square root and schoolfield models for predicting bacterial growth rate. *Applied and Environmental Microbiology, 58*(10), 3337-3342.

Andersen, J. K., Hald, T., Nielsen, N. L., Fiedler, C. S., Nørrung, B. (2007). New strategies for the use of microbiological examinations in food control in Denmark. *Food Control, 18*(3), 273-277.

Baranyi, J., McClure, P. J., Sutherland, J. P., Roberts, T. A. (1993). Modeling bacterial growth responses. *Journal of Industrial Microbiology, 12*(3-5), 190-194.

Baranyi, J., Roberts, T. A., McClure, P. (1993). A non-autonomous differential equation to model bacterial growth. *Food Microbiology, 10*(1), 43-59.

Baranyi, J., Roberts, T. A. (1995). Mathematics of predictive food microbiology. *International Journal of Food Microbiology, 26*(2), 199-218.

Baranyi, J. (1998). Comparison of stochastic and deterministic concepts of bacterial lag. *Journal of Theoretical Biology, 192*(3), 403-408.

Baranyi, J. (2002). Stochastic modelling of bacterial lag phase. *International Journal of Food Microbiology, 73*(2-3), 203-206.

Bassett, J., Jackson, T., Jewell, K., Jongenburger, I., Zwietering, M. (2010). Impact of microbial distributions on food safety. ILSI Europe Report, Brussels.

Bassett, J., Nauta, M., Lindqvist, R., Zwietering, M. (2012). Tools for Microbiological Risk Assessment. ILSI Europe Report, Brussels.

Baty, F., Delignette-Muller, M. (2004). Estimating the bacterial lag time: Which model, which precision? *International Journal of Food Microbiology, 91*(3), 261-277.

Boysen, L., Nauta, M., Duarte, A. S. R., Rosenquist, H. (2013). Human risk from thermotolerant *Campylobacter* on broiler meat in Denmark. *International Journal of Food Microbiology, 162*(2), 129-134.

Bratchell, N., Gibson, A. M., Truman, M., Kelly, T. M., Robets, T. A. (1989). Predicting microbial growth: The consequences of quantity of data. *International Journal of Food Microbiology, 8*(1), 47-58.

Broughall, J. M., Anslow, P. A., Kilsby, D. C. (1983). Hazard analysis applied to microbial-growth in foods - development of mathematical-models describing the effect of water activity. *Journal of Applied Bacteriology, 55*(1), 101-110.

Broughall, J. M., Brown, C. (1984). Hazard analysis applied to microbial growth in foods: Development and application of three-dimensional models to predict bacterial growth. *Food Microbiology, 1*(1), 13-22.

Brul, S., Bassett, J., Kathariou, S., Mcclure, P., Jasti, P., Betts, R. (2012). 'Omics' technologies in quantitative microbiological risk assessment. ILSI Europe Report, Brussels.

Busschaert, P., Geeraerd, A. H., Uyttendaele, M., Van Impe, J. F. (2010). Estimating distributions out of qualitative and (semi)quantitative microbiological contamination data for use in risk assessment. *International Journal of Food Microbiology, 138*(3), 260-269.

Busschaert, P., Geeraerd, A. H., Uyttendaele, M., Van Impe, J. F. (2011). Hierarchical bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. *Food Microbiology, 28*(4), 712-719.

CAC (Codex Alimentarius Commission) (1999). Principles and guidelines for the conduct of microbiological risk assessment. Document CAC/GL-30. FAO, Rome. [online] Available at: www.codexalimentarius.netownload/standards/357/CXG_030e.pdf [Accessed: May 2013].

Cassin, M. H., Lammerding, A. M., Todd, E. C. D., Ross, W., McColl, R. S. (1998). Quantitative risk assessment for *Escherichia coli* O157:H7 in ground beef hamburgers. *International Journal of Food Microbiology, 41*(1), 21-44.

Christensen, B. B., Rosenquist, H., Sommer, H. M., Nielsen, N. L., Fagt, S., Andersen, N. L., Nørrung, B. (2005). A model of hygiene practices and consumption patterns in the consumer phase. *Risk Analysis, 25*(1), 49-60.

