Statistical Analysis of Bioassay Data With Dependent Replicates

by

Liam Cann

Bachelor of Science, UNB, 2022

A Report Submitted in Partial Fulfilment of
the Requirements for the Degree of

Master of Science

In the Graduate Academic Unit of Mathematics and Statistics

Supervisor:	Connie Stewart, PhD, Department of Mathematics & Statistics Matthew Stephenson, PhD, Department of Mathematics & Statistics, Adjunct
Examining Board:	Dylan Spicker, PhD, Department of Mathematics & Statistics Chris Gray, PhD, Department of Biological Sciences

This report is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

March, 2025

Abstract

Bioassays play a key role in ensuring that every batch of drug produced is safe and effective for release. They play a critical role in testing the potency of biologic drugs, such as vaccines or monoclonal antibodies. Due to the importance of bioassays in ensuring a safe and effective product, it is critical that the statistical methods used are appropriate. However, it is the current common practice to treat the replicate responses at each dose group as if they are independent despite the fact they are often correlated. In this research, we look at quantitatively assessing the risks of the conventional analysis methods using a simulation study to investigate the impact of correlation on the statistical analysis of bioassays. Specifically, we consider parallelism assessment, model goodness-of-fit, and relative potency estimation. We also make recommendations within the constraints of the current available commercial bioassay analysis software to provide valid statistical inference.

Acknowledgements

I would like to express my deep appreciation to my supervisors, Dr. Connie Stewart and Dr. Matthew Stephenson, for their guidance, feedback, and patience. Dr. Connie Stewart introduced me to statistics and provided invaluable guidance as I wrote this report. Dr. Matthew Stephenson brought the weight of his considerable experience and knowledge of bioassays to this project. Without their guidance and help, this report would not have been possible.

I would like to thank Quantics Biostatistics for providing me with the software required to analyze the data, and ultimately completing the report.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

When testing a new batch of a biological drug, it is important to ensure that the batch is safe and effective through the use of bioassays. Bioassays, or biological assays, are essential for the determination of potency and the analysis of these biologics. The potency of a biologic is related both to the concentration (or amount) in preparation, but also its biological activity. The biological activity is related to the structure of the product being tested.

A bioassay is an analytical method based on a measurable response from a test organism, as a result of the biological action of the compound of interest [1]. The biological products, or compounds, could be a biologic drug, a toxin, or any substance that can be induced into a biological system [1, 2]. Examples of the test organisms for these compounds could be animals, such as mice and rats, but more commonly are cell-based. Testing in a biological system is carried out to measure the biological activity of a drug. Depending on the bioassay, the measured response will vary. For example, a possible response could be a change in luminescence that depends on the dosage and the biological activity of the drug administered. Another example could be the survival status of animals at the end of a study. Ascertaining the potency of

the compound is critical to high quality drug manufacturing.

The methods for assessing potency are complex. The potency of a drug can be defined as the amount of the drug needed to produce a certain effect [3]. However, the measurements collected from biological systems are highly variable. For example, the results from one day can be different from the next day, even when using the same set up. Therefore potency is commonly measured relative to a standard to account for this intrinsic variability. The reference standard is often a sample with known characteristics, such as a batch of the drug that has already proven to be effective in humans in a clinical trial. However, the reference standard could also be a batch with little information or understanding of the product, as is common early in a product's lifecycle.

The objective of the assay is to find the potency of the test sample relative to the reference standard. Relative potency (RP) is the ratio of the doses required for the test sample and the reference sample to produce the same effect [3, 4]. Since the potency of the reference sample is typically known, the RP gives us a potency measurement of the test sample. The RP is often used to determine if a batch of drugs is suitable for release; the RP value changes depending on the tested drug [4]. For a drug to be suitable for release the RP value must be unique across all tested doses [4]. The European Pharmacopoeia (8th edition) and the United States Pharmacopia (Chapters 1032, and 1034), help describe the procedures for measuring RP [4, 5, 6, 7].

Let the dose of the reference standard be d_S and the dose of the test sample be d_T, where d_S and d_T yield the same response. The RP is defined as:

$$RP = \frac{d_S}{d_T}. \tag{1.1}$$

Typically, the reference and test samples are tested across a range of doses and the

dose-response relationship is modeled. The dose-response relationship describes how the response to the drug changes with respect to the dose administered, and is used to estimate potency. Figure 1.1 plots the dose-response relationship for an example reference and test sample pair. For the RP calculation to be unique across doses, the RP must be independent of the response level. In practice, the change in response with respect to log(dose) is typically examined. For the reference standard it would be $\log(d_S)$, and for the test sample it would be $\log(d_T)$. Thus, we obtain

$$\log(RP) = \log(d_S) - \log(d_T). \tag{1.2}$$

The log of the RP corresponds to the horizontal distance between the test sample and the reference standard curves at the same response level; see Figure 1.1 (red lines) [4]. If the horizontal distance between the test sample and reference standard is constant, then the samples are parallel. When parallelism occurs between the reference standard curve and the test sample curve, the value of the RP will be the same at all dose levels. This results in the two curves being the same horizontal distance apart at all response levels and thus the test sample has the same RP regardless of response [4]. If parallelism is not achieved, there will not be a unique RP. The two samples being parallel means that the samples act as dilutions of one another [8]. An example depicting a parallel dose-response relationship is shown in Figure 1.1.

When testing the RP, it is important to determine if the dose response curves for the test sample and reference standard are parallel. If the two curves are reasonably parallel, then it is assumed that the test sample is biologically similar to the reference standard. If the test sample does not have the same biologically active components as the reference standard, parallelism may not not achieved, and the RP will vary over the dose range, as shown in Figure 1.2 [4].

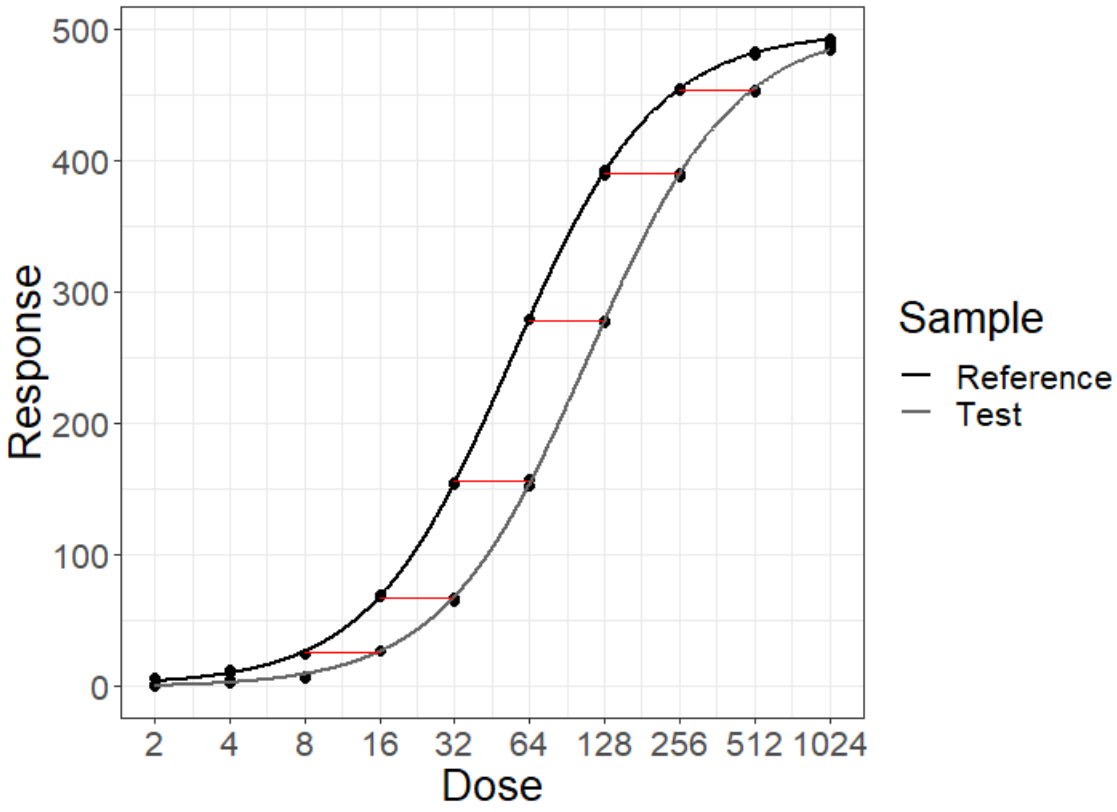

Figure 1.1: A depiction of dose-response relationship over multiple doses showing parallel curves, where there is constant horizontal distance between the curves, and the relative potency is independent of the response level.

