Experimental design and statistical analyses of fish growth studies

Helgi Thorarensen[a*], Godfrey Kawooya Kubiriza[a,b,c], Albert Kjartansson Imsland[d,e]

[a]Department of Aquaculture and Fish Biology, Hólar University College, 550 Sauðárkrókur, Iceland

[b]Department of Biological Sciences, College of Natural Sciences, Makerere University, P. O. Box, 7062, Kampala, Uganda

[c]United Nations University - Fisheries Training Programme, Skulagata 4, 101 Reykjavik, Iceland

[d]Department of Biology, University of Bergen, High Technology Centre, 5020 Bergen, Norway

[e]Akvaplan-niva Iceland Office, Akralind 4, 201 Kópavogur, Iceland,

Keywords: growth studies; statistical power; minimum detectable difference; number of replicates; sample size; ANOVA; polynomial; non-linear models

*****Corresponding author**.** Tel: +354 455 6300; fax +354 455 6301

E-mail address: helgi@holar.is (H. Thorarensen)

Abstract

Every year, numerous studies are published that compare the effects of different factors on the growth of aquaculture fish. However, comparatively little attention has been given to the experimental designs of these studies - in how many rearing units should each treatment be replicated, how many fish should be in each tank (*n*) and how should the data be analysed. The reliability of the results increases with increased replication and *n*. In reality, however, the experimental design must strike a balance between limited resources and the reliability of the statistical analysis. A survey of recent publications in Aquaculture suggests, that most (83%) aquaculture growth studies apply each treatment in triplicates with an average of 26 fish in each tank (range: 4 to 100). The minimum difference that can reliably be detected with statistical analyses is determined by the number of replications of each treatment, *n*, the variance of the data and the number of treatments applied. In the present study, we accumulated information on the variance of data in aquaculture growth studies on different species to estimate the minimum detectable difference and to assist researchers in designing experiments effectively. These results suggest that the variance is similar for different aquaculture species and, therefore, the same designs (level of replication and *n*) are suitable for studies on different species of fish.

The minimum difference (MDD) in mean body-mass of different treatment groups that can be detected in a typical aquaculture study (triplicates, 25 fish in each tank and average variance) with 80% statistical power (less than 20% chance of Type II error) is around 26% of the grand mean. Increasing the *n* from 25 to 100 will reduce the MDD to 19% of the grand mean, while a further increase in *n* will have comparatively lesser effect. Increasing replication to quadruplicates or sextuplicates (with *n* as 100), will further reduce the MDD to 16% and 12% of the grand mean respectively. MDD under 10% of the grand mean is only possible when fish for the experiment are selected within a narrow size range to reduce variance.

Simulations were performed, where samples (experiments) were repeatedly drawn from artificial populations with identical distribution and with the same experimental design as is commonly used in growth studies. Two of the populations had dose-dependent responses to treatment while one population showed no response to treatment. The resulting data was analysed with a mixed model ANOVA and by fitting either polynomials or asymptotic models to the data. Contrary to earlier suggestions, the critical treatment (minimum treatment to generate maximum response) estimated with the ANOVA approached more closely the population responses than did the critical treatments estimated with the non-linear models.

Keywords: Growth studies, statistical power, minimum detectable difference, number of replicates, sample size, ANOVA, polynomial, non-linear models

1. Introduction

Information on the effect of feed ingredients, physical environment and other factors on the growth of fish are important for the development of aquaculture. Therefore, growth studies are common in aquaculture research where the mean sizes of different groups are compared following various treatments; the objective being to predict the performance of populations (all fish of the same species/strain) under different conditions.

The design of aquaculture growth experiments usually includes replication of treatments in two or more rearing units (e.g. tanks, ponds or net pens) where the replicates are considered independent samples from the populations. How accurately the results of experiments reflect the mean responses of the populations depends primarily on the number of fish sampled (within each replicated unit), the number of replicates and the variance of responses, both among individual fish within a replicated unit and among replicates.

A number of approaches have been used to analyse the results of growth studies, but the method most commonly used is analysis of variance (ANOVA). A cursory examination of growth studies (Table 3) published during the last year in the journal Aquaculture (29 in total) suggests that ANOVA is used in in some capacity in all studies although 24% of the studies complement the analysis of dose response data with linear or non-linear methods.