Christensen, B. B., Nauta, M., Korsgaard, H., Sørensen, A., Irene Vedel, Rosenquist, H., Boysen, L., Nørrung, B. (2013). Case-by-case risk assessment of broiler meat batches: An effective control strategy for campylobacter. *Food Control, 31*(2), 485-490.

Clough, H. E., Clancy, D., O'Neill, P. D., Robinson, S. E., French, N. P. (2005). Quantifying uncertainty associated with microbial count data: A bayesian approach. *Biometrics, 61*(2), 610-616.

Cochran, W. G. (1950). Estimation of bacterial densities by means of the most probable number. *Biometrics, 6*(2), 105-116.

Cole, M.B. (1991). Opinion: predictive modelling - yes it is! *Letters of Applied Microbiology,* 13, 218-219.

Coleman, M. E., Tamplin, M. L., Phillips, J. G., Marmer, B. S. (2003). Influence of agitation, inoculum density, pH, and strain on the growth parameters of *Escherichia coli* O157:H7-relevance to risk assessment. *International Journal of Food Microbiology, 83*(2), 147-160.

Commeau, N., Parent, E., Delignette-Muller, M., Cornu, M. (2012). Fitting a lognormal distribution to enumeration and absence/presence data. *International Journal of Food Microbiology, 155*(3), 146-152.

Corkrey, R., Olley, J., Ratkowsky, D., McMeekin, T., Ross, T. (2012). Universality of thermodynamic constants governing biological growth rates. *PLoS One, 7*(2)

Crepet, A., Albert, I., Dervin, C., Carlin, F. (2007). Estimation of microbial contamination of food from prevalence and concentration data: Application to *Listeria monocytogenes* in fresh vegetables. *Applied and Environmental Microbiology, 73*(1), 250-258.

Currie, L. A. (1968). Limits for qualitative detection and quantitative determination - application to radiochemistry. *Analytical Chemistry, 40*(3), 586-&.

Dalgaard, P., Jørgensen, L.V. (1998). Predicted and observed growth of *Listeria monocytogenes* in seafood challenge tests and in naturally contaminated cold-smoked salmon. *International Journal of Food Microbiology, 40*(1-2), 105-115.

Dalgaard, P., Buch, P., Silberg, S. (2002). Seafood Spoilage Predictor - development and distribution of a product specific application software. *International Journal of Food Microbiology, 73*(2-3), 343-349.

Dalgaard, P. (2004). *Assessment and Management of Seafood Safety and Quality: Predictive microbiology*. FAO Fisheries Technical Paper, Rome.

Davey, K. R. (1989). A predictive model for combined temperature and water activity on microbial-growth during the growth-phase. *Journal of Applied Bacteriology, 67*(5), 483-488.

Davey, K. R. (1991). Applicability of the davey (linear arrhenius) predictive model to the lag phase of microbial growth. *Journal of Applied Bacteriology, 70*(3), 253-257.

Delignette-Muller, M., Rosso, L., Flandrois, J. P. (1995). Accuracy of microbial growth predictions with square root and polynomial models. *International Journal of Food Microbiology, 27*(2-3), 139-146.

Delignette-Muller, M., Pouillot, R., Denis, J., Dutang, C. (2010) *Use of the package 'fitdistrplus' to specify a distribution from non-censored or censored data*. [online] Available at: https://r-forge.r-project.org/scm/viewvc.php/*checkout*/www/intro2fitdistrplus.pdf?revision=82&root=riskassessment&pathrev=82 [Accessed: January 2011].

Dens, E. J., Bernaerts, K., Standaert, A. R., Impe, J. F. (2005). Cell division theory and individual-based modeling of microbial lag. part I. the theory of cell division. *International Journal of Food Microbiology, 101*(3), 303-318.

Dens, E. J., Van Impe, J. F. (2001). On the need for another type of predictive model in structured foods. *International Journal of Food Microbiology, 64*(3), 247-260.

Donald, M., Cook, A., Mengersen, K. (2009). Bayesian network for risk of diarrhea associated with the use of recycled water. *Risk Analysis, 29*(12), 1672-1685.