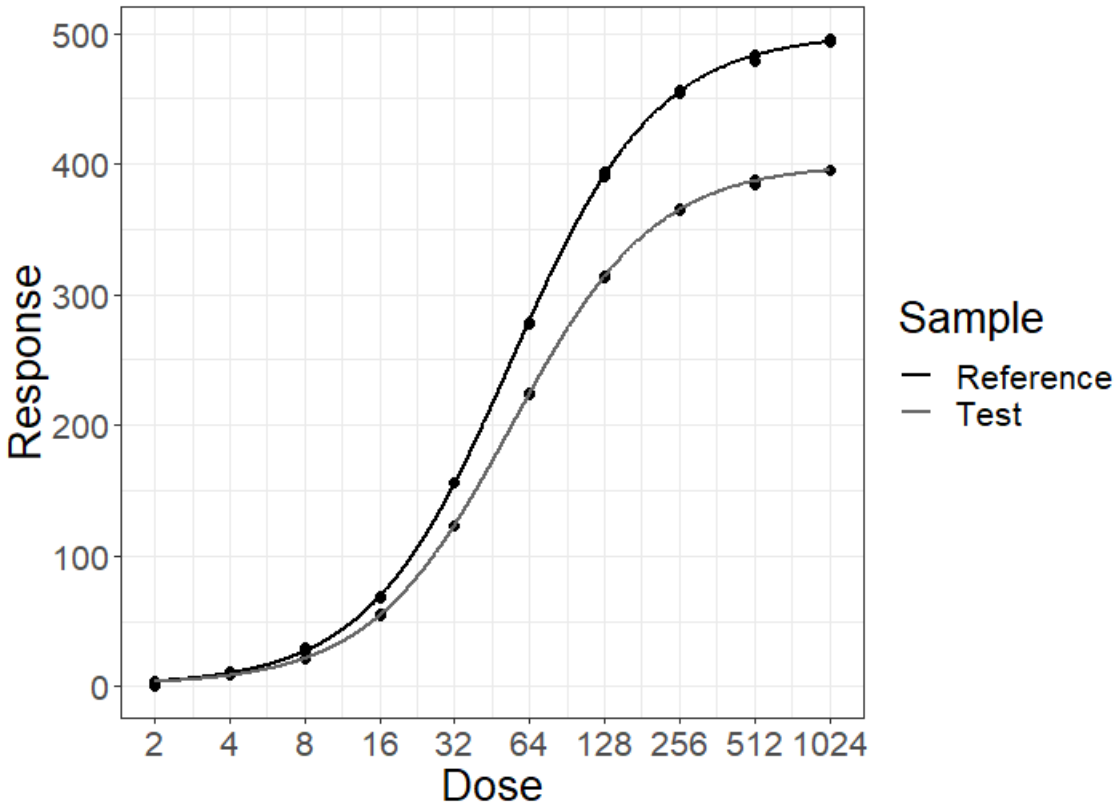

Figure 1.2: A depiction of dose-response relationship over multiple doses showing non-parallel curves, where the horizontal distance between the curves varies, and the relative potency is dependent on the response level.

Various statistical models are used to estimate the RP and these models require certain assumptions to be met. For bioassay analysis, it is conventional to assume that the residuals of the dose-response models are normally distributed, independent, and that there is homogeneity of variance [9].

There are two common methods of preparing replicates for a bioassay: researchers can obtain true replicates by taking separate samples of the stock solution to prepare dilutions for each dilution series, or they can obtain pseudo replicates by using a single sample taken from the stock solution to make up all the dilution series for that sample. The second method will produce correlated results as pseudo replicates will necessarily be dependent, as the replicates come from the same sample. This violates the independence assumption and could impact the results depending on the degree of correlation. This report will investigate the impact of violations to the independence assumption on analyzing the results in bioassays.

In practice, it is common for scientists to ignore structure and analyze the replicates as independent, even when prepared as pseudo-replicates. However, a alternative that may be more appropriate would be to average the dose replicates before the analysis. That is, if multiple test samples with dose replicates are used, then each replicate is averaged with respect to their test sample. This means that for both the reference sample and the test sample, the researcher would average the replicates at each dose level prior to statistical modelling. Averaging the replicates removes any correlation from the data as there would be one response at each dose level per sample instead of multiple. This report will compare the estimates of the RP of assays containing independent replicates, with replicates having varying degrees of correlation but treated as independent, as well as the effect of averaging replicates.

In Chapter 2, we introduce various methods for analyzing dose-response data and the different methods to test for parallelism. In Chapter 3, we present the results of a

simulation study on one of the models to analyze the effect of correlation and impact of averaging. Finally, in Chapter 4 we conclude with a discussion of the results, as well as limitations of the study and alternative approaches that could be explored to handle correlation in the replicates.

Chapter 2

Methodology

In bioassay analysis, various statistical models are used to characterize how the response changes with respect to the dose of the drug. The estimated parameters of the dose-response relationship are then used to estimate relative potency. There are two classes of assay responses, namely continuous and categorical, and the choice of statistical model depends on the class of response [1, 9]. For categorical data the response of the assay is typically a binary response [9]. A common example occurs in the testing for the lethal dose of a toxin where it is recorded if the test animal is dead or alive at the end of study [1]. In this case, researchers would typically use logistic regression and maximum likelihood estimation to fit the binary response, but for this report we do not expand on these models further as the focus is on modeling continuous response.

Least-squares methods are used to characterize the dose-response relationship of continuous assay data, where the sum of the squared vertical distances of the response points to the regression line (residuals) is minimized [9, 10, 11]. In other words, the least squares fit provides the curve where the sum of the residual sum of squares (RSS) is minimized [10]. An example of a continuous response is optical density mea-

sured in nanometers. There are two common possible shapes for the dose response relationship: straight line and S-shaped curves.

In the conduct of a bioassay, it is common to report confidence intervals and perform various hypothesis tests. To ensure valid statistical inference, three assumptions regarding the data are required: the responses are independent, the residuals are normally distributed, and the variability is constant, that is, the variability in the response data is the same for each dose group [9, 11, 12, 13].

There are different methods for dealing with violations to these assumptions.Weighted least-squares can be used for addressing a lack of variance homogeneity and transforming the data can be used in the case of non-normally distributed data. However, in this report our focus is on violations of the independence assumption [10, 11].

2.1 Methods

The dose-response relationship can be modeled as a linear relationship or as a non-linear curve. A fundamental assumption often made in bioassay analysis is that the dose-response relationship typically results in a non-linear S-shaped curve provided the dose range is sufficient [6]. In cases where a drug is made for a specific individual, often very small, limited amounts of material are available to both test and use on the individual. In these instances, it is not practical to observe a full dose-response curve due to limited material. Thus a simplified version of analyzing the data would be to focus on the central linear part of the non-linear curve [6]. We will examine the linear model and slope-ratio model for linear relationships, and the 4 parameter logistic (4PL) and 5 parameter logistic (5PL) models for non-linear S-shaped curves.

2.1.1 Linear Model

The linear regression model is appropriate when the log(dose)-response relationship follows an approximately straight line [7]. In this case, the relationship between the logarithm of the dose and the response is estimated using ordinary least squares regression, often after restricting the dose range to only a few central doses (linear portion) [7]. The linear dose-response model is given by the pair of equations:

$$\text{Reference: } \mathrm{E}(Y_S) = \beta_{0S} + \beta_1 \log(\text{dose}) \tag{2.1}$$

$$\text{Test: } \mathrm{E}(Y_T) = (\beta_{0S} + \Delta_{\beta_0}) + \beta_1 \log(\text{dose}) \tag{2.2}$$

where the subscript "S" denotes the reference standard, Y is the response at the log(dose) value, β_{0S} is the intercept for the reference standard, β_{0T} is the intercept for the test sample, β_1 is the common slope, and Δ_{β_0} is the intercept difference between the test and standard samples [7]. The horizontal distance between the two lines is equal to the log of the RP, and so the RP is expressed by

$$RP = \exp\left(\frac{\Delta_{\beta_0}}{\beta_1}\right). \tag{2.3}$$

2.1.2 Slope-Ratio Model

If the response is linear with respect to the dose, rather than the log(dose), the slope-ratio model can be used. The differences between the linear model and the slope-ratio model is that the slope-ratio model requires that the reference and test samples have the same intercept but different slopes and the relationship between the dose and the response, and not that with the log of the dose [7]. The slope-ratio

model is given by the following equations

$$\text{Reference: } \mathbb{E}(Y_{\text{S}}) = \beta_0 + \beta_{1S}(\text{dose}) \tag{2.4}$$

$$\text{Test: } \mathbb{E}(Y_{\text{T}}) = \beta_0 + \beta_{1T}(\text{dose}) \tag{2.5}$$

where y is the response at the dose value, β_0 is the intercept, β_{1S} is the slope of the reference standard, and β_{1T} is the slope of the test sample [7]. The ratio of the slopes gives the RP, which can then be found by:

