Politecnico di Torino

DEPARTMENT OF MATHEMATICAL SCIENCES "G.L. LAGRANGE"
Master of Science in Mathematical Engineering

Double Degree with the MSc in Indutrial and Applied Mathematics
at Eindhoven University of Technology (TU/e)

MASTER THESIS

Comparison of Parametric and Non-parametric Approaches for Accuracy of Quantitative Microbiological Methods

at Eindhoven University of Technology

Candidate:
Simone De Rosa

Supervisors:
Prof. Mauro Gasparini
Prof. Edwin R. van den Heuvel (TU/e)
Dr. Pieta C. IJzerman - Boon (TU/e)

Academic Year 2018/2019

To Francesco and Valerio,
I love you.

Abstract

Quantitative microbiological methods are aimed at counting the number of microorganisms in a sample. They are extremely important in the pharmaceutical industry to ensure drug safety: indeed, bacteria and toxins produced by microorganisms could contaminate medicines, which may harm humans if the contamination remains undetected.

Many new technologies are being developed to measure the microbial content in a sample: they are generally called rapid methods because they provide much quicker results than compendial methods, which are standardized methods and need up to 14 days to produce a result. Rapid methods need to be validated before being practically used and accuracy is one of the parameters which need to be evaluated during validation. Essentially, the accuracy of the rapid method is evaluated by comparing its expected measurement to the expected measurement of a compendial method.

This Master thesis consists of an in-depth comparison of statistical methods to assess the accuracy of a microbiological method. One approach is parametric, since it is based on the estimation of two generalized linear models, while the other one is non-parametric, namely it does not require estimation of a model. The former is referred to as model-based approach and the latter is referred to as non model-based approach.

A simulation study is performed to compare the performances of the two approaches in terms of ability to correctly assess the accuracy of the rapid method. The results show that it is not possible to definitely declare that one of the two approaches is preferable to the other one in any situations, but some interesting patterns can be derived in the performances of the two approaches.

Finally, the design used to estimate the linear models in the model-based approach is optimized to improve the performance of this approach.

Some interesting conclusions can be derived from this analysis. However, many questions remain unresolved and could be the basis for future work, especially with respect to the use of the model-based approach and the optimal design.

Keywords: Microbiological method, validation, accuracy, equivalence, count data, confidence interval, model-based approach, non model-based approach, simulation, optimal design.

Executive Summary

This Master thesis is the result of a five-month work at Eindhoven University of Technology. Previously, I did a three-month internship at the Center for Mathematical Sciences of MSD, which is one of the leading pharmaceutical companies in the world: my work during that period has also served as a preparation to deal with the topic of this thesis.

Quantitative microbiological methods are measurement methods used to count the number of microorganisms in a sample. New technologies are being proposed to quantify the microbial presence in a sample. They are generally called rapid methods because they can provide results much faster than the standardized methods currently used, called compendial methods. Before being practically used, rapid methods need to be validated, which means that their performance parameters have to satisfy specified requirements.

One of the validation parameters is accuracy, which is essentially the closeness of the test result of the method to a reference value. The accuracy of the rapid method is evaluated by comparing its results to those of a compendial method, using an equivalence test: the null hypothesis states that the ratio between the expected numbers of microorganisms counted by the two methods is not included between the equivalence bounds 0.7 and 1.3. The Two-One Sided Tests procedure is used to execute the equivalence test: it consists of rejecting the null hypothesis of non-equivalence at a significance level α if and only if the two-sided $100(1-2\alpha)\%$ confidence interval for the ratio between the two expected counts is included in the equivalence range [0.7,1.3].

Two statistical methods to assess accuracy are proposed. The *model-based* approach is based on the estimation of a generalized linear model between expected measurement and theoretical concentration of the analysed sample, for each of the two microbiological methods; once these models have been fitted, the estimated coefficients and standard errors are used to construct the confidence interval at each concentration. Two types of linear models have been considered: one is linear in the original scale, since it is represented by a straight line in a plot of expected count versus concentration; the other one is linear in the log scale, since it is represented by a straight line in a plot of log expected count versus concentration. On the other hand, the *non model-based* approach does not require estimation of a model and permits to compute the confidence interval using asymptotic statistical results: with this approach, only the repeated measurements at a specific concentration are used to build the confidence interval at that concentration.

A simulation study is performed in order to compare the performances of the two approaches in terms of ability to correctly evaluate the accuracy of the rapid method. The true expected numbers of microorganisms counted by the two microbiological methods are known in the context of this simulations study, so the decision about declaration of equivalence of each approach can be labelled as correct or wrong. In the simulations, the model-based approach is tested only using a specific design (homogeneous design) to estimate the linear models for the two microbiological methods, while the non model-based approach is tested with different experimental designs, changing each time the number of repeated measurements per concentration, to allow a more fair comparison of different scenarios.

The results of this simulation study do not permit to conclude in general that one of the two approaches is preferable to the other one. Even if the model-based approach seems to give better results, it should be considered that it is based on the assumption that a specific relation between expected measurements and concentrations exists, and this is not known to be true in reality. Even if the non model-based approach needs many repeated measurements per concentration to

obtain better results than those of the model-based approach, it is possible to think of a situation in which the non model-based approach is more advisable.

Finally, the design used to estimate the linear models for the two microbiological methods in the model-based approach is optimized in order to maximize the power of the equivalence test. The results permit to derive some features of the optimal design. In addition, when testing the model-based approach with the optimal design using simulations, the results show that the performance of the model-based approach substantially improves, especially at concentrations where the model-based approach with the homogeneous design has very poor performance.

The results of this analysis are not claimed to be valid in general and many further developments are possible, especially with respect to the importance of showing goodness of fit when using the model-based approach and with respect to the optimal design.

Acknowledgements

Thank you very much to Prof. Mauro Gasparini, for supervising this thesis from Italy and for sharing with me opportunities for my future career. Thank you for your teachings in Statistics, they were necessary for working on this topic.

Thank you very much to Prof. Edwin van den Heuvel, your supervision and your expertise during these months were invaluable for me, the suggestions you gave me during our meetings allowed me to find a solution when confusion was prevailing in my mind.

Thank you very much to Pieta IJzerman, you allowed me to join the CMS and you always gave me precious suggestions, both during my internship and while I was working on this thesis. I really appreciate what you did for me, especially how you always involved me in different activities at the company, even when I was not working there, trying to show me that what I was doing had a practical relevance.

Thank you to Erik, Jos, Michiel, Nazanin and Maryam, my colleagues during my amazing experience at MSD. I started working on this thesis while I was in the office with you and your suggestions were extremely useful. Thank you for giving me the opportunity to present my work during one of your meetings: it was challenging and the comments you made were useful to improve what I was doing.

I will never be able to express how grateful I am to my family. I left you more than five years ago, but I have always felt you close to me, despite thousands of kilometers between us. Thank you to my parents, you gave me the opportunity to study, to follow my ambitions far from home, you left me always free to make my decisions. I understand how difficult it is to let a son go away and I appreciate all the sacrifices you made for me. I owe you everything I achieved during these five years. Thanks Mamma, thanks Papà. Thank you to my brothers, Francesco and Valerio, for improving my life from the moment you were born. You can not imagine how important you are for me and how difficult it was for me being far from you during these years while you were growing up. I am sorry for leaving you in such an important period of your growth, I hope that you will always know that you can count of me, despite the distance.

Thank you to my new family in Torino: Roberto, Raffaele, Gianluca, Sergio, Francesco, Vincenzo. Thank you for teaching me that friendship can exist despite huge distances which separate us and despite not being in touch with each other for months. I know there is something that will connect us forever, wherever each of us will be in the future, and your friendship is the most important thing I achieved during these years. Thank you to Tony, who has been with me since the first day of the first year of my studies and who supported me during the defense of this thesis in Eindhoven. I know that I will always be able to count on you, thanks my friend.

Thank you to Marina, you were always willing to help me with my English when I had doubts, and your suggestions were invaluable for improving the style of my thesis.

Last, but not least, thank you Rita. Sharing my life with you is the best thing I can imagine. During these years, you were a constant and invaluable presence in my life: sharing my experiences with you allowed me to see things from a different point of view and broaden my horizons. Thank you for listening to me at any time I needed, thanks for always giving me your sincere opinion, both if you agreed with me or you did not. Your criticism allowed me to understand many times that I had to change my behaviour, especially when I was complaining about something, without appreciating the great opportunities I had. I grew up a lot with you and thanks to you, and I want to keep on sharing my way with you. Thanks for changing my life. I love you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Brief history of microbiology

Microbiology has a long and rich history, initially centered in the causes of infectious diseases but now including practical applications of the science. Historians are unsure about who made the first observations of microorganisms. In the 1670s and the decades thereafter, a Dutch merchant named Antoni van Leeuwenhoek made careful observations of microscopic organisms, which he called animalcules.

Louis Pasteur worked in the middle and late 1800s and he called attention to the importance of microorganisms in everyday life and encouraged scientists to think that bacteria could cause human illnesses. Pasteur postulated the germ theory of disease, which states that microorganisms are the causes of infectious diseases. Pasteur's attempts to prove the germ theory were unsuccessful. However, the German scientist Robert Koch provided the proof by cultivating anthrax bacteria apart from any other type of organism. He then injected pure cultures of the bacilli into mice and showed that the bacilli invariably caused anthrax.

In the late 1800s and the first decades of the 1900s, many of the microorganisms causing diseases were discovered, leading to the ability to halt epidemics by interrupting the spread of microorganisms. Then, after World War II, the antibiotics were introduced in medicine.

Work with viruses could not be effectively performed until instruments were developed to help scientists see these disease agents. In the 1940s, the electron microscope was developed and perfected. In that decade, cultivation methods for viruses were also introduced, and the knowledge of viruses developed rapidly.

Modern microbiology reaches into many fields of human endeavour, including the development of pharmaceutical products, the use of quality-control methods in food and dairy products production, the control of disease-causing microorganisms in consumable waters, and the industrial applications of microorganisms. This information and much more details about the history of microbiology and microbiological methods are described by Guardino [1].

1.2 Microbiological methods

The brief history of microbiology described in the previous section is useful to get a feel for the importance of microbiology and, as a consequence, for the importance of microbiological methods, which are measurement methods used to study microorganisms. Within pharmaceutical industry, different microbiological methods are used for different purposes. The main purpose of microbiological testing is to ensure drug safety: pathogenic bacteria, fungi and toxins produced by microorganisms are all possible contaminants of medicines, which may harm human beings when the contamination remains undetected. Microbiological methods can be essentially divided into three categories.

- **Qualitative or sterility methods:** provide a result on presence/absence of microorganisms in a sample.
- **Quantitative or enumeration methods:** provide a result on the number of microorganisms in a sample.
- **Identification methods:** provide a result on the type of microorganism (e.g., E. coli, Mycoplasma) when a microorganism is found.

This Master thesis is focused on quantitative methods only and in the remainder, the phrase *microbiological method* will always refer to a quantitative method.

Nowadays, the microbiological methods most commonly used and accepted are referred to as compendial methods: they are standardized methods provided in the pharmacopoeias, that is, they have been formally accepted and can be considered as references for any new method. Most compendial methods count colony-forming units (CFUs). A CFU is a unit used to estimate the number of culturable microorganisms in a sample, where *culturable* means that the microorganism can grow and divide in a suitable environment. This way of counting microorganisms requires a lot of time because the microorganisms need to be cultured in order to multiply and visual appearance of a colony requires a significance growth. In addition, these methods do not permit to count dead cells in the sample. Finally, there is uncertainty because a colony can arise from one cell or a group of cells, but these methods are not able to distinguish between the two cases.

Before being practically used, microbiological methods need to be validated. **Validation** is a necessary step to demonstrate that a microbiological method is fit for its intended use. In order to deem a method validated, it is necessary to show that different performance characteristics of the method satisfy specified requirements. The type of methods that are currently being validated in pharmaceutical industry are referred to as rapid microbiological methods (RMMs), since they are meant to replace the compendial microbiological methods by providing much faster results.

The European Pharmacopoeia (EP) [2] and the United States Pharmacopoeia (USP) [3] describe different types of rapid microbiological methods: they show potential for real-time or near real-time results and they may or may not use the same underlying technology of the compendial methods. Three techniques for measuring microorganisms are mentioned below.

- Methods based on bioluminescence deploy the ability of some microorganisms to release adenosine triphosphate (ATP), which is a signal of cell viability and can be detected by light emission, making a quantitative determination possible.
- Methods may also be based on vital staining: cells are stained or exhibit autofluorescence and then can be directly counted, either microscopically or instrumentally.
- Flow cytometry is a technique which allows to count microorganisms based on the wavelengths of the light scattered by the microorganisms when they are subject to a laser beam.

This means that RMMs do not necessarily obtain CFUs, like the compendial methods, and may therefore result in different counts than those obtained by the compendial method.

1.3 Validation of microbiological methods

The validation of a quantitative microbiological method consists of verifying that certain performance parameters meet the specified requirements in order to ensure that the method works properly. This Master thesis is focused on the evaluation of **accuracy**, which is one of the validation parameters. In general (not only in microbiology), accuracy is the ability of a measurement method to produce results which are close to or in agreement with known reference values [4].

In microbiology, precise reference values are not available due to the intrinsic uncertainty in the number of microorganisms present in the samples. Even if samples are created by spiking a "known" quantity of microorganisms or by diluting a sample with a specific theoretical and known concentration, the true number of microorganisms in the sample can be substantially different from

the expected or anticipated number of microorganisms. Thus, the accuracy of a RMM is evaluated by means of a comparison between its results and those produced by a compendial method [2].

Another validation parameter which has an important role in this work is **linearity**. Generally speaking, it is the ability of the method to produce results which can be expressed as a linear function of the theoretical concentration of the samples on which the method is performed. Usually, the guidelines define linearity as the ability of the method to produce results which are *proportional* to the theoretical concentration of microorganisms in the sample, but in practice also non-proportional relations are accepted to declare the method linear [2]. Linearity is not the focus of this thesis, but it plays a role because one of the described approaches to assess accuracy is based on the estimation of a linear model between expected results and theoretical concentrations.

1.3.1 Mathematical definitions of accuracy and linearity

Before proceeding with a more precise description of this work and its goals, it is appropriate to provide clear mathematical definitions of accuracy and linearity. These definitions have been formalized during my internship at MSD [5]. They are based on a study of the literature, but are not present in the literature in this format as far as it is known. In the remainder, linearity and accuracy will always be used as specified in these definitions and in the remarks reported in this subsection.

Definition 1 (Accuracy)**.** *Let Y_r and Y_c be two random variables representing the number of microorganisms counted by the rapid microbiological method to be validated and the compendial method, respectively, in samples with the same theoretical concentration X. The rapid microbiological method is defined* **accurate at concentration x** *if*

$$\mathbb{E}[Y_r|X = x] = \mathbb{E}[Y_c|X = x].$$

From a practical point of view, it is really too strict to require that the expected numbers of microorganisms counted by the two methods are exactly equal in order to declare that the rapid method is accurate. Indeed, the statistical methods described in the next chapters are all based on the following **equivalence formulation**, where d is some prescribed equivalence margin.

$$\begin{aligned} \mathrm{H}_0 &: \left| \frac{\mathbb{E}[Y_r|X = x]}{\mathbb{E}[Y_c|X = x]} - 1 \right| \geq d \\ \mathrm{H}_1 &: \left| \frac{\mathbb{E}[Y_r|X = x]}{\mathbb{E}[Y_c|X = x]} - 1 \right| < d \end{aligned} \tag{1.1}$$

The new method is declared accurate at concentration x if and only if the null hypothesis in (1.1) is rejected at a specified confidence level. Essentially, according to this formulation, the rapid method is accurate if the ratio between the expected measurements by the two methods does not differ too much from 1. The equivalence margin d will be set at 0.3 in the remainder of this thesis [6].

The **Two One-Sided Tests (TOST)** procedure described by Schuirmann [7] is used in this work to execute the equivalence test (1.1): it consists in declaring equivalence at a significance level α if and only if the $100(1 - 2\alpha)\%$ two-sided confidence interval for the ratio between the expected values is included in the interval $[1 - d, 1 + d]$.

Definition 2 (Linearity)**.** *Let Y be the number of microorganisms counted by the microbiological method to be validated in a sample with theoretical concentration X. The microbiological method is defined* **g-linear in $f(x)$** *if the relation between the expected measurement $\mathbb{E}[Y|X = x]$ and the concentration x has the form*

$$g(\mathbb{E}[Y|X = x]) = \alpha + \beta f(x)$$

where $g, f : \mathbb{R} \to \mathbb{R}$ are continuous invertible functions and $\alpha, \beta \in \mathbb{R}$.

This definition is extremely general and many relations between expected count and theoretical concentration satisfy the requirements needed to define the method linear, for appropriate choices of the functions g and f. The focus in the following chapters will mainly be on methods which are **log-linear in log (x)**: in this case

$$\log(\mathbb{E}[Y|X=x]) = \alpha + \beta \log(x) \quad (\text{i.e.,} \quad \mathbb{E}[Y|X=x] = e^{\alpha}x^{\beta}) \tag{1.2}$$

is the form of the relation between expected count of microorganisms and theoretical concentration. Moreover, **identity-linearity in x** will be considered an alternative, with expected count and theoretical concentration linked by the relation

$$\mathbb{E}[Y|X=x] = \alpha + \beta x. \tag{1.3}$$

It should be noted that both forms are strongly related to the concept of proportionality. If a concentration x is diluted with factor ρ (i.e., the new concentration is x/ρ), then

$$\mathbb{E}[Y|X=x/\rho] = \mathbb{E}[Y|X=x]/\rho$$

when $\beta = 1$ for the "log-linear in $\log(x)$" relation and when $\alpha = 0$ for the "identity-linear in x" relation. Thus, linearity is less strong than proportionality since the latter is a particular case of the former.

Essentially, a linearity study consists of estimating a model having the form reported in Definition 2 between expected number of microorganisms counted by the microbiological method and theoretical concentration. The specific form of the model to be estimated should be chosen according to a visual analysis of the data or to a specific form which is needed to be verified between measurements and concentrations. Finally, acceptable goodness of fit should be observed to declare that linearity holds.

1.3.2 Remarks about the equivalence test

Testing equivalence between the two microbiological methods by looking at the ratio between the expected numbers of microorganisms is suggested by the guidelines, which usually specify acceptance criteria in terms of percentage number of microorganisms detected by the rapid method [2, 6]. The equivalence formulation (1.1) makes no sense when $\mathbb{E}[Y_c|X=x]$ is equal to 0: this is likely to happen for $x = 0$, since usually the compendial methods are growth-based and therefore they do not detect microorganisms in blank samples. On the other hand, rapid methods may count something in blank samples, but the detection of microorganisms in blank samples by the rapid method is an issue which is already covered by the evaluation of another validation parameter, namely the **specificity**, defined as the ability of the method to quantify only the required microorganisms, i.e. does not generate false positive results [2]. Therefore, accuracy at concentration 0 is not needed.

It is appropriate to specify how the analysis in this thesis will deal with null expected values. The expected number of microorganisms counted by a microbiological method will always be assumed to have one of the forms (1.2) or (1.3) as a function of the theoretical concentration x. In the first case, the model assumes that the expected value is 0 at concentration 0, so the blank samples will not be taken into account to estimate the models for the two microbiological methods when model (1.2) will be estimated. On the other hand, when using the model (1.3), the blank samples will be used for the estimation in order to obtain a proper estimate of the intercept α. Equivalence at $x = 0$ will be assessed only when using the model (1.3), even if this evaluation concerns the specificity of the rapid method rather than its accuracy.

1.4 Problem description

This work compares essentially two different approaches to evaluate the accuracy in validation of microbiological methods. As specified in Subsection 1.3.1, it is necessary to compute a two-sided confidence interval for the ratio between the expected numbers of microorganisms counted

by the two microbiological methods. The two approaches differ because of how they lead to the computation of the confidence interval.

- **Model-based approach:** this is a parametric approach based on the estimation of a linear model between expected measurements and theoretical concentrations, for each of the two microbiological methods. Once these two models have been estimated, it is possible to build the confidence interval for the ratio between the two expected counts by using the estimates of the model coefficients and their standard errors. Thus, this approach deploys the models used to verify linearity in order to evaluate the accuracy of the rapid method at different concentrations.

- **Non model-based approach:** this is a non-parametric approach, because no parameter estimation is required. For each concentration, the measurements obtained by the two methods are used to build a confidence interval for the ratio between the expected counts at that specific concentration, so data obtained at different theoretical concentrations are not used simultaneously, contrary to the model-based approach.

Both methods have of course advantages and disadvantages. A clear advantage of the model-based approach is that it theoretically permits to evaluate accuracy also at concentrations at which no measurements have been performed, assuming that the models describe the expected numbers of microorganisms counted by the two methods correctly. Indeed, once the models between expected counts and concentrations have been estimated, the confidence interval for the ratio between the expected counts can be theoretically computed at each concentration in the analysed range. On the other hand, the non model-based approach permits to compute the confidence interval only at concentrations at which measurements are available. A drawback of the model-based approach is that it leads to confidence intervals which are reliable under the hypothesis that the relation between expected count and theoretical concentration has a specific functional form, while the non model-based approach does not depend on any particular assumptions.

One of the **goals** of this work is to find out under which conditions it is possible to say that one approach can be preferred to the other one in terms of ability to lead to the correct decision about the accuracy of the rapid method. It is possible to judge the performances of each approach because simulations are used throughout the work: in this way, the true values of the expected counts of each method are known and the conclusion to which each approach leads can be judged as correct or wrong. In case the model-based approach should be declared superior to the non model-based approach, this would mean that the evaluation of accuracy can be linked to the study of linearity, otherwise the two validation parameters should be studied independently of each other. Using a study of linearity as starting point in the assessment of accuracy is uncommon and the guidelines do not suggest any link between the two validation parameters, leaving the evaluations of accuracy and linearity independent of each other [2, 6], even though they may often be based on the same data. Another important objective is to try to understand how the performances of the model-based approach can be maximized by using an appropriate design to estimate the linear models for the rapid method and the compendial method.

1.5 Distributional assumptions and statistical framework

Quantitative microbiological methods measure the number of microorganisms in a sample, so they provide **count data**. Count data are such that the observations can take only non-negative integer values. The most immediate way of modelling count data is to use the **Poisson distribution**. The distributional assumptions and the notation that will be used in the remainder are described in this subsection.

As in Definition 1, Y_r and Y_c denote the number of microorganisms counted by the rapid method and the compendial method, respectively, in samples with the same theoretical concentration. It should be noted that the two methods do not necessarily measure the same sample, but, as it is usually the case, they can measure the concentration of microorganisms in different samples taken

from the same stock solution, thus having the same mean concentration in the stock solution. The two measurements are assumed to have the following distributions:

$$Y_r \sim Poisson(\lambda_r) \quad Y_c \sim Poisson(\lambda_c)$$

where $\lambda_r, \lambda_c > 0$ are the expected values and depend on the theoretical concentration of the samples on which the measurements are performed.

The dependencies $\lambda_r = F_r(x)$ and $\lambda_c = F_c(x)$ between the expected counts and the theoretical concentrations are not known exactly in practice. The model-based approach is based on the hypothesis that the functions F_r and F_c have a specific form and they need to be estimated, while the non model-based approach does not take these functional dependencies into account.

The Poisson distribution is a discrete probability distribution which expresses the number of events occurring in a fixed interval of time or space, under the assumption that these events occur with a constant rate and independently of the time since the last event. In addition, the Poisson distribution can also be used to model the number of events in other specified intervals such as distance, area or volume. This last usage is the most appropriate in the context of microbiological methods, since here the Poisson distribution is used to model the number of microorganisms measured in a sample. The assumption of Poisson distributed data is very common in microbiology [8, 9]. The most important features of the Poisson distribution are summarized in Table 1.1.

Table 1.1: Main features of the Poisson distribution.

$Y \sim Poisson(\lambda)$	
Support	$k \in \{0, 1, 2, \ldots\}$
Probability density function	$\mathbb{P}[Y = k] = e^{-\lambda}\frac{\lambda^k}{k!}$
Expected value	λ
Variance	λ

1.5.1 Generalized linear models

The statistical models used to link the expected number of microorganisms counted by each of the two microbiological methods to the theoretical concentration are part of the **Generalized Linear Models (GLM)** family. The basic definition of a GLM is provided in this subsection, following McCullagh and Nelder [10].

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top$ be a random vector (the response variable) with independent components and expectation $\mathbb{E}[\boldsymbol{Y}] = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^\top$. A Generalized Linear Model is a statistical model linking the expected value $\boldsymbol{\mu}$ to a linear combination of parameters $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^\top$ in the following way:

$$g(\mu_i) = (\boldsymbol{Z\beta})_i \quad \forall i = 1, \ldots, N.$$

- $\boldsymbol{Z} \in \mathbb{R}^{N \times k}$ is the model matrix: its i-th row contains the values of the k covariates $z_1, \ldots, z_k$ for the i-th component of the random vector $\boldsymbol{Y}$. The covariates are the variables used to explain the behaviour of the expectation of the response variable. $(\boldsymbol{Z\beta})_i$ denotes the i-th component of the vector $\boldsymbol{Z\beta} \in \mathbb{R}^N$.
- $g : \mathbb{R} \to \mathbb{R}$ is the link function and can be any monotonic differentiable function linking each component of the linear predictor $(\boldsymbol{Z\beta})$ to the corresponding component of the expected response.

An important feature of a GLM is that each component of the random vector $\boldsymbol{Y}$ has a distribution in the exponential family. This means that the probability density function of the generic Y_i must have the following form:

$$f_{Y_i}(y_i; \theta_i, \phi) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}$$

where $\theta_i \in \mathbb{R}$ is a parameter related to the expected value μ_i, ϕ is a a nuisance parameter related to the variance of Y_i and a, b and c are some functions of the parameters. Many common distributions are part of the exponential family (e.g., Normal, Poisson, Binomial).

Estimates for the coefficients in Generalized Linear Models are obtained by means of maximum likelihood estimation. Since the focus of this work is not on GLM, the details are not reported. The second chapter of the book written by McCullagh and Nelder [10] describes an algorithm to fit these kinds of models.

In particular, the GLM used in this work is **Poisson regression**. The Poisson regression is based on the assumption that each component Y_i follows a Poisson distribution and that a function g of its expected value can be modeled as a linear combination of unknown parameters $\beta_1, \ldots, \beta_k$, with the weights of the linear combination represented by the covariates $z_1, \ldots, z_k$:

$$g(\mathbb{E}[Y_i]) = \beta_1 z_1 + \ldots + \beta_k z_k.$$

The models actually used in this thesis depend on two parameters, and the covariates are $z_1 = 1$ and a function of the theoretical concentration x (either $z_2 = x$ or $z_2 = \log(x)$). In most cases, the following models will be considered in the next chapters:

$$\log(\lambda_r) = \delta_r + \beta_r \log(x) \quad \log(\lambda_c) = \delta_c + \beta_c \log(x)$$

where the subscripts identify the microbiological method to which the model refers (r for the rapid method and c for the compendial method). In this case the link function is $g(x) = \log(x)$. As already stated in Subsection 1.3.1, if these relations hold, the two methods are said to be log-linear in $\log(x)$; in the next chapters, these models will often be referred to as **linear in the log scale** since they are represented by a straight line in a plot of log expected counts versus log theoretical concentrations. In the original scale, the dependencies between expected counts and theoretical concentrations are represented by the following expressions:

$$\lambda_r = e^{\delta_r} x^{\beta_r} = \alpha_r x^{\beta_r} \quad \lambda_c = e^{\delta_c} x^{\beta_c} = \alpha_c x^{\beta_c}. \tag{1.4}$$

As described in Subsection 1.3.2, according to this model, λ_r and λ_c are 0 when $x = 0$: thus, in the remainder, these models will be estimated only for concentrations greater than 0, avoiding issues caused by taking the logarithm of 0.

Identity-linearity in x will also be considered as an alternative. In this case, the models for the expected counts by the two methods are

$$\lambda_r = \alpha_r + \beta_r x \quad \lambda_c = \alpha_c + \beta_c x. \tag{1.5}$$

These models will be referred to as **linear in the original scale**. In this case, the natural logarithm, which is known as the *canonical link function* for the Poisson regression, is replaced with the identity.

Both models (1.4) and (1.5) have the advantage of permitting to express the ratio between the expected values in a form which is independent of the theoretical concentration under specific conditions.

- When using linear models in the log scale, $\frac{\lambda_r}{\lambda_c} = \frac{\alpha_r}{\alpha_c} x^{\beta_r - \beta_c}$ is constant when $\beta_r = \beta_c$.
- When using linear models in the original scale, $\frac{\lambda_r}{\lambda_c} = \frac{\alpha_r + \beta_r x}{\alpha_c + \beta_c x}$ is constant when $\alpha_r = \alpha_c = 0$.

Models with these characteristics are taken into account in microbiology when the ratio between the expected values is believed to be constant over the analysed range of concentrations: some situations like these will be analysed in Chapter 3.

1.5.2 Delta method

Both the model-based approach and the non model-based approach make strong use of asymptotic distributions in order to compute the two-sided confidence interval for the ratio between the

expected numbers of microorganisms counted by the rapid method and the compendial method at a certain concentration. The main statistical result used to compute this confidence interval when using the non model-based approach is the Delta Method.

The Delta Method permits to find a convergence in distribution for a function of a random variable, which could be a scalar or a vector. In particular, the **Multivariate Delta Method**[1] is used in this work: it permits to find a convergence in distribution for a function of a random vector. In the next chapter, the Multivariate Delta Method in the form described in Theorem 1 will be used: this theorem is a particular case[2] of Theorem 8.22 in Lehmann and Casella [11]. In Theorem 1 (and in the whole thesis), $\mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ denotes a k-dimensional multivariate Normal distribution with expected value $\boldsymbol{\mu} \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$; when the subscript k is omitted, the notation refers to an univariate Normal distribution. The symbol ∇ denotes the gradient: given a function $h : \mathbb{R}^k \to \mathbb{R}$, the gradient of h is defined as

$$\nabla h(x_1, \ldots, x_k) = \left(\frac{\partial h}{\partial x_1}(x_1, \ldots, x_k), \ldots, \frac{\partial h}{\partial x_k}(x_1, \ldots, x_k) \right)^\top \in \mathbb{R}^k.$$

Theorem 1 (Multivariate Delta Method)**.** *Let* $\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_n} \in \mathbb{R}^k$ *be a sequence of random vectors such that*

$$\sqrt{n}(\boldsymbol{Y_n} - \boldsymbol{\theta}) \xrightarrow{D} \mathcal{N}_k(0, \Sigma) \quad \textit{with} \quad n \to \infty$$

for a vector of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k) \in \mathbb{R}^k$ *and a covariance matrix* $\Sigma \in \mathbb{R}^{k \times k}$*. Given a function* $h : \mathbb{R}^k \to \mathbb{R}$ *continuosly differentiable in a neighbourhood* ω *of* $\boldsymbol{\theta}$ *and such that* $\nabla h(\boldsymbol{x}) \neq \boldsymbol{0}$ *for any* $\boldsymbol{x} \in \omega$*, the following holds:*

$$\sqrt{n}(h(\boldsymbol{Y_n}) - h(\boldsymbol{\theta})) \xrightarrow{D} \mathcal{N}(0, \nabla h(\boldsymbol{\theta})^\top \cdot \Sigma \cdot \nabla h(\boldsymbol{\theta})) \quad \textit{with} \quad n \to \infty.$$

In the remainder, the Delta Method as described in Theorem 1 will be used to find a convergence in distribution for the (log) ratio between the average numbers of microorganisms counted by the two microbiological methods at a specific concentration. Thus, the number n in Theorem 1 will be the number of repeated measurements performed by each method at a specific concentration, while the number k in Theorem 1 will always be equal to 2. Indeed, the convergence in distribution appearing in the hypothesis of Theorem 1 will be the result of the Central Limit Theorem [11] applied to the vector having as components the average numbers of microorganisms counted by the two methods at a specific concentration.