Duffy, G., Sheridan, J. J., Buchanan, R. L., McDowell, D. A., Blair, I. S. (1994). The effect of aeration, initial inoculum and meat microflora on the growth kinetics of *Listeria monocytogenes* in selective enrichment broths. *Food Microbiology, 11*(5), 429-438.

European Food Safety Authority (EFSA) (2010).Quantitative microbiological risk assessment on *Salmonella* in slaughter and breeder pigs: Final report. The EFSA Journal.

EFSA Panel on Biological Hazards (BIOHAZ), 2011. Scientific opinion on Campylobacter in broiler meat production: control options and performance objectives and/or targets at different stages of the food chain. EFSA Journal 9 (4), 2105.

Evers, E. G., Post, J., Putirulan, F. F., Wal, F. J. (2010). Detection probability of *Campylobacter*. *Food Control, 21*(3), 247-252.

Evers, E. G., Chardon, J. E. (2010). A swift quantitative microbiological risk assessment (sQMRA) tool. *Food Control, 21*(3), 319-330.

FDA. (2006). Bacteriological Analytical Manual. Appendix 2: Most Probable Number from serial dilutions.

Fischer, A. R. H., Frewer, L. J., De Jong, Aarieke E. I., De Jonge, R., Nauta, M. J., Fischer, A. R. H. (2005). Perspective: Improving food safety in the domestic environment: The need for a transdisciplinary approach. *Risk Analysis, 25*(3), 503-517.

Garthright, W. E., Blodgett, R. J. (2003). FDA's preferred MPN methods for standard, large or unusual tests, with a spreadsheet. *Food Microbiology, 20*(4), 439-445.

Genigeorgis, C., Martin, S., Franti, C. E., Riemann, H. (1971). Initiation of staphylococcal growth in laboratory media. *Applied Microbiology, 21*(5), 934-939.

Gerwen, S. J. C., Zwietering, M. H. (1998). Growth and inactivation models to be used in quantitative risk assessments. *Journal of Food Protection, 61*(11), 1541-1549.

Gibson, A. M., Bratchell, N., Roberts, T. A. (1988). Predicting microbial-growth - growth-responses of salmonellae in a laboratory medium as affected by ph, sodium-chloride and storage-temperature. *International Journal of Food Microbiology, 6*(2), 155-178.

Gilchris, J.E., Campbell, J.E., Donnelly, C. B., Peeler, J. T., Delaney, J. M. (1973). Spiral plate method for bacterial determination. *Applied Microbiology, 25*(2), 244-252.

Giménez, B., Dalgaard, P. (2004). Modelling and predicting the simultaneous growth of *Listeria monocytogenes* and spoilage micro-organisms in cold-smoked salmon. *Journal of Applied Microbiology, 96*(1), 96-109.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London, 115*, 513-583.

Gonzales-Barron, U., Butler, F. (2011). A comparison between the discrete poisson-gamma and poisson-lognormal distributions to characterise microbial counts in foods. *Food Control, 22*(8), 1279-1286.

Gonzales-Barron, U., Kerr, M., Sheridan, J. J., Butler, F. (2010). Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. *International Journal of Food Microbiology, 136*(3), 268-277.

Greiner, M., Smid, J., Havelaar, A. H., Müller-Graf, C. (2013). Graphical models and bayesian domains in risk modelling: Application in microbiological risk assessment. *Preventive Veterinary Medicine, 110*(1), 4-11.

Havelaar, A. H., Nauta, M. J. (2007). Second-order modeling of variability and uncertainty in microbial hazard characterization. *Journal of Food Protection, 70*(10), 2228-2229.

Havelaar, A. H., Evers, E. G., Nauta, M. J. (2008). Challenges of quantitative microbial risk assessment at EU level. *Trends in Food Science and Technology, 19*(1)

Helsel, D. R. (2005). Insider censoring: Distortion of data with nondetects. *Human and Ecological Risk Assessment, 11*(6), 1127-1137.

Helsel, D. R. (2006). Fabricating data: How substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere, 65*(11), 2434-2439.

Hills, B. P., Wright, K. M. (1994). A new model for bacterial growth in heterogeneous systems. *Journal of Theoretical Biology, 168*(1), 31-41.