In growth studies where treatments are replicated, individual fish should not be considered the experimental units. The fish within a tank are all exposed to the same "tank effects" (differences between tanks independent of treatment effects) and complicated interactions among the fish may contribute to variability within the tank that are not caused by the treatment (Gardeur et al. 2001; Imsland 2001; Koslow & Hurlbert 2006). In fact, it can be argued that because of the common "tank effect", individual fish within a tank are not independent samples from the population but are instead "pseudoreplicates" as defined by Hurlbert (1984). A better approach is to perform ANOVA based on the total biomass or mean

body-mass in each tank (Cowey 1992; Smart et al., 1998) or, better still, to use a mixed model ANOVA where treatments are fixed factors and tanks are nested as random factors within treatments. With the latter method, the information on individual fish is modelled to fully account for the data structure (Ruohonen 1998, Ling and Cotter 2003). If the design of the experiment is balanced, i.e. the number of fish in all tanks and the number of tanks in all treatments is the same, the results of the simple and mixed model ANOVA will be the same. However, in long term growth studies the design may not be balanced, since mortality can vary among rearing units and all fish from single rearing units may be lost due to mishaps. When the design is not balanced, a mixed model should be used since the risk of type I error (rejecting a correct hypothesis) is increased when a simple ANOVA is used for the analysis of unbalanced data (Ruohonen 1998).

In recent years, methods for mixed model analysis have developed rapidly and now many software packages such as SAS (SAS Institute Inc., Cary, NC, USA) and R (R Core Team 2014) offer the possibility of linear mixed models with the Kenward-Roger modification of *F*-tests (Kenward and Roger, 1997, 2009). The Kenward-Roger modification adjusts the *F* values and degrees of freedom depending on the size of the "tank effect" and thus increases statistical power when the "tank effect" is small. The method has been used in aquaculture growth studies (Tobin et al. 2006; Schram et al. 2012). Over 83% of the growth studies published last year in Aquaculture use the mean body-mass or total biomass in each tank as the unit of analysis while only 11% used a mixed model analysis (Table 3).

In ANOVA, the null hypothesis of no effect of experimental treatments is tested and the means of the treatment groups are considered significantly different when the test statistics (*p*-value) indicates that the probability of the null hypothesis being true is less than 5% (α level less than 0.05). In other words, the probability of rejecting a correct null hypothesis (type I

105 error) is less than 5%. However, it is also possible that an incorrect hypothesis is not rejected
106 and differences among means are not detected where they truly exist. Failing to reject an
107 incorrect hypothesis is called Type II error. The probability of Type II error is β and the
108 power of a statistical test is defined as 1-β. There is no conventional criterion for statistical
109 power as there is for α, although a minimum of 80% is commonly regarded as suitable
110 (Araujo & Frøyland 2007). Statistical power is rarely reported in aquaculture growth studies
111 (Searcy-Bernal 1994) indicating that researchers are less concerned with Type II error than
112 they are with α and Type I error.
113 The statistical power of mixed models depends on five factors: (1) The difference among
114 means caused by the treatment (effect size), (2) the variance of the data, both among fish
115 within a tank and among tanks receiving identical treatments, (3) the number of replicate
116 tanks, (4) the number of fish within each tank and (5) the number of treatments tested (Ling
117 and Cotter 2003, Sokal and Rolf 2012). Statistical power increases with increased effect size,
118 the number of replicate tanks and the number of fish within each replicate tank while
119 statistical power is reduced with increased variance and number of treatments tested (Ling and
120 Cotter 2003). Hence, to secure acceptable statistical power, replications and sample size per
121 replicate should be maximized. However, the number of tanks available and the cost of
122 resources for aquaculture growth studies are usually limited. Therefore, experimental design
123 must strike a balance between acceptable power and the available resources.
124 The issue of the minimum detectable difference (MDD) in aquaculture studies, i.e. the
125 minimum difference that is likely to be detected with 80% statistical power, has received little
126 attention. Ling and Cotter (2003) shed important light on this subject when they compiled
127 information on the coefficient of variation within tanks (CV_{ε}) and the coefficient of variation
128 among tanks within treatment (CV_{β}) for triploid Atlantic salmon. In the present study, we
129 compiled information on variance in body-mass in growth studies on different fish species to

be able to estimate statistical power and the MDD. This information was then used to calculate the expected statistical power and effect size for experimental designs with different levels of replication and number of fish in each replicate tank.