1.6 General notation and assumptions

This section shows the main notation that will be used throughout the next chapters. This notation reflects the experimental setting that has been used in the simulations described in this work.

Suppose that the rapid method and the compendial method measure the number of microorganisms in n samples coming from a stock solution with theoretical concentration x. Actually, in general, there are $2n$ samples, n of which measured by the rapid method and n by the compendial method. The number n is referred to as **number of replicates**. Let $Y_1^{(r)}, \ldots, Y_n^{(r)}$ and $Y_1^{(c)}, \ldots, Y_n^{(c)}$ be the number of microorganisms counted by the rapid method and the compendial method, respectively, in the n samples each of them has to analyse. All the computations and remarks in the next chapters will be based on the following three assumptions:

1. $Y_i^{(r)} \sim Pois(\lambda_r)$ and $Y_i^{(c)} \sim Pois(\lambda_c) \quad \forall \quad i = 1, \ldots, n$;
2. $Y_i^{(r)} \perp\!\!\!\perp Y_j^{(r)}$ and $Y_i^{(c)} \perp\!\!\!\perp Y_j^{(c)} \quad \forall \quad i, j = 1, \ldots, n$ with $i \neq j$, namely the different measurements by the same method are independent of each other;

[1] In the remainder, the phrase *Delta Method* will always refer to *Multivariate Delta Method*, if not specified. There will be no confusion since the univariate version of the method will never be used in this work.

[2] Theorem 8.22 in Lehmann and Casella is more general because it deals with a function $h : \mathbb{R}^k \to \mathbb{R}^m$, with $m \geq 1$.

3. $Y_i^{(r)} \perp\!\!\!\perp Y_j^{(c)} \quad \forall \quad i, j = 1, \ldots, n$, namely all measurements by the two different methods are independent of each other.

These assumptions of independence between the different measurements by the same method or between the measurements by different methods are realistic [9] and permit to substantially simplify the computations, which even under these assumptions can lead to very complex expressions.

It is important to keep in mind that $Y_1^{(r)}, \ldots, Y_n^{(r)}$ and $Y_1^{(c)}, \ldots, Y_n^{(c)}$ are all measurements performed on samples coming from the same stock solution with theoretical concentration x. Thus, the expected values λ_r and λ_c depend on the concentration x and the assumptions 1, 2 and 3 hold for each possible value of x. The number of replicates n could in principle be different per each concentration: however, it is assumed constant when not specified.

There is one last assumption which will be used in the remainder and which is necessary for the estimation of the linear models used in the model-based approach: the measurements performed by a microbiological method at different concentrations (i.e., in samples coming from different stock solutions with different theoretical concentrations) are independent of each other.

Chapter 2

Confidence intervals for the ratio between the expected measurements by the two methods

As described in Subsection 1.3.1, the TOST procedure used for executing the equivalence test (1.1) requires the computation of a **two-sided confidence interval** for the ratio between the expected number of microorganisms counted by the rapid method and the expected number of microorganisms counted by the compendial method at a certain concentration. In this chapter, different model-based and non model-based approaches to compute a confidence interval for the ratio λ_r/λ_c are described. It should be noted that all these intervals are approximate since all the illustrated approaches are mainly based on approximate distributions derived from asymptotic results.

2.1 Non model-based approaches

This section shows three different non model-based approaches to compute the confidence interval of interest. The first one is based on a theoretical result which links the Poisson distribution to the Binomial distribution, while the others are based on the Delta Method and differ from each other because of the transformations they use.

2.1.1 Approach based on the Binomial distribution

The non model-based approach described in this subsection is referred to as **binomial approach** because it is based on the following theoretical result which links the Poisson distribution to the Binomial distribution.

Theorem 2. *Let $X_1 \sim Pois(\lambda_1)$ and $X_2 \sim Pois(\lambda_2)$ be independent Poisson distributed random variables and let $S = X_1 + X_2$ denote their sum. Then the following holds:*

$$X_i|S \sim Bin\left(S, \frac{\lambda_i}{\lambda_1 + \lambda_2}\right) \quad i = 1, 2.$$

Proof. The proof deals with the case $i = 1$; the proof for the other case is obtained by switching the subscripts 1 and 2 in the formulas.
Because of the independence between X_1 and X_2, it is possible to conclude that $S \sim Pois(\lambda_1+\lambda_2)$ [12]. Let $k, n \in \mathbb{N}, k \leq n$.

$$\mathbb{P}(X_1 = k|S = n) = \frac{\mathbb{P}(X_1 = k, S = n)}{\mathbb{P}(S = n)} = \frac{\mathbb{P}(X_1 = k, X_1 + X_2 = n)}{\mathbb{P}(S = n)} =$$

$$= \frac{\mathbb{P}(X_1 = k, X_2 = n-k)}{\mathbb{P}(S=n)} = \frac{\mathbb{P}(X_1=k)\cdot\mathbb{P}(X_2=n-k)}{\mathbb{P}(S=n)} =$$

$$= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\cdot\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\cdot\frac{n!}{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n} =$$

$$= \binom{n}{k}\cdot\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\cdot\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k} =$$

$$= \binom{n}{k}\cdot\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\cdot\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k}.$$

This is the probability density function of a Binomial random variable with parameters n and $\pi = \lambda_1/(\lambda_1+\lambda_2)$. So $X_1|S \sim Bin(S, \lambda_1/(\lambda_1+\lambda_2))$. □

Theorem 2 can be used to construct a confidence interval for $\lambda_i/(\lambda_1+\lambda_2)$, which can then be transformed into a confidence interval for λ_1/λ_2 or λ_2/λ_1.

There are different methods for determining a confidence interval for the probability of success of a Binomial random variable. In this work, the **Wilson score interval** [13] is used: it allows to find an approximate confidence interval for the probability of success based on an asymptotic approximation of the Binomial distribution by the Normal distribution.

In general, a Binomial random variable $B \sim Bin(N,q)$ can be defined as the sum of N independent random variables $Z_1,\ldots,Z_N$ identically distributed as a Bernoulli with parameter q. Then $\hat{q} = (\sum_{i=1}^{N} Z_i)/N$ is an unbiased estimator of q. The two-sided $100(1-\alpha)\%$ Wilson score confidence interval for q is given by

$$\frac{\hat{q} + \frac{z^2_{1-\alpha/2}}{2N} \pm z_{1-\alpha/2}\sqrt{\frac{z^2_{1-\alpha/2}}{4N^2} + \frac{\hat{q}}{N}(1-\hat{q})}}{1+\frac{z^2_{1-\alpha/2}}{N}} \tag{2.1}$$

where $z_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$-quantile of the Standard Normal distribution[1].

Going back to the ratio between the expected values of two Poisson random variables X_1 and X_2, the expression (2.1) can be used to build a confidence interval for $\pi = \lambda_1/(\lambda_1+\lambda_2)$, the probability of success of $X_1|S$ (notation in Theorem 2), by using the following substitutions:

$$\hat{q} = \frac{X_1}{S} \quad N = S.$$

The ratio between the two expected values can be expressed as a strictly monotonic function of π:

$$\frac{\lambda_1}{\lambda_2} = \frac{\pi}{1-\pi}.$$

Thus, if $[\pi_L, \pi_U]$ is a two-sided $100(1-\alpha)\%$ confidence interval for π, a two-sided $100(1-\alpha)\%$ confidence interval for the ratio between the two expected values λ_1 and λ_2 is

$$\left[\frac{\pi_L}{1-\pi_L}, \frac{\pi_U}{1-\pi_U}\right]. \tag{2.2}$$

It should be noted that the coverage probability remains the same because a monotonic function is used to transform the bounds of the confidence interval: the same procedure is used for instance by Breslow and Day [14].

The procedure just described can be used to find a two-sided confidence interval for the ratio λ_r/λ_c between the expected numbers of microorganisms counted by the rapid method and the

[1] $Z \sim \mathcal{N}(0,1) \Rightarrow \mathbb{P}[Z < z_{1-\alpha/2}] = 1-\alpha/2$.

compendial method at a certain theoretical concentration x. Using the notation described in Section 1.6, X_1 and X_2 used so far are respectively replaced with

$$X_r = \sum_{i=1}^{n} Y_i^{(r)} \sim Pois(n\lambda_r) \quad \text{and} \quad X_c = \sum_{i=1}^{n} Y_i^{(c)} \sim Pois(n\lambda_c),$$

namely the sums of the repeated measurements for the rapid method and the compendial method at concentration x. Performing these substitutions in (2.1) and then transforming the interval into (2.2), the following two-sided $100(1-\alpha)\%$ confidence interval for λ_r/λ_c is found: the lower confidence limit is obtained by using the minus sign in the numerator and the plus in the denominator, and vice versa for the upper bound, and z denotes the quantile $z_{1-\alpha/2}$.

$$\frac{2X_r + z^2 \pm z\sqrt{\dfrac{z^2(X_r+X_c)+4X_rX_c}{X_r+X_c}}}{2X_c + z^2 \mp z\sqrt{\dfrac{z^2(X_r+X_c)+4X_rX_c}{X_r+X_c}}} \tag{2.3}$$

2.1.2 First approach based on the Delta Method

Both this approach and the following one are based on two important asymptotic results in Statistics: the Multivariate Central Limit Theorem [11] and the Delta Method (Theorem 1). In the remainder, these two approaches will be referred to as **delta approaches**.

A confidence interval for the ratio λ_r/λ_c between the two expected counts can also be computed by applying the Delta Method to the random vector $(\overline{Y}_r, \overline{Y}_c)$, whose components are the average numbers of microorganisms counted by the rapid method and the compendial method at a specific concentration:

$$\overline{Y}_r = \frac{1}{n}\sum_{i=1}^{n} Y_i^{(r)} \quad \overline{Y}_c = \frac{1}{n}\sum_{i=1}^{n} Y_i^{(c)}.$$

Using the Central Limit Theorem and the independence between the measurements by the two methods, the following convergence in distribution can be derived:

$$\sqrt{n}\left(\begin{pmatrix}\overline{Y}_r\\ \overline{Y}_c\end{pmatrix} - \begin{pmatrix}\lambda_r\\ \lambda_c\end{pmatrix}\right) \xrightarrow{D} \mathcal{N}_2\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\lambda_r & 0\\ 0 & \lambda_c\end{pmatrix}\right). \tag{2.4}$$

Now the Delta Method with the function $h(x_1, x_2) = x_1/x_2$ can be applied to conclude that

$$\sqrt{n}\left(\frac{\overline{Y}_r}{\overline{Y}_c} - \frac{\lambda_r}{\lambda_c}\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

where σ^2 is computed as

$$\begin{aligned}\sigma^2 &= \left(\tfrac{\partial h}{\partial x_1}(\lambda_r,\lambda_c) \quad \tfrac{\partial h}{\partial x_2}(\lambda_r,\lambda_c)\right) \cdot \begin{pmatrix}\lambda_r & 0\\ 0 & \lambda_c\end{pmatrix} \cdot \begin{pmatrix}\frac{\partial h}{\partial x_1}(\lambda_r,\lambda_c)\\ \frac{\partial h}{\partial x_2}(\lambda_r,\lambda_c)\end{pmatrix} = \\ &= \begin{pmatrix}\frac{1}{\lambda_c} & -\frac{\lambda_r}{\lambda_c^2}\end{pmatrix} \cdot \begin{pmatrix}\lambda_r & 0\\ 0 & \lambda_c\end{pmatrix} \cdot \begin{pmatrix}1/\lambda_c\\ -\lambda_r/\lambda_c^2\end{pmatrix} = \frac{\lambda_r}{\lambda_c^2} + \frac{\lambda_r^2}{\lambda_c^3}.\end{aligned} \tag{2.5}$$

In conclusion, the distribution of the ratio between the two averages can be approximated by a Normal distribution with expected value λ_r/λ_c and variance σ^2/n. From this approximate distribution, a confidence interval for the ratio λ_r/λ_c can be computed. Since the standard deviation is a function of λ_r and λ_c, it is necessary to use their estimates to build the confidence interval: in particular, the averages $\overline{Y}_r$ and $\overline{Y}_c$ are used to estimate λ_r and λ_c, respectively. In conclusion, the two-sided $100(1-\alpha)\%$ confidence interval obtained by using this approach is

$$\frac{\overline{Y}_r}{\overline{Y}_c} \pm z_{1-\alpha/2}\sqrt{\frac{1}{n}\left(\frac{\overline{Y}_r}{\overline{Y}_c^2} + \frac{\overline{Y}_r^2}{\overline{Y}_c^3}\right)}. \tag{2.6}$$

It should be noted that the computed interval makes no sense when $\overline{Y}_c = 0$: in order to deal with this situation in the simulations described in Chapter 3, the confidence interval will not be computed when $\overline{Y}_c = 0$ and equivalence will be declared only if $\overline{Y}_r = 0$.

2.1.3 Second approach based on the Delta Method

In this subsection, another confidence interval for the ratio between λ_r and λ_c is proposed. The interval is obtained again by means of the Delta Method, but using the function $h(x_1, x_2) = \log(x_1/x_2) = \log(x_1) - \log(x_2)$ in order to get a confidence interval for the the logarithm of the ratio λ_r/λ_c. This is useful in order to compare this interval to the one obtained in Subsection 2.2.4 using a model-based approach: the comparison between the intervals will be described in Section 3.4.

Applying the Delta Method to the convergence in distribution (2.4) using the function h defined in this subsection, it is possible to conclude that

$$\sqrt{n}((\log(\overline{Y}_r) - \log(\overline{Y}_c)) - (\log(\lambda_r) - \log(\lambda_c))) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

where

$$\sigma^2 = \begin{pmatrix} \frac{1}{\lambda_r} & -\frac{1}{\lambda_c} \end{pmatrix} \cdot \begin{pmatrix} \lambda_r & 0 \\ 0 & \lambda_c \end{pmatrix} \cdot \begin{pmatrix} 1/\lambda_r \\ -1/\lambda_c \end{pmatrix} = \frac{1}{\lambda_r} + \frac{1}{\lambda_c}. \tag{2.7}$$

Thus, a two-sided $100(1-\alpha)\%$ confidence interval for the logarithm of the ratio between λ_r and λ_c is

$$\log(\overline{Y}_r) - \log(\overline{Y}_c) \pm z_{1-\alpha/2} \sqrt{\frac{1}{n}\left(\frac{1}{\overline{Y}_R} + \frac{1}{\overline{Y}_C}\right)}. \tag{2.8}$$

If $[L_{log}, U_{log}]$ is the interval in the logarithmic scale, the interval for the ratio is $[e^{L_{log}}, e^{U_{log}}]$: the confidence level remains the same because $f(x) = e^x$ is a strictly monotonic function. In conclusion, the two-sided $100(1-\alpha)\%$ confidence interval for λ_r/λ_c computed by using this approach is

$$\left[\frac{\overline{Y}_r}{\overline{Y}_c} \cdot e^{-z_{1-\alpha/2}\sqrt{\frac{1}{n}\left(\frac{1}{\overline{Y}_R}+\frac{1}{\overline{Y}_C}\right)}}, \frac{\overline{Y}_r}{\overline{Y}_c} \cdot e^{z_{1-\alpha/2}\sqrt{\frac{1}{n}\left(\frac{1}{\overline{Y}_R}+\frac{1}{\overline{Y}_C}\right)}}\right]. \tag{2.9}$$

Also in this case, the confidence interval makes no sense when $\overline{Y}_c = 0$: this situation will be handled in the same way as described in the previous subsection.

2.1.4 Patterns in the length of the confidence intervals

The length of a confidence interval is one of the features which is taken into account for an assessment of the quality of a confidence interval [15, 16]. This section shows how the theoretical concentration x and the number of replicates n affect the length of the confidence intervals computed by using the non model-based approaches.

In particular, Figures 2.1, 2.2 and 2.3 show how the mean lengths of the confidence intervals vary when changing the number of replicates or the theoretical concentration. The mean length has been computed over 1000 confidence intervals, each of them constructed using a different simulated dataset. The confidence intervals used for Figure 2.1 and Figure 2.3 have been computed using datasets simulated according to the models

$$\lambda_r = 1.2x \quad \lambda_c = 1.1x,$$

while the confidence intervals used for Figure 2.2 have been computed using datasets simulated according to the models

$$\lambda_r = 0.35 + 0.8x \quad \lambda_c = 0.94 + 0.7x.$$

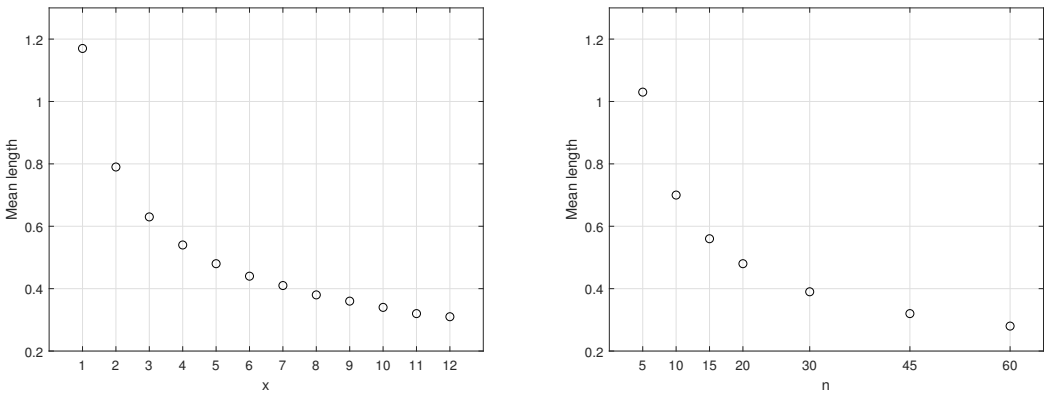
These models represent the assumed true relations between expected values and theoretical concentrations: the data used for computing the confidence intervals have been simulated as realizations of Poisson random variables with these expected values. Section 3.1 will explain in more details how the confidence intervals are computed using simulated datasets.

Clear patterns can be observed in the figures: at a certain concentration, the lengths of the intervals decrease when the number of replicates increases; for a certain number of replicates, the higher the theoretical concentration is, the shorter the confidence intervals are. There is only an exception: in the left panel of Figure 2.2, the mean length increases from $x = 0$ to $x = 1$, but the main pattern is still observable.

These patterns in the intervals computed by using the delta approaches can be explained by noting that the higher the number of replicates n is, the smaller the standard deviations of the approximate distributions used to construct these intervals are. The dashed lines in Figures 2.2 and 2.3 represent the standard deviation σ^2/n (multiplied by 2) of the approximate distribution used to compute the confidence interval, where σ^2 is given by the formula (2.5) for Figure 2.2 and by the formula (2.7) for Figure 2.3. The right panels of Figure 2.2 and Figure 2.3 show that the standard deviation and the mean length of the computed intervals have the same pattern also as functions of the concentration x.

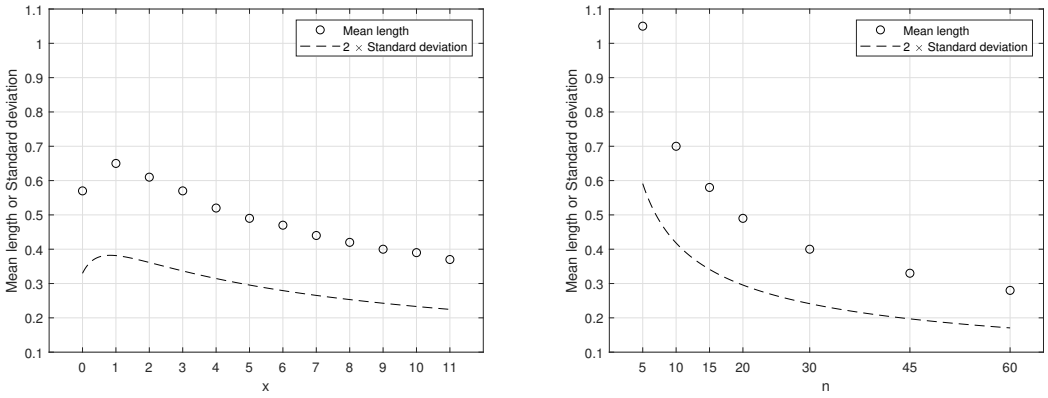
The average expected length of the Wilson score interval (2.1) is decreasing with the number of trials N [17]. The number of trials N used to derive the interval (2.3) is given by $N = X_r + X_c$ and $\mathbb{E}[N] = \mathbb{E}[X_r] + \mathbb{E}[X_c] = n\lambda_r + n\lambda_c$ is increasing both in the number of replicates and the theoretical concentration since the expected values are increasing functions of x. Thus, the observed patterns seem justified also for the binomial approach.

In this subsection, only two possible choices of true linear models have been taken into account. However, since the true expected values λ_r and λ_c will always be modelled as increasing functions of the concentration x, the observed trends should not depend on the choice of the true relations. The remarks reported in this subsection will be useful to explain some results which will be described in Chapter 3.



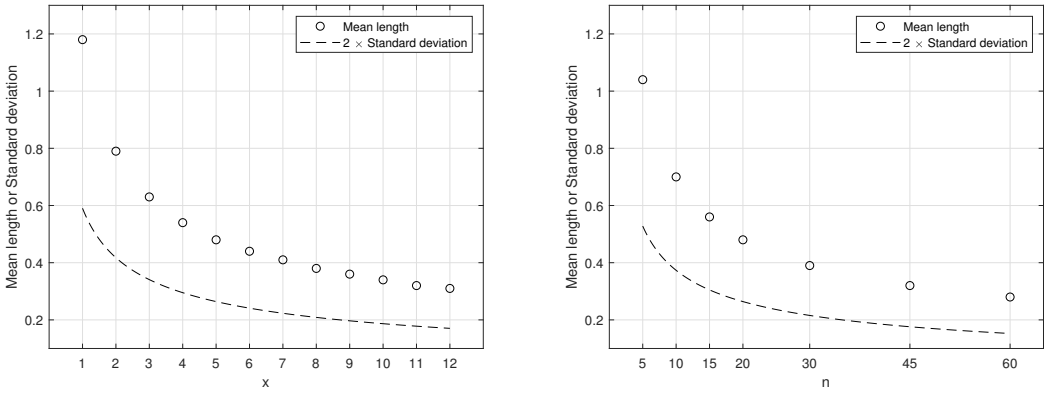
(a) $n = 20$ replicates per concentration. **(b)** Concentration $x = 5$.

Figure 2.1: Mean length of the confidence interval (2.3) computed by the binomial approach.



(a) $n = 20$ replicates per concentration. **(b)** Concentration $x = 5$.

Figure 2.2: Mean length of the confidence interval (2.6) computed by the first delta approach.



(a) $n = 20$ replicates per concentration. **(b)** Concentration $x = 5$.

Figure 2.3: Mean length of the confidence interval (2.9) computed by the second delta approach.

2.2 Model-based approaches

The confidence intervals for the ratio between the expected numbers of microorganisms counted by the two methods proposed in this section are computed using a generalized linear model estimated between expected count and theoretical concentration, both for the rapid method and the compendial method. As described in the first chapter, two possible models for linking the expected count to the theoretical concentration are considered:

- the linear model in the original scale $\mathbb{E}[Y] = \alpha + \beta x$;
- the linear model in the log scale $\mathbb{E}[Y] = \alpha x^{\beta}$.

The general idea underlying the model-based approach is the following. Once a model between expected count and theoretical concentration has been estimated for the rapid method and the compendial method, asymptotic normality of the estimated coefficients and a variation of the Delta Method are used to derive an asymptotic Normal distribution for the (log) ratio between the estimates of the two expected counts. From these asymptotic distributions, an approximate confidence interval can be immediately computed at any concentration. Before proceeding with

the description of the computations leading to the confidence intervals, the next two subsections describe the two theoretical results used to calculate the confidence intervals.

2.2.1 Approximate Normal distribution for a function of an approximately Normal random variable

Let $X \in \mathbb{R}$ be a random variable having an approximate Normal distribution with expected value μ and variance σ^2:

$$X \overset{approx.}{\sim} \mathcal{N}(\mu, \sigma^2).$$

The symbol $\overset{approx.}{\sim}$ means that the distribution of X is not exactly $\mathcal{N}(\mu, \sigma^2)$, but it can be approximated by a Normal distribution: in the remainder, the approximation will be the consequence of a convergence in distribution.

Let $h : \mathbb{R} \to \mathbb{R}$ be a function such that $h'(\mu)$ exists and $h'(\mu) \neq 0$: the intention is to derive an approximate distribution for $h(X)$. Using a Taylor expansion of $h(X)$ in a neighbourhood of $h(\mu)$ truncated to the first order, $h(X)$ can be approximated by an affine function[2] of X:

$$h(X) \approx h(\mu) + (X - \mu)h'(\mu).$$

Assuming that X is normally distributed, then also $h(\mu) + (X - \mu)h'(\mu)$ is normally distributed because it is an affine transformation of a Normal random variable, and it has the following expectation and variance [18]:

$$\mathbb{E}[h(\mu) + (X - \mu)h'(\mu)] = h(\mu) \quad Var(h(\mu) + (X - \mu)h'(\mu)) = (h'(\mu))^2\sigma^2.$$

In conclusion, the distribution of $h(X)$ can be approximated by a Normal distribution with expectation $h(\mu)$ and variance $(h'(\mu))^2\sigma^2$:

$$h(X) \overset{approx.}{\sim} \mathcal{N}(h(\mu), (h'(\mu))^2\sigma^2). \tag{2.10}$$

This approach to compute the approximate distribution of a function of a random variable has the same rationale behind the proof of the (univariate) Delta Method (Theorem 8.12 in Lehmann and Casella [11]): indeed, the proof of the Delta Method makes a similar use of the Taylor expansion of $h(X)$ truncated to the first order and then uses some asymptotic results, such as the Central Limit Theorem, to derive its conclusion.

The result (2.10) can be generalized to multivariate distributions. Suppose $\boldsymbol{X} \in \mathbb{R}^k$ is a random vector whose approximate distribution is

$$\boldsymbol{X} \overset{approx.}{\sim} \mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$$

with $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\Sigma \in \mathbb{R}^{k \times k}$ (the vectors are column vectors). Let $h : \mathbb{R}^k \to \mathbb{R}$ be a function continuously differentiable in a neighbourhood of $\boldsymbol{\mu}$ and such that $\nabla h(\boldsymbol{\mu}) \neq \mathbf{0}$. Then, the distribution of $h(\boldsymbol{X})$ can be approximated by the following Normal distribution. The similarities to the result of the Delta Method (Theorem 1) are evident.

$$h(\boldsymbol{X}) \overset{approx.}{\sim} \mathcal{N}(h(\boldsymbol{\mu}), \nabla h(\boldsymbol{\mu})^\top \cdot \Sigma \cdot \nabla h(\boldsymbol{\mu})) \tag{2.11}$$

2.2.2 Asymptotic normality of MLE in Poisson regression

As illustrated in Subsection 1.5.1, Poisson regression, either with log link function or with identity link function, is used in this work to model the relation between expected number of microorganisms counted by a microbiological method and theoretical concentration.

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top$ be the random vector with the N observations used to estimate the model. Each component Y_i has a Poisson distribution:

$$Y_i \sim Pois(\lambda_i) \quad \text{with} \quad g(\lambda_i) = \beta_0 + \beta_1 z_i \quad \forall i = 1, \ldots, N$$

[2] A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if $f(x) = Ax + b$ for a certain matrix $A \in \mathbb{R}^{m \times n}$ and a certain vector $b \in \mathbb{R}^m$.

where β_0 and β_1 are the parameters which have to be estimated ($\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$), g is the link function and the model matrix is

$$\boldsymbol{Z} = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_N \end{pmatrix} \in \mathbb{R}^{N \times 2}.$$

The log likelihood for this model is [10]

$$L_N(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{N} (y_i \log(\lambda_i) - \lambda_i - \log(y_i!))$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ is a realization of the response $\boldsymbol{Y} = (Y_1, \ldots, Y_N)^\top$. The Fisher information matrix $\mathcal{I}_N(\boldsymbol{\beta})$ is defined as the covariance matrix of the score function $\boldsymbol{s}_N(\boldsymbol{\beta})$ [19]:

$$\mathcal{I}_N(\boldsymbol{\beta}) = Cov(\boldsymbol{s}_N(\boldsymbol{\beta})) \quad \text{with} \quad \boldsymbol{s}_N(\boldsymbol{\beta}) = \left(\frac{\partial L_N}{\partial \beta_0}(\boldsymbol{\beta}; \boldsymbol{Y}), \frac{\partial L_N}{\partial \beta_1}(\boldsymbol{\beta}; \boldsymbol{Y}) \right)^\top$$

It is easy to see that $\mathcal{I}_\mathrm{N}(\boldsymbol{\beta}) = \mathbb{E}[-\mathrm{H}_\mathrm{N}(\boldsymbol{\beta}; \boldsymbol{Y})]$ [19] where $\mathrm{H}_\mathrm{N}(\boldsymbol{\beta}; \boldsymbol{Y})$ is the Hessian of the log likelihood and the expectation has to be taken over the distribution of $\boldsymbol{Y}$.