ISO10272-1. (2006). Horizontal method for detection and enumeration of Campylobacter spp. Part 1: detection method.

ISO10272-2. (2006). Horizontal method for detection and enumeration of Campylobacter spp. Part 2: colony-count technique.

Jameson, J. E. (1962). A discussion of the dynamics of *Salmonella* enrichment. *The Journal of Hygiene, 60*(2), 193-207.

Jongenburger, I., Bassett, J., Jackson, T., Gorris, L. G. M., Jewell, K., Zwietering, M. H. (2012a). Impact of microbial distributions on food safety II. quantifying impacts on public health and sampling. *Food Control, 26*(2), 546-554.

Jongenburger, I., Reij, M. W., Boer, E. P. J., Zwietering, M. H., Gorris, L. G. M. (2012b). Modelling homogeneous and heterogeneous microbial contaminations in a powdered food product. *International Journal of Food Microbiology, 157*(1), 35-44.

Kelly, E. J., Campbell, K. (2000). Separating variability and uncertainty in environmental risk assessment - making choices. *Human and Ecological Risk Assessment, 6*, 1-14.

Kennedy, M., Hart, A. (2009). Bayesian modeling of measurement errors and pesticide concentration in dietary risk assessments. *Risk Analysis, 29*(10), 1427-1442.

Kilsby, D. C., Pugh, M. E. (1981). The relevance of the distribution of micro-organisms within batches of food to the control of microbiological hazards from foods. *Journal of Applied Bacteriology, 51*(2), 345-354.

Kjeldgaard, N. O., Maaloe, O., Schaechter, M. (1958). The transition between different physiological states during balanced growth of *Salmonella typhimurium*. *Journal of General Microbiology, 19*(3), 607-616.

Krämer, N., Løfstrøm, C., Vigre, H., Hoorfar, J., Bunge, C., Malorny, B. (2011). A novel strategy to obtain quantitative data for modelling: Combined enrichment and real-time PCR for enumeration of salmonellae from pig carcasses. *International Journal of Food Microbiology, 145*(1)

Le Marc, Y., Huchet, V., Bourgeois, C. M., Guyonnet, J. P., Mafart, P., Thuault, D. (2002). Modelling the growth kinetics of listeria as a function of temperature, pH and organic acid concentration. *International Journal of Food Microbiology, 73*(2-3), 219-237.

Le Marc, Y., Pin, C., Baranyi, J. (2005). Methods to determine the growth domain in a multidimensional environmental space. *International Journal of Food Microbiology, 100*(1-3), 3-12.

Legan, J.D. (2007). Application of models and other quantitative microbiology tools in predictive microbiology. In S. Brul, S. van Gerwen, M. Zwietering (Eds.), *Modelling microorganisms in food* (83-109). Cambridge: Woodhead Publishing Ltd.

Lindroth, S. E., Genigeorgis, C. A. (1986). Probability of growth and toxin production by nonproteolytic *Clostridium botulinum* in rockfish stored under modified atmospheres. *International Journal of Food Microbiology, 3*(3), 167-181.

Lopez, S., Prieto, M., Dijkstra, J., Dhanoa, M. S., France, J. (2005). Statistical evaluation of mathematical models for microbial growth. *International Journal of Food Microbiology, 96*(3)

Lorimer, M. F., Kiermeier, A. (2007). Analysing microbiological data: Tobit or not tobit? *International Journal of Food Microbiology, 116*(3), 313-318.

Marks, H. M., Coleman, M. E., Lin, C.-T.J., Roberts, T. (1998). Topics in microbial risk assessment: Dynamic flow tree process. *Risk Analysis, 18*(3), 309-328.

Marks, H. M., Coleman, M. E. (2005). Accounting for inherent variability of growth in microbial risk assessment. *International Journal of Food Microbiology, 100*(1-3), 275-287.

McDonald, K., Sun, D. (1999). Predictive food microbiology for the meat industry: A review. *International Journal of Food Microbiology, 52*(1-2), 1-27.

McKellar, R. C. (1997). A heterogeneous population model for the analysis of bacterial growth kinetics. *International Journal of Food Microbiology, 36*(2-3), 179-186.