$$RP = \frac{\beta_{1T}}{\beta_{1S}}. \tag{2.6}$$

2.1.3 S-Shaped Models

Dose-response data typically form an S-shaped curve. When the S-shaped curve is

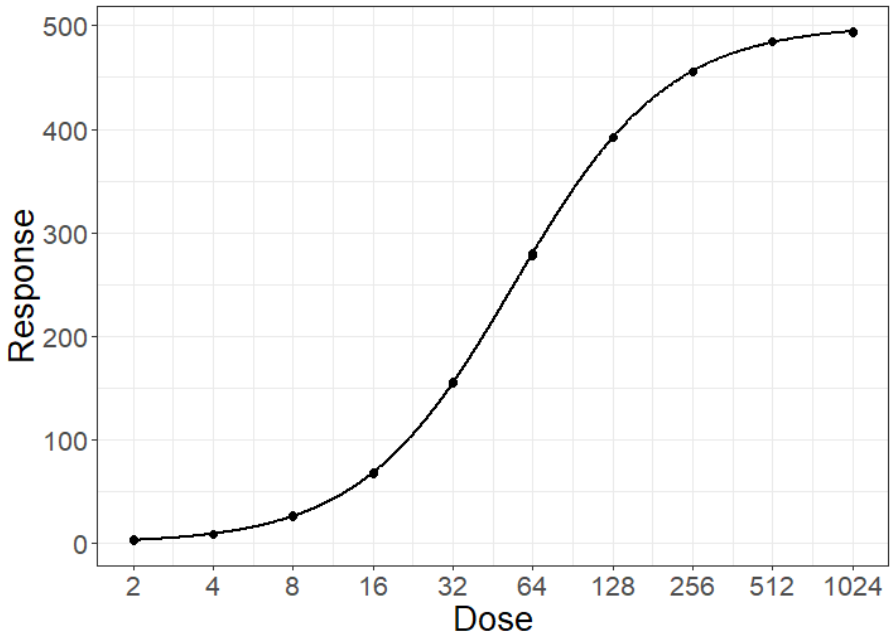

Figure 2.1: A depiction of the 4PL model, a symmetric S-shaped curve.

symmetrical, as shown in Figure 2.1, it can be modeled with the 4PL model, which is

the most common non-linear model used in assay data analysis [9]. The 4PL model is given by

$$\text{Reference: } \mathrm{E}(Y_S) = D + \frac{A - D}{1 + e^{B(log(dose) - C_S)}} \tag{2.7}$$

$$\text{Test: } \mathrm{E}(Y_T) = D + \frac{A - D}{1 + e^{B(log(dose) - C_T)}} \tag{2.8}$$

where y is the response at the log(dose) value [9]. The 4 parameters in the model are: the lower asymptote A, the upper asymptote D, the inflection point (i.e., the point on the curve where the slopes changes concavity) of the curve C, and the slope of the inflection point B [9, 14, 15]. The 4PL model is widely used for its ease in fitting symmetric dose-response curves to the data and reflects the underlying biology [7, 10, 15]. The underlying biology being the cellar mechanism that results in that particular response when the drug is administered [7, 10, 15].

The 5PL model is a commonly used alternative to the 4PL model when the S-shaped dose-response curve is asymmetrical [10, 16]. We define the 5PL model with the equation:

$$\mathbb{E}(Y) = D + \frac{A - D}{[1 + e^{B(log(dose) - C)}]^E} \tag{2.9}$$

where the A, D, C, and B parameters are the same as the 4PL and the additional parameter, E, controls for the degree of asymmetry [10, 15, 16]. If $E = 1$, then the curve is symmetric and identical to the 4PL model, as shown in Figure 2.2. As E gets further from 1, the asymmetry increases. As E increases, the curve near the lower asymptote becomes more pronounced while the curve near the upper asymptote becomes less pronounced, and vice versa as shown in Figure 2.3. Both the 4PL and the 5PL models are fitted using a nonlinear least squares regression. Despite the 5PL model's ability to fit asymmetric data, there is a limit to the amount of asymmetry it can handle, where extreme values of asymmetry will result in a poor fit. The RP

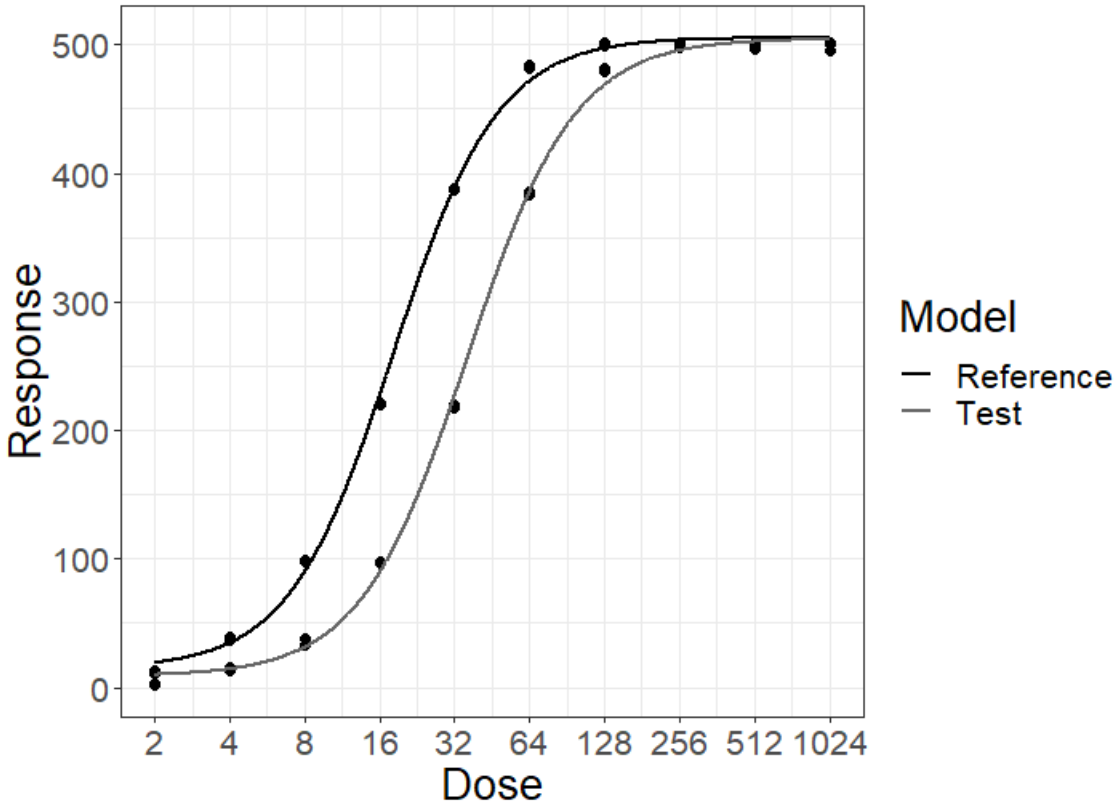

Figure 2.2: A depiction of the 5PL model where $E = 1$, where the standard and test curves are parallel. With $E = 1$, the curves are symmetric.