As proved by Fahrmeir and Kaufmann [19], the following convergence in distribution for the vector of coefficients' estimators $\hat{\boldsymbol{\beta}}_N$ holds: the subscript N underlines that the estimates depend on the number of observations. As described in the paper, this convergence holds for any GLM, not only for Poisson regression.

$$(\mathcal{I}_N(\boldsymbol{\beta})^{1/2})^\top (\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \xrightarrow{D} \mathcal{N}_2\left(\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \quad with \quad N \to \infty. \tag{2.12}$$

$\mathcal{I}_N(\boldsymbol{\beta})^{1/2}$ denotes the square root of $\mathcal{I}_N(\boldsymbol{\beta})$. Given a positive definite matrix A, a square root of A is a matrix, denoted as $A^{1/2}$, such that $A^{1/2}(A^{1/2})^\top = A$ [19]. In addition, if the matrix A is symmetric (it should be noted that this is the case for the Fisher information matrix), then a symmetric square root exists and is given by the spectral decomposition of A [20] after replacing the diagonal matrix containing the eigenvectors with the diagonal matrix containing the square roots of the eigenvectors. The convergence in distribution (2.12) allows to approximate - for big N - the distribution of the vector of estimated coefficients by the following Normal distribution:

$$\hat{\boldsymbol{\beta}}_N \overset{approx.}{\sim} \mathcal{N}_2(\boldsymbol{\beta}, \mathcal{I}_N(\boldsymbol{\beta})^{-1}). \tag{2.13}$$

In conclusion, the entries of the inverse of the Fisher information matrix represent the asymptotic variances and covariance of the estimated coefficients:

$$\mathcal{I}_N(\boldsymbol{\beta})^{-1} = \begin{pmatrix} \sigma^2_{\beta_0} & \sigma_{\beta_0\beta_1} \\ \sigma_{\beta_0\beta_1} & \sigma^2_{\beta_1} \end{pmatrix}.$$

In the next subsections, the number of observations is equal to $N = np$, where p is the number of concentrations used to estimate the model and n is the number of replicates at each concentration[3]. Thus, the convergence in distribution (2.12) holds in any of the following cases:

- n fixed and $p \to \infty$;
- p fixed and $n \to \infty$;
- both $n \to \infty$ and $p \to \infty$.

[3]It is assumed here that the same number of replicates is used at each concentration. In a more general situation, np should be replaced with $\sum_{i=1}^{p} n_i$, where n_i is the number of replicates used at concentration x_i to estimate the model $\forall i = 1, \ldots, p$.

2.2.3 Linear models in the original scale

In this case, the linear model in the original scale $\mathbb{E}[Y] = \alpha + \beta x$ is used to express the relation between expected number of microorganisms counted by each microbiological method and theoretical concentration. Since Y is distributed as a Poisson random variable, this model is known as Poisson regression with identity link function.

The coefficients α and β are estimated by maximizing the log likelihood function, as usual in the estimation of generalized linear models. In this case, the approximate distribution (2.13) of the coefficients' estimators becomes

$$\begin{pmatrix}\hat{\alpha}\\ \hat{\beta}\end{pmatrix} \overset{approx.}{\sim} \mathcal{N}_2\left(\begin{pmatrix}\alpha\\ \beta\end{pmatrix}, \begin{pmatrix}\sigma_\alpha^2 & \sigma_{\alpha\beta}\\ \sigma_{\alpha\beta} & \sigma_\beta^2\end{pmatrix}\right).$$

Starting from this approximate distribution and using the approach described in Subsection 2.2.1 with the function $h(a, b) = a + bx$, it is possible to find the following approximate Normal distribution for the estimated expected value at concentration x:

$$\hat{\alpha} + \hat{\beta}x \overset{approx.}{\sim} \mathcal{N}(\alpha + \beta x, \sigma_x^2) \tag{2.14}$$

where $\sigma_x^2 = \sigma_\alpha^2 + 2x\sigma_{\alpha\beta} + x^2\sigma_\beta^2$.

Now suppose that the expected numbers of microorganisms counted by the rapid method and the compendial method at concentration x are estimated as

$$\hat{\lambda}_r = \hat{\alpha}_r + \hat{\beta}_r x \quad \hat{\lambda}_c = \hat{\alpha}_c + \hat{\beta}_c x.$$

Since (2.14) holds for both methods and because of the assumption of independence between the measurements by the two methods, it is possible to derive the following approximate distribution:

$$\begin{pmatrix}\hat{\alpha}_r + \hat{\beta}_r x\\ \hat{\alpha}_c + \hat{\beta}_c x\end{pmatrix} \overset{approx.}{\sim} \mathcal{N}_2\left(\begin{pmatrix}\alpha_r + \beta_r x\\ \alpha_c + \beta_c x\end{pmatrix}, \begin{pmatrix}\sigma_{r,x}^2 & 0\\ 0 & \sigma_{c,x}^2\end{pmatrix}\right) \tag{2.15}$$

where $\sigma_{r,x}^2$ and $\sigma_{c,x}^2$ are the variances of the estimated expected counts at concentration x for the rapid method and the compendial method, respectively.

Starting from (2.15), the last step consists of using the approach described in Subsection 2.2.1 with the function $h(a, b) = a/b$ to derive an approximate Normal distribution for the ratio between the two estimated expected counts:

$$\frac{\hat{\alpha}_r + \hat{\beta}_r x}{\hat{\alpha}_c + \hat{\beta}_c x} \overset{approx.}{\sim} \mathcal{N}\left(\frac{\alpha_r + \beta_r x}{\alpha_c + \beta_c x}, \frac{\sigma_{r,x}^2}{(\alpha_c + \beta_c x)^2} + \frac{(\alpha_r + \beta_r x)^2}{(\alpha_c + \beta_c x)^4}\sigma_{c,x}^2\right). \tag{2.16}$$

A confidence interval for the ratio between the two expected values can be immediately derived using this approximate distribution. Since the coefficients $\alpha_r, \beta_r, \alpha_c, \beta_c$ appear in the expression of the standard deviation, it is necessary to replace them with their estimates to compute the confidence interval. In conclusion, the following two-sided $100(1-\alpha)\%$ confidence interval for the ratio between the two expected counts is derived.

$$\frac{\hat{\alpha}_r + \hat{\beta}_r x}{\hat{\alpha}_c + \hat{\beta}_c x} \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_{r,x}^2}{(\hat{\alpha}_c + \hat{\beta}_c x)^2} + \frac{(\hat{\alpha}_r + \hat{\beta}_r x)^2}{(\hat{\alpha}_c + \hat{\beta}_c x)^4}\sigma_{c,x}^2} \tag{2.17}$$

Figure 2.4 shows the confidence intervals computed using a dataset simulated according to the true models $\lambda_r = 0.35 + 0.8x$ and $\lambda_c = 0.94 + 0.7x$. It should be noted that the shortest confidence interval occurs at a concentration around the middle of the range ($x = 5$) and the length of the interval increases when the concentration approaches the boundaries of the range. However, this pattern is not symmetric: concentrations 0 and and 10 are equidistant from $x = 5$, but the interval at $x = 0$ is much wider than the interval at $x = 10$. This pattern is due to two main factors. First, the standard errors $\sigma_{r,x}^2$ and $\sigma_{c,x}^2$ in (2.17) are minimum at a concentration $\widetilde{x}$ towards the middle

of the range and the further from $\widetilde{x}$ the concentration is, the higher $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ are. Second, the intervals at $x = 0$ and $x = 1$ are much wider than in the rest of the range due to a combination of rather high standard errors and small estimated expected count for the compendial method $\hat{\alpha}_c + \hat{\beta}_c x$, which appears in the denominator in the length of the interval (2.17). Even if $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ are higher at concentrations towards the upper bound of the range rather than at $x = 0$ or $x = 1$, $\hat{\alpha}_c + \hat{\beta}_c x$ is much bigger at high concentrations than at $x = 0$ or $x = 1$ and this contributes to make the intervals at high concentrations much shorter than the intervals at low concentrations.

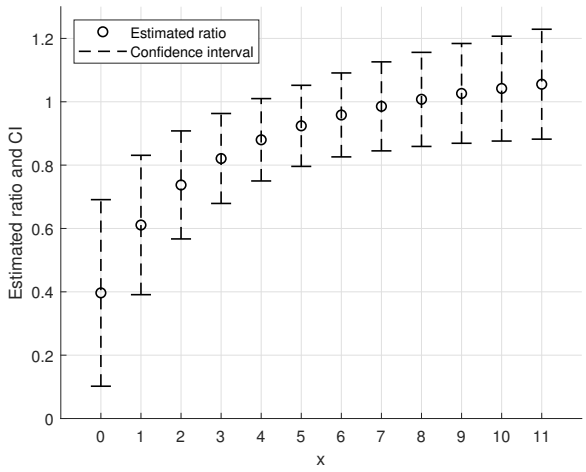


Figure 2.4: Confidence intervals obtained with the model-based approach with linear models in the original scale.

2.2.4 Linear models in the log scale

The relation $\mathbb{E}[Y] = \alpha x^\beta$ is represented by a straight line in the logarithmic scale: $\log(\mathbb{E}[Y]) = \log(\alpha) + \beta \log(x) = \delta + \beta \log(x)$. With Y distributed as a Poisson random variable, this model is known as Poisson regression with canonical (log) link function. In this case, the approximate distribution (2.13) of the coefficients' estimators becomes

$$\begin{pmatrix} \hat{\delta} \\ \hat{\beta} \end{pmatrix} \overset{approx.}{\sim} \mathcal{N}_2\left(\begin{pmatrix} \delta \\ \beta \end{pmatrix}, \begin{pmatrix} \sigma^2_\delta & \sigma_{\delta\beta} \\ \sigma_{\delta\beta} & \sigma^2_\beta \end{pmatrix} \right).$$

Starting from this approximate distribution and using the approach described in Subsection 2.2.1 with the function $h(a, b) = a + b \log(x)$, it is possible to find the following approximate Normal distribution for the estimated log expected value at concentration x:

$$\hat{\delta} + \hat{\beta} \log(x) \overset{approx.}{\sim} \mathcal{N}(\delta + \beta \log(x), \sigma^2_x) \tag{2.18}$$

where $\sigma^2_x = \sigma^2_\delta + 2 \log(x) \sigma_{\delta\beta} + \log^2(x) \sigma^2_\beta$.

Now suppose that the linear models in the log scale are estimated for the rapid method and the compendial method as

$$\widehat{\log(\lambda_r)} = \hat{\delta}_r + \hat{\beta}_r \log(x) \quad \widehat{\log(\lambda_c)} = \hat{\delta}_c + \hat{\beta}_c \log(x).$$

Since (2.18) holds for both methods and because of the assumption of independence between the measurements by the two methods, it is possible to derive the following approximate distribution:

$$\begin{pmatrix} \hat{\delta}_r + \hat{\beta}_r \log(x) \\ \hat{\delta}_c + \hat{\beta}_c \log(x) \end{pmatrix} \overset{approx.}{\sim} \mathcal{N}_2\left(\begin{pmatrix} \delta_r + \beta_r \log(x) \\ \delta_c + \beta_c \log(x) \end{pmatrix}, \begin{pmatrix} \sigma^2_{r,x} & 0 \\ 0 & \sigma^2_{c,x} \end{pmatrix} \right) \tag{2.19}$$

where $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ are the variances of the estimated log expected counts at concentration x for the rapid method and the compendial method, respectively.

Starting from (2.19), the last step consists of using the approach described in Subsection 2.2.1 with the function $h(a,b) = a - b$ to derive an approximate Normal distribution for the difference between the two estimated log expected counts:

$$(\hat{\delta}_r + \hat{\beta}_r \log(x)) - (\hat{\delta}_c + \hat{\beta}_c \log(x)) \overset{approx.}{\sim} \mathcal{N}((\delta_r + \beta_r \log(x)) - (\delta_c + \beta_c \log(x)), \sigma^2_{r,x} + \sigma^2_{c,x}). \quad (2.20)$$

The following two sided $100(1-\alpha)\%$ confidence interval for the logarithm of the ratio between the two expected values can be found using this approximate distribution.

$$(\hat{\delta}_r + \hat{\beta}_r \log(x)) - (\hat{\delta}_c + \hat{\beta}_c \log(x)) \pm z_{1-\alpha/2}\sqrt{\sigma^2_{r,x} + \sigma^2_{c,x}} \quad (2.21)$$

As already done in Subsection 2.1.3, this interval can be transformed into an interval for the ratio λ_r/λ_c by applying the function $f(x) = e^x$ to the two confidence bounds. Thus, the following two-sided $100(1-\alpha)\%$ confidence interval for λ_r/λ_c is obtained:

$$\left[\frac{e^{\hat{\delta}_r} x^{\hat{\beta}_r}}{e^{\hat{\delta}_c} x^{\hat{\beta}_c}} \cdot e^{-z_{1-\alpha/2}\sqrt{\sigma^2_{r,x}+\sigma^2_{c,x}}}, \frac{e^{\hat{\delta}_r} x^{\hat{\beta}_r}}{e^{\hat{\delta}_c} x^{\hat{\beta}_c}} \cdot e^{z_{1-\alpha/2}\sqrt{\sigma^2_{r,x}+\sigma^2_{c,x}}}\right]. \quad (2.22)$$

Figure 2.5 shows the confidence intervals computed using a dataset simulated according to the true models $\lambda_r = 1.2x$ and $\lambda_c = 1.1x$. Also in this case, the shortest confidence interval occurs at a concentration around the middle of the range ($x = 7$) and the length of the interval increases when the concentration approaches the boundaries of the range. As in the previous situation, this pattern is not symmetric: concentrations 2 and 12 are equidistant from $x = 7$, but the interval at $x = 2$ is much wider than the interval at $x = 12$. This pattern is due to the trend of the standard errors $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ in (2.22): they are minimum at a concentration towards the middle of the range, then they increase when the concentration approaches the boundaries of the range, being much higher at very low concentrations than in the rest of the range.

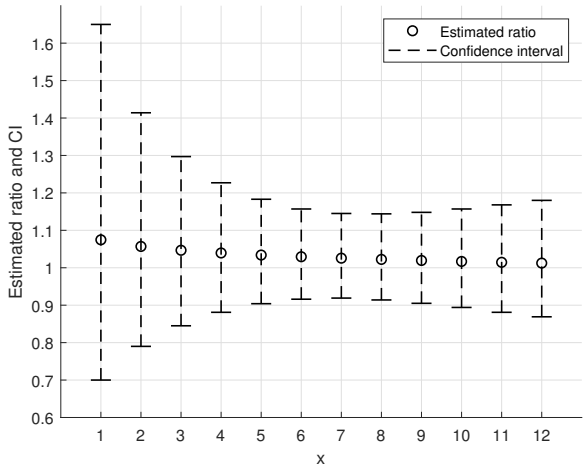


Figure 2.5: Confidence intervals obtained with the model-based approach with linear models in the log scale.

2.3 Estimation of coverage probability

This section deals with the estimation of the **coverage probability** of the confidence intervals described in this chapter. The coverage probability is defined as the probability that the confidence

interval contains the true value of the ratio between the two expected counts [21]: it can be estimated as the proportion of times that the computed confidence interval contains the true ratio. The intention is to verify that the estimated coverage probability of these intervals is actually close to the theoretical value 90% that they should have (90% is the value used in the simulation described in Chapter 3).

Given 1000 confidence intervals for the ratio λ_r/λ_c at a certain concentration, each of them computed on a different simulated dataset, suppose that S is a random variable denoting the number of times that the computed confidence interval includes the ratio between the true expected counts. Then S can be modelled as a random variable $S \sim Bin(1000, \pi)$, where π represents the probability that the true ratio is included in the confidence interval. This probability should be equal to the coverage probability of the confidence interval, set at 90%. In order to verify the coverage probability, the following procedure has been followed: 1000 datasets are simulated (see Section 3.1 for more details about the simulations), each time the confidence interval is computed at every concentration and a check is done to see if the true ratio is included in the computed confidence interval. At the end, a two-sided 95% confidence interval for π is computed[4]: if 0.90 is included in the confidence interval, the simulations support that the confidence interval has a coverage probability of 90%. This procedure used to estimate the coverage probability follows the approach described by Wicklin [22]: he actually suggests to compute the confidence interval 10000 times, but this would have required too much time.

The coverage probability has been checked both for the intervals computed using the model-based approach and the intervals computed using the non model-based approaches. Some of the configurations of linear models that will be used in Chapter 3 have been taken into account. For the model-based approach, the models for the two microbiological methods have been estimated using the homogeneous design which will be described in Section 3.1 (p concentrations, each with N/p replicates). On the other hand, the coverage probabilities of the intervals computed by the binomial approach and the delta approach using $n = 20$ replicates per concentration have been checked. When the true relations between expected count and concentration for the two microbiological methods used to simulate the datasets are linear in the original scale, the model-based approach leading to the interval (2.17) and the delta approach leading to the interval (2.6) have been used; on the other hand, the intervals (2.22) and (2.9) have been used with linear in the log scale relations. For each concentration and for each approach, the following tables show the lower and upper 95% confidence limits for the coverage probability.

The results obtained for all configurations confirm that the coverage probabilities of the computed intervals are reasonably close to their theoretical value: 0.90 is almost always included in the confidence interval for the coverage probability. This does not happen only at concentration $x = 0$ in Table 2.5, where the lower confidence limit is above 0.90, suggesting that the coverage probability is likely to be higher than its theoretical value 0.90. In conclusion, this analysis shows that the proposed approaches lead to confidence intervals with a coverage probability very close to the desired value, despite all the approximate distributions used to compute these intervals.

[4]The Wilson score interval for the binomial proportion [13] is computed.

Table 2.1: Coverage probability for the first configuration.

Configuration 1: $\lambda_r = 1.2x \quad \lambda_c = 1.1x$						
	Model		**Binomial ($n = 20$)**		**Delta ($n = 20$)**	
x	95% LCL	95% UCL	95% LCL	95% UCL	95% LCL	95% UCL
1	0.878	0.915	0.883	0.920	0.887	0.923
2	0.880	0.917	0.886	0.923	0.886	0.923
3	0.881	0.918	0.888	0.924	0.891	0.926
4	0.881	0.918	0.888	0.924	0.888	0.924
5	0.881	0.918	0.871	0.910	0.871	0.910
6	0.885	0.922	0.871	0.910	0.872	0.911
7	0.884	0.921	0.882	0.919	0.882	0.919
8	0.886	0.923	0.858	0.900	0.858	0.900
9	0.879	0.916	0.884	0.921	0.884	0.921
10	0.872	0.911	0.883	0.920	0.883	0.920
11	0.871	0.910	0.877	0.914	0.877	0.914
12	0.867	0.906	0.885	0.922	0.885	0.922

Table 2.2: Coverage probability for the second configuration.

Configuration 2: $\lambda_r = 0.8x \quad \lambda_c = 0.9x$						
	Model		**Binomial ($n = 20$)**		**Delta ($n = 20$)**	
x	95% LCL	95% UCL	95% LCL	95% UCL	95% LCL	95% UCL
1	0.880	0.917	0.870	0.909	0.883	0.920
2	0.879	0.916	0.884	0.921	0.887	0.923
3	0.884	0.921	0.890	0.925	0.891	0.926
4	0.892	0.927	0.885	0.922	0.887	0.923
5	0.887	0.923	0.869	0.908	0.869	0.908
6	0.877	0.914	0.873	0.912	0.874	0.913
7	0.870	0.909	0.864	0.903	0.864	0.903
8	0.865	0.904	0.868	0.907	0.868	0.907
9	0.862	0.901	0.881	0.918	0.882	0.919
10	0.862	0.901	0.881	0.918	0.881	0.918
11	0.868	0.907	0.899	0.933	0.899	0.933
12	0.864	0.903	0.873	0.912	0.874	0.913

Table 2.3: Coverage probability for the third configuration.

Configuration 3: $\lambda_r = 1.05x^{1.3} \quad 1.31x^{1.1}$						
	Model		**Binomial ($n = 20$)**		**Delta ($n = 20$)**	
x	95% LCL	95% UCL	95% LCL	95% UCL	95% LCL	95% UCL
1	0.864	0.903	0.883	0.920	0.891	0.926
2	0.866	0.905	0.895	0.930	0.896	0.931
3	0.865	0.904	0.893	0.928	0.894	0.929
4	0.866	0.905	0.887	0.923	0.887	0.923
5	0.865	0.904	0.879	0.916	0.879	0.916
6	0.866	0.905	0.865	0.904	0.866	0.905
7	0.877	0.914	0.865	0.904	0.868	0.907
8	0.880	0.917	0.881	0.918	0.881	0.918
9	0.885	0.922	0.871	0.910	0.872	0.911
10	0.876	0.913	0.877	0.914	0.878	0.915
11	0.871	0.910	0.897	0.932	0.897	0.932
12	0.868	0.907	0.871	0.910	0.871	0.910

Table 2.4: Coverage probability for the fourth configuration.

Configuration 4: $\lambda_r = 1.3x^{1.1} \quad 1.1x^{1.32}$						
	Model		**Binomial ($n = 20$)**		**Delta ($n = 20$)**	
x	95% LCL	95% UCL	95% LCL	95% UCL	95% LCL	95% UCL
1	0.881	0.918	0.884	0.921	0.887	0.923
2	0.874	0.913	0.893	0.928	0.894	0.929
3	0.873	0.912	0.896	0.931	0.896	0.931
4	0.879	0.916	0.883	0.920	0.884	0.921
5	0.870	0.909	0.867	0.906	0.867	0.906
6	0.869	0.908	0.864	0.903	0.865	0.904
7	0.873	0.912	0.862	0.901	0.862	0.901
8	0.874	0.913	0.890	0.925	0.893	0.928
9	0.879	0.916	0.881	0.918	0.881	0.918
10	0.874	0.913	0.882	0.919	0.882	0.919
11	0.870	0.909	0.895	0.930	0.895	0.930
12	0.867	0.906	0.876	0.913	0.876	0.913

Table 2.5: Coverage probability for the seventh configuration.

Configuration 7: $\lambda_r = 0.35 + 0.8x \quad \lambda_c = 0.94 + 0.7x$						
	Model		**Binomial ($n = 20$)**		**Delta ($n = 20$)**	
x	95% LCL	95% UCL	95% LCL	95% UCL	95% LCL	95% UCL
0	0.935	0.962	0.890	0.925	0.881	0.918
1	0.900	0.934	0.886	0.923	0.885	0.922
2	0.881	0.918	0.892	0.928	0.886	0.923
3	0.882	0.919	0.882	0.919	0.880	0.917
4	0.879	0.916	0.880	0.917	0.874	0.913
5	0.885	0.922	0.877	0.914	0.879	0.916
6	0.879	0.916	0.866	0.905	0.862	0.901
7	0.887	0.923	0.878	0.915	0.873	0.912
8	0.887	0.923	0.878	0.915	0.882	0.919
9	0.899	0.933	0.859	0.900	0.866	0.905
10	0.900	0.934	0.885	0.922	0.894	0.929
11	0.899	0.933	0.876	0.913	0.874	0.913

Chapter 3

Simulations

As described in Subsection 1.3.1, the evaluation of the accuracy of a rapid microbiological method is done by means of an equivalence test, which consists in a comparison between the results of the rapid method and those of a compendial method. Essentially, the rapid method is accurate at a specific theoretical concentration if the expected number of microorganisms counted by the rapid method at that concentration is between 70% and 130% of the expected number of microorganisms counted by the compendial method at the same concentration.

Let $Y_r \sim Poisson(\lambda_r)$ and $Y_c \sim Poisson(\lambda_c)$ denote the numbers of microorganisms counted by the rapid method and the compendial method, respectively, in two samples coming from the same stock solution with theoretical mean concentration x. The equivalence formulation (1.1) can be written in the following form.

$$\begin{aligned} \mathrm{H}_0 : &\quad \frac{\lambda_r}{\lambda_c} \leq 0.7 \quad \text{or} \quad \frac{\lambda_r}{\lambda_c} \geq 1.3 \\ \mathrm{H}_1 : &\quad 0.7 < \frac{\lambda_r}{\lambda_c} < 1.3 \end{aligned} \tag{3.1}$$

According to the TOST procedure [7], the null hypothesis of non equivalence is rejected at the significance level α if and only if the two-sided $100(1-2\alpha)\%$ confidence interval for λ_r/λ_c is included in the equivalence range $[0.7, 1.3]$. In this work, the significance level is set at $\boldsymbol{\alpha = 0.05}$, so **90% confidence intervals** for the ratio between the two expected counts are used to make the decision about the accuracy of the rapid method.

The confidence interval for the ratio λ_r/λ_c can be computed by using a model-based approach or a non model-based approach, as described in Chapter 2. This chapter describes a **simulation study** aimed at investigating the results of the model-based approach and the non model-based approach. In particular, the main intention is to compare the performances of the two approaches in terms of ability to correctly evaluate accuracy in order to at least get a feel for which of the two approaches should be preferred under specific conditions. The SAS programs used for the simulations described in this chapter are reported in Appendix C.

3.1 Description of the simulations

The numbers of microorganisms counted by the two methods are simulated from Poisson distributions with expected value depending on the theoretical concentration x through one of the following relations:

- **linear models in the log scale** $\lambda_r = \alpha_r x^{\beta_r}$ and $\lambda_c = \alpha_c x^{\beta_c}$;
- **linear models in the original scale** $\lambda_r = \alpha_r + \beta_r x$ and $\lambda_c = \alpha_c + \beta_c x$.

Choosing the coefficients of the models relating the expected count to the theoretical concentration permits to know the true ratio λ_r/λ_c at any concentration x, so the decision that the computed

confidence interval leads to take can be considered correct or wrong. In particular, according to the true values, equivalence should be declared if the ratio between the true expected values is strictly greater than 0.7 and strictly lower than 1.3.

The simulations illustrated in this chapter have been performed following the procedure described below.

1. A configuration of models and parameters is chosen. A configuration consists of a model between linear in the original scale and linear in the log scale and of the coefficients $\alpha_r, \alpha_c, \beta_r, \beta_c$ which characterise the relations for the two microbiological methods. Once this configuration has been established, the measurements performed by each method can be simulated as realizations of Poisson random variables with expected values depending on the concentration x through the chosen true models.

2. A set of concentrations $\mathcal{X} \subset \mathbb{N}$ [1] at which equivalence should be evaluated is chosen and $p = |\mathcal{X}|$ denotes the number of concentrations. In addition, a number of total available measurements per method N is fixed.

3. 1000 datasets with p concentrations and $n = N/p$ replicates per concentration are simulated. This means that each dataset contains $2N$ rows (N for the rapid method and N for the compendial method). An example dataset is shown in Figure 3.1. For each of these 1000 datasets, the model-based approach and the non model-based approaches (binomial approach and delta approach) are performed. Each time, the decision that each approach leads to take is labelled as correct or wrong and the number of correct decisions per each approach is updated.

4. For each possible value of number of replicates n equal to a factor of N and greater than N/p ($N/p < n \leq N$), 1000 datasets are simulated as before and only the non model-based approaches are performed. As in step 3, the number of correct decisions per each approach is updated after each simulation.

Obs	X	TRUE_RATIO	CORRECT_DECISION	METHOD	MEASURED
1	1	1.12500	1	1	2
2	1	1.12500	1	1	3
3	1	1.12500	1	2	1
4	1	1.12500	1	2	2
5	2	1.20575	1	1	2
6	2	1.20575	1	1	4
7	2	1.20575	1	2	4
8	2	1.20575	1	2	1
9	3	1.25564	1	1	4
10	3	1.25564	1	1	5
11	3	1.25564	1	2	2
12	3	1.25564	1	2	3

Figure 3.1: An example simulated dataset. The set of concentrations is $\mathcal{X} = \{1, 2, 3\}$ and the number of replicates per concentration is equal to 2. Method 1 is the rapid method, method 2 is the compendial method. Correct decision is 1 if the true ratio $\lambda_r/\lambda_c \in (0.7, 1.3)$, otherwise it is 0. The last column contains the simulated number of microorganisms measured by the microbiological method.

The tables reported in the following section will show the number of times (out of 1000 simulations) that each approach leads to the correct decision about declaring equivalence; for the

[1] In this thesis, $\mathbb{N}$ denotes the set of non-negative integer numbers, including 0.

non model-based approaches, the results are reported for any numbers of replicates used. As the tables will show, the non model-based approaches give terrible results when using N/p replicates per concentration. The purpose of testing the non model-based approaches with different values of number of replicates per concentration is to investigate how many samples are needed for these approaches to give results comparable to those obtained by using the model-based approach. It should be noted that the model-based approach is run only with a specific design used to estimate the linear models for the rapid method and the compendial method: in the remainder, this design will be referred to as **homogeneous design** since it consists of all p concentrations in $\mathcal{X}$, with N/p replicates per concentration.

The idea behind testing the non model-based approaches with numbers of replicates equal to a factor of the total number of experiments N is to investigate if it is possible to find a value $\widetilde{n}$ such that using the non model-based approach only on $N/\widetilde{n}$ concentrations (with the same number of replicates per concentration) permits to obtain results comparable to the model-based approach. If this is the case, then the non model-based approach could be used when accuracy needs to be evaluated only at some concentrations in the set $\mathcal{X}$.

The number of available experiments is always $\boldsymbol{N = 60}$ and the number of concentrations is always $\boldsymbol{p = 12}$. When linear models in the log scale are estimated, the set of concentration is $\mathcal{X} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, while the set is $\mathcal{X} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ when linear models in the original scale are used. The blank sample ($x = 0$) is used only when linear models in the original scale are estimated because these models do not assume that the expected number of microorganisms counted at concentration 0 is equal to 0.

Two non model-based approaches are used for each of the simulated datasets. One approach is the binomial approach leading to the interval (2.3), while the other one is one of the two delta approaches leading to the intervals (2.6) or (2.9): the former is used when the chosen true model is linear in the original scale ($\lambda = \alpha + \beta x$), while the latter is used when the true model is linear in the log scale ($\lambda = \alpha x^{\beta}$). Also the model-based approach which is used depends on the chosen true linear model: if linear models in the original scale are used, the interval (2.17) is computed, otherwise the interval (2.22) is used with linear models in the log scale.

The simulations have been performed for seven different configurations of linear models and parameters, reported in Table 3.1. In the table, $\mathcal{X}_{eq}$ denotes the subset of concentrations where equivalence is theoretically true: $\mathcal{X}_{eq} = \{x \in \mathcal{X} \quad | \quad 0.7 < \lambda_r(x)/\lambda_c(x) < 1.3\}$. Graphical representations of the models used for the simulations are shown in Appendix A. The first and the second configurations used in the simulations (see Table 3.1) are characterised by models which are linear both in the log scale and in the original scale. In these cases, linear models in the log scale are estimated because Poisson regression with canonical link function (log) seems the most appropriate choice for Poisson distributed data since it prevents the expected value from being potentially modelled as a negative number. The sixth configuration ($\lambda_r = 0.2 + 0.8x$ and $\lambda_c = 0.01 + 0.9x$) has been analysed to take into account a situation in which the rapid method counts something in the blank sample, while the compendial method does not. However, since an intercept for the compendial method exactly equal to 0 causes numerical problems in the simulations, α_c has been set equal to a very small value. The following legend applies to the tables in the next section:

- M: model-based approach;
- B: binomial approach;
- D: delta approach.
- Eq: theoretical equivalence (Y stands for "yes" and N stands for "no"), according to the ratio between the true expected counts.