Dose-response designs, where treatments are applied at incrementing levels of e.g. nutrient content or water quality, are common in aquaculture growth studies. These data can be analysed either with ANOVA or by using different linear and non-linear methods. The latter include: Broken line analyses, where two straight lines are fitted to the data, polynomial regression or non-linear regression models that fit asymptotic curves to the data (Baker 1986, Cowey 1992, Shearer 2000). When the results are analysed with ANOVA, the critical response is usually determined as the lowest treatment level that gives a response that is not significantly different from the maximum response. However, this approach has been criticised by Baker (1986) and then later by Cowey (1992) and Shearer (2000). After reviewing a number of published growth studies with dose-dependent relationship, Shearer (2000) concluded that ANOVA may underestimate the critical treatment level by as much as 50% due to the inability of the method to detect small differences. Instead several authors (Baker 1986, Cowey 1992, Shearer 2000) recommend the use of linear or non-linear methods and suggested that they provided more accurate results. However, fitting lines of different shape assumes that there is a certain underlying structure to the data. Moreover, due to the inherent variability in aquaculture growth data it may be difficult to determine visually if the response is polynomial or asymptotic. Therefore, it is questionable if this approach is more appropriate than ANOVA. A second objective of this study was to use simulation studies to compare the fidelity of different methods of statistical analysis to the true underlying responses of populations and the conclusions drawn based on their results.

2. Methods

2.1. Data acquisition

Original raw data from 24 independent growth studies on Arctic charr (*Salvelinus alpinus*), Atlantic halibut (*Hippoglossus hippoglossus*), Atlantic cod (*Gadus morhua*), turbot (*Scophthalmus maximus*) and tilapia (*Oreochromis shiranus*) were analysed in this study. Data on Arctic charr (Ólafur Sigurgeirsson and Jón Árnason, unpublished.), Atlantic halibut (Thorarensen *et al.*, 2010), Atlantic cod (Edelsparre, Pálsson and Steingrímsson, unpublished; Thorarensen, unpublished), and *turbot* (Imsland et al. 2013) were from growth studies conducted at Verið research station, Sauðárkrókur, Iceland. The studies examined different treatment effects (dietary ingredients, oxygen saturation, light regimes and temperature) on the growth performance of fish. Rearing conditions and fish size varied between experiments (Table 1). The data for tilapia were from a study conducted at Bunda College, University of Malawi on the effect of temperature on *Oreochromis shiranus* (Ssebisubi, 2008).

2.2. Data analysis

Data were analyzed using mixed model ANOVA in SPSS to obtain the mean sums of square for tanks nested within treatments (MS_{within}) and the error mean square (MS_{error}), which constituted the error variance ($\hat{\sigma}_{\varepsilon}^2$). The coefficient of variation of the error term (CV_{ε}) was calculated as $CV_{\varepsilon} = \frac{\hat{\sigma}_{\varepsilon}}{\overline{\overline{X}}}$ where $\overline{\overline{X}}$ is the grand mean. The variance among tanks within treatments ($\hat{\sigma}_{\beta}^2$) was calculated as $\hat{\sigma}_{\beta}^2 = \frac{(MS_{within}) - \hat{\sigma}_{\varepsilon}^2}{n}$, where *n* is the number of fish in each tank. The coefficient of variation for tanks within treatments (CV_{β}) was calculated as $CV_{\beta} = \frac{\hat{\sigma}_{\beta}}{\overline{\overline{X}}}$. The statistical power was estimated as described by Ling and Cotter (2003). Briefly, the mean variance of treatment groups ($s_{\bar{Y}}^2$) was estimated as: $s_{\bar{Y}}^2 = \frac{MS_{within}}{nb}$, where *b*

is the number of replicate tanks within treatments. The $s^2_{\bar{Y}}$ was used to compute Tang's parameter (ϕ) (Tang, 1938) as $\phi = \sqrt{\frac{d^2}{2as^2_{\bar{Y}}}}$; where *d* is the difference between means and *a* is the number of treatments tested. This value was then used to compute the non-centrality parameter (λ) as: $\lambda = a\phi^2$.

The statistical power of each study was then calculated with the program G*Power (Faul *et al.*, 2007) using the λ and degrees of freedom with the α-level set at 0.05. This protocol was repeated to model the MDD for different values of CV_ε and CV_β (Table 2) using levels of replications (*b*) from 2 to 6 and number of fish in each tank (*n*) from 10 to 1000.