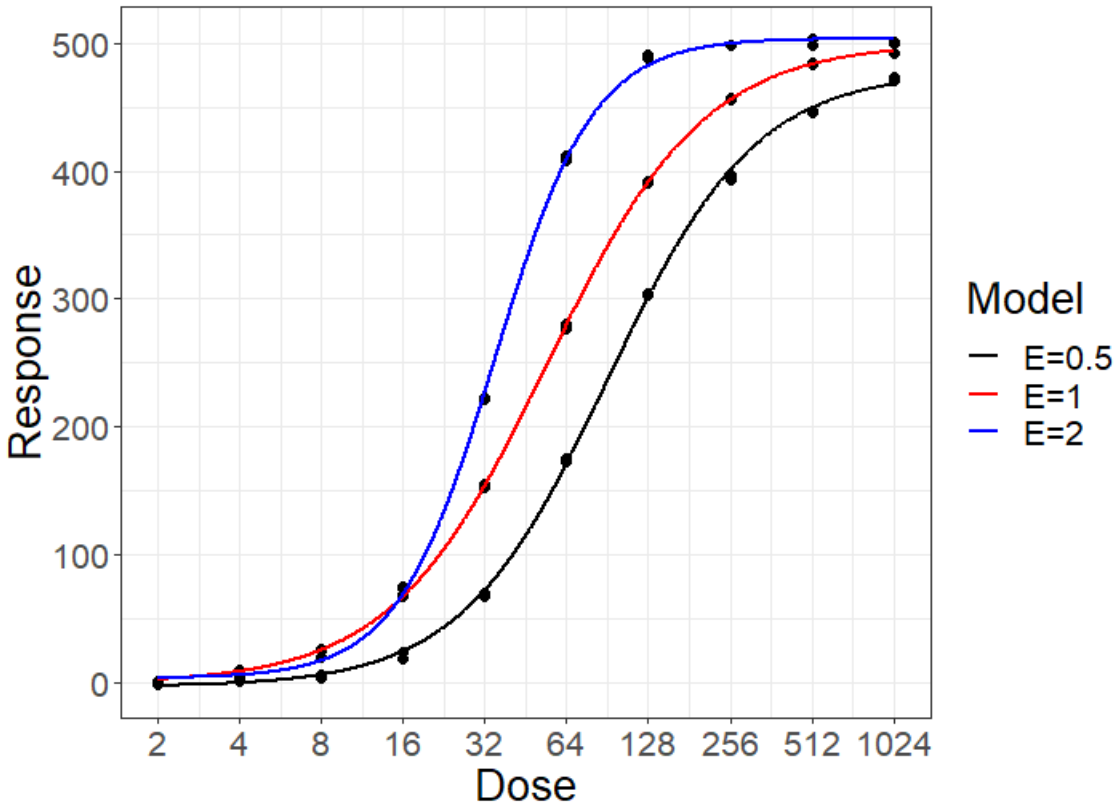

Figure 2.3: Examples of the 5PL model for three different values of the parameter E (degree of symmetry). The center red line in Figure 2.3 shows the fitted curve for $E = 1$, the blue line on top is for when $E = 2$ and the black line on the bottom is for when $E = 0.5$

for both the 4PL and 5PL models is given by the equation:

$$\log(RP) = C_S - C_T \tag{2.10}$$

where C_S is the inflection point of the reference standard, and C_T is the inflection point of the test sample [4].

2.2 Suitability Tests

In this section, we review methods for testing the goodness-of-fit and parallelism of the model. For the goodness-of-fit testing, we will examine how to find the test statistic to determine if model fits the data well. For parallelism testing, we will examine what "difference tests" and "equivalence tests" are and how they are used to determine if the reference standard and test sample have parallel curves.

2.2.1 Goodness-of-fit

Since the RP is a function of the model parameters it is necessary to examine how well the model fits the data using a goodness-of-fit test, namely a lack-of-fit F-test. The lack-of-fit F-test assumes that the observations are independent, the variance is constant across all groups within the assay, and that the residuals are normally distributed [11]. The test statistic for this test depends on the RSS. Given that the total variability of the observations about the fitted curve can be quantified by the RSS, if we denote y_{ij} as the j^{th} response at dose group i, $\hat{y}_i$ the fitted response for dose group i, d the number of dose groups, and n_i the number of replicates in dose

group i, then the RSS is given by:

$$RSS = \sum_{i=1}^{d} \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2$$

where $i = 1, ..., d$, and $j = 1, ..., n_i$ [11].

The total variability of the fitted dose-response relationship can be shown to be the sum of two parts: the lack-of-fit error, and the pure error. The lack-of-fit error represents the variability of the means of each dose group around the fitted dose-response relationship, while the pure error however represents the variability of the response values about their dose group mean [11]. If the model fits the dose-response relationship well, then the distance between the mean of the dose group and the fitted response should be small [11]. Letting $\bar{y}_i$ be the mean of dose group i, then the sum of squares for the lack-of-fit (SSLoF) is:

$$\text{SSLoF} = \sum_{i=1}^{d} n_i \times (\bar{y}_i - \hat{y}_i)^2 \tag{2.11}$$

with $d - p$ degrees of freedom [11]. The error, however, exists entirely due to chance and not from the dose-response relationship [11]. The sum of squares for the pure error (SSPE) is:

$$\text{SSPE} = \sum_{i=1}^{d} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \tag{2.12}$$

with $n - d$ degrees of freedom [11].

Given the two parts of the RSS, we can find the test statistic for the lack-of-fit test corresponding to the null hypothesis of no lack-of-fit. The mean square of the lack-of-fit error (MSLoF) can be found by dividing the SSLoF by its degrees of freedom, and the mean square of the pure error (MSPE) can be found by diving the SSPE by

its degrees of freedom [11]. From this, the test statistic for the lack-of-fit F test is:

$$F = \frac{\sum_{i=1}^{d} n_i \times (\bar{y}_i - \hat{y}_i)^2/(d-p)}{\sum_{i=1}^{d}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2/(n-d)} = \frac{\text{MSLoF}}{\text{MSPE}}. \tag{2.13}$$

The lack-of-fit is considered significant at level α if the test statistic F is greater than the critical value $F_{critical} = F_{1-\alpha,d-p,n-d}$, where $F_{1-\alpha,d-p,n-d}$ is the $(1-\alpha)$ percentile of the F-distribution with degrees of freedom $(d-p)$ and $(n-d)$ [11]. Therefore, if $F > F_{critical}$ then there is sufficient evidence to concluded that the model provides a poor fit to the data. Otherwise, the fitted data passes the goodness-of-fit test, indicating there is insufficient evidence to suggest a lack of fit.

The goodness-of-fit F-test requires there to be replicates at each dose group. However, if the replicates are averaged, for example, to account for replicates that are not independent, the goodness-of-fit test can no longer be used.

2.2.2 Parallelism

It is assumed for bioassays that the reference standard and the test sample are made of the same biologically active components, an assumption known as "similarity" [6]. Since the new batch of drug from which the test sample is derived may become compromised during its production or while in storage, biological similarity is an assumption rather than a guarantee [6]. Should the components be dissimilar, RP will change over dose, resulting in non-parallel curves. While biological similarity can neither be observed nor proven, the curves can be assessed for statistical similarity whereby the parallelism of the dose-response models is examined [6]. However, it stands to recognize that statistical similarity does not equate to biological similarity [6].

If the reference and test samples are statistically similar, then the curves for those

samples will be parallel. For the RP to be unique, the horizontal distance between the test sample and reference sample must be equal at every response level. Parallelism is not always easy to verify nor is it always visually apparent in practice. Two existing approaches for testing for parallelism exist with the first being to assume the curves are parallel and seek contradictory evidence, while the second assumes the curves are not parallel and seeks evidence to prove that the curves are parallel [8]. Tests that assume curves are parallel are known as "difference tests", and tests that assume curves are not parallel are known as "equivalence tests" [8].

A difference test makes use of both a parallel and non-parallel model. The parallel model is a constrained model, where the curves are fit assuming that some of the parameters are shared between the reference and test samples For the 4PL model, parallelism occurs when the A, D, and B parameters are shared and the curves differ only in the parameter C. Thus, only 5 parameters need to be estimated. The non-parallel model is an unconstrained model, where the curves are fit to the reference and test samples separately. For the non-parallelism 4PL model, parameters A, D, and B can vary, and thus requires 8 parameters.

Two examples of difference tests for assessing parallelism are the F-test and chi-squared test. Both tests use the same hypothesis, where the alternative hypothesis is that the curves are not parallel. Thus, we look to see if there is sufficient or insufficient evidence to conclude that the curves are not parallel. The F-test is a well known statistical test for comparing two models when the assumptions of independent observations, constant variance across all dose groups and normally distributed reference and test sample responses are met [8]. The F-statistic is defined as:

$$F = \left(\frac{\mathrm{RSS}_p - \mathrm{RSS}_q}{q-p}\right) / \left(\frac{\mathrm{RSS}_q}{n-q}\right) \sim F_{q-p,n-p} \tag{2.14}$$

where RSS_p is the RSS of the parallel constrained model with p parameters, RSS_q

is the RSS of the non-parallel unconstrained model with q parameters, and n is the total number of observations [8, 11]. It is common to set the significance level to 1% [8].

The chi-squared test has the same normality assumption as the F-test, but the model is weighted using historical assays that estimate the relationship between the response and its variance [8]. Since in practice, the chi-squared test is seen as an alternative approach and not often used, no further details of this test will be discussed.

Unlike difference tests, equivalence testing framework is based on demonstrating that any differences between model parameters associated with parallelism are negligible. This hypothesis testing framework states a null hypothesis that the curves are not parallel, and an alternative hypothesis that the curves are parallel. For example, using the linear model, this would correspond to assessing if the difference between the slopes of the reference and test samples is less than or equal to a pre-specified value [8, 17]. In the context of parallelism testing, the goal is to state that parallelism exists. This aligns with the alternative hypothesis of an equivalence test and thus equivalence tests are often preferred. Difference tests are often used early in assay development when little is known about assay performance.

For equivalence tests, separate models are fitted to the reference and test data. For a tested parameter, such as the slope or an asymptote, a 95% confidence interval is estimated for the difference in parameters between the reference and test models. The confidence interval for the parameter difference is then compared to a set of pre-specified limits to determine if the confidence interval falls inside these limits, the parameters are considered equivalent and therefore parallelism is demonstrated.