Table 3.1: Configurations used in the simulations.

Configuration	Linear models	$\mathcal{X}_{eq}$	Results
1	$\lambda_r = 1.2x \quad \lambda_c = 1.1x$	$\{1, \ldots, 12\}$	Table 3.2
2	$\lambda_r = 0.8x \quad \lambda_c = 0.9x$	$\{1, \ldots, 12\}$	Table 3.3
3	$\lambda_r = 1.05x^{1.3} \quad \lambda_c = 1.31x^{1.1}$	$\{1, \ldots, 11\}$	Table 3.4
4	$\lambda_r = 1.3x^{1.1} \quad \lambda_c = 1.1x^{1.32}$	$\{1, \ldots, 10\}$	Table 3.5
5	$\lambda_r = 0.6x^{1.1} \quad \lambda_c = 0.7x^{1.1}$	$\{1, \ldots, 12\}$	Table 3.6
6	$\lambda_r = 0.2 + 0.8x \quad \lambda_c = 0.01 + 0.9x$	$\{1, \ldots, 11\}$	Table 3.7
7	$\lambda_r = 0.35 + 0.8x \quad \lambda_c = 0.94 + 0.7x$	$\{1, \ldots, 11\}$	Table 3.8

3.2 Results of the simulations

For each configuration, a table reporting the results of the performed simulations is shown on the next pages. In these tables, the blue horizontal line separates the concentrations where equivalence is theoretically true from those where the true ratio is outside the interval (0.7,1.3).

It is clear from the results of the simulations that the two non model-based approaches (binomial and delta) are essentially equivalent in terms of number of correct decisions. In the majority of cases, either they lead to the correct decision the same number of times or the differences are very small. Some more evident differences between the performances of the binomial approach and the delta approach can be observed in the results for the sixth and seventh configurations (Table 3.7 and Table 3.8). However, the highest difference is of around 70 correct decisions ($x = 10$ and $n = 15$ in Table 3.7) and the results in general do not suggest that one of the two approaches has always higher performances than the other one (the binomial approach seems superior in Table 3.7, but in Table 3.8 there are cases in which the delta approach performs better). In the sixth configuration (Table 3.7), the delta approach has much higher performances than the binomial approach at $x = 0$. This is due to numerical issues in the binomial approach caused by very small simulated measurements, but since equivalence at blank samples does not directly concerns the accuracy of the rapid method (see Subsection 1.3.2), this situation will not be investigated further.

It should be noted that when the true ratio λ_r/λ_c is outside the interval $(0.7, 1.3)$, all the approaches lead to correctly not rejecting the null hypothesis of non-equivalence in almost all of the cases. On the other hand, when the true ratio is between 0.7 and 1.3 but close to one of the two equivalence bounds (0.7 and 1.3), it is very difficult to correctly reject the null hypothesis using any approach. Examples of this situation are concentrations $x = 10$ or $x = 11$ in the third configuration (Table 3.4), $x = 9$ or $x = 10$ in the fourth configuration (Table 3.5) and $x = 1$ in the seventh configuration (Table 3.8).

Usually, the higher the **number of replicates** used at a specific concentration is, the better the performances of the non model-based approaches at that concentration are. However, when the true ratio is close to one of the two equivalence bounds 0.7 and 1.3, sometimes no improvements in the performances of the non model-based approaches are observed when increasing the number of replicates: an example is $x = 11$ in the third configuration (Table 3.4).

When the ratio between the true expected counts is constant over the analysed range of concentrations (Tables 3.2, 3.3 and 3.6), for a given number of replicates, the higher the **theoretical concentration** is, the better the performances of the non model-based approaches are. This pattern can be observed also for the sixth and the seventh configurations (Table 3.7 and Table 3.8), where the true ratio is not constant.

The results of the model-based approach follow a very specific pattern: the best performance is obtained at a certain concentration $\widetilde{x}$ towards the **middle of the range**, and the more the concentration moves away from $\widetilde{x}$, the more the performance declines. This pattern is especially clear for the first, second, fifth and sixth configurations (Tables 3.2, 3.3, 3.6 and 3.7, respectively), where the true ratio is either constant or far from the equivalence bounds 0.7 and 1.3. The best performance can be observed at a concentration between 5 and 8 and the results get worse when the theoretical concentration gets closer to the boundaries of the analysed range. The same pattern

can be observed also for the other configurations in the range of concentrations where theoretical equivalence holds. It should be noted that the decrease in performances is not always symmetric around $\widetilde{x}$: for instance, in the fifth configuration (Table 3.6), even if concentrations 4 and 12 are equidistant from $\widetilde{x} = 8$, where the best performance is observed, the numbers of correct decisions at these concentrations are substantially different (338 and 557, respectively).

The last remark concerns very **low concentrations**: the model-based approach hardly ever succeeds in correctly declaring equivalence at $x = 1$ and at $x = 2$.

Table 3.2: Results of the simulations for configuration 1 ($\lambda_r = 1.2x$ and $\lambda_c = 1.1x$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
1	Y (1.091)	0	0	0	0	0	0	0	0	0	0	0	0	0	97	93
2	Y (1.091)	3	0	0	0	0	0	0	0	0	0	0	105	100	406	406
3	Y (1.091)	243	0	0	0	0	0	0	4	3	117	107	242	241	567	567
4	Y (1.091)	465	0	0	0	0	13	12	94	91	240	235	370	370	662	662
5	Y (1.091)	615	0	0	25	19	105	100	195	191	354	354	480	480	752	752
6	Y (1.091)	750	0	0	110	105	183	183	278	277	370	370	545	545	829	829
7	Y (1.091)	822	0	0	182	180	246	243	368	366	461	459	609	609	842	842
8	Y (1.091)	820	1	1	209	205	315	315	417	417	484	484	685	685	898	898
9	Y (1.091)	767	3	1	281	280	356	352	472	472	550	550	704	704	915	915
10	Y (1.091)	702	18	15	347	343	410	410	458	458	590	590	759	759	941	941
11	Y (1.091)	642	54	51	383	382	449	449	526	524	637	636	775	775	963	963
12	Y (1.091)	581	97	91	395	395	453	453	553	553	667	667	815	815	982	982

Table 3.3: Results of the simulations for configuration 2 ($\lambda_r = 0.8x$ and $\lambda_c = 0.9x$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
1	Y (0.889)	0	0	0	0	0	0	0	0	0	0	0	0	0	3	3
2	Y (0.889)	0	0	0	0	0	0	0	0	0	0	0	2	1	363	363
3	Y (0.889)	149	0	0	0	0	0	0	0	0	6	6	200	198	621	619
4	Y (0.889)	467	0	0	0	0	0	0	7	7	134	128	400	399	751	750
5	Y (0.889)	721	0	0	0	0	7	5	98	88	236	234	508	506	853	852
6	Y (0.889)	850	0	0	3	3	56	52	182	181	407	405	618	617	901	901
7	Y (0.889)	902	0	0	39	34	145	141	285	279	459	455	700	700	940	940
8	Y (0.889)	909	0	0	116	108	214	210	396	394	555	553	781	780	946	946
9	Y (0.889)	872	0	0	170	168	302	295	460	455	625	623	807	807	967	967
10	Y (0.889)	819	0	0	254	249	356	355	523	518	681	676	837	835	985	985
11	Y (0.889)	749	1	1	319	314	439	434	596	592	745	745	861	861	989	989
12	Y (0.889)	662	3	3	365	361	498	492	619	616	768	767	895	895	994	994

Table 3.4: Results of the simulations for configuration 3 ($\lambda_r = 1.05x^{1.3}$ and $\lambda_c = 1.31x^{1.1}$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
1	Y (0.802)	0	0	0	0	0	0	0	0	0	0	0	0	0	86	86
2	Y (0.921)	84	0	0	0	0	0	0	0	0	21	20	259	254	700	700
3	Y (0.998)	491	0	0	0	0	7	5	106	99	301	290	523	522	905	905
4	Y (1.058)	647	0	0	64	60	166	161	271	267	441	440	597	597	881	881
5	Y (1.106)	668	0	0	175	172	237	234	311	310	441	441	576	576	794	794
6	Y (1.147)	621	2	2	213	212	224	222	309	309	370	370	486	485	709	709
7	Y (1.183)	518	28	26	188	186	192	192	227	227	258	258	372	372	570	570
8	Y (1.215)	377	66	66	121	121	154	154	170	170	204	204	256	255	416	416
9	Y (1.244)	204	57	57	89	89	122	122	122	122	140	140	153	153	251	251
10	Y (1.270)	110	76	75	56	56	84	84	75	74	87	87	114	114	129	129
11	Y (1.295)	57	39	39	52	52	68	68	52	52	46	46	40	40	66	66
12	N (1.318)	965	950	950	958	958	964	964	971	971	959	959	967	967	973	973

Table 3.5: Results of the simulations for configuration 4 ($\lambda_r = 1.3x^{1.1}$ and $\lambda_c = 1.1x^{1.32}$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
1	Y (1.182)	0	0	0	0	0	0	0	0	0	0	0	0	0	72	69
2	Y (1.015)	94	0	0	0	0	0	0	0	0	16	12	224	219	695	695
3	Y (0.928)	531	0	0	0	0	8	8	133	121	344	339	629	627	929	929
4	Y (0.871)	702	0	0	88	80	178	175	326	322	490	490	701	700	913	912
5	Y (0.829)	704	0	0	185	183	276	274	329	327	421	418	581	580	854	854
6	Y (0.797)	659	5	5	207	206	265	263	282	282	373	373	474	474	738	738
7	Y (0.770)	515	51	45	180	179	211	207	216	216	297	296	346	346	561	561
8	Y (0.748)	327	58	55	140	137	147	147	154	153	205	205	251	251	386	386
9	Y (0.729)	180	64	62	99	99	115	115	104	104	120	120	160	160	212	212
10	Y (0.712)	87	67	67	83	82	80	80	72	72	92	92	88	87	131	131
11	N (0.697)	948	957	957	958	958	960	961	964	964	963	963	954	954	966	967
12	N (0.684)	974	964	966	966	966	963	963	968	968	975	975	985	985	991	991

Table 3.6: Results of the simulations for configuration 5 ($\lambda_r = 0.6x^{1.1}$ and $\lambda_c = 0.7x^{1.1}$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
1	Y (0.857)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Y (0.857)	0	0	0	0	0	0	0	0	0	0	0	0	0	209	205
3	Y (0.857)	35	0	0	0	0	0	0	0	0	0	0	92	87	436	432
4	Y (0.857)	338	0	0	0	0	0	0	0	0	42	36	250	245	569	567
5	Y (0.857)	556	0	0	0	0	0	0	34	27	146	146	375	373	697	697
6	Y (0.857)	696	0	0	1	1	23	22	137	129	297	297	486	484	774	774
7	Y (0.857)	779	0	0	20	20	97	93	218	212	358	355	541	539	831	829
8	Y (0.857)	809	0	0	64	58	146	144	290	286	435	430	635	635	860	860
9	Y (0.857)	772	0	0	141	138	231	226	387	384	497	494	674	674	906	906
10	Y (0.857)	712	0	0	216	213	306	304	434	431	539	538	755	753	943	943
11	Y (0.857)	635	1	1	262	256	359	359	483	479	615	615	745	745	956	956
12	Y (0.857)	557	5	5	300	298	404	401	507	507	628	628	806	806	963	963

Table 3.7: Results of the simulations for configuration 6 ($\lambda_r = 0.2 + 0.8x$ and $\lambda_c = 0.01 + 0.9x$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
0	N (20)	1000	783	647	619	888	568	932	474	968	383	987	318	999	535	1000
1	Y (1.100)	1	0	0	0	0	0	0	0	0	0	0	0	0	28	13
2	Y (0.994)	82	0	0	0	0	0	0	0	0	0	0	10	5	464	451
3	Y (0.959)	682	0	0	0	0	0	0	0	0	13	1	245	216	735	732
4	Y (0.942)	871	0	0	0	0	0	0	8	6	137	92	449	407	880	862
5	Y (0.931)	901	0	0	0	0	1	2	81	72	313	288	571	556	935	916
6	Y (0.924)	899	0	0	7	1	81	37	220	171	425	394	697	667	965	951
7	Y (0.919)	877	0	0	55	25	173	105	341	286	535	483	782	766	975	969
8	Y (0.915)	857	0	0	122	104	270	219	431	378	627	578	820	794	979	977
9	Y (0.912)	842	0	0	222	181	349	293	503	450	676	629	891	863	992	990
10	Y (0.910)	820	0	0	289	235	385	364	601	526	751	720	884	864	995	995
11	Y (0.908)	797	0	0	341	291	469	414	624	570	777	727	921	899	997	996

Table 3.8: Results of the simulations for configuration 7 ($\lambda_r = 0.35 + 0.8x$ and $\lambda_c = 0.94 + 0.7x$).

		Number of correct decisions (out of 1000)														
		$n = 5$			$n = 10$		$n = 12$		$n = 15$		$n = 20$		$n = 30$		$n = 60$	
x	Eq (λ_r/λ_c)	M	B	D	B	D	B	D	B	D	B	D	B	D	B	D
0	N (0.372)	1000	994	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
1	Y (0.701)	0	0	0	0	0	0	0	0	0	0	0	0	0	33	21
2	Y (0.833)	214	0	0	0	0	0	0	0	0	0	0	70	50	351	279
3	Y (0.905)	704	0	0	0	0	0	0	0	0	33	12	275	229	726	683
4	Y (0.949)	914	0	0	0	0	0	0	14	9	166	117	479	438	894	881
5	Y (0.980)	954	0	0	0	0	7	1	95	63	300	294	583	579	933	944
6	Y (1.002)	936	0	0	4	2	77	39	211	194	394	400	675	695	924	939
7	Y (1.019)	902	0	0	38	19	158	108	292	288	481	483	715	732	953	962
8	Y (1.032)	848	0	0	98	70	233	228	328	335	545	571	734	761	948	959
9	Y (1.043)	815	0	0	177	148	255	262	411	428	550	583	710	752	955	965
10	Y (1.052)	779	0	0	181	207	319	329	435	461	600	631	748	783	951	962
11	Y (1.059)	740	0	0	259	249	374	397	473	511	608	647	755	785	951	957

3.3 Discussion of the results

This section is divided into two parts: first, an explanation of the results highlighted in Section 3.2 is provided; after that, a comparison between the model-based approach and the non model-based approach is discussed.

3.3.1 Explanation of the results

1. The extremely good performances of all the approaches when equivalence is not theoretically true are due to the fact that the probability of rejecting a true null hypothesis has been set equal to 0.05. Thus, when the true ratio is below 0.7 (resp. above 1.3), it is almost sure that the lower (resp. upper) confidence bound is below 0.7 (resp. above 1.3).

2. All the approaches have bad results at concentrations where the true ratio is close to one of the two equivalence bounds 0.7 and 1.3 because the confidence interval in these cases should be very small in order to entirely fall in the equivalence range.

3. The performances of the non model-based approaches at a certain concentration are strongly affected by the number of replicates. This is due to the effect of the number of replicates on the length of the intervals computed by using the binomial approach and the delta approach, as discussed in Subsection 2.1.4: the more replicates are used, the shorter the computed interval is. The length of the interval is of course a factor which affects the number of correct decisions: when equivalence is theoretically true, the smaller the interval is, the more likely to be included in the equivalence range it is. However, when the true ratio is close to 0.7 or 1.3, the performances of the non model-based approaches do not improve when increasing the number of replicates: in these cases, the differences in the performances for the different values of n are extremely low and more due to randomness than to changes in the number of replicates; more than 60 replicates would be needed in these cases to substantially improve the performances. The length of the intervals constructed by the non model-based approaches is also affected by the theoretical concentration, as described in Subsection 2.1.4: however, the effect of the theoretical concentration (the higher x, the better the performances) can not be observed when the true ratio approaches 0.7 or 1.3 because the ratio mainly affects the number of correct decisions in these situations.

4. Because of the bad performances due to a true ratio close to 0.7 or 1.3, it would be reasonable to observe that, for a given number of replicates, the closer to 1 the true ratio at a certain concentration x is, the better the performances of the non model-based approaches at that concentration are. However, this pattern can not be observed because the magnitude of the concentration must be taken into account too. For instance, in the third configuration (Table 3.4), only with $n = 60$ the best performances are obtained at $x = 3$, where the closest to 1 true ratio occurs. Using less replicates, the best results are observed at concentrations where the true ratio is further from 1 (e.g., the best performances are at $x = 4$ when $n = 30$). This is due to the effect of the concentration on the performances of the non model-based approaches. If the true ratio remains reasonably close to 1, namely approximately between 0.9 and 1.1, the number of correct decisions, for a given n, is mainly affected by the theoretical concentration rather than by the value of the ratio (see for example Table 3.7), at least up to a certain number of replicates (e.g., when using $n = 60$ in the third configuration, the best results are obtained at $x = 3$).

5. The pattern in the performances of the model-based approach is very similar to the trend of the lengths of the confidence intervals described in Section 2.2: the shortest interval occurs at a concentration $\widetilde{x}$ towards the middle of the range and the more the concentration moves away from $\widetilde{x}$, the more the length increases. However, this pattern in the lengths of the intervals is not symmetric around $\widetilde{x}$, and the same happens for the trend in the performances. In particular, the intervals are much wider at low concentrations than in the rest of the range and this is reflected in the poor performances of the model-based approach at low concentrations.

3.3.2 Comparison between the two approaches

The main purpose of this simulation study is to try to understand which approach (model-based or non model-based) is preferable in terms of ability to correctly evaluate the accuracy of the rapid method, at least under certain conditions. In the simulations described in this chapter, the model-based approach has been performed only using the homogeneous design to estimate the linear models for the two microbiological methods, while the non model-based approach has been tested with different values of number of replicates per concentration.

In the majority of cases, 30 or 60 replicates are needed for the non model-based approach to have performances comparable to those of the model-based approach. This could lead to the conclusion that the model-based approach should be preferred, but it is necessary to consider that the model-based approach is based on the assumption that the relation between expected count and concentration has a certain form. In this simulation study, this assumption holds because of how the measurements are simulated. However, in practice the true relation between expected counts and concentrations is not known. Since the non model-based approach is not based on such an assumption, it should be investigated if there are conditions in which the non model-based approach can be preferred to the model-based approach.

Suppose we are willing to believe that if accuracy holds at two concentrations x_L and x_U, then the rapid method is accurate at all concentrations between x_L and x_U. This could be a reasonable assumption, based for instance on the results of other analyses performed on the microbiological method, and can be considered true especially if the concentrations at which accuracy is evaluated are not too far from each other. It should be noted that this is the case in the described simulations, where a range of concentrations from 0 to 12 is taken into account, with two consecutive concentrations being two consecutive integer numbers. If we consider the first configuration (Table 3.2), we could choose $x_L \in \{2, 3\}$ and $x_U \in \{10, 11, 12\}$ and use the non model-based approach with 30 replicates for each of the two chosen concentrations: indeed, with 30 replicates, the non model-based approach gives better results than the model-based approach at these concentrations (or at least results very close to those of the model-based approach). Moreover, the non model-based approach has higher performance than the model-based approach at concentration 12 already with $n = 20$: choosing $x_U = 20$ and using imbalanced numbers of replicates per concentration would probably permit to further increase the performance of the non model-based approach at x_L (we may use $n = 20$ at $x_U = 12$ and $n = 40$ at $x_L = 2$ or $x_L = 3$). These remarks hold for many of the analysed configurations.

One of the results discussed in the previous subsection is the effect of the theoretical concentration on the performance of the non model-based approach: the higher the concentration is, the shorter the confidence interval at that concentration is and so the better the performance at that concentration is (unless there are issues due to the value of the true ratio). This could suggest that a comparison between the two approaches may give different results if performed on a range with higher concentrations than those used in the simulations. In order to investigate this, Table 3.9 reports the results of the simulations for the first configuration of models and coefficients, using as set of concentrations $\mathcal{X} = \{10, 20, 30, \ldots, 100, 110, 120\}$, containing the concentrations previously used multiplied by 10 (only the binomial approach has been tested, since the results of the delta approach are almost equivalent, as previously observed). The results show that a number of replicates between 10 and 15 is enough to obtain almost equal performances using the model-based approach or the non model-based approach at concentrations greater than 60. At lower concentrations, more replicates are needed for the non model-based approach to produce results comparable to those of the model-based approach: however, the performances are quite similar already with $n = 20$. These remarks permit to underline that the results discussed in this chapter should be considered valid when dealing with low concentrations, approximately in a range from 0 to 20; when dealing with higher values, substantial improvements can be observed in the performances of the non model-based approach.

Table 3.9: Results of the simulations for configuration 1 ($\lambda_r = 1.2x$ and $\lambda_c = 1.1x$) with $\mathcal{X} = \{10, 20, \ldots, 120\}$.

		Number of correct decisions (out of 1000)							
		$n = 5$		$n = 10$	$n = 12$	$n = 15$	$n = 20$	$n = 30$	$n = 60$
x	Eq (λ_r/λ_c)	M	B	B	B	B	B	B	B
10	Y (1.091)	607	26	338	403	477	609	743	944
20	Y (1.091)	868	351	607	654	743	828	942	1000
30	Y (1.091)	987	482	753	810	899	949	987	1000
40	Y (1.091)	999	579	853	899	955	984	998	1000
50	Y (1.091)	1000	700	907	952	979	993	999	1000
60	Y (1.091)	1000	746	944	973	994	999	1000	1000
70	Y (1.091)	1000	788	974	983	996	999	1000	1000
80	Y (1.091)	1000	860	980	992	1000	999	1000	1000
90	Y (1.091)	1000	882	989	997	1000	1000	1000	1000
100	Y (1.091)	1000	904	994	998	1000	1000	1000	1000
110	Y (1.091)	1000	921	994	999	1000	1000	1000	1000
120	Y (1.091)	1000	945	1000	999	1000	1000	1000	1000

3.4 Sample size calculation for the delta approach

The confidence intervals (2.6) and (2.8) obtained by using the non model-based approach based on the delta method have the following form, in the original scale or in the log scale, respectively:

$$\hat{T} \pm z_{1-\alpha/2}\sqrt{\sigma^2} \tag{3.2}$$

where $\hat{T}$ is an estimate of the (log) ratio and $\sqrt{\sigma^2}$ is its standard error. The intervals (2.17) and (2.21) computed by the model-based approaches also appear in the form (3.2), the former in the original scale and the latter in the log scale. A comparison between the **lengths** of the intervals computed by the model-based approach and the delta approach can give an estimate of the number of replicates necessary for the delta approach to produce a confidence interval shorter than the interval computed by the model-based approach. The intention of this analysis is to check if the results are in agreement with what has been observed in the results of the simulations.

For these computations, it is assumed that the models between expected count and theoretical concentration for the two microbiological methods have been estimated, so the standard errors and the estimated coefficients used in (2.17) and (2.21) are known. In particular, the models have been estimated using the homogeneous design with p concentrations and N/p replicates per concentration used in the previous simulations.

When the linear model in the log scale $\lambda = \alpha x^{\beta}$ is used, the comparison between the two intervals is performed in the log scale, where both intervals are in the form (3.2).

- Model-based approach: $(\hat{\delta}_r + \hat{\beta}_r \log(x)) - (\hat{\delta}_c + \hat{\beta}_c \log(x)) \pm z_{1-\alpha/2}\sqrt{\sigma^2_{r,x} + \sigma^2_{c,x}}$.
- Delta approach: $\log(\overline{Y}_r) - \log(\overline{Y}_c) \pm z_{1-\alpha/2}\sqrt{\frac{1}{n}\left(\frac{1}{\overline{Y}_R} + \frac{1}{\overline{Y}_C}\right)}$.

It should be noted that the average measurements $\overline{Y}_r$ and $\overline{Y}_c$ in the standard error in the interval computed by the delta approach have been used to estimate the expected values λ_r and λ_c, respectively, as explained in Subsection 2.1.2. In the following computations for the minimum sample size, the true expected values are used, since they are known in the context of this simulation study. In order to obtain a shorter confidence interval by using the delta approach rather than by using the model-based approach, the number of replicates n must be such that

$$\sqrt{\frac{1}{n}\left(\frac{1}{\lambda_r} + \frac{1}{\lambda_c}\right)} \leq \sqrt{\sigma^2_{r,x} + \sigma^2_{c,x}}$$

and the following condition on the number of replicates is found:

$$n \geq \frac{1}{\sigma_{r,x}^2 + \sigma_{c,x}^2}\left(\frac{1}{\lambda_r} + \frac{1}{\lambda_c}\right).$$

Thus, the minimum number of replicates n_{min} such that the previous condition is satisfied is given by the following expression, where $\lceil t \rceil$ denotes the rounding of t to the first integer greater than t.

$$n_{min} = \left\lceil \frac{1}{\sigma_{r,x}^2 + \sigma_{c,x}^2}\left(\frac{1}{\lambda_r} + \frac{1}{\lambda_c}\right) \right\rceil \tag{3.3}$$

When the linear model in the original scale $\lambda = \alpha + \beta x$ is used, the comparison between the two intervals is performed in the original scale.

- Model-based approach: $\dfrac{\hat{\alpha}_r + \hat{\beta}_r x}{\hat{\alpha}_c + \hat{\beta}_c x} \pm z_{1-\alpha/2}\sqrt{\dfrac{\sigma_{r,x}^2}{(\hat{\alpha}_c + \hat{\beta}_c x)^2} + \dfrac{(\hat{\alpha}_r + \hat{\beta}_r x)^2}{(\hat{\alpha}_c + \hat{\beta}_c x)^4}\sigma_{c,x}^2}$.
- Delta approach: $\dfrac{\overline{Y}_r}{\overline{Y}_c} \pm z_{1-\alpha/2}\sqrt{\dfrac{1}{n}\left(\dfrac{\overline{Y}_r}{\overline{Y}_c^2} + \dfrac{\overline{Y}_r^2}{\overline{Y}_c^3}\right)}$.

Also in this case, the average counts appearing in the standard error of the interval computed by the delta approach are replaced with the true expected counts in order to compute the minimum sample size. The condition to be satisfied in order to obtain a shorter confidence interval by using the delta approach rather than by using the model-based approach is

$$\sqrt{\frac{1}{n}\left(\frac{\lambda_r}{\lambda_c^2} + \frac{\lambda_r^2}{\lambda_c^3}\right)} \leq \sqrt{\frac{\sigma_{r,x}^2}{(\hat{\alpha}_c + \hat{\beta}_c x)^2} + \frac{(\hat{\alpha}_r + \hat{\beta}_r x)^2}{(\hat{\alpha}_c + \hat{\beta}_c x)^4}\sigma_{c,x}^2}$$

which becomes the following condition for the number of replicates:

$$n \geq \frac{\lambda_r/\lambda_c^2 + \lambda_r^2/\lambda_c^3}{\frac{\sigma_{r,x}^2}{(\hat{\alpha}_c+\hat{\beta}_c x)^2} + \frac{(\hat{\alpha}_r+\hat{\beta}_r x)^2}{(\hat{\alpha}_c+\hat{\beta}_c x)^4}\sigma_{c,x}^2}.$$

Thus, the minimum number of replicates n_{min} such that the previous condition is satisfied is given by

$$n_{min} = \left\lceil \frac{\lambda_r/\lambda_c^2 + \lambda_r^2/\lambda_c^3}{\frac{\sigma_{r,x}^2}{(\hat{\alpha}_c+\hat{\beta}_c x)^2} + \frac{(\hat{\alpha}_r+\hat{\beta}_r x)^2}{(\hat{\alpha}_c+\hat{\beta}_c x)^4}\sigma_{c,x}^2} \right\rceil. \tag{3.4}$$

Table 3.10 contains the values of the minimum number of replicates n_{min} per each value of the theoretical concentration for each of the seven configurations used in the simulations. The intention is to see if there is correspondence between the theoretical value of the minimum sample size and the minimum number of replicates needed for the non model-based approach to obtain results comparable to the model-based approach in the simulations.

The results show that the minimum sample size is maximum at a specific concentration around the middle of the range and decreases for values x approaching the boundaries of the range. This trend is in agreement with what has been observed in the results of the simulations: in order to get results comparable to those of the model-based approach, the non model-based approach needs a number of replicates which is lower for high concentrations and higher for concentrations in the middle of the range, while it is difficult to compare the performances for low concentrations because both approaches give bad results. This trend is observed especially for the configurations with a constant true ratio (configurations 1, 2, 5) or with a variable true ratio which remains far from 0.7 or 1.3 (configuration 6). When the true ratio changes and approaches 0.7 or 1.3, it is more difficult to compare the results in Table 3.10 to the results of the simulations because the true ratio substantially affects the number of correct decisions.

Table 3.10: Minimum sample size to achieve the same length of the confidence interval with the delta approach and the model-based approach for each of the seven configurations when $N = 60$.

x	Conf. 1	Conf. 2	Conf. 3	Conf. 4	Conf. 5	Conf. 6	Conf. 7
0	-	-	-	-	-	-	17
1	26	26	30	30	28	8	41
2	28	29	28	28	28	28	61
3	35	35	32	31	33	53	76
4	43	44	38	38	40	68	80
5	53	53	47	46	49	68	73
6	58	58	53	53	55	60	61
7	56	56	54	55	55	51	49
8	47	47	48	48	47	43	40
9	37	37	38	38	37	37	33
10	29	29	28	29	28	32	27
11	22	22	21	22	21	28	23
12	17	17	16	16	16	-	-

It should be noted that in most cases more than 30 replicates are needed for the delta approach to give a shorter interval than that computed by the model-based approach. This result seems to confirm that, with 60 available experiments, the non model-based approach can give results comparable to those of the model-based approach only if used at 2 concentrations, as discussed in Subsection 3.3.2.