2.3. Simulation studies

Simulations were performed to compare three different methods for statistical analysis of growth studies with a graded response: ANOVA, a second order polynomial and a three parameter logistic growth model. The simulations were performed with R (R Core Team 2014). The datasets used for the analysis represent random samples from three different populations:

Res45%: A population with a saturation type relationship to treatment where the response increased with treatment level until it plateaued with a response of 100% at treatment levels over 100%. The response to the minimum treatment was 45% lower than the maximum response (100%) (Fig. 1).

Res 11%: A population with saturation type relationship to treatment where the minimum response was 11% lower than the maximum response. The maximum response was 100% and reached when the treatment level was 100% (Fig. 1).

Res0%: A population with no response to treatment (Fig. 1).

198 The population responses to the treatments were normally distributed at each treatment level
199 and the same variance was assumed for all responses regardless of treatment level.
200 The simulations were performed on 1000 datasets generated from each population. The
201 simulations were made for experiments with 18 tanks and 50 fish in each tank. The datasets
202 were random samples, generated based on the mean responses of the population at different
203 treatment levels with equal variance for the means of tanks within all treatment levels. The
204 means of tanks within treatments were normally distributed with a standard deviation equal to
205 4.5% of the grand mean for tanks within treatments. The residual variance within each tank
206 was normally distributed with a standard deviation equal to 30.6% of the grand mean. These
207 standard deviations are the same as the mean CV_β and CV_ε for all species found in this study
208 (Table 1). In the data sets generated, the treatment levels tested were in arbitrary units
209 expressed in percentages and could range between 85% and 121%. To reflect the strengths of
210 different statistical approaches, tanks were allocated differently for mixed model ANOVA,
211 polynomial models and non-linear models. In the mixed model simulations, six levels of
212 treatments were tested, each in triplicate. In each sample, the lowest treatment levels tested
213 ranged at random between 85% and 90% and then successive treatment levels were applied in
214 5% increments. The samples for the polynomial and non-linear models were in duplicate at
215 nine treatment levels. In each sample, the lowest treatment level tested ranged at random
216 between 85% and 89% and then successive treatment levels were applied in 4% increments
217 covering a range of treatment levels of 32%.
218 Three methods were used to analyse the data:
219 1) Mixed model ANOVA with tanks as random factors nested within treatments and
220 measurements of individual fish in each tank using the lme function within the nlme
221 package (Pinheiro et al. 2014) in R. All designs were balanced with the treatment

222 degrees of freedom as 5 (treatment levels - 1) and the residual degrees of freedom as
223 12 (treatment levels × (tanks within treatments - 1)).
224 2) Second order polynomial using the lm function in R.
225 3) Non-linear three parameter logistic growth model using a self-starting logistic function
226 in R (SSlogis)
227 Three approaches were used to compare the analysis methods:
228 1. The critical treatment levels, the minimum treatment level required to generate a
229 maximum response were estimated for all the models:
230 a. For the ANOVA, the highest treatment level that did not generate a response
231 significantly different from those of the two highest treatment levels.
232 b. For the polynomial model, the critical level was the estimated treatment level
233 that caused the maximum response.
234 c. In the logistic growth model, the treatment level causing a response that was
235 98% of the asymptote was arbitrarily chosen as the critical treatment.
236 2. The residual variance of the predicted values for each model from the population
237 values: $\frac{1}{t}\sum_{i=1}^{t}(\hat{Y}_i - Y_i)^2$ where t are the treatment levels tested, $\hat{Y}$ is the predicted
238 response and Y is the population response.
239 3. The maximum responses, estimated from the predicted values of the ANOVA and the
240 second order polynomial and for the asymptote of the logistic regression model.
241

3. Results

3.1. Coefficient of variation for fish within tanks (CV_{ε})

In most studies, CV_{ε} increased as the experiments progressed but tended to stabilize when the factorial increase in body mass (mean body-mass / mean initial body-mass) was about 1.5 (Fig. 2a,b,c). However, this pattern was not entirely consistent: In the study on Atlantic halibut, the CV_{ε} was nearly constant throughout and in the study on tilapia the CV_{ε} increased progressively (Fig. 1a). At the end of the experiments, the mean CV_{ε} was 30.6 ± 4.5% (mean ± SD) and ranged from 15% to 56% (Table 1). There were no clear differences in final CV_{ε} for different species and the CV_{ε} varied between different studies on a single species. Thus the final CV_{ε} for Atlantic cod ranged from 32 to 56% (Fig. 2b; Table 1) and from 15 to 39% for Arctic charr (Fig. 2c; Table 1).