2.3 Assumptions

Certain assumptions must be met for valid inference. Since the three models above are estimated using least-squares regression, it must be assumed that the observations are independent, the residuals are normally distributed, and the variability in the residuals is constant [9, 11, 12]. The equal variance assumption states that the residual variance from the fitted model is constant across the entirety of the dataset [6]. Should this assumption be violated, the estimated RP may appear accurate, but the model will fail to provide accurate estimates for the within-assay variance [6]. The constant variance assumption can be visually examined using residual plots [6]. To address violations of this assumption, the data can either be transformed (eg. using a log transformation) or weighted [6, 7]. In addition, if the residuals are not normally distributed, the suitability tests and the confidence interval for RP may be incorrect [6]. While the models in Section 2.1 can handle some departure from normality, extreme departures can be problematic [6]. QQ plots can be used to visually assess the normality assumption; if the plot forms a approximately straight line, then the residuals are normal, otherwise normality cannot be assumed [6]. A transformation of the data to address a lack of constant variance can often resolve the issue of non-normal residuals [6, 7].

In practice, the independence assumption is often violated and must be accounted for in the analysis [7]. Non-independence typically arises from the design or conduct of the assay, and can be addressed by using an appropriate statistical model, like a mixed effects model or, more simply, by averaging the replicates [7]. In this report, we will simulate replicate data that are dependent but treat them as if they are independent. Then, the dependency will then be addressed by considering the mean replicate value at each dose, and comparing the results from the goodness-of-fit and parallelism tests.

Chapter 3

Simulation Study

In this section, we describe the results of a simulation study investigating the impact of violating the independence assumption when conducting a bioassay analysis. To do this, dependent replicate data were generated from the 4PL model discussed in Chapter 2 and were then analyzed as independent data to mimic common practice [10, 15]. The dependency was then addressed by averaging the replicates.

3.1 Model and Data Generation

Additional considerations are required to induce dependency among replicates when simulating data from the 4PL model. Let i index the substances used for simulating the data, where $i = 1$ corresponds to the reference and $i = 2$ corresponds to the test. Let j index the dose level for each substance, where $j = 1, ..., 10$ in our simulations. Finally, let k represent the replicate, where $k = 1$ represents the first replicate and $k = 2$ represents the second replicate. Let y_{ijk} represent the value of the response for the k^{th} replicate at the j^{th} dose group for the i^{th} substance with corresponding dose x_{ijk}. Let z_{ijk} be an indicator variable whose value is 0 if the corresponding

observation is derived from reference material, and 1 if the observation is derived from a test sample. Lastly, let τ_{ij} and ϵ_{ijk} represent the shared random effect and the random error, respectively. The shared random effect is shared by all replicates in a given dose group. The random error is specific to individual responses. For the random effect, let $\tau_{ij} \sim N(0, \sigma_\tau)$, and for the random error, let $\epsilon_{ijk} \sim N(0, \sigma_\epsilon)$. We can then define a random effects 4PL model as:

$$y_{ijk} = D + \frac{A - D}{1 + \exp[B(\log(x_{ijk})) - C + \log(RP) \times z_{ijk}]} + \tau_{ij} + \epsilon_{ijk}. \tag{3.1}$$

where $i = 1, 2$, $j = 1, ..., 10$, and $k = 1, 2$. As before, parameters A and D are the lower and upper asyptotes respectively, B is the slope of the curve at its inflection point, and C is the inflection point [9, 14, 15].

3.1.1 Correlation

The model in equation (3.1) results in a correlation among the replicates. That is, observations within the same dose group are more similar than those across dose groups. This similarity can be characterized by the intraclass correlation coefficient for the population [18, 19].The intraclass correlation coefficient is found by simplifying equation (3.1) as:

$$y_{ijk} = \mu + \tau_{ij} + \epsilon_{ijk}. \tag{3.2}$$

where τ_{ij} and ϵ_{ijk} are the random shared effect and random error terms respectively, and μ is the overall mean response given by the 4PL dose-response [18, 19]. Using the covariance properties and assuming that τ_{ij} is independent of ϵ_{ijk}, the intraclass correlation can then be defined as [19]:

$$ICC = \frac{\sigma_\tau^2}{(\sigma_\tau^2 + \sigma_\epsilon^2)} \tag{3.3}$$

where the sum of σ_τ^2 and σ_ϵ^2 is the total variability of an individual observation. This is because the total variability of an individual observation is made up of a shared error component between all observations within a dose group, and an error component specific to an individual observation. As the variance of the shared error component gets larger, relative to the variance of the random error component, the correlation within the cluster increases.

A correlation of 0 occurs when $\sigma_\tau^2 = 0$, while a correlation of 1 occurs when $\sigma_\epsilon^2 = 0$. When $\sigma_\tau^2 = \sigma_\epsilon^2$, the correlation is 0.5. For the simulation study, we examined correlation values of 0, 0.25, 0.5, 0.75, and 0.95. To obtain the desired correlation values, the following pairs of σ_τ and σ_ϵ were used: (0, 10), (5, $\sqrt{75}$), ($\sqrt{50}$, $\sqrt{50}$), ($\sqrt{75}$, 5), and ($\sqrt{95}$, $\sqrt{5}$). These values were obtained by setting the denominator of equation (3.3) equal to 100, and solving for σ_τ and σ_ϵ for each correlation value we examined. The values of σ_ϵ and σ_τ have an inverse relationship with the correlation, where high values of σ_ϵ and low values of σ_τ (e.g., (0, 10)) result in a low correlation, and low values of σ_ϵ and high values of σ_τ (e.g., ($\sqrt{95}$, $\sqrt{5}$)) result in high correlation.

In practice, a statistician will typically rely on knowledge of how the assay was conducted to determine if the data are correlated. One option used in practice for removing correlation in the data is to average the replicates.

3.1.2 Data Generation

The data were generated in R [20] and then analyzed using a bioassay analysis software called QuBAS [21]. Using equation (3.1), responses for each dose level j were simulated for both the reference and test data. The initial parameters in the 4PL model were set to $A = 0$, $D = 500$, $B = 1.5$, and $C = 4$, and RP values of 0.5, 0.71, 1.141, and 2 were considered. The 4PL model is used as most data follow the 4PL shape and the 5PL model is used only when necessary. The parameters were chosen

to represent a well characterized plausible assay. For each of the reference and test samples, we considered a series of 10 doses of $2^1, 2^2, ..., 2^{10}$ each with $k = 2$ replicates. Since $k = 2$, both the generated reference and test data have two responses for each dose level, one for each replicate. Upon initializing the parameters, the reference and test samples for the dependent data that were treated as independent data set were then generated. Given the different RP values, a test sample was generated for each considered RP value. Shared and random error were added to the reference and test sample generated per equation (3.1) using pairs of σ_τ and σ_ϵ previously specified.

For each of the 5 correlation levels, 250 datasets were generated, where each dose group consisted of two replicates. This resulted in a total of 1250 datasets. The data were first analyzed as generated, where the replicates were treated as independent. The replicates for each dose group were then averaged and the datasets reanalyzed using the averaged responses.

3.1.3 QuBAS

After the data were generated in R, QuBAS was used to perform the goodness-of-fit F-test and a parallelism F-test for the 4PL model on each simulated data set. The data were generated from a parallel 4PL model, and both the goodness-of-fit test and the parallelism test were performed at the 5%-level. That is, when the assumption of independence is met (i.e., correlation is zero), we would expect approximately 5% of the datasets will have insufficient evidence to conclude that they pass the goodness-of-fit and parallelism tests.

Additionally, 95% confidence intervals were calculated for the RP for the test samples. For the RP, approximately 95% of the confidence intervals should contain the RP used to generate the data. In other words, when the data are independent, we expect approximately 5% of the confidence intervals will not contain the true RP

value.

For each RP, the geometric mean and geometric coefficient of variation were calculated. These values are used to summarize RPs (rather than the arithmetic mean and standard deviations) as RPs are typically normally distributed on the log scale. Summary statistics were calculated for the width of the confidence intervals where the replicates were treated as independent (Table A.1), and where the replicates were averaged (Table A.2).

3.2 Results

3.2.1 Non-Averaged Replicates

Table 3.1 displays the proportion of 95% confidence intervals that contain the true RP value (denoted as "Coverage of True RP"), the geometric mean of the RP (denoted as "Geometric Mean") and the geometric coefficient of variation for the RP (denoted as "GCV") measured across the iterations of the simulations study, and the percentage of datasets that passed the goodness-of-fit (denoted as "GOF") and parallelism tests for each combination of correlation and RP in the analyses where data were assumed to be independent.