The results in Table 3.10 may suggest that if the number of total experiments N increases, the non model-based approach could be used at more concentrations and provide better performances than the model-based approach. For instance, if $N = 120$, looking at the results in Table 3.10 for the first configuration, the non model-based approach could be used at concentrations 1, 6 and 11 with 30, 60 and 30 replicates, respectively, and provide shorter intervals than those constructed by the model-based approach. However, it should be noted that the standard errors of the estimated expected counts at a certain concentration $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ in (3.3) and (3.4) depend on N, which is the number of observations used to estimate the linear models for the two microbiological methods. Thus, if $N = 120$ and the homogeneous design is used, the linear models are estimated using 12 concentrations and 10 replicates per concentration. As a consequence, the minimum sample size should be computed again after estimating the models using 120 observations. Table 3.11 shows that n_{min} substantially increases at all concentrations when the linear models are estimated with 120 observations (only the first configuration is analysed). The results suggest that, as in the previous case, the non model-based approach can give better results than the model-based approach when used at only 2 concentrations, located towards the boundaries of the range.

Table 3.11: Minimum sample size to achieve the same length of the confidence interval with the delta approach and the model-based approach for configuration 1 when $N = 120$.

x	1	2	3	4	5	6	7	8	9	10	11	12
n_{min}	45	48	59	75	93	107	108	94	75	58	44	34

3.5 Main conclusions

The simulations illustrated in this chapter have been performed to investigate the performances of the model-based approach and the non model-based approach described in Chapter 2 in terms of number of correct decisions about declaration of equivalence between the rapid method and the compendial method at a specific theoretical concentration. In this section, no distinction

among the different non model-based approaches is specified since they have shown approximately equivalent performances. The model-based approach has been tested only using a homogeneous design to estimate the linear models for the two microbiological methods, while the non model-based approach has been used with different values of number of replicates per concentration. The following list summarizes the main patterns observed in the results of the simulations.

- Both approaches correctly lead to not declare accuracy when theoretical equivalence does not hold. On the other hand, issues in declaration of equivalence using any approaches occur when equivalence holds but the true ratio is very close to one of the two equivalence bounds 0.7 and 1.3.

- The performances of both approaches are poor at low concentrations when equivalence is theoretically true: at $x = 1$ and $x = 2$, the model-based approach hardly ever leads to the correct decision, while the non model-based approach needs 30 or 60 replicates to obtain a non-negligible percentage of correct declarations of equivalence.

- The performances of the model-based approach follow a specific pattern: the number of correct decisions is maximum at a certain concentration $\widetilde{x}$ towards the middle of the range and it decreases when the concentration moves away from $\widetilde{x}$.

- In general, the higher the number of replicates used at a certain concentration is, the better the performance of the non model-based approach at that concentration is. For a given number of replicates, the higher the theoretical concentration is, the better the performance of the non model-based approach is.

As far as the comparison between the two approaches is concerned, in most cases the non model-based approach needs 30 or 60 replicates to obtain results comparable to those of the model-based approach: this has been observed in the results of the simulations and confirmed by the calculation of the minimum sample size described in Section 3.4. However, this does not mean that the model-based approach should be considered superior to the non model-based approach, because the model-based approach is based on the assumption that the relation between the expected count and the theoretical concentration has a certain functional form, and this assumption is not known to be true in a real validation study. Thus, in Subsection 3.3.2, a situation in which the non model-based approach could be preferred to the model-based approach has been described: under the assumption that if accuracy holds at two concentrations x_L and x_U then it holds at all concentrations between x_L and x_U, the non model-based approach could be used just at these two concentrations to assess accuracy. The results of the simulations have shown that the non model-based approach with $n = 30$ performs better than the model-based approach at concentrations towards the boundaries of the analysed range of concentrations, thus x_L and x_U could be chosen in this way.

The remarks reported in Subsection 3.3.2 have highlighted that the conclusions about the comparison between the two approaches should be carefully considered to be valid only with concentrations in a range approximately between 0 and 20. Indeed, when dealing with higher concentrations, the performances of the non model-based approach substantially improve and are comparable to those of the model-based approach even with numbers of replicates per concentration between 10 and 15.

It should be always taken into account that the model-based approach in this simulation study has been tested only with a homogeneous design used to estimate the linear models for the two microbiological methods, while the non model-based approach has been tested with different values of number of replicates per concentration. Both the results of the simulations and the computation of the minimum sample size have shown that the concentrations where the model-based approach has the worst performances are located towards the boundaries of the analysed range: these are the concentrations where the non model-based approach has better results even with 30 or less replicates. This suggests that it would be appropriate to investigate how to improve

the performances of the model-based approach, especially at concentrations where the non model-based approach could be preferred. Chapter 4 will focus on optimizing the design used to estimate the linear models in order to improve the performance of the model-based approach.

Chapter 4

Optimal design for the model-based approach

The simulations described in the previous chapter have been performed to investigate the performance of each approach in terms of ability to correctly evaluate equivalence between the rapid method and the compendial method at a certain concentration. More specifically, the interest is in computing the **statistical power** of the equivalence test, that is, the probability of correctly rejecting the null hypothesis of non-equivalence. Indeed, problems in the evaluation of accuracy arise when equivalence is theoretically true, namely the ratio between the true expected counts is included between the equivalence bounds 0.7 and 1.3. On the other hand, when the null hypothesis is not true (i.e., there is no equivalence), both approaches correctly lead to not reject the null hypothesis.

The simulations have allowed to investigate how the performance of the non model-based approach changes for different values of number of replicates, while the model-based approach has been performed only once for each configuration, using a homogeneous design made of p concentrations and N/p replicates per concentration to estimate the linear models for the two microbiological methods. In this chapter, the objective is to find out how the performance of the model-based approach can improve by changing the design used to estimate the linear models for the rapid method and the compendial method. In particular, the **optimal design** is found by maximizing the statistical power of the equivalence test over the analysed range of concentrations.

First, the computation of the statistical power of the equivalence test when using the model-based approach to compute the two-sided confidence interval is described. After that, the procedure used to find the optimal design is illustrated, focusing on the optimization problem which needs to be solved. Finally, the results for different configurations of parameters are reported and discussed.

4.1 Power calculation

This section shows an approach to the computation of the statistical power of the equivalence test (3.1) when the model-based approach is used to compute the confidence interval, both with linear models in the log scale and linear models in the original scale. The described computations follow the approach used by Zhu [23].

In order to compute the statistical power of the test, the equivalence formulation (3.1) is split in the following two hypothesis tests.

$$\begin{aligned} \mathrm{H}_0^{(1)} &: \quad \frac{\lambda_r}{\lambda_c} \leq 0.7 \\ \mathrm{H}_1^{(1)} &: \quad \frac{\lambda_r}{\lambda_c} > 0.7 \end{aligned} \tag{4.1}$$

$$\begin{aligned} \mathrm{H}_0^{(2)}: &\quad \frac{\lambda_r}{\lambda_c} \geq 1.3 \\ \mathrm{H}_1^{(2)}: &\quad \frac{\lambda_r}{\lambda_c} < 1.3 \end{aligned} \tag{4.2}$$

The TOST procedure used for executing the equivalence test consists in rejecting the null hypothesis of non-equivalence if and only if both $\mathrm{H}_0^{(1)}$ and $\mathrm{H}_0^{(2)}$ are rejected at a chosen significance level.

4.1.1 Linear models in the log scale

The confidence interval (2.22) is computed using the models

$$\lambda_r = \alpha_r x^{\beta_r} \quad \lambda_c = \alpha_c x^{\beta_c}$$

estimated between expected count and theoretical concentration for the rapid method and the compendial method, respectively. These relations can be expressed in the log scale as

$$\log(\lambda_r) = \log(\alpha_r) + \beta_r \log(x) = \delta_r + \beta_r \log(x) \quad \log(\lambda_c) = \log(\alpha_c) + \beta_c \log(x) = \delta_c + \beta_c \log(x) \tag{4.3}$$

Assuming these expressions of the log expected counts, the logarithm of the ratio λ_r/λ_c can be expressed as

$$\log(\lambda_r/\lambda_c) = \delta_r + \beta_r \log(x) - \delta_c - \beta_c \log(x) = \delta + \beta \log(x),$$

where $\delta = \delta_r - \delta_c$ and $\beta = \beta_r - \beta_c$. Using this notation, the equivalence test (3.1) can be written in the log scale as

$$\begin{aligned} \mathrm{H}_0: &\quad \delta + \beta \log(x) \leq \log(0.7) \quad \text{or} \quad \delta + \beta \log(x) \geq \log(1.3) \\ \mathrm{H}_1: &\quad \log(0.7) < \delta + \beta \log(x) < \log(1.3) \end{aligned} \tag{4.4}$$

and the two hypothesis tests (4.1) and (4.2) can be expressed in the log scale as

$$\begin{aligned} \mathrm{H}_0^{(1)}: &\quad \delta + \beta \log(x) \leq \log(0.7) \\ \mathrm{H}_1^{(1)}: &\quad \delta + \beta \log(x) > \log(0.7) \end{aligned} \tag{4.5}$$

$$\begin{aligned} \mathrm{H}_0^{(2)}: &\quad \delta + \beta \log(x) \geq \log(1.3) \\ \mathrm{H}_1^{(2)}: &\quad \delta + \beta \log(x) < \log(1.3) \end{aligned} \tag{4.6}$$

It is now necessary to derive the rejection rule for each of the two hypothesis tests. The probability of type 1 error is set at 0.05 because 90% confidence intervals have been used to make the decision about accuracy in the simulations. The estimates of δ and β are given by

$$\hat{\delta} = \hat{\delta}_r - \hat{\delta}_c \quad \hat{\beta} = \hat{\beta}_r - \hat{\beta}_c$$

where $\hat{\delta}_r, \hat{\delta}_c, \hat{\beta}_r$ and $\hat{\beta}_c$ are the estimates of the coefficients in the models (4.3).

Given $\hat{\delta}$ and $\hat{\beta}$ estimates of δ and β, the rejection rules are derived using the approximate distribution (2.20) which has been found in Subsection 2.2.4:

$$\hat{\delta} + \hat{\beta} \log(x) \overset{approx.}{\sim} \mathcal{N}(\delta + \beta \log(x), \sigma^2(x))$$

where the variance $\sigma^2(x)$ depends on the standard errors of the estimated coefficients in the models for the rapid method and the compendial method. The variance $\sigma^2(x)$ has the following expression:

$$\sigma^2(x) = Var(\hat{\delta} + \hat{\beta} \log(x)) = Var(\hat{\delta}_r + \hat{\beta}_r \log(x) - (\hat{\delta}_c + \hat{\beta}_c \log(x))) =$$

$$= Var(\hat{\delta}_r + \hat{\beta}_r \log(x)) + Var(\hat{\delta}_c + \hat{\beta}_c \log(x)) = \sigma^2_{r,x} + \sigma^2_{c,x}$$

where the variances of the estimated log expected values $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ are given by the expression (B.2) reported in Appendix B.

To derive the rejection rule in (4.5), a value K_L needs to be found such that

$$\mathbb{P}[\text{Reject } H_0^{(1)} | H_0^{(1)} \text{true}] = \mathbb{P}[\hat{\delta} + \hat{\beta} \log(x) > K_L | \delta + \beta \log(x) = \log(0.7)] \approx \mathbb{P}\left[Z > \frac{K_L - \log(0.7)}{\sqrt{\sigma^2(x)}}\right] = 0.05$$

where $Z \sim \mathcal{N}(0,1)$. Thus, $\mathbf{H_0^{(1)}}$**in** (4.5) **is rejected when** $\boldsymbol{\hat{\delta} + \hat{\beta} \log(x) > \log(0.7) - z\sqrt{\sigma^2(x)}}$ where z is the quantile of the Standard Normal distribution $z = z_{0.05}$.

Similarly, to derive the rejection rule in (4.6), a value K_U needs to be found such that

$$\mathbb{P}[\text{Reject } H_0^{(2)} | H_0^{(2)} \text{true}] = \mathbb{P}[\hat{\delta} + \hat{\beta} \log(x) < K_U | \delta + \beta \log(x) = \log(1.3)] \approx \mathbb{P}\left[Z < \frac{K_U - \log(1.3)}{\sqrt{\sigma^2(x)}}\right] = 0.05.$$

Thus, $\mathbf{H_0^{(2)}}$**in** (4.6) **is rejected when** $\boldsymbol{\hat{\delta} + \hat{\beta} \log(x) < \log(1.3) + z\sqrt{\sigma^2(x)}}$ where $z = z_{0.05}$.

In order to compute the power of the equivalence test, it is necessary to calculate the probability of rejecting at the same time $H_0^{(1)}$ in (4.5) and $H_0^{(2)}$ in (4.6): this corresponds to the probability that $K_L < \hat{\delta} + \hat{\beta} \log(x) < K_U$. However, it can happen in principle that $K_L > K_U$: this is due to the fact that the rejection rules for the two hypothesis tests (4.5) and (4.6) have been derived independently of each other. When $K_L > K_U$, the power of the test is equal to 0, since the null hypothesis H_0 in (4.4) would never be rejected. Thus, when $K_L < K_U$, the power of the equivalence test (4.4) is given by:

$$\begin{aligned}
&\mathbb{P}[\text{Reject } H_0 | H_0 \text{false}] = \mathbb{P}[K_L < \hat{\delta} + \hat{\beta} \log(x) < K_U | \delta, \beta \quad \text{such that the null hypothesis is false}] = \\
&= \mathbb{P}[\hat{\delta} + \hat{\beta} \log(x) < K_U | \delta, \beta] - \mathbb{P}[\hat{\delta} + \hat{\beta} \log(x) < K_L | \delta, \beta] = \\
&= \mathbb{P}[\hat{\delta} + \hat{\beta} \log(x) < \log(1.3) + z\sqrt{\sigma^2(x)} | \delta, \beta] - \mathbb{P}[\hat{\delta} + \hat{\beta} \log(x) < \log(0.7) - z\sqrt{\sigma^2(x)} | \delta, \beta] \approx \\
&\approx \Phi\left(\frac{\log(1.3) + z\sqrt{\sigma^2(x)} - (\delta + \beta \log(x))}{\sqrt{\sigma^2(x)}}\right) - \Phi\left(\frac{\log(0.7) - z\sqrt{\sigma^2(x)} - (\delta + \beta \log(x))}{\sqrt{\sigma^2(x)}}\right)
\end{aligned}$$

where Φ is the cumulative distribution function of the Standard Normal distribution.

In conclusion, the power of the equivalence test when using the model-based approach with the linear models in the log scale is given by the following formula. The formula contains a maximum between the just computed expression and 0 in order to ensure that the power is non-negative even when $K_L > K_U$: as previously explained, the power of the equivalence test is 0 in this case, but the expression computed above would be negative.

$$P(x) = \max\left\{\Phi\left(\frac{\log(1.3) + z\sqrt{\sigma^2(x)} - (\delta + \beta \log(x))}{\sqrt{\sigma^2(x)}}\right) - \Phi\left(\frac{\log(0.7) - z\sqrt{\sigma^2(x)} - (\delta + \beta \log(x))}{\sqrt{\sigma^2(x)}}\right), 0\right\}$$

4.1.2 Linear models in the original scale

The confidence interval (2.17) is computed using the models

$$\lambda_r = \alpha_r + \beta_r x \quad \lambda_c = \alpha_c + \beta_c x$$

estimated between expected count and theoretical concentration for the rapid method and the compendial method, respectively. In this case, the computations are performed in the original scale, so it is necessary to derive the rejection rule in each of the two hypothesis tests (4.1) and (4.2) in which the equivalence formulation (3.1) can be split. The computations are based on the approximate distribution (2.16) described in Subsection 2.2.3

$$\frac{\hat{\lambda}_r}{\hat{\lambda}_c} = \frac{\hat{\alpha}_r + \hat{\beta}_r x}{\hat{\alpha}_c + \hat{\beta}_c x} \overset{approx.}{\sim} \mathcal{N}\left(\frac{\lambda_r}{\lambda_c}, \sigma^2(x)\right)$$

where

$$\sigma^2(x) = \frac{\sigma^2_{r,x}}{\lambda^2_c} + \frac{\lambda^2_r}{\lambda^4_c}\sigma^2_{c,x} \tag{4.7}$$

with $\sigma^2_{r,x} = Var(\hat{\lambda}_r)$ and $\sigma^2_{c,x} = Var(\hat{\lambda}_c)$. The formula for the variance $\sigma^2(x)$ can be found by replacing $\sigma^2_{r,x}$ and $\sigma^2_{c,x}$ in (4.7) with the expression (B.1) reported in Appendix B.

To derive the rejection rule in (4.1), a value K_L needs to be found such that

$$\mathbb{P}[\text{Reject } \mathrm{H}_0^{(1)} | \mathrm{H}_0^{(1)}\text{true}] = \mathbb{P}[\hat{\lambda}_r/\hat{\lambda}_c > K_L | \lambda_r/\lambda_c = 0.7] \approx \mathbb{P}\left[Z > \frac{K_L - 0.7}{\sqrt{\sigma^2(x)}}\right] = 0.05.$$

Thus, $\mathbf{H_0^{(1)}}$**in** (4.1) **is rejected when** $\boldsymbol{\hat{\lambda}_r/\hat{\lambda}_c > 0.7 - z\sqrt{\sigma^2(x)}}$ where z is the quantile of the Standard Normal distribution $z = z_{0.05}$.

Similarly, to derive the rejection rule in (4.2), a value K_U needs to be found such that

$$\mathbb{P}[\text{Reject } \mathrm{H}_0^{(2)} | \mathrm{H}_0^{(2)}\text{true}] = \mathbb{P}[\hat{\lambda}_r/\hat{\lambda_c} < K_U | \lambda_r/\lambda_c = 1.3] \approx \mathbb{P}\left[Z < \frac{K_U - 1.3}{\sqrt{\sigma^2(x)}}\right] = 0.05.$$

Thus, $\mathbf{H_0^{(2)}}$**in** (4.2) **is rejected when** $\boldsymbol{\hat{\lambda}_r/\hat{\lambda}_c < 1.3 + z\sqrt{\sigma^2(x)}}$ where $z = z_{0.05}$.

In order to compute the power of the equivalence test, it is necessary to calculate the probability of rejecting at the same time $\mathrm{H}_0^{(1)}$ in (4.1) and $\mathrm{H}_0^{(2)}$ in (4.2): this corresponds to the probability that $K_L < \hat{\lambda}_r/\hat{\lambda}_c < K_U$. As explained in the previous subsection, it can happen that $K_L > K_U$: in this case, the power is equal to 0. When $K_L < K_U$, the power of the equivalence test (3.1) is given by:

$$\begin{aligned}
&\mathbb{P}[\text{Reject } \mathrm{H}_0 | \mathrm{H}_0\text{false}] = \mathbb{P}[K_L < \hat{\lambda}_r/\hat{\lambda}_c < K_U | \lambda_r, \lambda_c \quad \text{such that the null hypothesis is false}] = \\
&= \mathbb{P}[\hat{\lambda}_r/\hat{\lambda}_c < K_U | \lambda_r, \lambda_c] - \mathbb{P}[\hat{\lambda}_r/\hat{\lambda}_c < K_L | \lambda_r, \lambda_c] = \\
&= \mathbb{P}[\hat{\lambda}_r/\hat{\lambda}_c < 1.3 + z\sqrt{\sigma^2(x)} | \lambda_r, \lambda_c] - \mathbb{P}[\hat{\lambda}_r/\hat{\lambda}_c < 0.7 - z\sqrt{\sigma^2(x)} | \lambda_r, \lambda_c] \approx \\
&\approx \Phi\left(\frac{1.3 + z\sqrt{\sigma^2(x)} - \lambda_r/\lambda_c}{\sqrt{\sigma^2(x)}}\right) - \Phi\left(\frac{0.7 - z\sqrt{\sigma^2(x)} - \lambda_r/\lambda_c)}{\sqrt{\sigma^2(x)}}\right).
\end{aligned}$$

In conclusion, the power of the equivalence test when using the model-based approach with the linear models in the original scale is given by the following formula, where the maximum is used to avoid that the power is negative when $K_L > K_U$, as in the previous subsection.

$$P(x) = \max\left\{\Phi\left(\frac{1.3 + z\sqrt{\sigma^2(x)} - \lambda_r/\lambda_c}{\sqrt{\sigma^2(x)}}\right) - \Phi\left(\frac{0.7 - z\sqrt{\sigma^2(x)} - \lambda_r/\lambda_c)}{\sqrt{\sigma^2(x)}}\right), 0\right\}$$

4.2 Optimization problem

This section illustrates the procedure used to find the optimal design, focusing in particular on the description of the optimization problem which needs to be solved. Optimality in this context is with respect to the mean power of the equivalence test over the analysed range of concentrations. More precisely, given a set $\mathcal{X}$ of p concentrations that could be potentially used to estimate the linear models, the intention is to find a subset $\mathcal{X}_k \subseteq \mathcal{X}$, containing k concentrations ($2 \leq k \leq p$), and numbers of replicates $n_1, \ldots, n_k$, such that estimating the linear models for the two microbiological methods using concentrations in $\mathcal{X}_k$ with the determined number of replicates per concentration permits to maximize the power of the equivalence test at the concentrations in $\mathcal{X}$ where theoretical equivalence holds. In order to describe the procedure used to find the optimal design, it is necessary to first illustrate the notation used.

- $\mathcal{X} \subset \mathbb{N}$: set of concentrations that could be potentially used to estimate the linear models for the rapid method and the compendial method. $p = |\mathcal{X}|$ denotes the number of concentrations.

- N: number of total available experiments to estimate the linear model for each of the two microbiological methods.
- k: number of concentrations used to estimate the linear models. Then $k \in \{2, \dots, p\}$ because each linear model has two coefficients, so at least 2 different concentrations are needed to estimate them.
- $\mathcal{X}_k \subseteq \mathcal{X}$: given a value k, $\mathcal{X}_k$ denotes the subset of $\mathcal{X}$ including the k concentrations used to estimate the models.
- n_i: if $\mathcal{X}_k = \{x_1, \dots, x_k\}$, n_i denotes the number of replicates at concentration $x_i \in \mathcal{X}_k$ used to estimate the models. This holds for any $i = 1, \dots, k$.
- $\mathcal{X}_{eq} \subseteq \mathcal{X}$: subset of $\mathcal{X}$ including the concentrations at which equivalence is theoretically true. In formulas, $\mathcal{X}_{eq} = \{x \in \mathcal{X} \quad | \quad 0.7 < \lambda_r(x)/\lambda_c(x) < 1.3\}$. For concentrations in the set $\mathcal{X}_{eq}$ it is actually appropriate to talk about statistical power, since at these concentrations the null hypothesis of non-equivalence is false.
- $P(n_1, \dots, n_k; \mathcal{X}_k, x)$: power of the equivalence test (when the model-based approach is used to compute the confidence interval) computed at concentration x when the linear models for the two microbiological methods are estimated using concentrations in $\mathcal{X}_k = \{x_1, \dots, x_k\}$ with n_i replicates at concentration $x_i \quad \forall i = 1, \dots, k$. This is considered as a function of the numbers of replicates $n_1, \dots, n_k$, while the set $\mathcal{X}_k$ and the concentration x are treated as fixed parameters. This power is computed using one of the two formulas derived in the previous section, depending on the type of model that is chosen for the two microbiological methods.

For each possible value of k, there are $\binom{p}{k}$ possible subsets of the p concentrations that can be used to estimate the models. For each possible value of k and for each possible subset $\mathcal{X}_k$, the following optimization problem is solved.

$$\begin{aligned} \max_{n_1,\dots,n_k} & \sum_{x \in \mathcal{X}_{eq}} P(n_1, \dots, n_k; \mathcal{X}_k, x) \\ \text{s.t.} & \sum_{i=1}^{k} n_i = N \\ & 1 \leq n_i \leq N - k + 1 \quad \forall i = 1, \dots, k \\ & n_i \in \mathbb{N} \quad \forall i = 1, \dots, k \end{aligned} \tag{4.8}$$

The solution of this optimization problem consists in the allocation of replicates to the k concentrations in the set $\mathcal{X}_k$ used to estimate the linear models which maximizes the mean power of the equivalence test over the set $\mathcal{X}_{eq}$. According to the first constraint, the experiments used at each concentration must sum up to the total number of available experiments. The second constraint specifies bounds on each number of replicates: the lower bound is necessary to ensure that each concentration in $\mathcal{X}_k$ is used at least once to estimate the models, while the upper bound is actually redundant because the first constraint and the lower bound on each n_i imply that each number of replicates can not exceed the value $N - k + 1$.

Overall, $\sum_{k=2}^{p} \binom{p}{k}$ optimization problems like (4.8) are solved for each configuration. Each of these problems is solved by means of a Genetic Algorithm in MATLAB and the programs used are reported in Appendix C. Finally, the combination of concentrations and numbers of replicates giving the best value of the objective function is identified as optimal design.

4.3 Optimal designs for the different configurations

The optimal design has been determined for each of the seven configurations of linear models and parameters used in the simulations described in Chapter 3. The number of available experiments

is always equal to $\boldsymbol{N = 60}$. The set of possible concentrations is $\mathcal{X} = \{1, \ldots, 12\}$ when linear models in the log scale are used, while it is $\mathcal{X} = \{0, \ldots, 11\}$ when linear models in the original scale are used. The following subsections illustrate the optimal designs for the different configurations. Moreover, for each configuration, once the optimal design has been determined, simulations to evaluate the number of correct decisions when using the model-based approach with the optimal design have been performed in order to observe the improvements in the performance of the model-based approach. In the tables referring to the results of these simulations, the most substantial improvements in comparison to the results obtained with the homogeneous design used in the simulations described in the previous chapter are highlighted in red. These tables also report the theoretical power computed using the formulas derived in Section 4.1, both using the homogeneous design and using the optimal design to estimate the linear models.

4.3.1 First configuration

This configuration is characterised by the linear models in the log scale

$$\lambda_r = 1.2x \quad \lambda_c = 1.1x.$$

With this choice of models and parameters, the optimal design consists of 3 concentrations. Almost half of the experiments are used at concentration 12 and the remaining experiments are approximately equally distributed between concentrations 2 and 3. As far as the power of the equivalence test is concerned (Table 4.2), there is a substantial improvement in the performances at concentration 2 and at high concentrations (10, 11, 12). The number of correct decisions at $x = 2$ increases from 3 to 113: of course, 11.3% is not an acceptable power, but the improvement in comparison to the previous case is considerable. The optimal design permits to reach a power of around 80% at $x \in \{10, 11, 12\}$, while with the homogeneous design the maximum value at these concentrations is 70% ($x = 10$). When the power is lower with the optimal design rather than with the homogeneous design ($x \in \{5, 6, 7\}$), the differences are not so remarkable as for the observed improvements.

Table 4.1: Optimal design for configuration 1 ($\lambda_r = 1.2x$ and $\lambda_c = 1.1x$).

Concentration	**Number of replicates**
2	16
3	17
12	27

Table 4.2: Results of the model-based approach with the two designs for configuration 1 ($\lambda_r = 1.2x$ and $\lambda_c = 1.1x$).

		Number of correct decisions (out of 1000)		**Theoretical power**	
$\boldsymbol{x}$	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
1	Y (1.091)	0	0	0	0
2	Y (1.091)	113	3	0.12	0
3	Y (1.091)	342	243	0.35	0.25
4	Y (1.091)	490	465	0.51	0.46
5	Y (1.091)	613	615	0.64	0.63
6	Y (1.091)	727	750	0.74	0.76
7	Y (1.091)	807	822	0.81	0.83
8	Y (1.091)	836	820	0.84	0.83
9	Y (1.091)	837	767	0.85	0.80
10	Y (1.091)	822	702	0.84	0.75
11	Y (1.091)	797	642	0.81	0.67
12	Y (1.091)	758	581	0.77	0.60

4.3.2 Second configuration

This configuration is characterised by the linear models in the log scale

$$\lambda_r = 0.8x \quad \lambda_c = 0.9x.$$

The optimal design consists of 3 concentrations also for this configuration: however, while the allocation of replicates is nearly homogeneous between concentrations 3 and 12, only 10 replicates are used at concentration 2. The results of the simulations (Table 4.4) show again evident improvements for high concentrations; in addition, the power considerably increases at concentrations 3 and 4. It should also be noted that in this case the optimal design has permitted to obtain better results at all concentrations in the range (in the simulations).

Table 4.3: Optimal design for configuration 2 ($\lambda_r = 0.8x$ and $\lambda_c = 0.9x$).

Concentration	Number of replicates
2	10
3	25
12	25

Table 4.4: Results of the model-based approach with the two designs for configuration 2 ($\lambda_r = 0.8x$ and $\lambda_c = 0.9x$).

		Number of correct decisions (out of 1000)		Theoretical power	
x	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
1	Y (0.889)	0	0	0	0
2	Y (0.889)	12	0	0	0
3	Y (0.889)	342	149	0.31	0.14
4	Y (0.889)	605	467	0.58	0.48
5	Y (0.889)	775	721	0.76	0.73
6	Y (0.889)	872	850	0.86	0.86
7	Y (0.889)	912	902	0.91	0.92
8	Y (0.889)	928	909	0.92	0.92
9	Y (0.889)	939	872	0.92	0.89
10	Y (0.889)	915	819	0.91	0.84
11	Y (0.889)	895	749	0.88	0.77
12	Y (0.889)	860	662	0.85	0.69

4.3.3 Third configuration

This configuration is characterised by the linear models in the log scale

$$\lambda_r = 1.05x^{1.3} \quad \lambda_c = 1.31x^{1.1}.$$

With this configuration of models and parameters, the optimal design consists of concentrations 3 and 12. In addition to the number of concentrations, the most remarkable difference in comparison to the optimal designs obtained for the previous configurations is in the allocation of experiments to $x = 12$: only 14 replicates are used at this concentration, with more than 75% of the available experiments used at $x = 3$. A possible explanation, which is also confirmed by the results obtained for the next configurations, is that the ratio between the true expected counts at $x = 3$ is between 0.7 and 1.3, while this does not happen at $x = 12$. As far as the results of the simulations are concerned (Table 4.6), increases in power can be observed especially at $x \in \{2, 3, 4\}$, while in this case it is not possible to compare the results at high concentrations since the true ratios are too close to the equivalence bound 1.3 or theoretical equivalence does not hold. At concentrations 7, 8 and 9, the results seem considerably worse using the optimal design than using the homogeneous design: however, the power is already low with the homogeneous design, probably due to the values of the true ratios, and the differences in the performances are not so evident as at concentrations 2, 3 and 4.

Table 4.5: Optimal design for configuration 3 ($\lambda_r = 1.05x^{1.3}$ and $\lambda_c = 1.31x^{1.1}$).

Concentration	Number of replicates
3	46
12	14

Table 4.6: Results of the model-based approach with the two designs for configuration 3 ($\lambda_r = 1.05x^{1.3}$ and $\lambda_c = 1.31x^{1.1}$).