3.2. Coefficient of variation for tanks within treatments (CV_{β})

The mean CV_{β} at the end of all studies was 4.5 ± 0.4% (Mean ± SD; range: 0 – 12). The CV_{β} increased initially in many studies but stabilised as the experiments progressed (Fig. 3a,c). However, this pattern was not consistent in all studies and in some, the CV_{β} decreased as the experiments progressed (Fig. 3a,b). Of the 24 studies investigated, eight had a final CV_{β} of zero; five had CV_{β} ranging from 2% to 5%, while 11 had CV_{β} of above 5%, the highest being 11% (Table 1).

3.3 Correlation between initial and final *CV* and body mass.

In 20 studies (Table 1), information was available on both initial and final variance in body-mass. The final CV_{ε} in these studies was significantly correlated with initial CV_{ε} ($r = 0.621$; $p<0.003$; $N = 20$). Similarly, final CV_{β} in different studies was significantly correlated with the initial CV_{β} ($r = 0.657$; $p<0.002$; $N = 20$).

Information was available from several studies on Arctic charr and Atlantic cod (Table 1.). These data were used to compare the variance in studies on the two species. The final CV_ε and CV_β in experiments on both species ($P < 0.05$) - decreased with increasing final body mass (Fig. 4a, b). Adjusting for body mass, CV_ε was significantly lower ($P < 0.0001$) in Arctic charr than in Atlantic cod (Fig4a); while CV_β were not significantly different (Fig. 4b). However, the initial CV_ε in the studies on Atlantic cod were higher than in the studies on Arctic charr and, when the initial CV_ε is included as a variable in the model, the difference between the species was no longer significant.

3.4. Statistical power and minimum detectable difference with 80% statistical power.

When experiments are designed it is recommended that statistical power is 80%. In the experiments analysed (Table 1), the mean statistical power estimated post hoc was 53.9±0.3% (mean±SD) and ranged from 12% to 100%. The MDD was 18.1 ± 12.8% (range: 4% to 56%) of the grand mean.

To show how experimental design is likely to affect the MDD, we modelled MDD using different number of replications and numbers of fish within each tank. The MDD was modelled for medium, high or low CV_ε and CV_β using the average, maximum and minimum CV_ε and CV_β encountered (Table 1). For the purpose of the modelling, it was assumed that five different treatments were being tested.

The level of replication and the number of fish in each tank affects the MDD (Fig. 5a,b,c). For all levels of replication, the MDD decreases markedly with increasing n until it reaches about 100. There is comparatively little gained in reduced MDD by increasing n over 100. For average CV_ε and CV_β, designs in triplicate are required for reaching an MDD of 20% or less.

Similarly, four to six replications can give a MDD of 10-14% (Fig. 5a). A MDD under 10% is only possible when both CV_ε and CV_β are low (Fig. 6c); reaching 4 to 10% when n is 100.

3.5. Comparison of different methods to analyse graded treatment growth data

Datasets were generated from random samplings of three different populations (Fig. 1) based on the average CV_ε and CV_β (Table 1). In total, 1000 datasets were generated for each population and analysed using a mixed model ANOVA, a second order polynomial and logistic regression. The logistic regression failed to converge on average in 0.1%, 20% and 67% of trials for the Res45%, Res11% and Res0% populations respectively.

With the ANOVA, the estimated mean treatment level required to create a 100% response for the Res45% population was 99.7%, matching closely the critical treatment of the population (100%) with 95% of estimated values being between 96% and 104% (Table 2). The second order polynomial overestimated the critical treatment of the population with more than 95% of the estimates being higher than 107% (Table 2). The critical treatment estimated through the logistic regression (Table 2) was 101% (95% range 97%-107%). However, it should be stressed that the critical treatment was arbitrarily chosen to be where the response reached 98% of the estimated maximum. Obviously the response level chosen will affect the estimate of the critical treatment value.

Analysis of the Res11% population showed a significant treatment effect in 36% of tests with ANOVA and 51% with the polynomial tests. The mean critical treatment estimate from the ANOVA was 95% (Range 90%-103%) while statistical analysis with the polynomial estimated the critical treatment values as 109% (range: 102%-113%) (Table 2). The mean critical treatment estimate from the logistic regression was 103% (range: 93%-113%).