As the correlation values increased, the pass rate decreased as shown in Table 3.1.For each RP value when the correlation was 0, we see that the confidence intervals that contain the true RP value hovers around 95%, and the number of datasets that pass the goodness-of-fit (GOF) and parallelism tests are both roughly 95%. As the correlation value increased, the percentage of datasets where the true RP value falls within a 95% confidence interval decreased, and the number of datasets that pass the goodness-of-fit and parallelism tests decrease as well. When the correlation was

Table 3.1: Summary of the coverage of the confidence interval containing the true RP, the geometric mean and geometric coefficient of variation for the RP, and the empirical pass rates for the goodness of fit and parallelism F tests for various combinations of true relative potency and within dose-group correlation for treated as independent replicates.

Correlation	RP	Coverage of True RP	Geometric Mean	GCV	GOF	Parallelism
	0.50	0.976	0.502	3.185	0.972	0.976
	0.71	0.936	0.710	3.435	0.952	0.952
0	1.00	0.952	1.000	3.306	0.956	0.960
	1.41	0.956	1.408	3.458	0.980	0.952
	2.00	0.968	2.001	3.211	0.940	0.956
	0.50	0.880	0.499	4.072	0.848	0.924
	0.71	0.932	0.708	4.006	0.824	0.880
0.25	1.00	0.920	0.997	3.734	0.820	0.880
	1.41	0.900	1.410	4.017	0.852	0.896
	2.00	0.880	1.993	4.269	0.792	0.900
	0.50	0.884	0.499	4.129	0.560	0.832
	0.71	0.880	0.712	3.973	0.588	0.860
0.5	1.00	0.908	0.999	3.980	0.580	0.840
	1.41	0.876	1.415	4.390	0.564	0.852
	2.00	0.900	2.007	4.234	0.560	0.844
	0.50	0.856	0.497	4.307	0.172	0.728
	0.71	0.824	0.706	4.479	0.160	0.720
0.75	1.00	0.856	0.993	4.343	0.152	0.728
	1.41	0.864	1.402	4.518	0.176	0.692
	2.00	0.824	1.989	4.623	0.136	0.752
	0.50	0.752	0.499	5.177	0.008	0.660
	0.71	0.796	0.711	4.849	0.004	0.640
0.95	1.00	0.752	0.997	5.214	0.000	0.644
	1.41	0.768	1.403	5.147	0.008	0.672
	2.00	0.796	2.003	5.163	0.000	0.600

0.95, less than 80% of the confidence intervals contain the true RP, virtually none of the datasets passed the goodness-of-fit test, and less than 70% of the datasets passed the parallelism test. The change in RP does not have an impact on the probability of passing the goodness-of-fit and parallelism tests. There is no relationship between the RP and failure rates of both tests and the coverage of the true RP because we have a well characterized 4PL curve with doses covering the entire curve.

From the results in Table 3.1, we see that the geometric means of the RP remained consistently similar to the true RP values across all correlation values. Thus, the geometric means for the RP did not depend on correlation. Therefore, we can conclude that the difference in coverage of the RP is due to the confidence intervals

being falsely too narrow and not because of bias. As the correlation increases, we also see that the geometric coefficient of variation increases. As the correlation is increased, the pair of dose-replicates become more similar to one another. This in effect will decrease the effective sample size. Since we kept the total variability the same for each correlation value, this results in our RP estimate being more variable among the repeated iterations of the simulation study.

The results found in Table 3.1 show that correlation had a substantial impact on the results. However, as the correlation increases, so does the difference in the percentage of confidence intervals that contain the true RP compared to the percentage of datasets that pass the parallelism test. For example, when the correlation was 0 both the percentage of 95% confidence intervals that contain the true RP and the percentage of datasets that passed the parallelism test were between 90-95%. When the correlation was 0.95 however, the percentage of 95% confidence intervals that contain the true RP was above 75%, and the percentage of datasets that passed the parallelism test were under 70%. In the simulation study we looked at the coverage of the RP confidence interval regardless of the result of the goodness-of-fit and parallelism test. However, in practice, the RP would not be reported in the case of a failure of the goodness-of-fit/parallelism test. In our simulations, the data were generated under the assumption of parallelism and the purpose was to investigate the effects of correlation on inference procedures such as the RP confidence intervals. Therefore, there were more datasets that contain the true RP in a 95% confidence interval than the number of datasets that passed the parallelism test.

For the summary statistics of the non-averaged datasets in Table A.1, the width of the confidence intervals for all tested RP values tended to decrease as the correlation increased. In addition, the medians of each confidence interval width for each tested RP value tended to decreased. As an example, when the correlation was 0 and the RP was 2, the median of the width of the confidence intervals of the RP was 0.1346.

When the correlation was 0.95 with an RP of 2, the median was 0.1243.

3.2.2 Averaged Replicates

The results of the analysis of the simulated datasets where the replicates were averaged are given in Table 3.2 and are similar to the results of the analysis where the replicates were treated as independent when the correlation was 0. From Table 3.2, for each RP value, when the correlation was 0 for each RP value, the coverage probability of the confidence intervals for RP hovers around 95%, and the number of datasets that pass the parallelism tests are also roughly 95%.

Table 3.2: Summary of the coverage of the confidence interval containing the true RP, the geometric mean and geometric coefficient of variation for the RP, and the empirical failure rates for the parallelism F test for various combinations of true relative potency and within dose-group correlation for averaged replicates.

Correlation	RP values	Coverage of True RP	Geometric Mean	GCV	Parallelism
	0.50	0.976	0.502	3.185	0.960
	0.71	0.956	0.710	3.435	0.936
0	1.00	0.968	1.000	3.306	0.964
	1.41	0.964	1.408	3.458	0.952
	2.00	0.964	2.001	3.211	0.968
	0.50	0.908	0.499	4.072	0.968
	0.71	0.944	0.708	4.006	0.964
0.25	1.00	0.952	0.997	3.734	0.940
	1.41	0.936	1.410	4.017	0.944
	2.00	0.920	1.993	4.269	0.948
	0.50	0.948	0.499	4.129	0.952
	0.71	0.956	0.712	3.972	0.968
0.5	1.00	0.976	0.999	3.980	0.964
	1.41	0.964	1.415	4.390	0.972
	2.00	0.948	2.007	4.234	0.948
	0.50	0.968	0.502	4.287	0.952
	0.71	0.940	0.712	4.645	0.940
0.75	1.00	0.956	1.003	4.559	0.944
	1.41	0.952	1.420	4.469	0.948
	2.00	0.956	2.011	4.378	0.944
	0.50	0.920	0.499	5.177	0.980
	0.71	0.952	0.711	4.849	0.976
0.95	1.00	0.944	0.997	5.214	0.976
	1.41	0.920	1.403	5.147	0.972
	2.00	0.952	2.003	5.163	0.964

Unlike the results found with the non-averaged replicates, the results remained relatively consistent as the correlation between replicates increases. For instance, in Table 3.2, as the correlation increases to 0.95, for each RP value, the confidence intervals that contain the true RP value still hovered around 95%, and the number of datasets that pass the parallelism tests were slightly above 95%. This is evidence that averaging the replicates addresses the violation of the independence assumption as the results do not decline as the correlation increases. However, with averaging we cannot determine if the 4PL model was a good fit for the data due to the inability to perform the goodness-of-fit test.

Given the geometric mean and GCV results in Table 3.2, compared to the results found in Table 3.1, the estimates of the RP are identical whether the replicates are averaged, or not. Therefore, our conclusions regarding the (lack of) bias and variability of the RP estimates remain unchanged in the scenario where the replicates are averaged.

When the replicates were averaged, the summary statistics shows slightly different results than dependent replicates that were treated as independent. Similar to the case where replicates were treated as independent, the width of confidence intervals for the averaged datasets for all parameter values tended to increase as the correlation increased. Unlike the independent replicate datasets however, the median confidence interval width increased as the correlation increased for all RP values. As an example, when the correlation was 0 and the RP value was 1.41, the median confidence interval width was 0.1388. When the correlation was 0.95 with an RP of 1.41, the median width was found to be 0.1971. With higher correlation values, the confidence intervals for the RP of the averaged replicates got wider, and thus increased the likelihood of covering the true RP.

Chapter 4

Conclusion

When running bioassays to estimate the potency of biological drugs, the replicates at each dose group are often not independent due to the design and conduct of the assay [7]. When all model assumptions are met, including uncorrelated observations, we obtained results consistent with expectations. However, we found that when replicates are correlated but treated as independent, estimation of the potency of a new batch of drug is impacted. More specifically, this resulted in 95% confidence intervals for the RP being falsely too narrow, and failed to capture the true RP as often as would be expected. Correlated replicates also resulted in parallelism and goodness-of-fit tests failing inappropriately. When not accounting for correlation in the model, the results got increasingly poor with the magnitude of the correlation.