		Number of correct decisions (out of 1000)		Theoretical power	
x	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
1	Y (0.802)	3	0	0	0
2	Y (0.921)	509	84	0.51	0.10
3	Y (0.998)	808	491	0.82	0.52
4	Y (1.058)	812	647	0.82	0.67
5	Y (1.106)	755	668	0.75	0.67
6	Y (1.147)	627	621	0.63	0.63
7	Y (1.183)	446	518	0.45	0.52
8	Y (1.215)	273	377	0.28	0.35
9	Y (1.244)	149	204	0.16	0.20
10	Y (1.270)	93	110	0.09	0.10
11	Y (1.295)	56	57	0.06	0.06
12	N (1.318)	961	965	-	-

4.3.4 Fourth configuration

This configuration is characterised by the linear models in the log scale

$$\lambda_r = 1.3x^{1.1} \quad \lambda_c = 1.1x^{1.32}.$$

The optimal design for this configuration is identical to the optimal design obtained for the third configuration. Concentrations 3 and 12 are selected, with 46 and 14 replicates, respectively. As before, equivalence at $x = 3$ is theoretically true, while the ratio between the true expected counts at concentration 12 is lower than 0.7. Using the optimal design permits to substantially increase the power at concentrations 2, 3, 4 and 5, while it is not possible to see improvements for high concentrations because of the true ratios being very close to 0.7 or because theoretical equivalence does not hold (Table 4.8).

Table 4.7: Optimal design for configuration 4 ($\lambda_r = 1.3x^{1.1}$ and $\lambda_c = 1.1x^{1.32}$).

Concentration	Number of replicates
3	46
12	14

4.3.5 Fifth configuration

This configuration is characterised by the linear models in the log scale

$$\lambda_r = 0.6x^{1.1} \quad \lambda_c = 0.7x^{1.1}.$$

The optimal design consists of concentrations 3 and 12 also for this configuration. However, there is an evident difference in the number of replicates in comparison to the previous two configurations: 33 replicates are used at $x = 3$ and 27 replicates at $x = 12$. This gives a further confirmation of the effect of theoretical equivalence on the allocation of experiments to the concentrations selected for the optimal design: equivalence holds at both concentrations in this case and the available experiments are almost equally split between the two concentrations. As the results of the simulations show (Table 4.10), using the optimal design permits to substantially

Table 4.8: Results of the model-based approach with the two designs for configuration 4 ($\lambda_r = 1.3x^{1.1}$ and $\lambda_c = 1.1x^{1.32}$).

		Number of correct decisions (out of 1000)		**Theoretical power**	
$\boldsymbol{x}$	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
1	Y (1.182)	10	0	0.01	0
2	Y (1.015)	513	94	0.50	0.11
3	Y (0.928)	870	531	0.85	0.56
4	Y (0.871)	892	702	0.87	0.72
5	Y (0.829)	827	704	0.80	0.72
6	Y (0.797)	673	659	0.67	0.66
7	Y (0.770)	453	515	0.47	0.54
8	Y (0.748)	288	327	0.28	0.35
9	Y (0.729)	150	180	0.15	0.18
10	Y (0.712)	80	87	0.08	0.09
11	N (0.697)	954	948	-	-
12	N (0.684)	973	974	-	-

increase the power at concentrations 3, 10, 11 and 12, confirming the results observed for the other configurations.

Table 4.9: Optimal design for configuration 5 ($\lambda_r = 0.6x^{1.1}$ and $\lambda_c = 0.7x^{1.1}$).

Concentration	**Number of replicates**
3	33
12	27

Table 4.10: Results of the model-based approach with the two designs for configuration 5 ($\lambda_r = 0.6x^{1.1}$ and $\lambda_c = 0.7x^{1.1}$).

		Number of correct decisions (out of 1000)		**Theoretical power**	
$\boldsymbol{x}$	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
1	Y (0.857)	0	0	0	0
2	Y (0.857)	0	0	0	0
3	Y (0.857)	130	35	0.13	0.03
4	Y (0.857)	384	338	0.38	0.32
5	Y (0.857)	582	556	0.56	0.55
6	Y (0.857)	705	696	0.69	0.71
7	Y (0.857)	784	779	0.78	0.79
8	Y (0.857)	833	809	0.83	0.81
9	Y (0.857)	851	772	0.84	0.78
10	Y (0.857)	830	712	0.83	0.71
11	Y (0.857)	793	635	0.80	0.64
12	Y (0.857)	742	557	0.76	0.56

4.3.6 Configurations 1, 2, 3, 4, 5 with different sets of concentrations

The optimal designs found for the five configurations just described never include the lowest concentration $x = 1$ and only for the first and second configurations $x = 2$ is included. In order to understand if the lowest concentration in the range is never included in the optimal design or if there are other reasons behind these results, some other analyses have been performed with these configurations, changing the set of concentrations which could be potentially used to estimate the models. In particular, concentration 1 has not been included in the set $\mathcal{X}$ for the first two configurations, while concentrations 2 and 3 have not been included in the set $\mathcal{X}$ for the third, fourth and fifth configurations. The results are summarised in Table 4.11.

It should be noted that, with a restricted set of concentrations that could be potentially used to estimate the linear models, the lowest concentration in the range is always included in the

Table 4.11: Optimal design for configurations 1, 2, 3, 4, 5 with restricted sets of concentrations.

		Optimal design	
Configuration	$\mathcal{X}$	**Concentrations**	**Numbers of replicates**
First	$\{2,\dots,12\}$	$x_1 = 2, x_2 = 3, x_3 = 12$	$n_1 = 16, n_2 = 17, n_3 = 27$
Second	$\{2,\dots,12\}$	$x_1 = 2, x_2 = 3, x_3 = 12$	$n_1 = 10, n_2 = 25, n_3 = 25$
Third	$\{3,\dots,12\}$	$x_1 = 3, x_2 = 4, x_3 = 12$	$n_1 = 17, n_2 = 28, n_3 = 15$
Fourth	$\{3,\dots,12\}$	$x_1 = 3, x_2 = 4, x_3 = 12$	$n_1 = 15, n_2 = 30, n_3 = 15$
Fifth	$\{3,\dots,12\}$	$x_1 = 3, x_2 = 12$	$n_1 = 33, n_2 = 27$

optimal design. For the first, second and fifth configurations, the designs are exactly the same as those obtained using the set $\mathcal{X} = \{1,\dots,12\}$. The designs obtained for the third and fourth configurations are very similar: they are made of three concentrations, with around 50% of all the available experiments used at concentration 4 and the other half almost equally split between concentrations 3 and 12.

Since in the first, second and fifth configurations the true ratio λ_r/λ_c does not depend on x, it is possible to look for the optimal design also on other ranges of concentrations, without obtaining results that are affected by values of the ratio between the true expected counts (it probably has an effect, as described in Subsection 4.3.3).

For example, using the first configuration of models and parameters and allowing for concentrations in the set $\mathcal{X} = \{5,\dots,16\}$, the optimal design consists of concentrations $x_1 = 5$ and $x_2 = 16$, with numbers of replicates $n_1 = 33$ and $n_2 = 27$. With the second configuration and the set $\mathcal{X} = \{20,\dots,31\}$, the optimal design includes concentrations $x_1 = 20$ and $x_2 = 31$ with numbers of replicates $n_1 = 35$ and $n_2 = 25$. Using the fifth configuration and the set $\mathcal{X} = \{30,\dots,41\}$, the optimal design consists of concentrations $x_1 = 30$ and $x_2 = 41$, with 34 and 26 replicates, respectively.

Overall, the results illustrated in this subsection aim to show that the lowest concentration in the range is usually included in the optimal design. This seems in agreement with what happens for the D-optimal design in linear regression, thus under the assumption of Normally distributed observations. The D-optimal design maximizes the determinant of the Fisher information matrix; with an intercept and a slope to be estimated, the D-optimal design consists of the lowest and highest values of the covariate [24].

A possible explanation why the lowest concentration is not included in the optimal design for the analysed five configurations when the set of possible concentrations is $\mathcal{X} = \{1,\dots,12\}$ could be that the power of the test at concentrations $x = 1$ and $x = 2$ is very low, due to high standard errors of the estimated log expected counts at these concentrations. As a result of such a low power, it could happen that the optimization solver ignores these concentrations, focusing on a range where a higher power can be achieved.

4.3.7 Sixth configuration

This configuration is characterised by the linear models in the original scale

$$\lambda_r = 0.2 + 0.8x \quad \lambda_c = 0.01 + 0.9x.$$

It should be noted that $x = 0$ is included in the optimal design even if equivalence is not theoretically true at this concentration. This can be justified by noting that with a linear model in the original scale it is necessary to have observations at $x = 0$ in order to obtain a proper estimate of the intercept. The resulting allocation of experiments suggests again that theoretical equivalence has an effect on the optimal number of replicates per concentration: around 75% of the experiments are used at $x = 11$, where theoretical equivalence holds. Outstanding improvements are observed in the performances of the model-based approach at concentrations 1 and 2 when using the optimal design (Table 4.13).

Table 4.12: Optimal design for configuration 6 ($\lambda_r = 0.2 + 0.8x$ and $\lambda_c = 0.01 + 0.9x$).

Concentration	Number of replicates
0	16
11	44

Table 4.13: Results of the model-based approach with the two designs for configuration 6 ($\lambda_r = 0.2 + 0.8x$ and $\lambda_c = 0.01 + 0.9x$).

		Number of correct decisions (out of 1000)		Theoretical power	
x	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
0	N (20)	1000	1000	-	-
1	Y (1.100)	384	1	0.35	0
2	Y (0.994)	805	82	0.85	0.15
3	Y (0.959)	884	682	0.91	0.72
4	Y (0.942)	864	871	0.90	0.91
5	Y (0.931)	849	901	0.90	0.91
6	Y (0.924)	836	899	0.86	0.89
7	Y (0.919)	827	877	0.85	0.88
8	Y (0.915)	821	857	0.85	0.87
9	Y (0.912)	811	842	0.85	0.86
10	Y (0.910)	802	820	0.84	0.85
11	Y (0.908)	801	797	0.84	0.83

4.3.8 Seventh configuration

This configuration is characterised by the linear models in the original scale

$$\lambda_r = 0.35 + 0.8x \quad \lambda_c = 0.94 + 0.7x.$$

As in the previous configuration, $x = 0$ is included in the optimal design even if there is no power at this concentration (the null hypothesis of non-equivalence is true). This confirms the need to have data at $x = 0$ when linear models in the original scale are estimated. The optimal number of replicates per concentration is almost identical to that of the previous configuration, confirming that theoretical equivalence at a certain concentration seems to determine how many replicates should be used at that concentration. The number of correct decisions increases at all concentrations using the optimal design: the most substantial improvements are observed at concentrations 2, 3, 8, 9, 10 and 11 (Table 4.15).

Table 4.14: Optimal design for configuration 7 ($\lambda_r = 0.35 + 0.8x$ and $\lambda_c = 0.94 + 0.7x$).

Concentration	Number of replicates
0	17
11	43

Table 4.15: Results of the model-based approach with the two designs for configuration 7 ($\lambda_r = 0.35 + 0.8x$ and $\lambda_c = 0.94 + 0.7x$).

		Number of correct decisions (out of 1000)		Theoretical power	
x	Eq (λ_r/λ_c)	Optimal design	Homogeneous design	Optimal design	Homogeneous design
0	N (0.372)	1000	1000	-	-
1	Y (0.701)	3	0	0.05	0.03
2	Y (0.833)	367	214	0.39	0.30
3	Y (0.905)	836	704	0.80	0.71
4	Y (0.949)	969	914	0.95	0.90
5	Y (0.980)	980	954	0.98	0.95
6	Y (1.002)	981	936	0.99	0.95
7	Y (1.019)	972	902	0.98	0.92
8	Y (1.032)	962	848	0.97	0.88
9	Y (1.043)	943	815	0.96	0.84
10	Y (1.052)	915	779	0.95	0.80
11	Y (1.059)	890	740	0.93	0.75

4.4 Main conclusions

The results illustrated in this chapter permit to derive some conclusions about the optimal design which should be used to estimate the linear models for the rapid method and the compendial method in order to improve the performance of the model-based approach. It should be noted that nothing about the optimal design has been really proved mathematically, but based on the performed tests it is at least possible to suppose some characteristics of the optimal design.

- The optimal design consists of **2 or 3 concentrations**. When using linear models in the original scale, 2 concentrations are included in the optimal design. On the other hand, sometimes 3 concentrations are selected for linear models in the log scale. When this happens, two consecutive concentrations towards the beginning of the range are selected, while the third one is the highest concentration in the range. Maybe in this case allowing for non-integer concentrations could permit to obtain an optimal design with two concentrations. For instance, using the first configuration of models and parameters and using $\{1, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ as set of possible concentrations, the optimal design which comes out from the optimization consists of concentrations 2.5 and 12, with 34 and 26 replicates, respectively. In addition, this design permits to achieve a higher average theoretical power than that obtained with the optimal design using only integer concentrations.

- The **highest concentration** in the range is always included in the optimal design.

- The **lowest concentration** in the range is usually included in the optimal design. This does not happen for linear models in the log scale when the lowest concentration in the set is 1. A possible explanation is that, as shown by the results of the simulations in the third chapter, the power at this concentration is very low, due to high standard errors of the estimated log expected counts. Thus, it could happen that the optimization solver does not consider that including this concentration in the design can be useful for increasing the mean power and focuses only on the range of concentrations higher than $x = 2$ (or higher than $x = 3$ for some configurations). Concentration $x = 0$ is always included in the optimal design when linear models in the original scale are estimated, even if the test has no power at this concentration. When estimating such a model, data at $x = 0$ are necessary in order to get a reasonable estimate of the intercept.

- The **distribution of experiments** among the concentrations depends on theoretical equivalence. When the optimal design consists of 2 concentrations, the following trend is observed: if equivalence holds at both concentrations x_1^* and x_2^* included in the design, then the number of replicates at x_1^* is approximately equal to the number of replicates at x_2^*; on the other hand, if equivalence holds only at one of the two concentrations, around 75% of the available experiments (N) are allocated to the concentration where equivalence is theoretically true. In the

first and second configurations, where the optimal design consists of concentrations 2, 3 and 12 and equivalence holds at all concentrations in $\mathcal{X}$, the sum of the numbers of replicates used at concentrations 2 and 3 is roughly equal to the number of replicates used at concentration 12. As described in the first point of this list, allowing to use the non-integer concentration 2.5 in the first configuration would result in an optimal design made of concentrations 2.5 and 12, with 34 and 26 replicates, respectively: this would be another example of approximately equal distribution of experiments when equivalence holds at both concentrations included in the design.

For each of the seven configurations, after determining the optimal design, the model-based approach with the optimal design has been tested on 1000 simulated datasets in order to investigate its performance. The results have shown substantial improvements in the power of the test at concentrations far from the value $\widetilde{x}$ where the highest power is reached using the homogeneous design. Of course, these improvements can be observed only when the true ratios at these concentrations are not too close to the equivalence bounds 0.7 and 1.3.

Overall, using the optimal design permits to achieve **higher power at extreme concentrations in the range** in comparison to what happens when using the homogeneous design. On the other hand, sometimes the numbers of correct decisions obtained using the optimal design are lower in the middle of the range, but the differences are not as evident as for the observed improvements. It should also be noted that the concentrations at which the optimal design permits to achieve a substantially higher power than that obtained using the homogeneous design are the concentrations at which the non model-based approach needs less replicates (20 or 30) to reach the same performance of the model-based approach with the homogeneous design. Thus, a consequence of using the optimal design is that the number of replicates which the non model-based approach needs to obtain results comparable to those of the model-based approach increases at concentrations towards the boundaries of the range.

In conclusion, the following design to estimate the linear models for the rapid method and the compendial method in the model-based approach is proposed. Given a set of possible concentrations $\mathcal{X} = \{x_1, \ldots, x_p\} \subset \mathbb{N}$ $(x_1 < x_2 < \ldots < x_p)$ and a number of available experiments N, the optimal design consists of concentrations x_1^* and x_2^* given by

$$x_1^* = x_1 \quad x_2^* = x_p$$

with n_1^* replicates at x_1^* and n_2^* replicates at x_2^* determined in the following way.

$$\begin{cases} n_1^* = \left\lceil \frac{N}{2} \right\rceil \quad \text{and} \quad n_2^* = N - n_1^* \quad \text{if} \quad \frac{\lambda_r(x_1^*)}{\lambda_c(x_1^*)} \in (0.7, 1.3) \quad \text{and} \quad \frac{\lambda_r(x_2^*)}{\lambda_c(x_2^*)} \in (0.7, 1.3) \\ n_1^* = \left\lceil \frac{3}{4}N \right\rceil \quad \text{and} \quad n_2^* = N - n_1^* \quad \text{if} \quad \frac{\lambda_r(x_1^*)}{\lambda_c(x_1^*)} \in (0.7, 1.3) \quad \text{and} \quad \frac{\lambda_r(x_2^*)}{\lambda_c(x_2^*)} \notin (0.7, 1.3) \\ n_1^* = \left\lceil \frac{1}{4}N \right\rceil \quad \text{and} \quad n_2^* = N - n_1^* \quad \text{if} \quad \frac{\lambda_r(x_1^*)}{\lambda_c(x_1^*)} \notin (0.7, 1.3) \quad \text{and} \quad \frac{\lambda_r(x_2^*)}{\lambda_c(x_2^*)} \in (0.7, 1.3) \end{cases}$$

Obviously, the concentrations at which theoretical equivalence holds are not known in a real validation study. However, previous knowledge of the method may be used to suppose that equivalence holds in a certain range of concentrations. The proposed design is similar to the optimal designs found for configurations 3, 4, 5, 6 and 7, while the differences with the optimal designs found for the first two configurations are more evident, since the optimal design consists of 3 concentrations in these cases. The proposed design has been derived as a synthesis of the features highlighted in this section. Even if this design is different from the optimal one, it permits to achieve a higher mean power (according to the formulas derived in Section 4.1) than that obtained with the homogeneous design, increasing the power especially at concentrations towards the boundaries of the range. Thus, this design may be used as a starting point and then further optimized by testing the model-based approach with designs similar to this one using simulations.

Chapter 5

Conclusions

This Master thesis has dealt with the evaluation of accuracy in validation of quantitative microbiological methods. As described in Chapter 1, the accuracy of a rapid microbiological method has been evaluated by comparing its results to those of a compendial method.

The assessment of accuracy at a certain theoretical concentration x has been performed by means of the equivalence test (1.1). In this thesis, the TOST procedure has been used to execute this equivalence test. In particular, with the significance level and the equivalence margin used throughout the text, the rapid method has been declared accurate at concentration x if the two-sided 90% confidence interval for the ratio between the expected numbers of microorganisms counted by the rapid method and the compendial method at concentration x was included in the equivalence range [0.7,1.3].

In Chapter 1, two types of approaches to compute the confidence interval have been presented: the model-based approach is based on estimation of linear models between the expected number of microorganisms counted by each of the two microbiological methods and the theoretical concentration of the samples on which the measurements are performed, while the non model-based approach does not require any parameter estimation. In Chapter 2, a description of the computations used to build the confidence interval using the different approaches has been provided.

Chapter 3 has illustrated a simulation study aimed at comparing the performances of the model-based approach and the non model-based approach in order to understand if one of them can be considered preferable to the other in terms of probability of leading to the correct decision about accuracy of the rapid method.

Finally, in Chapter 4, the focus has been on the model-based approach and on how to improve its performance. In particular, the purpose has been optimizing the design used to estimate the linear models for the two microbiological methods in order to increase the power of the equivalence test when using the model-based approach to compute the two-sided confidence interval for the ratio between the two expected values λ_r and λ_c.

5.1 Main results

The simulations described in Chapter 3 have allowed to compare the results of the model-based approach and the non model-based approach. The main purpose has been trying to understand how many measurements per concentration are needed for the non model-based approach in order to obtain results comparable to those produced by the model-based approach. In this simulation study, the model-based approach has always been tested with a homogeneous design used to estimate the linear models for the two microbiological methods: this design consists of all concentrations in the analysed set, with the same number of replicates at each concentration.

The simulations have been run for seven different configurations of linear models and coefficients used to represent the true relations between the expected number of microorganisms counted by each method and the theoretical concentration. The results have shown some interesting features

in the performances of the two approaches.

- As far as the model-based approach is concerned, the main result that can be derived from the simulations is the pattern in the performance. There is a concentration $\widetilde{x}$ around the middle of the analysed range where the model-based approach reaches the highest number of correct decisions and then the performance decreases when the theoretical concentration moves away from $\widetilde{x}$.
- The higher the number of replicates used at a certain concentration is, the better the performance of the non model-based approach at that concentration is. The performance of this approach is also affected by the theoretical concentration: the higher it is, the better the performance is.
- Both approaches lead almost always to the correct decision when equivalence between the two methods is not theoretically true.
- Both approaches give bad results when equivalence is theoretically true, but the ratio between the true expected numbers of microorganisms counted by the rapid method and the compendial method is close to one of the two equivalence bounds 0.7 and 1.3.

As far as the comparison between the two approaches is concerned, at almost all of the concentrations 30 or more replicates are needed for the non model-based approach to give results comparable to those produced by the model-based approach. Having 60 available measurements, this means that accuracy can be evaluated at most at 2 concentrations in the range using the non-model based approach to get results as good as or better than those of the model-based approach. This may lead to the conclusion that the model-based approach should be preferred to the non model-based approach to assess accuracy. However, the model-based approach is based on the assumption that the relation between expected number of microorganisms counted by each microbiological method and theoretical concentration has a specific form, and this assumption is not known to be true in reality. Thus, Subsection 3.3.2 has described a situation in which the non model-based approach could be preferred. If we are willing to believe that if accuracy holds at two concentrations x_L and x_U, then the rapid method is accurate at all concentrations between x_L and x_U, then the non model-based approach could be used at x_L and x_U, choosing these concentrations towards the boundaries of the analysed range, where the non model-based approach needs less replicates to reach better performances than the model-based approach.

Since the results of the simulations have shown that the theoretical concentration affects the performances of the non model-based approach, a comparison between the two approaches has been done at higher concentrations than those used in the simulations. This analysis has shown that the non model-based approach needs much less replicates to give results comparable to those of the model-based approach when accuracy is evaluated at concentrations higher than 20, approximately. Thus, the main results highlighted in this thesis about the comparison between the two approaches have to be considered valid when dealing with low concentrations, approximately up to 20.

Chapter 4 has described a general procedure to find an optimal design to estimate the linear models for the rapid method and the compendial method in the model-based approach. Optimality in this context has been considered with respect to the mean power of the equivalence test over the analysed range of concentrations when using the model-based approach to compute the confidence interval.

Given a set of concentrations $\mathcal{X}$ that could be potentially used to estimate the linear models and a total number of available experiments N, a design consists of a subset of $\mathcal{X}$ containing the concentrations used to estimate the models and the number of replicates that should be used at each concentration in this subset; the numbers of replicates at the different concentrations must sum up to N. The optimal designs have been computed for the seven configurations of models and parameters used in the simulations described in Chapter 3. The optimal design has been found by solving many optimization problems using a heuristic algorithm, so the results do not allow to prove exactly that the optimal design for the model-based approach has a certain structure. However, the results obtained for the different configurations permit to hypothesise some features of the optimal design.

- The optimal design is made of 2 or 3 concentrations.
- The highest concentration in the range is always included in the optimal design.
- The lowest concentration is usually included in the optimal design.
- The distribution of the experiments among the concentrations included in the optimal design depends on theoretical equivalence at these concentrations. More replicates are used at concentrations where the ratio between the true expected numbers of microorganisms counted by the two microbiological methods is included between 0.7 and 1.3. When equivalence holds at all concentrations included in the optimal design, the available experiments are approximately equally split among the different concentrations (especially when the optimal design consists of 2 concentrations).

For each configuration, once the optimal design has been determined, new simulations have been performed in order to compare the performances of the model-based approach using the optimal design to those of the model-based approach using the homogeneous design. The results have shown that the optimal design permits to substantially increase the number of correct declarations of equivalence between the two microbiological methods at concentrations located towards the boundaries of the analysed range. It should be noted that the non model-based approach needs less replicates at these concentrations rather than in the rest of the range to obtain better results than the model-based approach with the homogeneous design. Thus, the optimal design permits to increase the number of replicates that the non model-based approach needs to have performances comparable to those of the model-based approach at these concentrations. Sometimes, the performances of the model-based approach with the optimal design are lower at concentrations towards the middle of the range, where the model-based approach with the homogeneous design gives the best results, but the differences are not as evident as for the observed improvements.

5.2 Further developments

Although some interesting conclusions can be derived from the analysis described in this thesis, there is still a lot of work to do on this topic. The main possible developments listed below concern the verification of the assumptions behind the model-based approach and the optimal design for the model-based approach.

- The model-based approach is based on the estimation of linear models between the expected number of microorganisms counted by each microbiological method and the theoretical concentration. In order to consider the results of the model-based approach reliable, it should be proved that the relations between expected counts and concentrations are reasonably represented by the estimated linear models. What really matters to evaluate accuracy is that the ratio between the expected numbers of microorganisms counted by the two methods is estimated correctly and without a very high standard error which would increase the length of the two-sided confidence interval used to make the decision about equivalence. Thus, two wrong models for the two methods could still provide a good estimate of the true ratio, but probably high standard errors in the estimates of the coefficients would lead to wide confidence intervals which would negatively affect the power of the equivalence test. In conclusion, a measure of goodness of fit of the linear models should be taken into account and it would be appropriate to specify for which values of this metric the results of the model-based approach can be considered reliable.
- Another possible development, which is strongly related to the previous one, could be investigating how the optimal design can be adapted to be at the same time as useful as possible both to maximize the power of the equivalence test and to understand how the estimated models fit the measurements. The optimal designs obtained in Chapter 4 consist of only 2 or 3 concentrations out of 12, located towards the boundaries of the range. However, it would be appropriate to also have measurements at other concentrations to show that the estimated models reasonably fit the data.

- In this thesis, only integer concentrations have been taken into account, but in practice it is also possible to have non-integer theoretical concentrations. Allowing for non-integer concentrations may permit to further optimize the design for the model-based approach, as briefly shown for only one case in Section 4.4.
- The optimal designs found in Chapter 4 have been obtained using a heuristic algorithm, without any mathematical proof of the highlighted features of the optimal design. A possible future step could be analysing if it is possible to prove any characteristic of the optimal design, trying to be as general as possible with respect to the used linear model and the analysed range of concentrations.

Overall, many developments are possible for the analysis described in this thesis. The main results are not claimed to be valid in general, but they have allowed to derive some interesting conclusions which have been reasonably justified and supported by comparing the results obtained for the different analysed cases. The hope is that this work can give at least some good insights and ideas to people involved in the validation of quantitative microbiological methods, providing at the same time a starting point for further and deeper analyses.

Bibliography

[1] R.F. Guardino. Early History of Microbiology and Microbiological Methods. In M.J. Miller, editor, *Encyclopedia of Rapid Microbiological Methods*, volume 1, chapter 1. PDA, 2005.

[2] *EP 9.2. 5.1.6 Alternative methods for control of microbiological quality.* European Pharmacopoeia, 2017.

[3] *USP < 1223 >: Validation of alternative microbiological methods.* United States Pharmacopoeia, 2017.

[4] *ICH Harmonised Tripartite Giudeline. Validation of Analytical Procedures: Text and Methodology Q2(R1).* International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, 1996.

[5] S. De Rosa, P.C. IJzerman-Boon, and E.R. van den Heuvel. Towards a general definition of linearity, accuracy and recovery as method validation parameters. Technical report, Unpublished.

[6] Evaluation, Validation and Implementation of New Microbiological Testing Methods. Technical Report 3, PDA, 2000.

[7] D.J. Schuirmann. A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. *Journal of Pharmacokinetics and Biopharmaceutics*, 15(6):657–680, 1987.

[8] A.H. El-Shaarawi, S.R. Esterby, and B.J. Dutka. Bacterial Density in Water Determined by Poisson or Negative Binomial Distribution. *Applied and Environmental Microbiology*, 41(1): 107–116, 1981.

[9] P.C. IJzerman-Boon and E.R. van den Heuvel. Validation of qualitative microbiological test methods. *Pharmaceutical Statistics*, 14(2):120–128, 2015.

[10] P. McCullagh and J.A. Nelder. *Generalized Linear Models.* Chapman and Hall, 1983.

[11] E.L. Lehmann and G. Casella. *Theory of Point Estimation*, chapter 1. New York: Springer-Verlag, 1998.

[12] E.L. Lehmann. *Testing Statistical Hypotheses*, chapter 2. Springer, 1986.

[13] E.B. Wilson. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22(158):209–212, 1927.

[14] N.E. Breslow and N.E. Day. *Statistical Methods in Cancer Reasearch: The Analysis of Case-Control Studies*, pages 165–166. IARC: Lyon, 1980.

[15] R.G. Newcombe. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Satistics in Medicine*, 17:857–872, 1998.

[16] S.E. Vollset. Confidence intervals for a binomial proportion. *Statistics in Medicine*, 12: 809–824, 1993.

[17] D.L. Brown, T.T. Cai, and A. DasGupta. Interval Estimation for a Binomial Proportion. *Statistical Science*, 16(2):101–117, 2001.

[18] D.J. Olive. *Statistical Theory and Inference*, chapter 2. Springer, 2014.

[19] L. Fahrmeir and H. Kaufmann. Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *The Annals of Statistics*, 13(1):342–368, 1985.

[20] J.E. Gentle. *Matrix Algebra. Theory, Computations, and Applications in Statistics*, chapter 3. Springer, 2007.

[21] G. Casella and R. L. Berger. *Statistical Inference*, chapter 9. Duxbury, 2002.

[22] R. Wicklin. *Simulating data with SAS®*, chapter 5. SAS Institute Inc., Cary, NC, 2013.

[23] H. Zhu. Sample Size Calculation for Comparing Two Poisson or Negative Binomial Rates in Noninferiority or Equivalence Trials. *Statistics in Biopharmaceutical Research*, 9(1):107–115, 2017.

[24] F. Pukelsheim. *Optimal Design of Experiments*, chapter 6,7. SIAM, 1948.