For the Res0% population, where treatment had no effect (all responses were 100%), the polynomial showed significant effects in 5% of tests while the mixed model ANOVA only

showed significant differences in 1% of the analyses. As described above, the logistic regression analysis did not converge in most of the analyses of samples from of the 0% population.

The estimated maximum responses were similar for all methods of analysis with the 95% range of responses covering the population maximum response of 100%. For the Res45% population, estimates from all statistical methods show a similar mean square residual deviation from the population response (Table 2), while at Res11% and Res0% the residual values for the ANOVA were slightly higher than for either the polynomial or the logistic regression. The mean MDD in the ANOVA was 18.3% and 13.1% for the Res45% and Res11% populations respectively.

4. Discussion

This is the first study to evaluate the variance, statistical power and MDD in growth studies of various aquaculture species. Earlier, Ling and Cotter (2003) evaluated the variance in growth studies of triploid Atlantic salmon, finding a mean CV_{ε} of 28 ± 8.6% (range: 14-41%) and CV_{β} of 3.2 ± 1.9% (range: 1-7%). In 29 growth studies on 24 species published during the last year in Aquaculture (Table 3), the estimated mean CV_{β} was 5% (range: 0-49%) while the mean CV_{ε}, was 28%. All these values are in accord with the results of the present study where the CV_{ε} and CV_{β} (mean ± SD) were 30.6 ± 4.5% (range: 15%-56%) and 4.5 ± 0.4% (range: 0-12%) respectively. Both the present study and that of Ling and Cotter (2003), show that CV_{ε} and CV_{β} for a single species can range widely among different studies. The only indication of species differences in variance in body mass is the apparent difference in CV_{ε} between the Atlandic cod and Arctic charr (Fig. 4a). However, this may not reflect species specific variance, but instead higher initial CV_{ε} in the former studies. Fish were selected for these

studies to be within certain size ranges and, therefore, the CV_{ε} does not reflect the natural variation of the species, but rather the abundance of fish available. Combined, these results suggest that the variance encountered in growth studies of different species of fish is similar, suggesting, that similar experimental designs are appropriate for all these species.

The model calculation conducted in this study show, as expected, that both the number of fish in each treatment and the level of replication affect the MDD. Increasing *n* up to 100 decreases the MDD considerably, while increasing *n* over 100 has a limited effect (Fig. 5a,b,c). Increasing the level of replication from duplicates to triplicates reduces the MDD by about 30%. Further increases in the level of replication will reduce the MDD further, although the gain in reduced MDD is progressively decreased with each increase in level of replication.

The MDD is of particular interests for researchers. The average expected MDD for mixed model ANOVA (for statistical power of 80%) in the experimental data analysed from the different growth studies (Table 1) was 23% of the mean (range: 6-55%). In studies published in Aquaculture during the last year (Table 3), treatments in triplicate were the most common (83% of studies), with duplicates (10%) and quadruplicates (3%) being less common. One study used six tanks per treatment. The mean number of fish in each tank in these studies was 25.7 (range: 4-100). For triplicates, *n* of 26 and statistical power of 80%, the expected minimum detectable difference is 26% when variance is average. These results suggest that in most growth studies published, differences smaller than about 25% of the grand mean are not reliably detected (i.e. in least 80% of trials) and half of studies will fail to detect true differences under 20%.

Researcher can take active measures to increase the resolution of statistical tests by increasing the level of replication and the *n*. Furthermore, when CV_{ε} and CV_{β} are low the MDD is also reduced. Both CV_{ε} and CV_{β} tend to increase as the experiments progressed (Fig. 2a,b) and this was also the case in 74% of the growth studies published in Aquaculture during the last year

(Table 3). However, the initial variance and final variance are positively correlated and, therefore, our results suggest that it is possible to reduce the MDD further by selecting fish for experiments within a narrow size range. By using stochastic models Imsland (2001) suggested, that there were two main causes for size variation seen in laboratory studies with turbot: (a) Individual genetical growth rate variation, this trait is stochastic in the population and changes with time (stochastic growth with memory) (b) Combination of individual genetical growth rate and size-related dominance hierarchies. By selecting fish within a narrow size range both a) and b) above will be minimized which makes it possible to reduce MDD. However, if the treatments are size specific, i.e. treatment effect depends on size, selection of fish within a narrow size range may produce a bias in the results.