McMeekin, T. A., Chandler, R. E., Doe, P. E., Garland, C. D., Olley, J., Putro, S., Ratkowsky, D. A. (1987). Model for combined effect of temperature and salt concentration/water activity on the growth rate of *Staphylococcus xylosus*. *Journal of Applied Microbiology, 62*(6), 543-550.

McMeekin, T. A., Ross, T., Olley, J. (1992). Application of predictive microbiology to assure the quality and safety of fish and fish products. *International Journal of Food Microbiology, 15*(1-2), 13-32.

McMeekin, T.A., Olley, J.N., Ross, T., Ratkowsky, D.A. (1993). Basic concepts and methods. In A.N. Sharpe (Ed.), Predictive microbiology: theory and application (pp. 11-86). Somerset: Research Studies Press.

McMeekin, T., Olley, J., Ratkowsky, D., Corkrey, R., Ross, T. (2013). Predictive microbiology theory and application: Is it all about rates? *Food Control, 29*(2), 290-299.

Mejlholm, O., Dalgaard, P. (2009). Development and validation of an extensive growth and growth boundary model for *Listeria monocytogenes* in lightly preserved and ready-to-eat shrimp. *Journal of Food Protection, 72*(10), 2132-2143.

Nørrung, B., Buncic, S. (2007). Microbial safety of meat in the European Union. *Meat Science, 78*(1-2), 14-24.

Nauta, M.J., Dufrenne, J.B. (1999). Variability in growth characteristics of different *E. coli* O157:H7 isolates, and its implications for predictive microbiology. *Quantitative microbiology, 1*, 137-155.

Nauta, M. J. (2000). Separation of uncertainty and variability in quantitative microbial risk assessment models. *International Journal of Food Microbiology, 57*(1-2), 9-18.

Nauta, M. (2001). A modular process risk model structure for quantitative microbiological risk assessment and its application in an exposure assessment of *Bacillus cereus* in a REPFED. RIVM Report, Bilthoven.

Nauta, M. J. (2002). Modelling bacterial growth in quantitative microbiological risk assessment: Is it possible? *International Journal of Food Microbiology, 73*(2-3), 297-304.

Nauta, M. J., Litman, S., Barker, G. C., Carlin, F. (2003). A retail and consumer phase model for exposure assessment of bacillus cereus. *International Journal of Food Microbiology, 83*(2), 205-218.

Nauta, M.J., Jacobs-Reitsma, W.F., Evers, E.G.; van Pelt, W., Havelaar, A.H. (2005). Risk assessment of *Campylobacter* in the Netherlands via broiler meat and other routes. RIVM report: Bilthoven.

Nauta, M.J. (2007). Uncertainty and variability in predictive models of microorganisms in food. In S. Brul, S. van Gerwen, M. Zwietering (Eds.), *Modelling microorganisms in food* (110-125). Cambridge: Woodhead Publishing Ltd.

Nauta, M.J. (2008). The Modular Process Risk Model (MPRM): a structured approach to food chain exposure assessment. In D.W. Schaffner (Ed.), *Microbial risk analysis of foods* (pp.99-136). Washington, D.C.: ASM Press.

Nauta, M., Hill, A., Rosenquist, H., Brynestad, S., Fetsch, A., van, d. Logt, P., Fazil, A., Christensen, B., Katsma, E., Borck, B., Havelaar, A. (2009a). A comparison of risk assessments on *Campylobacter* in broiler meat. *International Journal of Food Microbiology, 129*(2), 107-123.

Nauta, M. J., van, d. W., Putirulan, F. F., Post, J., van, d. K., Bolder, N. M. (2009b). Evaluation of the testing and scheduling strategy for control of *Campylobacter* in broiler meat in the Netherlands. *International Journal of Food Microbiology, 134*(3), 216-222.

Nauta, M., Christensen, B. (2011). The impact of consumer phase models in microbial risk analysis. *Risk Analysis, 31*(2), 255-265.

Neumeyer, K., Ross, T., McMeekin, T. A. (1997). Development of a predictive model to describe the effects of temperature and water activity on the growth of spoilage pseudomonads. *International Journal of Food Microbiology, 38*(1), 45-54.