A large correlation value tended to underestimate the variance of the residuals due to the decreased effective sample size, resulting in a large F statistic and failing the lack-of-fit test, as demonstrated in Table 3.1 [11]. From the results, the goodness-of-fit test showed a roughly inverse relationship with the correlation. As the correlation increased, responses within a dose group were more similar to one another, and thus the mean of those individual values were very similar as a result. Therefore

the calculated SSPE for the goodness-of-fit F test was very small [11]. With a small SSPE, the resulting F-test statistic was much greater than the critical value, resulting in a significant lack-of-fit [11].

This report investigated a method for addressing dependence in the data by average the replicates within each dose group, for each sample, before modelling. Using the 4PL model, we generated 1250 datasets using R that were analyzed using QuBAS. From the results of the simulation study, averaging the replicates showed to be effective in removing the issue of dependent data at the cost of additionally removing information about the variance between the replicates. Further research should be conducted to examine the effects of averaging the replicates has when using different models and parameter values.

Averaging the replicates prior to modeling to account for the violation of the independence assumption was shown to work, as the degree of correlation had virtually no affect on the results. The simulated datasets with averaged replicates had a pass rate consistent with expectations for all correlation values, and likelihood the 4PL model contains the true RP value remained relatively consistent. However, the goodness-of-fit test cannot be performed on data with averaged replicates and thus we lose information on how well the model fits the datasets. From the summary statistics table for non-averaged results, we found that the width of the confidence intervals for the true RP values were smaller when both the correlation was high, and the median widths were shown to decrease as correlation increased. Testing with smaller or larger error terms may potentially shed more light on the relation between variance and the width of the confidence intervals for the true RP. Despite the drawback of no goodness-of-fit test, the removal of correlation by averaging results in a higher proportion of samples appropriately passing and may increase laboratory throughput.

Additional approaches can be made to further explore the relation between non-averaged and averaged replicates. The only assumption that was violated during testing was the independence assumption by adding correlation to the data. We then compared the results to the same datasets where the replicates were averaged, addressing the independence violation. One approach is to look at the results when each of the three assumptions (independence, normality, and equal variance) are violated. Transformed datasets are typically used to account for the violation of the equal variance and normality assumptions. Simulating and testing datasets that violate multiple or all three assumptions can also be approached to test the effects of how a combination of averaging the replicates, and transforming the data affects inference.

Another future study would be to test different variance parameters. For all generated data sets, the parameters A, D, B, and C remained the same and only the RP and the two variance components, σ_ϵ and σ_τ were manipulated. We kept $\sigma_\tau^2 + \sigma_\epsilon^2 = 100$ constant for all simulated datasets, and chose the values of σ_ϵ and σ_τ from this constant to achieve the desired correlation. The signal of the data is defined by the difference in the asymptotes $(D - A)$. Since we kept the asymptote parameters A and D constant, the signal also remained constant. By examining larger or smaller values of σ_ϵ and σ_τ beyond the restraints of the constant, we can assess what happens if the sum of the error terms are small or large relative to the overall signal. During testing, the data was simulated to be symmetrical and thus the 4PL model was used. By using the 5PL model instead and incorporating the asymmetry parameter E, further research can be conducted to identify how asymmetry affects the results when the data are correlated.

There are other statistical techniques that can be used to explicitly model correlation such as a mixed model. However, these complex techniques are not implemented in off the shelf bioassay analysis software, whereas the 4PL and 5PL models are readily

accessible. As such, it would be practically impossible to make use of these techniques in the context of bioassay analysis. Averaging the replicates prior to any statistical modelling provides a simple technique to provide valid statistical inference that can be readily implemented for bioassay analysis.

Bibliography

[1] Finney, D. J. (1947). The principles of biological assay. Supplement to the Journal of the Royal Statistical Society, 9(1), 46-91.

[2] Center for Biologics Evaluation and Research. What are "biologics" Questions and answers. U.S. Food and Drug Administration.

[3] Tallarida, R.J., Murray, R.B. (1987). Relative Potency I. In: Manual of Pharmacologic Calculations. Springer, New York, NY.

[4] Bursa, F. (2022, June 14). *What is relative potency?* Quantics Biostatistics.

[5] Ph. Eur. Chapter 5.3: Statistical analysis of results of biological assays and tests. *European Pharmacopoeia: Version 11.0*, 2020.

[6] USP. Design and development of biological assays <1032>. *USP-NF*, 2013.

[7] USP. Analysis of biological assays <1034>. *USP-NF*, 2013.

[8] Fleetwood K, Bursa F, Yellowlees A. Parallelism in practice: approaches to parallelism in bioassays. PDA J Pharm Sci Technol. 2015 Mar-Apr;69(2):248-63.

[9] Bursa, F. (2022, April 25). *Choosing a statistical model: Continuous response data.* Quantics Biostatistics.

[10] Wild, D. (2013). Chapter 3.6 Calibration Curve Fitting. In The immunoassay handbook (pp. 323–336). essay, Elsevier.

[11] Montgomery, D. C. (2012). Introduction to Linear Regression Analysis. John Wiley Sons, Incorporated.

[12] Dylan Z. Childs, B. J. H. and P. H. W. (2022, December 7). Introductory biostatistics with R. Chapter 24 Non-linear regression in R.

[13] Weisberg, S. (2013). Applied Linear Regression. John Wiley Sons, Incorporated.

[14] Drummond, J. E., Four parameter logistic regression. MyAssays.

[15] Davis, D., Zhang, A., Etienne, C., Huang, I., Malit, M. (n.d.). Principles of Curve Fitting For Multiplex Sandwich Immunoassays. Bio-Plex suspension array system tech note 2861.

[16] Bursa, F. (2022, April 25). *Complications of fitting 4PL and 5PL models to bioassay data.* Quantics Biostatistics.

[17] Baljé-Volkers, C., Mzolo, T., Talens, E., IJzerman-Boon, P., Van den Heuvel, E. (2018). Equivalence testing for similarity in bioassays using bioequivalence criteria on the relative bioactivity.

[18] Donner, A., Koval, J. J. (1980). The estimation of intraclass correlation in the analysis of family data. Biometrics, 36(1), 19–25.

[19] Stanish, W. M., Taylor, N. (1983). Estimation of the Intraclass Correlation Coefficient for the Analysis of Covariance Model. The American Statistician, 37(3), 221–224.

[20] R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[21] Quantics Biostatistics. QuBAS Bioassay Software version 3.1.2, 2024.

Appendix A

Summary Statistics

Table A.1: Summary statistics for various combinations of relative potency and within dose-group correlation for treated as independent replicates.

Correlation	RP Values	Width of RP Confidence Interval					
		Min	Q1	Median	Mean	Q3	Max
0.00	0.50	0.087	0.125	0.137	0.136	0.147	0.176
	0.71	0.088	0.121	0.135	0.134	0.146	0.175
	1.00	0.092	0.124	0.135	0.134	0.144	0.175
	1.41	0.093	0.122	0.134	0.134	0.147	0.179
	2.00	0.092	0.123	0.135	0.135	0.145	0.188
0.25	0.50	0.089	0.120	0.131	0.131	0.141	0.177
	0.71	0.086	0.122	0.133	0.134	0.144	0.185
	1.00	0.096	0.121	0.133	0.133	0.143	0.173
	1.41	0.094	0.121	0.132	0.132	0.143	0.173
	2.00	0.093	0.120	0.130	0.131	0.142	0.170
0.50	0.50	0.087	0.120	0.132	0.132	0.143	0.195
	0.71	0.088	0.117	0.130	0.130	0.140	0.190
	1.00	0.092	0.120	0.130	0.130	0.141	0.188
	1.41	0.091	0.118	0.131	0.131	0.143	0.183
	2.00	0.092	0.117	0.128	0.130	0.142	0.177
0.75	0.50	0.080	0.113	0.125	0.127	0.140	0.179
	0.71	0.085	0.111	0.124	0.125	0.136	0.189
	1.00	0.078	0.115	0.127	0.127	0.140	0.181
	1.41	0.086	0.111	0.127	0.126	0.140	0.180
	2.00	0.077	0.113	0.126	0.125	0.138	0.173
0.95	0.50	0.073	0.109	0.122	0.124	0.137	0.187
	0.71	0.075	0.107	0.123	0.124	0.141	0.212
	1.00	0.070	0.110	0.125	0.125	0.138	0.183
	1.41	0.076	0.107	0.125	0.123	0.137	0.186
	2.00	0.067	0.109	0.124	0.124	0.137	0.194

Table A.2: Summary statistics for various combinations of relative potency and within dose-group correlation for averaged replicates.