Appendix A

Graphical representation of the true models used in the simulations

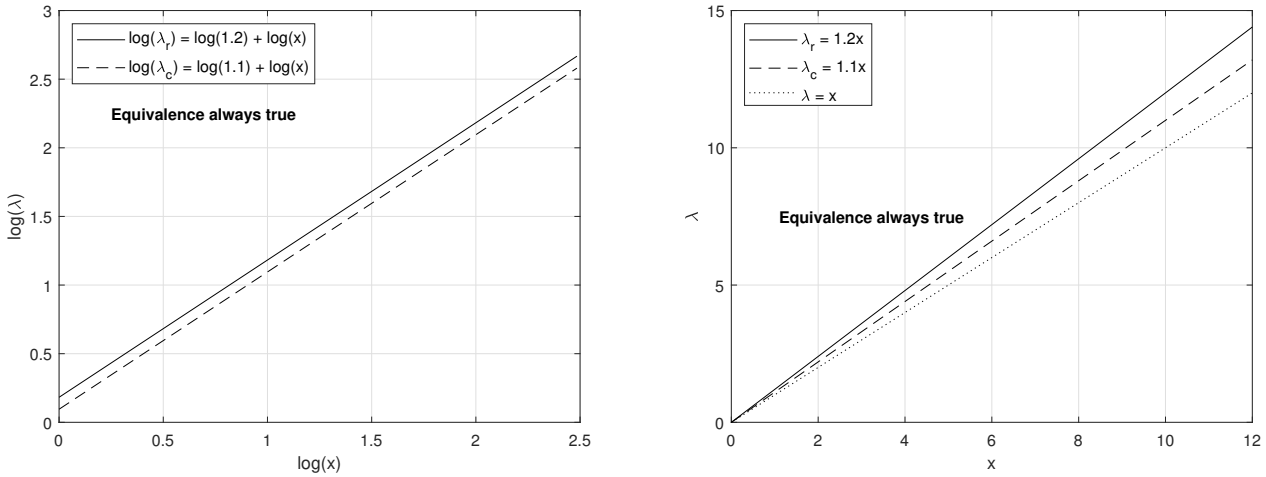
This appendix shows graphical representations of the true models used in the simulations described in Chapter 3. Each of the following sections is about one of the seven configurations of models and coefficients reported in Table A.1. For each configuration, a figure showing how the expected values λ_r and λ_c vary as a function of the concentration x is provided. When the models are linear in the log scale, a plot in the log scale is also shown. In addition, for each configuration, the true ratio λ_r/λ_c is plotted as a function of the concentration x.

Table A.1: Configurations used in the simulations.

Configuration	**Linear models**	$\boldsymbol{\mathcal{X}_{eq}}$
1	$\lambda_r = 1.2x \quad \lambda_c = 1.1x$	$\{1, \ldots, 12\}$
2	$\lambda_r = 0.8x \quad \lambda_c = 0.9x$	$\{1, \ldots, 12\}$
3	$\lambda_r = 1.05x^{1.3} \quad \lambda_c = 1.31x^{1.1}$	$\{1, \ldots, 11\}$
4	$\lambda_r = 1.3x^{1.1} \quad \lambda_c = 1.1x^{1.32}$	$\{1, \ldots, 10\}$
5	$\lambda_r = 0.6x^{1.1} \quad \lambda_c = 0.7x^{1.1}$	$\{1, \ldots, 12\}$
6	$\lambda_r = 0.2 + 0.8x \quad \lambda_c = 0.01 + 0.9x$	$\{1, \ldots, 11\}$
7	$\lambda_r = 0.35 + 0.8x \quad \lambda_c = 0.94 + 0.7x$	$\{1, \ldots, 11\}$

A.1 First configuration

$$\lambda_r = 1.2x \quad \lambda_c = 1.1x$$



(a) Log scale. **(b)** Original scale.

Figure A.1: Expected values versus concentration for the first configuration.

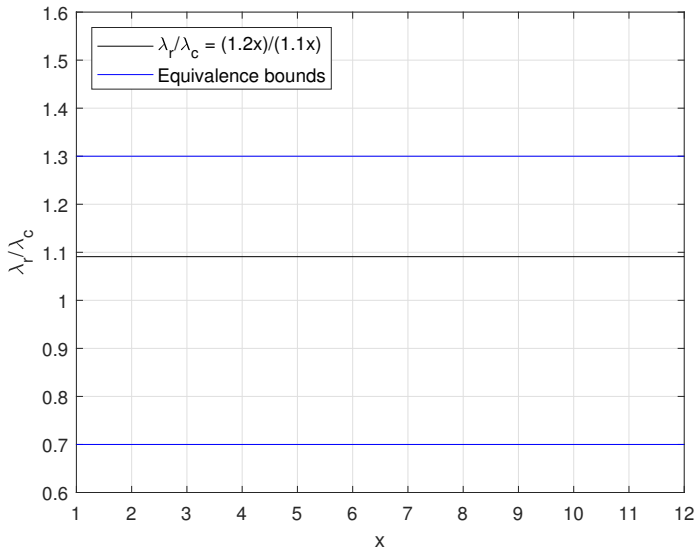
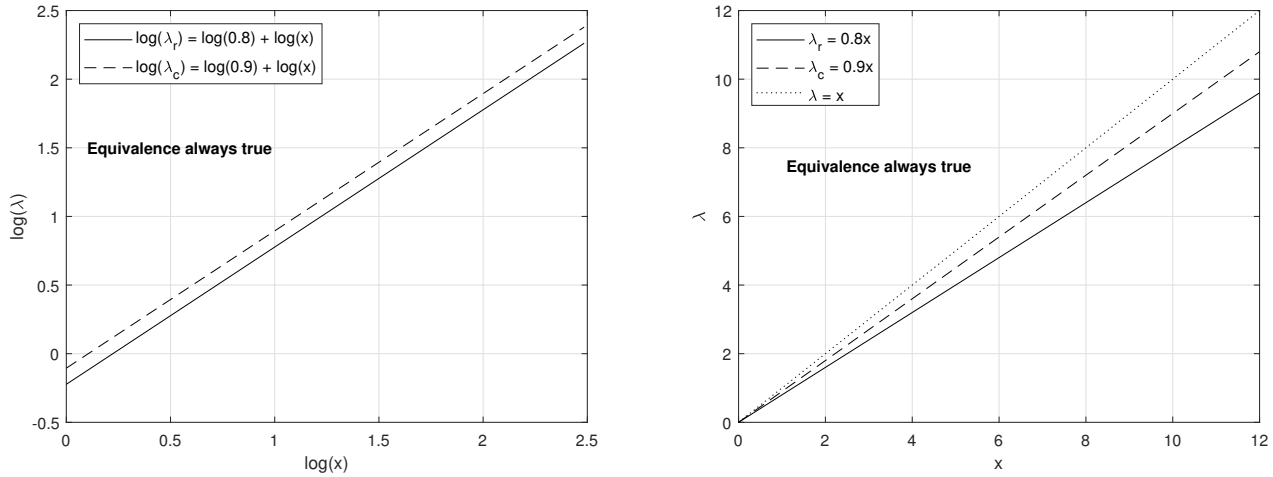


Figure A.2: λ_r/λ_c versus concentration for the first configuration.

A.2 Second configuration

$$\lambda_r = 0.8x \quad \lambda_c = 0.9x$$



(a) Log scale. **(b)** Original scale.

Figure A.3: Expected values versus concentration for the second configuration.

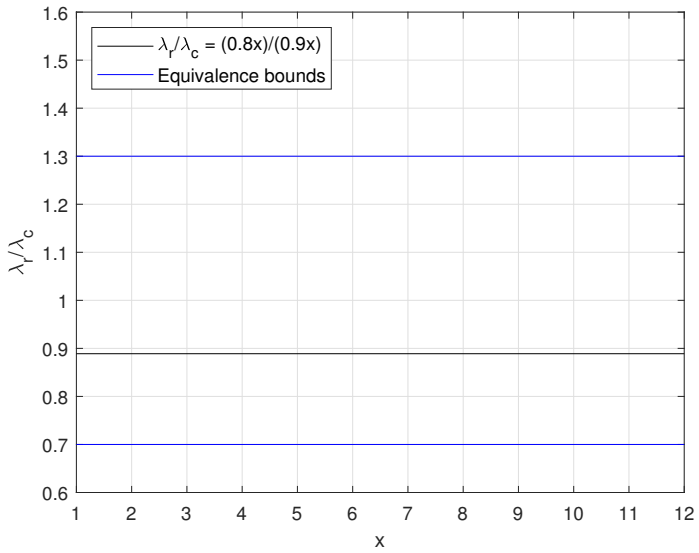
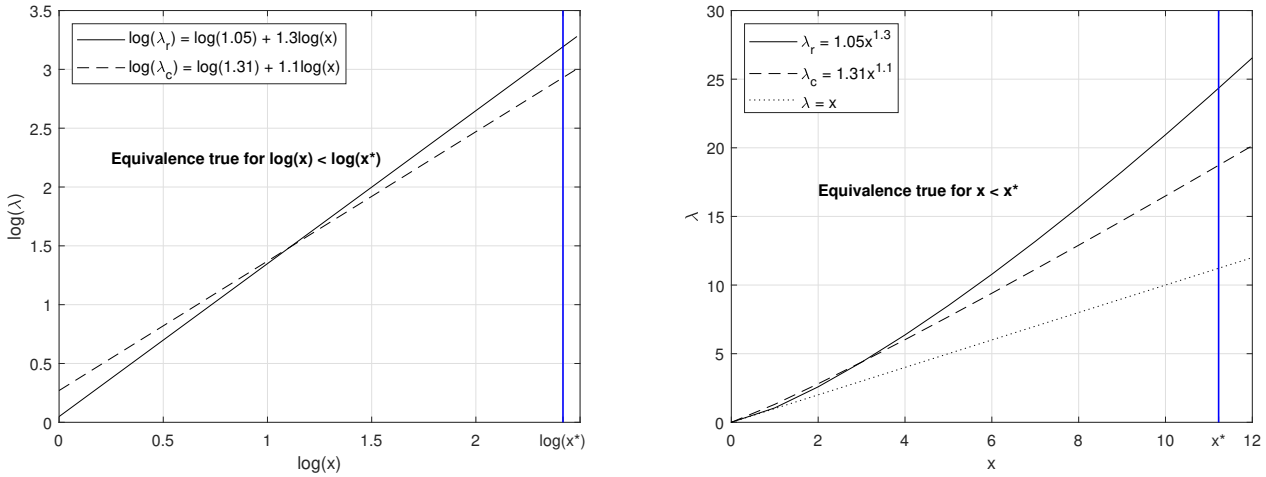


Figure A.4: λ_r/λ_c versus concentration for the second configuration.

A.3 Third configuration

$$\lambda_r = 1.05x^{1.3} \quad \lambda_c = 1.31x^{1.1}$$



(a) Log scale. **(b)** Original scale.

Figure A.5: Expected values versus concentration for the third configuration.

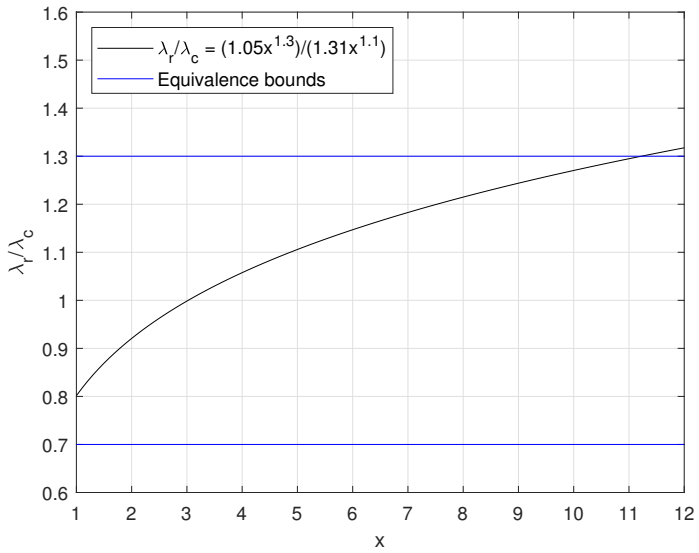
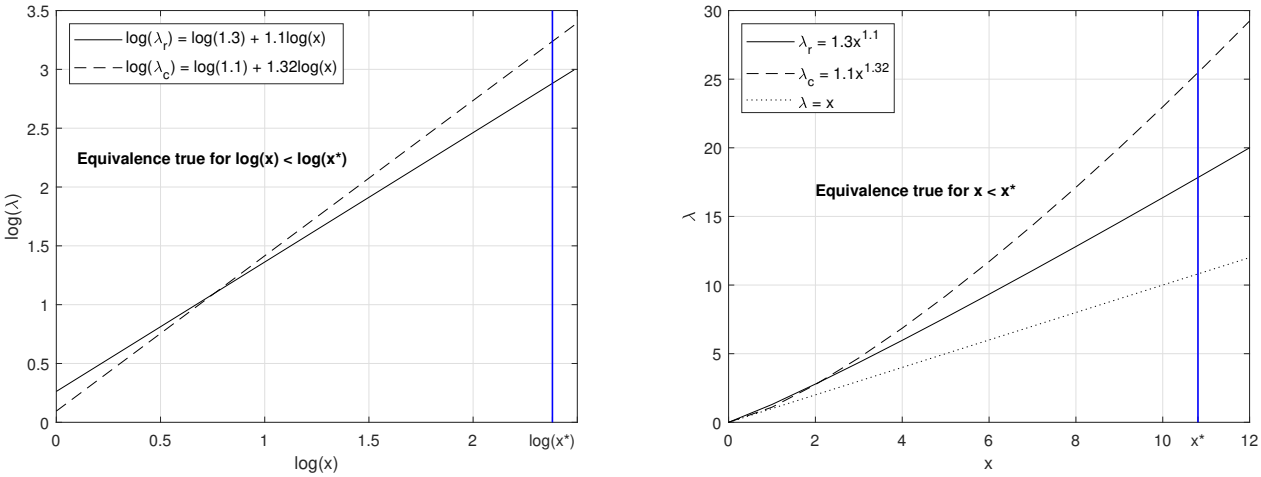


Figure A.6: λ_r/λ_c versus concentration for the third configuration.

A.4 Fourth configuration

$$\lambda_r = 1.3x^{1.1} \quad \lambda_c = 1.1x^{1.32}$$



(a) Log scale. **(b)** Original scale.

Figure A.7: Expected values versus concentration for the fourth configuration.

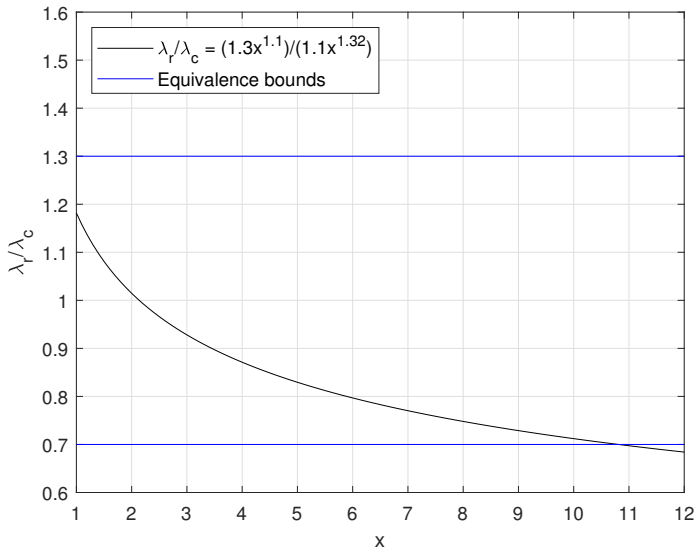
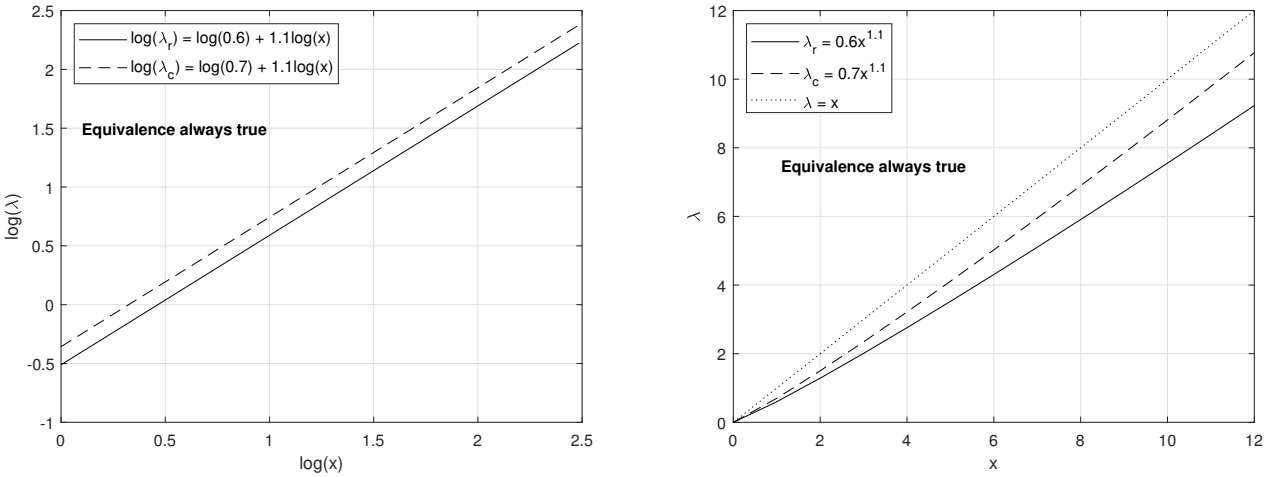


Figure A.8: λ_r/λ_c versus concentration for the fourth configuration.

A.5 Fifth configuration

$$\lambda_r = 0.6x^{1.1} \quad \lambda_c = 0.7x^{1.1}$$



(a) Log scale. **(b)** Original scale.

Figure A.9: Expected values versus concentration for the fifth configuration.

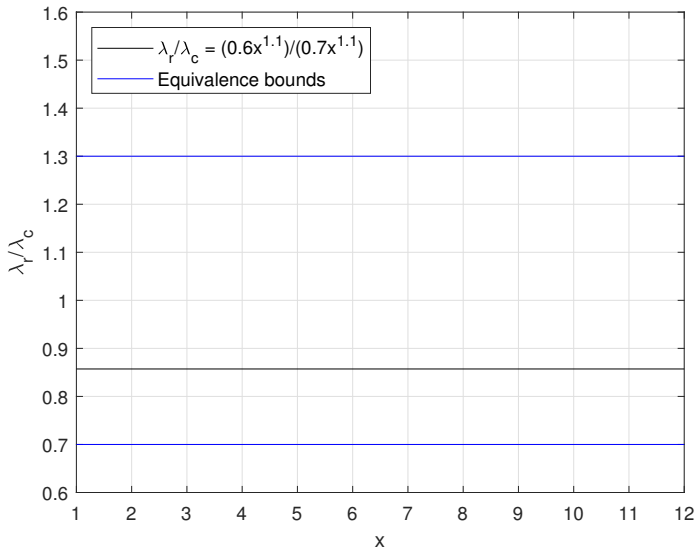


Figure A.10: λ_r/λ_c versus concentration for the fifth configuration.

A.6 Sixth configuration

$$\lambda_r = 0.2 + 0.8x \quad \lambda_c = 0.01 + 0.9x$$

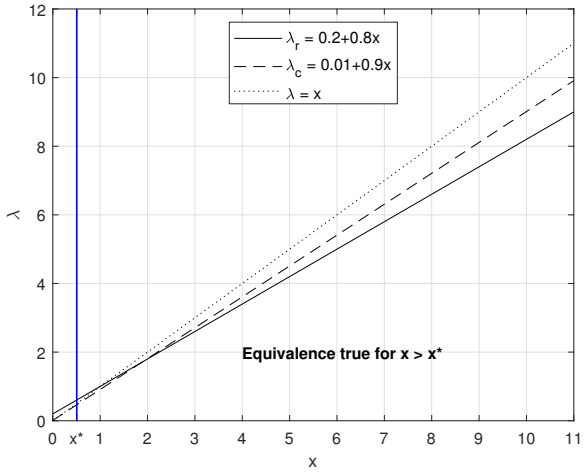


Figure A.11: Expected values versus concentration for the sixth configuration.

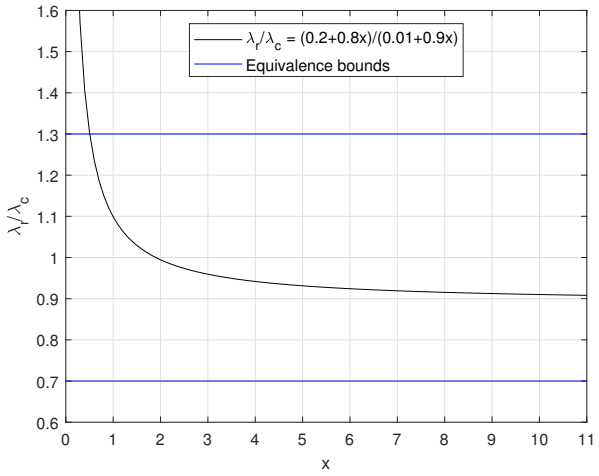


Figure A.12: λ_r/λ_c versus concentration for the sixth configuration.

A.7 Seventh configuration

$$\lambda_r = 0.35 + 0.8x \quad \lambda_c = 0.94 + 0.7x$$

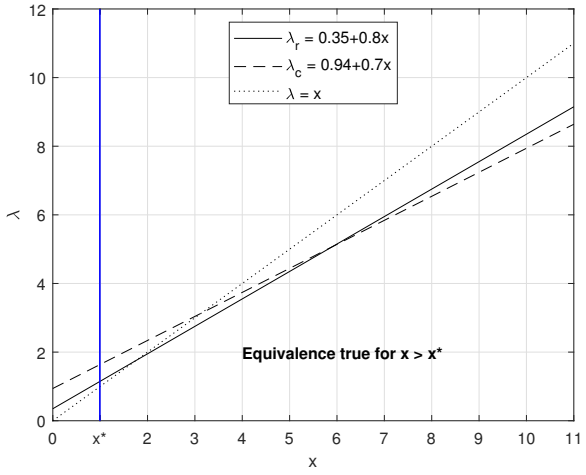


Figure A.13: Expected values versus concentration for the seventh configuration.

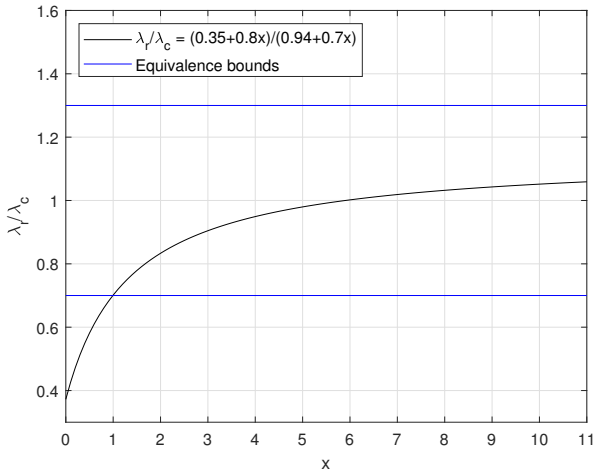


Figure A.14: λ_r/λ_c versus concentration for the seventh configuration.

Appendix B

Asymptotic covariance matrix of MLE in Poisson regression

This appendix contains detailed computations to derive the expression of the asymptotic covariance matrix of Maximum Likelihood estimators of model coefficients in the models used throughout the text. As described in Subsection 2.2.2, the vector $\hat{\boldsymbol{\beta}}_N$ of ML coefficients' estimators in Poisson regression has the approximate Normal distribution

$$\hat{\boldsymbol{\beta}}_N \overset{approx.}{\sim} \mathcal{N}_2(\boldsymbol{\beta}, \mathcal{I}_N(\boldsymbol{\beta})^{-1})$$

where $\mathcal{I}_N(\boldsymbol{\beta})$ is the Fisher information matrix. Both the vector of coefficients' estimators and the Fisher information matrix depend on the number of observations N used to estimate the model. This approximate distribution is a consequence of the convergence in distribution (2.12): for this reason, the inverse of the Fisher information matrix is referred to as *asymptotic* covariance matrix. This appendix essentially shows how to derive an expression for this matrix. In particular, the first subsection deals with the computations in the case of a linear model in the original scale, namely Poisson regression with identity link function, while the second section is about a linear model in the log scale, that is, Poisson regression with canonical (log) link function.

B.1 Linear model in the original scale

Using the linear model in the original scale, the number of microorganisms Y measured in a sample with theoretical concentration x is modelled in the following way:

$$Y \sim Pois(\lambda) \quad \text{with} \quad \lambda = \alpha + \beta x.$$

Suppose that p concentrations $x_1, \dots, x_p$ are used to estimate the model, with n_i replicates at concentration $x_i \quad \forall i = 1, \dots, p$, and let $\lambda_i = \alpha + \beta x_i$ denote the expected value at concentration x_i. The total number of observations used to estimate the model is $N = \sum_{i=1}^{p} n_i$. Finally, let $Y_i^{(j)}$ be the number of microorganisms measured at concentration x_i at the j-th experiment: according to this model, $Y_i^{(j)} \sim Pois(\lambda_i)$ for any $j = 1, \dots, n_i$ and for any $i = 1, \dots, p$, and all the measurements are independent of each other. The random vector $\boldsymbol{Y}$ collects all the replicated measurements at all concentrations.

The Fisher information matrix can be expressed as the expectation of minus the Hessian matrix of the log likelihood.

$$\mathcal{I}_N(\alpha, \beta) = \begin{pmatrix} \mathbb{E}\left[-\frac{\partial^2}{\partial \alpha^2} L_N(\alpha, \beta; \boldsymbol{Y})\right] & \mathbb{E}\left[-\frac{\partial^2}{\partial \alpha \partial \beta} L_N(\alpha, \beta; \boldsymbol{Y})\right] \\ \mathbb{E}\left[-\frac{\partial^2}{\partial \alpha \partial \beta} L_N(\alpha, \beta; \boldsymbol{Y})\right] & \mathbb{E}\left[-\frac{\partial^2}{\partial \beta^2} L_N(\alpha, \beta; \boldsymbol{Y})\right] \end{pmatrix}$$

where $L_N(\alpha, \beta; \boldsymbol{Y})$ is the log likelihood and the expectation is taken with respect to the probability distribution of $\boldsymbol{Y}$. Given $\boldsymbol{y}$ realization of $\boldsymbol{Y}$, the log likelihood function is

$$L_N(\alpha, \beta; \boldsymbol{y}) = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_i^{(j)} \log(\alpha + \beta x_i) - (\alpha + \beta x_i) - \log(y_i^{(j)}!)).$$

In order to compute the Fisher information matrix, it is necessary to calculate the second derivatives of L_N with respect to the model coefficients α and β.

$$\frac{\partial L_N}{\partial \alpha} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(\frac{y_i^{(j)}}{\alpha + \beta x_i} - 1 \right) \quad \frac{\partial L_N}{\partial \beta} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} \left(\frac{y_i^{(j)} x_i}{\alpha + \beta x_i} - x_i \right)$$

$$\frac{\partial^2 L_N}{\partial \alpha^2} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} -\frac{y_i^{(j)}}{(\alpha + \beta x_i)^2} \quad \frac{\partial^2 L_N}{\partial \beta^2} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} -\frac{y_i^{(j)} x_i^2}{(\alpha + \beta x_i)^2} \quad \frac{\partial^2 L_N}{\partial \alpha \partial \beta} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} -\frac{y_i^{(j)} x_i}{(\alpha + \beta x_i)^2}$$

Since $\mathbb{E}[Y_i^{(j)}] = \lambda_i$ for any $j = 1, \ldots, n_i$ and for any $i = 1, \ldots, p$, the Fisher information matrix is given by

$$\mathcal{I}_N(\alpha, \beta) = \sum_{i=1}^{p} \begin{pmatrix} n_i/\lambda_i & n_i x_i/\lambda_i \\ n_i x_i/\lambda_i & n_i x_i^2/\lambda_i \end{pmatrix}.$$

Thus, the asymptotic covariance matrix of the vector of ML coefficients' estimators $(\hat{\alpha}, \hat{\beta})$ is

$$\mathcal{I}_N(\alpha, \beta)^{-1} = \frac{1}{D} \sum_{i=1}^{p} \begin{pmatrix} n_i x_i^2/\lambda_i & -n_i x_i/\lambda_i \\ -n_i x_i/\lambda_i & n_i/\lambda_i \end{pmatrix}$$

where D is the determinant of the Fisher information matrix

$$D = det(\mathcal{I}_N(\alpha, \beta)) = \left(\sum_{i=1}^{p} \frac{n_i}{\alpha + \beta x_i} \right) \left(\sum_{i=1}^{p} \frac{n_i x_i^2}{\alpha + \beta x_i} \right) - \left(\sum_{i=1}^{p} \frac{n_i x_i}{\alpha + \beta x_i} \right)^2.$$

In conclusion, the asymptotic variances of the coefficients' estimators σ_α^2 and σ_β^2 and the asymptotic covariance between the two estimators $\sigma_{\alpha\beta}$ have the following expressions.

$$\sigma_\alpha^2 = \frac{1}{D} \sum_{i=1}^{p} \frac{n_i x_i^2}{\alpha + \beta x_i} \quad \sigma_\beta^2 = \frac{1}{D} \sum_{i=1}^{p} \frac{n_i}{\alpha + \beta x_i} \quad \sigma_{\alpha\beta} = -\frac{1}{D} \sum_{i=1}^{p} \frac{n_i x_i}{\alpha + \beta x_i}$$

The variance of $\hat{\lambda} = \hat{\alpha} + \hat{\beta} x$, which is the estimated expected value at concentration x, is given by the following expression, depending on the variances and covariance of the coefficients' estimators.

$$\sigma_x^2 = \sigma_\alpha^2 + 2x\sigma_{\alpha\beta} + x^2 \sigma_\beta^2 \tag{B.1}$$

B.2 Linear model in the log scale

Using the linear model in the log scale, the number of microorganisms Y measured in a sample with concentration x is modelled in the following way:

$$Y \sim Pois(\lambda) \quad \text{with} \quad \log(\lambda) = \delta + \beta \log(x).$$

Suppose that p concentrations $x_1, \ldots, x_p$ are used to estimate the model, with n_i replicates at concentration $x_i \quad \forall i = 1, \ldots, p$, and let $\lambda_i = e^\delta x_i^\beta$ denote the expected value at concentration x_i. The total number of observations used to estimate the model is $N = \sum_{i=1}^{p} n_i$. Finally, let $Y_i^{(j)}$ be the number of microorganisms measured at concentration x_i at the j-th experiment: according to

this model, $Y_i^{(j)} \sim Pois(\lambda_i)$ for any $j = 1, \dots, n_i$ and for any $i, \dots, p$, and all the measurements are independent of each other. The random vector $\boldsymbol{Y}$ collects all the replicated measurements at all concentrations.