Niemelä, S. (2002). Uncertainty of quantitative determinations derived by cultivation of microorganisms. Centre for Metrology and Accreditation, Helsinki.

Niemelä, S. I. (2003). Measurement uncertainty of microbiological viable counts. *Accreditation and Quality Assurance, 8*, 559-563.

NMKL119. (2007). Thermotolerant Campylobacter. Detection, semi-quantitative and quantitative determination in foods and drinking water.

Olley, J., Doe, P., Heruwati, E. (1988). The influence of drying and smoking on the nutritional properties of fish: an introductory overview. In J.R. Burt (Ed.), *Fish smoking and drying - the effect of smoking and drying on the nutritional properties of fish* (pp.1-17). Cambridge: Elsevier Science Publishers Ltd.

Parsons, D. J., Orton, T. G., D'Souza, J., Moore, A., Jones, R., Dodd, C. E. R. (2005). A comparison of three modelling approaches for quantitative risk assessment using the case study of salmonella spp. in poultry meat. *International Journal of Food Microbiology, 98*(1), 35-51.

Perez-Rodriguez, F., Asselt, E. D., Garcia-Gimeno, R., Zurera, G., Zwietering, M. H. (2007). Extracting additional risk mangers information from a risk assessment of *Listeria monocytogenes* in deli meats. *Journal of Food Protection, 70*(5), 1137-1152.

Pin, C., Baranyi, J. (1998). Predictive models as means to quantify the interactions of spoilage organisms. *International Journal of Food Microbiology, 41*(1), 59-72.

Pouillot, R., Delignette-Muller, M. (2010). Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *International Journal of Food Microbiology, 142*(3), 330-340.

Pouillot, R., Delignette-Muller, M.-L., Denis, Kelly, D.L., J.-B. (2011) *The mc2d package*. [online] Available at: http://cran.r-project.org/web/packages/mc2d/vignettes/docmcEnglish.pdf [Accessed: May 2013].

Pouillot, R., Hoelzer, K., Chen, Y., Dennis, S. (2013). Estimating probability distributions of bacterial concentrations in food based on data generated using the most probable number (MPN) method for use in risk assessment. *Food Control, 29*(2), 350-357.

Ratkowsky, D. A., Olley, J., McMeekin, T. A., Ball, A. (1982). Relationship between temperature and growth rate of bacterial cultures. *Journal of Bacteriology, 149*(1), 1-5.

Ratkowsky, D. A., Lowry, R. K., McMeekin, T. A., Stokes, A. N., Chandler, R. E. (1983). Model for bacterial culture growth rate throughout the entire bio kinetic temperature range. *Journal of Bacteriology, 154*(3), 1222-1226.

Ratkowsky, D. A., Ross, T., McMeekin, T. A., Olley, J. (1991). Comparison of Arrhenius-type and Belehradek-type models for prediction of bacterial growth in foods. *Journal of Applied Bacteriology, 71*(5), 452-459.

Ratkowsky, D.A. (1992). Predicting response times in predictive food microbiology. Occasional Paper 1992/1, Research and Development Unit, Biometrics Section, Department of Primary Industry, Fisheries and Energy, Tasmania, 38 pp.

Ratkowsky, D. A., Ross, T. (1995). Modelling the bacterial growth/no growth interface. *Letters in Applied Microbiology, 20*(1), 29-33.

Ratkowsky, D. A., Olley, J., Ross, T. (2005). Unifying temperature effects on the growth rate of bacteria and the stability of globular proteins. *Journal of Theoretical Biology, 233*(3), 351-362.

Reinders, R., Evers, E., De Jonge, R., Van Leusden, F. (2002). Variation in the numbers of Shiga toxin-producing *Escherichia coli* O157 in minced beef. RIVM report, Bilthoven.

Reinders, R. D., Jonge, R. D., Evers, E. G. (2003). A statistical method to determine whether micro-organisms are randomly distributed in a food matrix, applied to coliforms and *Escherichia coli* O157 in minced beef. *Food Microbiology, 20*(3), 297-303.

Ridout, M., Demétrio, C., Hinde, J. (1998). Models for count data with many zeros. *International Biometric Conference.* Cape Town.