Correlation	RP Values	Width of RP Confidence Interval					
		Min	Q1	Median	Mean	Q3	Max
0.00	0.50	0.085	0.125	0.139	0.141	0.157	0.204
	0.71	0.084	0.123	0.140	0.140	0.158	0.210
	1.00	0.069	0.125	0.141	0.141	0.158	0.231
	1.41	0.078	0.120	0.139	0.139	0.156	0.218
	2.00	0.059	0.124	0.141	0.142	0.159	0.207
0.25	0.50	0.067	0.135	0.153	0.155	0.172	0.247
	0.71	0.079	0.137	0.158	0.159	0.179	0.243
	1.00	0.087	0.136	0.154	0.157	0.178	0.250
	1.41	0.099	0.139	0.154	0.157	0.176	0.233
	2.00	0.092	0.136	0.155	0.157	0.174	0.230
0.50	0.50	0.094	0.155	0.175	0.176	0.193	0.273
	0.71	0.100	0.150	0.170	0.172	0.188	0.272
	1.00	0.102	0.153	0.175	0.174	0.192	0.269
	1.41	0.091	0.154	0.173	0.173	0.193	0.270
	2.00	0.098	0.151	0.170	0.172	0.191	0.259
0.75	0.50	0.107	0.164	0.180	0.184	0.203	0.283
	0.71	0.089	0.158	0.183	0.183	0.210	0.307
	1.00	0.093	0.159	0.183	0.183	0.207	0.269
	1.41	0.078	0.156	0.182	0.181	0.200	0.301
	2.00	0.119	0.159	0.188	0.187	0.208	0.291
0.95	0.50	0.111	0.171	0.193	0.195	0.217	0.297
	0.71	0.117	0.168	0.194	0.196	0.222	0.338
	1.00	0.105	0.172	0.198	0.197	0.219	0.290
	1.41	0.116	0.168	0.197	0.194	0.216	0.296
	2.00	0.097	0.171	0.196	0.195	0.217	0.309

Appendix B

R Code

```
library(ggplot2)
library(minpack.lm)
library(xlsx)
library(dplyr)
rm(list = ls())

###set parameters
#A = lower asymptote
#D = upper asymptote
#B = slope parameter
#C = log(EC50) reference standard
#E = asymmetry parameter (if 1 corresponds to 4PL)
#RP = relative potency
#sigma = variance for random error component
#n = number of reps per dose/sample
A = 0
D = 500
```

```
B = 1.5

C = 4

E = 1

RP = c(0.5,0.71,1,1.41,2) #RP = c(0.5,0.71,1,1.41,2)

n = 2 #variable for replicates k, #n=2

Ref_n = 2 #reference replicates

sigma_e = 10

sigma_t = 0

dose_vec<-2^seq(1,10)


Model<-function(A,D,B,C,E,RP,n,Ref_n,sigma_t,sigma_e,

 dose_vec, iter){

 ##########################

 for(I in 1:iter){

  dose_len<-length(dose_vec)

  ##########################

  #simulate doses

  len_RP<<-length(RP)

  Dose = rep(dose_vec,n*2)


  #Reference Matrix

  Ref <- NULL

  Ref_vec<- D + (A-D)/((1+exp(B*(log(Dose[1:dose_len])-C)))^E)

  Ref<-matrix(rep(Ref_vec,Ref_n),nrow=Ref_n,byrow=TRUE)


  #Test Matrix

  Test <- NULL
```

```
Test_vec<-NULL
for(i in 1:len_RP){
 for(j in 1:dose_len){
  Test_vec[j] <- D + (A-D)/((1+exp(B*(log(Dose[j])-C
   +log(RP[i])))))^E)
 }
 temp<-matrix(rep(Test_vec,n),nrow=n,byrow=TRUE)
 Test<- rbind(Test,temp)
}

#Combine Reference and Test Matricies
RefTest<- rbind(Ref,Test)

#Tau Matrix
Rtau_mat<-NULL #Reference tau
RTau <- rnorm(dose_len,0,sigma_t)
Rtemp<-matrix(rep(RTau,Ref_n),nrow=Ref_n,byrow=TRUE)
Rtau_mat<-rbind(Rtau_mat,Rtemp)
Ttau_mat<-NULL #test tau
for(i in 1:(nrow(Test)/n)){
 Tau <- rnorm(dose_len,0,sigma_t)
 temp<-matrix(rep(Tau,n),nrow=n,byrow=TRUE)
 Ttau_mat<-rbind(Ttau_mat,temp)
}
tau_mat<-rbind(Rtau_mat,Ttau_mat)

#Error Matrix
```

```
err_mat <- NULL
for(i in 1:(nrow(RefTest))){
 Error <- rnorm(dose_len,0,sigma_e)
 temp<-matrix(Error,nrow=1,byrow=TRUE)
 err_mat<-rbind(err_mat,temp)
}

#Response Matrix
Response_mat<-RefTest + tau_mat + err_mat
#Column Names
colnames(Response_mat)<-Dose[1:dose_len]
#Row Names
names<-NULL
for(i in 1:nrow(Response_mat)){
 if(i <= Ref_n){ #First half of rows
  names[i] <- paste("Reference rep",i)
   #First half of rows become Reference
 } else {             #Second half of rows
  names[i] <- paste("Test rep ",i-Ref_n)
   #Second half of rows become Test
 }
}
#var_nam<-c('A','D','B','C','E','RP','n',
 'Sigma Error','Sigma Tau')
rownames(Response_mat)<-names
df<-as.data.frame(Response_mat)
```

```
#average values
Ravg_mat<-NULL
Ravg<-(Response_mat[1,]+Response_mat[Ref_n,])/2
Ravg_mat<-rbind(Ravg_mat,Ravg)
Tavg_mat<-NULL
for (i in 1:(len_RP)){  #(n*len_RP+n)/n
 Tavg<-(Response_mat[n*i+Ref_n-1,]+Response_mat[n*i+Ref_n,])/2
 Tavg_mat<-rbind(Tavg_mat,Tavg)
}
avg_mat <- rbind(Ravg_mat,Tavg_mat)

#avg row names
names<-NULL
for(i in 1:nrow(avg_mat)){
 if(i == 1){ #First half of rows
  names[i] <- paste("Reference rep",i)
   #First half of rows become Reference
 } else {             #Second half of rows
  names[i] <- paste("Test rep ",i-1)
   #Second half of rows become Test
 }
}
rownames(avg_mat)<-names
df2<-as.data.frame(avg_mat)
```

```
#RP value column
#Non-averaged
RP_col<-NULL
for (i in 1:len_RP){
 col<-c(RP[i],rep(NaN,(n-1)))
 RP_col<-c(RP_col,col)
}
RP_col<-c(rep(NaN,Ref_n),RP_col)
df<-cbind(RP_col,df)
#Average
RP_col<-NULL
for (i in 1:len_RP){
 RP_col[i]<-RP[i]
}
RP_col<-c(NaN,RP_col)
df2<-cbind(RP_col,df2)

ogrow<-nrow(df) #original number of rows in data frame non-avg
ogrow2<-nrow(df2) #original number of rows in data frame avg



#Add variables
nan<-rep(NaN,ncol(df)-1)
var<-c(A,D,B,C,E,n,sigma_e,sigma_t)
var_nam<-c('A','D','B','C','E','n','Sigma Error','Sigma Tau')
for(i in 1:8){   #Non-averaged
 df[nrow(df)+1,]<-c(var[i],nan)
```

```
   rownames(df)[i+ogrow]<-var_nam[i]
  }
  for(i in 1:8){   #Average
   df2[nrow(df2)+1,]<-c(var[i],nan)
   rownames(df2)[i+ogrow2]<-var_nam[i]
  }

  #Create Excel file for Non-averaged data
  location<-paste0("C:/Qubas/Multi-Ref Non-Avg/SimStudy",
   I,"_tau",sigma_t,"_error",sigma_e,".csv")
  write.csv(df,location,row.names=TRUE)
  #Create Excel file for Averaged data
  location<-paste0("C:/Qubas/Multi-Ref Avg/AvgSimStudy",
   I,"_tau",sigma_t,"_error",sigma_e,".csv")
  write.csv(df2,location,row.names=TRUE)
 }
 return(df)
}

results<-Model(A,D,B,C,E,RP,n,5,20,5, dose_vec,250)
```

Vita

Candidate's full name: Liam Joseph Cann

University attended (with dates and degrees obtained): University of New Brunswick, Bachelor of Science, double major in Mathematics and Statistics, 2022

Non-Refereed Publications: Cann, L., Stephenson, M., Stewart, C., Segall, J. (2023) Should I Average My Pseudo-replicates? Quantics Biostatistics.

https://www.quantics.co.uk/blog/should-i-average-my-pseudo-replicates

Conference Presentations: Cann, L., Stephenson, M., and C. Stewart. June 2024. Impacts of correlation in bioassays. Canadian Statistics Student Conference (Memorial University of Newfoundland)