In this case, given $\boldsymbol{y}$ realization of $\boldsymbol{Y}$, the log likelihood function is

$$L_N(\delta, \beta; \boldsymbol{y}) = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_i^{(j)}(\delta + \beta \log(x_i)) - e^{\delta} x_i^{\beta} - \log(y_i^{(j)}!)).$$

The second derivatives of the log likelihood function are given by the following expressions.

$$\frac{\partial L_N}{\partial \delta} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_i^{(j)} - e^{\delta} x_i^{\beta}) \quad \frac{\partial L_N}{\partial \beta} = \sum_{i=1}^{p} \sum_{j=1}^{n_i} (y_i^{(j)} \log(x_i) - e^{\delta} x_i^{\beta} \log(x_i))$$

$$\frac{\partial^2 L_N}{\partial \delta^2} = \sum_{i=1}^{p} (-n_i e^{\delta} x_i^{\beta}) \quad \frac{\partial^2 L_N}{\partial \beta^2} = \sum_{i=1}^{p} (-n_i e^{\delta} x_i^{\beta} \log^2(x_i)) \quad \frac{\partial^2 L_N}{\partial \delta \partial \beta} = \sum_{i=1}^{p} (-n_i e^{\delta} x_i^{\beta} \log(x_i))$$

With this model, the expectation in the definition of the Fisher information matrix has no effect. Thus, the Fisher information matrix is

$$\mathcal{I}_N(\delta, \beta) = \sum_{i=1}^{p} \begin{pmatrix} n_i \lambda_i & n_i \lambda_i \log(x_i) \\ n_i \lambda_i \log(x_i) & n_i \lambda_i \log^2(x_i) \end{pmatrix}$$

and the asymptotic covariance matrix is given by

$$\mathcal{I}_N(\delta, \beta)^{-1} = \frac{1}{D} \sum_{i=1}^{p} \begin{pmatrix} n_i \lambda_i \log^2(x_i) & -n_i \lambda_i \log(x_i) \\ -n_i \lambda_i \log(x_i) & n_i \lambda_i \end{pmatrix}$$

where D is the determinant of the Fisher information matrix

$$D = det(\mathcal{I}_N(\delta, \beta)) = \left(\sum_{i=1}^{p} n_i e^{\delta} x_i^{\beta} \right) \left(\sum_{i=1}^{p} n_i e^{\delta} x_i^{\beta} \log^2(x_i) \right) - \left(\sum_{i=1}^{p} n_i e^{\delta} x_i^{\beta} \log(x_i) \right)^2$$

In conclusion, the asymptotic variances of the coefficients' estimators σ_δ^2 and σ_β^2 and the asymptotic covariance between the two estimators $\sigma_{\delta\beta}$ have the following expressions.

$$\sigma_\delta^2 = \frac{1}{D} \sum_{i=1}^{p} (n_i e^{\delta} x_i^{\beta} \log^2(x_i)) \quad \sigma_\beta^2 = \frac{1}{D} \sum_{i=1}^{p} (n_i e^{\delta} x_i^{\beta}) \quad \sigma_{\delta\beta} = -\frac{1}{D} \sum_{i=1}^{p} (n_i e^{\delta} x_i^{\beta} \log(x_i))$$

The variance of $\widehat{\log(\lambda)} = \hat{\delta} + \hat{\beta} \log(x)$, which is the estimated log expected value at concentration x, is given by the following expression, depending on the variances and covariance of the coefficients' estimators.

$$\sigma_x^2 = \sigma_\delta^2 + 2 \log(x) \sigma_{\delta\beta} + \log^2(x) \sigma_\beta^2 \tag{B.2}$$

Appendix C

Developed programs

This appendix contains the main programs developed to obtain the results described in this thesis. In particular, Section C.1 consists of the SAS code used for the simulations described in Chapter 3: the first subsection contains the code used when linear models in the original scale are chosen to represent the relation between expected numbers of microorganisms counted by the two methods and theoretical concentration, while linear models in the log scale are used in the code reported in the second subsection. Section C.2 contains the MATLAB code used to find the optimal design as described in Chapter 4: the first subsection deals with the case of linear models in the original scale, while the second subsection contains the code for the case of linear models in the log scale.

C.1 SAS programs used for the simulations

C.1.1 Linear models in the original scale

```
DM LOG "CLEAR";
DM OUTPUT "CLEAR";
DM ODSRESULTS "CLEAR";
proc datasets library=WORK kill; run; quit;
options nonotes nosource nosource2 errors=0;

%LET NSIMULATIONS = 1000;
%LET ALPHA = 0.1;  /*2 TIMES THE PROBABILITY OF TYPE 1 ERROR*/

%MACRO COMPUTE_CI_MODEL (DATASET = ); /*MODEL-BASED APPROACH*/
  DATA DATASET;
    SET &DATASET;
  RUN;

  ODS SELECT NONE;
  PROC NLMIXED DATA=DATASET;
    PARMS BR=0.5 BC=0.5 AR=0.5 AC=0.5;
    MU = (METHOD=1)*(AR+BR*X) + (METHOD=2)*(AC+BC*X);
    PREDICT AR+BR*X OUT = EXPECTED_RMM;
    PREDICT AC+BC*X OUT = EXPECTED_COMP;
    MODEL MEASURED~POISSON(MU);
  RUN;
  ODS SELECT ALL;

  PROC SORT DATA = EXPECTED_RMM(DROP = MEASURED METHOD) OUT = EXPECTED_RMM (KEEP = X
       CORRECT_DECISION PRED STDERRPRED) NODUPKEY;
    BY X;
  RUN;

  PROC SORT DATA = EXPECTED_COMP(DROP = MEASURED METHOD) OUT = EXPECTED_COMP (KEEP = X
       CORRECT_DECISION PRED STDERRPRED) NODUPKEY;
    BY X;
```

```
  RUN;

  DATA EXPECTED;
    MERGE EXPECTED_RMM (RENAME = (PRED = PRED1 STDERRPRED = STDE1))
        EXPECTED_COMP (RENAME = (PRED = PRED2 STDERRPRED = STDE2));
    BY X;
  RUN;

  DATA INTERVALS_RATIO;
    SET EXPECTED;
    LOWER = PRED1/PRED2 - QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(STDE1**2/PRED2**2+PRED1**2/PRED2**4*
         STDE2**2);
    UPPER = PRED1/PRED2 + QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(STDE1**2/PRED2**2+PRED1**2/PRED2**4*
         STDE2**2);
    IF LOWER>=0.7 AND UPPER<=1.3 THEN DO;
      DECISION = 1;
    END;
    ELSE DO;
      DECISION = 0;
    END;
    IF DECISION = CORRECT_DECISION THEN DO;
      CORRECT = 1;
    END;
    ELSE DO;
      CORRECT = 0;
    END;
    KEEP X CORRECT_DECISION LOWER UPPER DECISION CORRECT;
  RUN;
%MEND;

%MACRO COMPUTE_CI_BIN (DATASET = ); /*BINOMIAL APPROACH*/
  DATA DATASET;
    SET &DATASET;
  RUN;

  PROC SORT DATA = DATASET;
    BY X METHOD;
  RUN;

  PROC MEANS DATA = DATASET NOPRINT;
    VAR MEASURED;
    BY X METHOD TRUE_RATIO CORRECT_DECISION;
    OUTPUT OUT = SUMS_PER_CONC_METHOD SUM(MEASURED) = SUM_MEAS;
  RUN;

  DATA SUMS_PER_CONC_METHOD;
    SET SUMS_PER_CONC_METHOD;
    DROP _TYPE_ _FREQ_;
  RUN;

  PROC FREQ DATA = SUMS_PER_CONC_METHOD NOPRINT;
    BY X TRUE_RATIO CORRECT_DECISION;
    TABLES METHOD / BINOMIAL(WILSON) ALPHA = &ALPHA;
    WEIGHT SUM_MEAS / ZEROS;
    OUTPUT OUT = BINOMIAL_TEST BIN;
  RUN;

  DATA BINOMIAL_TEST;
    SET BINOMIAL_TEST;
    KEEP X TRUE_RATIO CORRECT_DECISION L_W_BIN U_W_BIN;
    RENAME L_W_BIN = LOW_WILS U_W_BIN = UP_WILS;
  RUN;

  DATA INTERVALS_RATIO;
    SET BINOMIAL_TEST;
    LOW_RATIO = LOW_WILS/(1-LOW_WILS);
    UP_RATIO = UP_WILS/(1-UP_WILS);
```

```
    IF LOW_RATIO>=0.7 AND UP_RATIO<=1.3 THEN DO;
      DECISION = 1;
    END;
    ELSE DO;
      DECISION = 0;
    END;
    IF DECISION = CORRECT_DECISION THEN DO;
      CORRECT = 1;
    END;
    ELSE DO;
      CORRECT = 0;
    END;
    KEEP X CORRECT_DECISION LOW_RATIO UP_RATIO DECISION CORRECT;
  RUN;
%MEND;

%MACRO COMPUTE_CI_DELTA (DATASET = ); /*DELTA APPROACH*/
  DATA DATASET;
    SET &DATASET;
  RUN;

  PROC SORT DATA = DATASET;
    BY X METHOD;
  RUN;

  PROC MEANS DATA = DATASET NOPRINT;
    VAR MEASURED;
    BY X METHOD CORRECT_DECISION;
    OUTPUT OUT = MEANS_PER_CONC_METHOD MEAN(MEASURED) = MEAN_MEAS N = N_MEAS;
  RUN;

  DATA MEANS_PER_CONC_METHOD;
    SET MEANS_PER_CONC_METHOD;
    DROP _TYPE_ _FREQ_;
  RUN;

  DATA MEANS_METHODS_PER_CONC;
    MERGE MEANS_PER_CONC_METHOD (WHERE = (METHOD = 1) RENAME = (MEAN_MEAS = MEAN1 N_MEAS = N))
        MEANS_PER_CONC_METHOD (WHERE = (METHOD = 2) RENAME = (MEAN_MEAS = MEAN2 N_MEAS = N2));
    DROP METHOD N2;
  RUN;

  DATA INTERVALS_RATIO;
    SET MEANS_METHODS_PER_CONC;
    IF MEAN2 ^= 0 THEN DO;
      LOWER = MEAN1/MEAN2 - QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(1/N*(MEAN1/MEAN2**2+MEAN1**2/MEAN2
          **3));
      UPPER = MEAN1/MEAN2 + QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(1/N*(MEAN1/MEAN2**2+MEAN1**2/MEAN2
          **3));
      IF LOWER>=0.7 AND UPPER <=1.3 THEN DO;
        DECISION = 1;
      END;
      ELSE DO;
        DECISION = 0;
      END;
    END;
    ELSE DO;
      LOWER = .;
      UPPER = .;
      IF MEAN1 = 0 THEN DO;
        DECISION = 1;
      END;
      ELSE DO;
        DECISION = 0;
      END;
    END;
    IF DECISION = CORRECT_DECISION THEN DO;
```

```
      CORRECT = 1;
    END;
    ELSE DO;
      CORRECT = 0;
    END;
  RUN;
%MEND;

DATA FINAL_INFO_MODEL;
  INPUT X COUNT;
  DATALINES;
  0 0
  1 0
  2 0
  3 0
  4 0
  5 0
  6 0
  7 0
  8 0
  9 0
  10 0
  11 0
  ;
RUN;
DATA FINAL_INFO_BIN;
  INPUT X COUNT;
  DATALINES;
  0 0
  1 0
  2 0
  3 0
  4 0
  5 0
  6 0
  7 0
  8 0
  9 0
  10 0
  11 0
  ;
RUN;
DATA FINAL_INFO_DELTA;
  INPUT X COUNT;
  DATALINES;
  0 0
  1 0
  2 0
  3 0
  4 0
  5 0
  6 0
  7 0
  8 0
  9 0
  10 0
  11 0
  ;
RUN;

/*SIMULATE THE DATASETS AND PERFORM THE DIFFERENT APPROACHES*/
/*LINEAR MODELS IN THE LOG SCALE: LAMBDA_R = AR + BR*x
                LAMBDA_C = AC + BC*x*/
%MACRO DECISION_EVALUATION(NSIM=, AR=, BR=, AC=, BC=,);
  %DO i = 1 %TO &NSIM;
    DATA DATASET (DROP = REP);
      CALL STREAMINIT(&i);
```

```
    DO X = 0 TO 11;
      TRUE_RATIO = (&AR+&BR*X)/(&AC+&BC*X);
      IF TRUE_RATIO > 0.7 AND TRUE_RATIO < 1.3 THEN DO;
        CORRECT_DECISION = 1;
      END;
      ELSE DO;
        CORRECT_DECISION = 0;
      END;
      DO METHOD = 1 TO 2;
        DO REP = 1 TO 5;
          IF METHOD = 1 THEN DO;
            MEASURED = RAND("POISSON",&AR+&BR*X);
          END;
          ELSE DO;
            MEASURED = RAND("POISSON",&AC+&BC*X);
          END;
          OUTPUT;
        END;
      END;
    END;
  RUN;

  /*COMPUTE CONFIDENCE INTERVALS AND UPDATE THE NUMBER OF CORRECT DECISIONS*/
  %COMPUTE_CI_MODEL(DATASET = DATASET);
  DATA FINAL_INFO_MODEL;
    MERGE FINAL_INFO_MODEL INTERVALS_RATIO;
    BY X;
    COUNT = COUNT + CORRECT;
    KEEP X COUNT;
  RUN;
  %COMPUTE_CI_BIN(DATASET = DATASET);
  DATA FINAL_INFO_BIN;
    MERGE FINAL_INFO_BIN INTERVALS_RATIO;
    BY X;
    COUNT = COUNT + CORRECT;
    KEEP X COUNT;
  RUN;
  %COMPUTE_CI_DELTA(DATASET = DATASET);
  DATA FINAL_INFO_DELTA;
    MERGE FINAL_INFO_DELTA INTERVALS_RATIO;
    BY X;
    COUNT = COUNT + CORRECT;
    KEEP X COUNT;
  RUN;
 %END;
%MEND;

%DECISION_EVALUATION(NSIM=&NSIMULATIONS,AR=0.35,BR=0.8,AC=0.94,BC=0.7);
PROC PRINT DATA = FINAL_INFO_MODEL;
 TITLE "MODEL";
RUN;
PROC PRINT DATA = FINAL_INFO_BIN;
 TITLE "BINOMIAL";
RUN;
PROC PRINT DATA = FINAL_INFO_DELTA;
 TITLE "DELTA";
RUN;
```

C.1.2 Linear models in the log scale

```
DM LOG "CLEAR";
DM OUTPUT "CLEAR";
DM ODSRESULTS "CLEAR";
proc datasets library=WORK kill; run; quit;
options nonotes nosource nosource2 errors=0;
```

```
%LET NSIMULATIONS = 1000;
%LET ALPHA = 0.1;  /*2 TIMES THE PROBABILITY OF TYPE 1 ERROR*/

%MACRO COMPUTE_CI_MODEL (DATASET = ); /*MODEL-BASED APPROACH*/
  DATA DATASET;
    SET &DATASET;
    LOG_X = LOG(X);
  RUN;

  PROC SORT DATA = DATASET OUT = DATASET;
    BY METHOD;
  RUN;

  ODS SELECT NONE;
  PROC GENMOD DATA = DATASET;
    BY METHOD;
    MODEL MEASURED = LOG_X / DIST = POISSON;
    OUTPUT OUT = ESTIMATES STDXBETA = STDE XBETA = PRED_LOG;
  RUN;
  ODS SELECT ALL;

  DATA ESTIMATES;
    SET ESTIMATES;
    DROP MEASURED;
  RUN;

  PROC SORT DATA = ESTIMATES OUT = ESTIMATES NODUPKEY;
    BY X METHOD;
  RUN;

  DATA ESTIMATES;
    MERGE ESTIMATES (WHERE = (METHOD = 1) RENAME = (PRED_LOG = PRED_LOG1 STDE = STDE1))
        ESTIMATES (WHERE = (METHOD = 2) RENAME = (PRED_LOG = PRED_LOG2 STDE = STDE2));
    BY X LOG_X;
    DROP METHOD;
  RUN;

  /*COMPUTE THE CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN THE LOG ESTIMATED EXPECTED COUNTS*/
  DATA INTERVALS_LOG_SCALE;
    SET ESTIMATES;
    LOWER = (PRED_LOG1 - PRED_LOG2) - QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(STDE1**2 + STDE2**2);
    UPPER = (PRED_LOG1 - PRED_LOG2) + QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(STDE1**2 + STDE2**2);
  RUN;

  /*MOVE CONFIDENCE INTERVALS TO THE ORIGINAL SCALE*/
  DATA INTERVALS_RATIO;
    SET INTERVALS_LOG_SCALE;
    LOWER_ORIG = EXP(LOWER);
    UPPER_ORIG = EXP(UPPER);
    IF LOWER_ORIG>=0.7 AND UPPER_ORIG<=1.3 THEN DO;
      DECISION = 1;
    END;
    ELSE DO;
      DECISION = 0;
    END;
    IF DECISION = CORRECT_DECISION THEN DO;
      CORRECT = 1;
    END;
    ELSE DO;
      CORRECT = 0;
    END;
    KEEP X CORRECT_DECISION LOWER UPPER LOWER_ORIG UPPER_ORIG DECISION CORRECT;
  RUN;
%MEND;

%MACRO COMPUTE_CI_BIN (DATASET = ); /*BINOMIAL APPROACH*/
```

```
  DATA DATASET;
    SET &DATASET;
  RUN;

  PROC SORT DATA = DATASET;
    BY X METHOD;
  RUN;

  PROC MEANS DATA = DATASET NOPRINT;
    VAR MEASURED;
    BY X METHOD TRUE_RATIO CORRECT_DECISION;
    OUTPUT OUT = SUMS_PER_CONC_METHOD SUM(MEASURED) = SUM_MEAS;
  RUN;

  DATA SUMS_PER_CONC_METHOD;
    SET SUMS_PER_CONC_METHOD;
    DROP _TYPE_ _FREQ_;
  RUN;

  PROC FREQ DATA = SUMS_PER_CONC_METHOD NOPRINT;
    BY X TRUE_RATIO CORRECT_DECISION;
    TABLES METHOD / BINOMIAL(WILSON) ALPHA = &ALPHA;
    WEIGHT SUM_MEAS / ZEROS;
    OUTPUT OUT = BINOMIAL_TEST BIN;
  RUN;

  DATA BINOMIAL_TEST;
    SET BINOMIAL_TEST;
    KEEP X TRUE_RATIO CORRECT_DECISION L_W_BIN U_W_BIN;
    RENAME L_W_BIN = LOW_WILS U_W_BIN = UP_WILS;
  RUN;

  DATA INTERVALS_RATIO;
    SET BINOMIAL_TEST;
    LOW_RATIO = LOW_WILS/(1-LOW_WILS);
    UP_RATIO = UP_WILS/(1-UP_WILS);
    IF LOW_RATIO>=0.7 AND UP_RATIO<=1.3 THEN DO;
      DECISION = 1;
    END;
    ELSE DO;
      DECISION = 0;
    END;
    IF DECISION = CORRECT_DECISION THEN DO;
      CORRECT = 1;
    END;
    ELSE DO;
      CORRECT = 0;
    END;
    KEEP X CORRECT_DECISION LOW_RATIO UP_RATIO DECISION CORRECT;
RUN;
%MEND;

%MACRO COMPUTE_CI_DELTA (DATASET = ); /*DELTA APPROACH*/
  DATA DATASET;
    SET &DATASET;
  RUN;

  PROC SORT DATA = DATASET;
    BY X METHOD;
  RUN;

  PROC MEANS DATA = DATASET NOPRINT;
    VAR MEASURED;
    BY X METHOD CORRECT_DECISION;
    OUTPUT OUT = MEANS_PER_CONC_METHOD MEAN(MEASURED) = MEAN_MEAS N = N_MEAS;
  RUN;
```

```
  DATA MEANS_PER_CONC_METHOD;
    SET MEANS_PER_CONC_METHOD;
    DROP _TYPE_ _FREQ_;
  RUN;

  DATA MEANS_METHODS_PER_CONC;
    MERGE MEANS_PER_CONC_METHOD (WHERE = (METHOD = 1) RENAME = (MEAN_MEAS = MEAN1 N_MEAS = N))
        MEANS_PER_CONC_METHOD (WHERE = (METHOD = 2) RENAME = (MEAN_MEAS = MEAN2 N_MEAS = N2));
    DROP METHOD N2;
  RUN;

  DATA INTERVALS_RATIO;
    SET MEANS_METHODS_PER_CONC;
    IF MEAN2 ^= 0 THEN DO;
      LOWER_LOG = LOG(MEAN1)-LOG(MEAN2) - QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(1/N*(1/MEAN1+1/MEAN2));
      UPPER_LOG = LOG(MEAN1)-LOG(MEAN2) + QUANTILE("NORMAL",1-&ALPHA/2)*SQRT(1/N*(1/MEAN1+1/MEAN2));
      LOWER = EXP(LOWER_LOG);
      UPPER = EXP(UPPER_LOG);
      IF LOWER>=0.7 AND UPPER<=1.3 THEN DO;
        DECISION = 1;
      END;
      ELSE DO;
        DECISION = 0;
      END;
    END;
    ELSE DO;
      LOWER = .;
      UPPER = .;
      IF MEAN1 = 0 THEN DO;
        DECISION = 1;
      END;
      ELSE DO;
        DECISION = 0;
      END;
    END;
    IF DECISION = CORRECT_DECISION THEN DO;
      CORRECT = 1;
    END;
    ELSE DO;
      CORRECT = 0;
    END;
  RUN;
%MEND;

DATA FINAL_INFO_MODEL;
  INPUT X COUNT;
  DATALINES;
  1 0
  2 0
  3 0
  4 0
  5 0
  6 0
  7 0
  8 0
  9 0
  10 0
  11 0
  12 0
  ;
RUN;
DATA FINAL_INFO_BIN;
  INPUT X COUNT;
  DATALINES;
  1 0
  2 0
  3 0
```

```
  4 0
  5 0
  6 0
  7 0
  8 0
  9 0
  10 0
  11 0
  12 0
  ;
RUN;
DATA FINAL_INFO_DELTA;
  INPUT X COUNT;
  DATALINES;
  1 0
  2 0
  3 0
  4 0
  5 0
  6 0
  7 0
  8 0
  9 0
  10 0
  11 0
  12 0
  ;
RUN;

/*SIMULATE THE DATASETS AND PERFORM THE DIFFERENT APPROACHES*/
/*LINEAR MODELS IN THE LOG SCALE: LAMBDA_R = AR*x^BR
                  LAMBDA_C = AC*x^BC*/
%MACRO DECISION_EVALUATION(NSIM=, AR=, BR=, AC=, BC=,);
  %DO i = 1 %TO &NSIM;
    DATA DATASET (DROP = REP);
      CALL STREAMINIT(&i);
      DO X = 1 TO 12;
        TRUE_RATIO = &AR*X**&BR/(&AC*X**&BC);
        IF TRUE_RATIO > 0.7 AND TRUE_RATIO < 1.3 THEN DO;
          CORRECT_DECISION = 1;
        END;
        ELSE DO;
          CORRECT_DECISION = 0;
        END;
        DO METHOD = 1 TO 2;
          DO REP = 1 TO 5;
            IF METHOD = 1 THEN DO;
              MEASURED = RAND("POISSON",&AR*X**&BR);
            END;
            ELSE DO;
              MEASURED = RAND("POISSON",&AC*X**&BC);
            END;
            OUTPUT;
          END;
        END;
      END;
    RUN;

    /*COMPUTE CONFIDENCE INTERVALS AND UPDATE THE NUMBER OF CORRECT DECISIONS*/
    %COMPUTE_CI_MODEL(DATASET = DATASET);
    DATA FINAL_INFO_MODEL;
      MERGE FINAL_INFO_MODEL INTERVALS_RATIO;
      BY X;
      COUNT = COUNT + CORRECT;
      KEEP X COUNT;
    RUN;
    %COMPUTE_CI_BIN(DATASET = DATASET);
```

```
    DATA FINAL_INFO_BIN;
      MERGE FINAL_INFO_BIN INTERVALS_RATIO;
      BY X;
      COUNT = COUNT + CORRECT;
      KEEP X COUNT;
    RUN;
    %COMPUTE_CI_DELTA(DATASET = DATASET);
    DATA FINAL_INFO_DELTA;
      MERGE FINAL_INFO_DELTA INTERVALS_RATIO;
      BY X;
      COUNT = COUNT + CORRECT;
      KEEP X COUNT;
    RUN;
  %END;
%MEND;

%DECISION_EVALUATION(NSIM=&NSIMULATIONS,AR=1.2,BR=1,AC=1.1,BC=1);
PROC PRINT DATA = FINAL_INFO_MODEL;
  TITLE "MODEL";
RUN;
PROC PRINT DATA = FINAL_INFO_BIN;
  TITLE "BINOMIAL";
RUN;
PROC PRINT DATA = FINAL_INFO_DELTA;
  TITLE "DELTA";
RUN;
```

C.2 MATLAB programs used for the optimal design

C.2.1 Linear models in the original scale

```
clear all
close all
clc
%NOTE: after pasting the code in MATLAB, cancel and rewrite the single quotes

%Coefficients in the models lambda_r = ar+br*x lambda_c = ac+bc*x
ar = 0.35;
br = 0.8;
ac = 0.94;
bc = 0.7;

N = 60;    %Total number of experiments
X = 0:11;   %Concentrations that can be used to estimate the model
p = length(X);  %Number of concentrations
X_eq = 1:11;   %Concentrations where equivalence holds
alpha = 0.05;  %Probability type 1 error

%Determinants of the Fisher information matrix in the models for the two methods
Dr = @(n,x) sum(n./(ar+br*x))*sum(n.*(x.^2)./(ar+br*x))-(sum(n.*(x./(ar+br*x))))^2;
Dc = @(n,x) sum(n./(ac+bc*x))*sum(n.*(x.^2)./(ac+bc*x))-(sum(n.*(x./(ac+bc*x))))^2;

%Variance of the estimator of lambda_r
Vr = @(n,x) 1/Dr(n,x)*(sum(n.*(x.^2)./(ar+br*x))+X_eq.^2*sum(n./(ar+br*x)) -...
                    2*X_eq*sum(n.*(x./(ar+br*x))));
%Variance of the estimator of lambda_c
Vc = @(n,x) 1/Dc(n,x)*(sum(n.*(x.^2)./(ac+bc*x))+X_eq.^2*sum(n./(ac+bc*x)) -...
                    2*X_eq*sum(n.*(x./(ac+bc*x))));
%Variance of the ratio between the two estimated expected counts
V = @(n,x) Vr(n,x)./((ac+bc*X_eq).^2) + ((ar+br*X_eq).^2)./((ac+bc*X_eq).^4).*Vc(n,x);

%Statistical power
P = @(n,x) -sum(max(normcdf((1.3+norminv(alpha)*sqrt(V(n,x))-(ar+br*X_eq)./(ac+bc*X_eq))./...
                                (sqrt(V(n,x)))) - ...
```

```
                normcdf((0.7-norminv(alpha)*sqrt(V(n,x))-(ar+br*X_eq)./(ac+bc*X_eq))./...
                                (sqrt(V(n,x)))),0));

results = cell(length(X)-1,3);
j = 1;
for k = 2:length(X)
   val = zeros(nchoosek(p,k),1);
   options = nchoosek(X,k);
   replicates = zeros(nchoosek(p,k),k);
   for i = 1:size(options,1)
        opts = optimoptions('ga','Display','off');
        lb = ones(1,k);
        ub = (N-1)*ones(1,k);
        rng(1);
        IntCon = 1:k;
        obj = @(n) P(n,options(i,:));
        [n_best,fval,exitflag] = ga(obj,k,[-ones(1,k);ones(1,k)],[-N;N],[],[],lb,ub,[],IntCon,opts);
        val(i) = -fval;
        replicates(i,:) = n_best;
   end
   [~,index] = max(val);
   results{j,1} = options(index,:);
   results{j,2} = replicates(index,:);
   results{j,3} = val(index);
   j = j + 1;
end
```

C.2.2 Linear models in the log scale

```
clear all
close all
clc
%NOTE: after pasting the code in MATLAB, cancel and rewrite the single quotes

%Coefficients in the models lambda_r = ar*x^br lambda_c = ac*x^bc
ar = 1.2;
br = 1;
ac = 1.1;
bc = 1;

N = 60;    %Total number of experiments
X = 1:12;   %Concentrations that can be used to estimate the model
p = length(X);  %Number of concentrations
X_eq = 1:12;   %Concentrations where equivalence holds
alpha = 0.05;  %Probability type 1 error

%Determinants of the Fisher information matrix in the models for the two methods
Dr = @(n,x) sum(n.*(ar*x.^br))*sum(n.*(ar*x.^br).*(log(x).^2))-(sum(n.*(ar*x.^br).*log(x)))^2;
Dc = @(n,x) sum(n.*(ac*x.^bc))*sum(n.*(ac*x.^bc).*(log(x).^2))-(sum(n.*(ac*x.^bc).*log(x)))^2;

%Variance of the difference between the estimated log expected counts
V = @(n,x) 1/Dr(n,x)*(sum(n.*(ar*x.^br).*(log(x).^2))+log(X_eq).^2*sum(n.*(ar*x.^br))-...
                    2*log(X_eq).*sum(n.*(ar*x.^br).*log(x))) + ...
        1/Dc(n,x)*(sum(n.*(ac*x.^bc).*(log(x).^2))+log(X_eq).^2*sum(n.*(ac*x.^bc))-...
                    2*log(X_eq).*sum(n.*(ac*x.^bc).*log(x)));

%Statistical power
P = @(n,x) -sum(max(normcdf((log(1.3)+norminv(alpha)*sqrt(V(n,x))-...
       (log(ar)-log(ac)+(br-bc)*log(X_eq))./(sqrt(V(n,x)))) - ...
             normcdf((log(0.7)-norminv(alpha)*sqrt(V(n,x))-...
         (log(ar)-log(ac)+(br-bc)*log(X_eq)))./(sqrt(V(n,x)))),0));

results = cell(length(X)-1,3);
j = 1;
for k = 2:length(X)
```

```
   val = zeros(nchoosek(p,k),1);
   options = nchoosek(X,k);
   replicates = zeros(nchoosek(p,k),k);
   for i = 1:size(options,1)
        opts = optimoptions('ga','Display','off');
        lb = ones(1,k);
        ub = (N-1)*ones(1,k);
        rng(1);
        IntCon = 1:k;
        obj = @(n) P(n,options(i,:));
        [n_best,fval,exitflag] = ga(obj,k,[-ones(1,k);ones(1,k)],[-N;N],[],[],lb,ub,[],IntCon,opts);
        val(i) = -fval;
        replicates(i,:) = n_best;
   end
   [~,index] = max(val);
   results{j,1} = options(index,:);
   results{j,2} = replicates(index,:);
   results{j,3} = val(index);
   j = j + 1;
end
```