Rieu, E., Duhem, K., Vindel, E., Sanaa, M. (2007). Food safety objectives should integrate the variability of the concentration of pathogen. *Risk Analysis, 27*(2), 373-386.

Rigaux, C., Ancelet, S., Albert, I., Carlin, F., Nguyen-thé, Albert, I. (2013). Inferring an augmented bayesian network to confront a complex quantitative microbial risk assessment model with durability studies: Application to bacillus cereus on a courgette Puree production chain. *Risk Analysis, 33*(5), 877-892.

Ritz, M., Nauta, M. J., Teunis, P. F. M., van Leusden, F., Federighi, M., Havelaar, A. H. (2007). Modelling of *Campylobacter* survival in frozen chicken meat. *Journal of Applied Microbiology, 103*(3), 594-600.

Rosenquist, H., Nielsen, N. L., Sommer, H. M., Nørrung, B., Christensen, B. B. (2003). Quantitative risk assessment of human campylobacteriosis associated with thermophilic *Campylobacter* species in chickens. *International Journal of Food Microbiology, 83*(1), 87-103.

Ross, T. (1993). Belehradek-type models. *Journal of Industrial Microbiology, 12*(3-5), 180-189.

Ross, T., McMeekin, T. A. (1994). Predictive microbiology. *International Journal of Food Microbiology, 23*(3-4), 241-264.

Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. *Journal of Applied Microbiology, 81*(5)

Ross, T., Dalgaard, P., Tienungoon, S. (2000). Predictive modelling of the growth and survival of *Listeria* in fishery products. *International Journal of Food Microbiology, 62*(3), 231-245.

Ross, T., Ratkowsky, D. A., Mellefont, L. A., McMeekin, T. A. (2003). Modelling the effects of temperature, water activity, pH and lactic acid concentration on the growth rate of *Escherichia coli*. *International Journal of Food Microbiology, 82*(1), 33-43.

Ross, T., Rasmussen, S., Fazil, A., Paoli, G., Sumner, J. (2009). Quantitative risk assessment of listeria monocytogenes in ready-to-eat meats in australia. *International Journal of Food Microbiology, 131*(2-3), 128-137.

Rosso, L., Lobry, J. R., Bajard, S., Flandrois, J. P. (1995). Convenient model to describe the combined effects of temperature and pH on microbial growth. *Applied and Environmental Microbiology, 61*(2), 610-616.

Rosso, L., Lobry, J. R., Flandrois, J. P. (1993). An unexpected correlation between cardinal temperatures of microbial-growth highlighted by a new model. *Journal of Theoretical Biology, 162*(4), 447-463.

Salter, M.A., Ross, T., McMeekin, T.A. (1998). Applicability of a model for non-pathogenic *Escherichia coli* for predicting the growth of pathogenic *Escherichia coli*. *Journal of Applied Microbiology, 85*, 357-364.

Sanaa, M., Coroller, L., Cerf, O. (2004). Risk assessment of listeriosis linked to the consumption of two soft cheeses made from raw milk: Camembert of normandy and brie of meaux. *Risk Analysis, 24*(2), 389-399.

Schaechter, M., Maaloe, O., Kjeldgaard, N. O. (1958). Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonellatyphimurium*. *Journal of General Microbiology, 19*(3), 592-606.

Schoolfield, R. M., Sharpe, P. J. H., Magnuson, C. E. (1981). Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. *Journal of Theoretical Biology, 88*(4), 719-731.

Shorten, P. R., Pleasants, A. B., Soboleva, T. K. (2006). Estimation of microbial growth using population measurements subject to a detection limit. *International Journal of Food Microbiology, 108*(3), 369-375.

Standaert, A.R., Poschet, F., Geeraerd, A.H., Uylbak, F.V., Kreft, J.U., Van Impe, J.F. (2004). A novel class of predictive microbial growth models: implementation in an individual-based framework. *Computer Applications in Biotechnology*, 183-188.

Straver, J. M., Janssen, A. F. W., Linnemann, A. R., Boekel, M. A. J. S., Beumer, R. R., Zwietering, M. H. (2007). Number of *Salmonella* on chicken breast fillet at retail level and its implications for public health risk. *Journal of Food Protection, 70*(9), 2045-2055.

Surette, M. G., Miller, M. B., Bassler, B. L. (1999). Quorum sensing in *Escherichia coli*, *Salmonella typhimurium*, and *Vibrio harveyi*: A new family of genes responsible for autoinducer production. *Proceedings of the National Academy of Sciences of the United States of America, 96*(4), 1639-1644.

Van Boekel, M.A.J.S., Zwietering, M.H. (2007). Experimental design, data processing and model fitting in predictive microbiology. In S. Brul, S. van Gerwen, M. Zwietering (Eds.), *Modelling microorganisms in food* (22-41). Cambridge: Woodhead Publishing Ltd.

Van Gerwen, S. J. C., Zwietering, M. H. (1998). Growth and inactivation models to be used in quantitative risk assessments. *Journal of Food Protection, 61*(11), 1541-1549.

Van Gerwen, S. J. C., te Giffel, M. C., van't Riet, K., Beumer, R. R., Zwietering, M. H. (2000). Stepwise quantitative risk assessment as a tool for characterization of microbiological food safety. *Journal of Applied Microbiology, 88*(6), 938-951.

Van Hauwermeiren, M., Vose, D. (2009). A compendium of distributions. Vose Software, Ghent, Belgium. [online] Available at www.vosesoftware.com. [Accessed May 2013].

Van Impe, J. F., Poschet, F., Geeraerd, A. H., Vereecken, K. M. (2005). Towards a novel class of predictive microbial growth models. *International Journal of Food Microbiology, 100*(1-3), 97-105.

Vicari, A. S., Mokhtari, A., Morales, R. A., Jaykus, L. A., Frey, H. C., Slenning, B. D., Cowen, P. (2007). Second-order modeling of variability and uncertainty in microbial hazard characterization. *Journal of Food Protection, 70*(2), 363-372.

Vold, L., Holck, A., Wasteson, Y., Nissen, H. (2000). High levels of background flora inhibits growth of *Escherichia coli* O157:H7 in ground beef. *International Journal of Food Microbiology, 56*(2-3), 219-225.

Wiens, B. L. (1999). Statistical practice - when log-normal and gamma models give different results: A case study. American Statistician, 53(2).

Williams, M. S., Ebel, E. D. (2012a). Methods for fitting a parametric probability distribution to most probable number data. *International Journal of Food Microbiology, 157*(2), 251-258.

Williams, M. S., Ebel, E. D. (2012b). Methods for fitting the poisson-lognormal distribution to microbial testing data. *Food Control, 27*(1), 73-80.

Whiting, R.C. (1995). Microbial modeling of foods. *Critical Reviews in Food Science and Nutrition*, 35(6), 467-494.

Wilson, P. D. G., Brocklehurst, T. F., Arino, S., Thuault, D., Jakobsen, M., Lange, M., Van Impe, J. F. (2000). Modelling microbial growth in structured foods: Towards a unified approach. *International Journal of Food Microbiology, 73*(2-3), 275-290.

Zhao, Y., Frey, H. C. (2004). Quantification of variability and uncertainty for censored data sets and application to air toxic emission factors. *Risk Analysis, 24*(4), 1019-1034.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M., van't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology, 56*(6), 1875-1881.

Zwietering, M. H., Wijtzes, T., Rombouts, F. M., van't Riet, K. (1993). A decision support system for prediction of microbial spoilage in foods. *Journal of Industrial Microbiology and Biotechnology, 12*(3-5)

Zwietering, M. H., de Wit, J. C., Notermans, S. (1996). Application of predictive microbiology to estimate the number of *Bacillus cereus* in pasteurised milk at the point of consumption. *International Journal of Food Microbiology, 30*(1-2), 55-70.

Zwietering, M.H., Nauta, M.J. (2007). Predictive models in microbiological risk assessment. In S. Brul, S. van Gerwen, M. Zwietering (Eds.), *Modelling microorganisms in food* (110-125). Cambridge: Woodhead Publishing Ltd.

Zwietering, M. H. (2009). Quantitative risk assessment: Is more complex always better? simple is not stupid and complex is not always more correct. *International Journal of Food Microbiology, 134*(1), 57-62.