When the differences among treatments in growth studies are small, the duration of the experiment is also important. As most of the growth experiments evaluated in this study progressed, both CV_{ε} and CV_{β} tended to level off (Fig. 2a,b,c). If CV_{ε} and CV_{β} are stable while the difference in mean size of treatment groups increases with time, statistical power will increase. Furthermore, both CV_{ε} and CV_{β} are reduced as size increases (Fig. 4a,b). Therefore, in order to avoid type II errors, the duration of experiments must be extended where differences between effects of different treatments are small for adequate time.

Another possibility to increase statistical power is to include data from the entire study rather than analysing only the final size of the fish. This can be done with mixed model ANOVA by including time either as a categorical factor (Ling 2007), as a covariate or using repeated measures ANOVA (Imsland 2001). When time is included as a covariate the growth performance is compared as the slopes of the growth curves rather than the final size. However, when there are large differences in the size of the fish at different times, the variances may not be equal and then one of the assumptions of the ANOVA may be violated.

Therefore, it may be necessary to use statistical procedures such as GLM in R which allows data with gamma distribution or PROC MIXED in SAS where variance and covariance structures can be directly modelled.

The results of the present study are an interesting contribution to the discussion of which is the most appropriate statistical method to analyse data from growth studies. Analysing published data on feed studies, Shearer (2000) suggested that ANOVA, in dose-response studies, might under-estimate the critical treatment effect required to produce a maximum response due to the inability of ANOVA to detect small differences. Instead he recommended using regression techniques, either polynomial or logistic. However, the results of the simulations performed in the present study directly contradict his conclusion. They suggest that ANOVA does not necessarily underestimate the critical treatment effect. In fact, the estimate of critical treatment with ANOVA most closely matched the critical value of the populations. Polynomials tended to overestimate the critical treatment level by 11% on average. With the logistic asymptotic function, it is difficult to decide when the maximum response is reached and this will limit its usefulness. Furthermore, the logistic regression procedure failed in many cases to fit the model, especially when the treatment effect was small. Moreover, the advantage of using ANOVA rather than the linear and nonlinear methods is that it does not presuppose the shape of the relationship between treatment and effect. Therefore, we suggest that a mixed model ANOVA is the most appropriate statistical method to analyse data from growth studies.

4.1. Conclusions

The results of this study suggest that the variance in aquaculture growth studies on different species is similar and, therefore, a similar experimental design (replication level and number of fish in each unit) can be employed in growth studies regardless of the species of fish. The results of the study suggest that most aquaculture growth studies cannot reliably (with 80%

power) detect a difference in weight that is less than 26%. However, researchers can take measures to reduce the minimum detectable difference by selecting fish within a narrow size range for experiments. This may reduce the MDD to 5% with adequate replication.

The results of the present study suggest, that in contrast to the suggestions of Baker (1986), Cowey (1992) and Shearer (2000), a mixed model ANOVA is the best approach to analyse growth data with graded responses and superior to non-linear models.

Acknowlegdements

This work was funded by the United Nation University, Fisheries Training Programme. The research team at Holar University College and Maurice Ssebisubi are thanked for providing their data. Parts of the paper were written while HT was a visiting scholar at the Department of Environmental and Health Studies at Telemark University College, Norway.

425 **References**

426 Araujo, P., Frøyland, L., 2007. Statistical power and analytical quantification. J. Chromotogr.
427 847, 305–308.

428 Baker, D.H. 1986. Problems and pitfalls in animal experiments designed to establish dietary
429 requirements for essential nutrients. J. Nutr. 116, 2339-2349.

430 Cowey C.B.1992. Nutrition: estimating the requirements of rainbow trout. Aquaculture 100,
431 177-189.

432 Faul, F., Erdfelder, E., Lang, A.G., Buchner, A., 2007. G*Power 3: A flexible statistical
433 power analysis program for the social, behavioral, and biomedical sciences. Behavior
434 Research Methods 39, 175-191.

435 Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiements.
436 Ecological Monographs 54: 187-211.

437 Imsland, A.K., Gústavsson, A., Gunnarsson, S., Foss, A., Árnason, J., Jónsson, A.,
438 Smáradóttir, H., Arnarson, I., Thorarensen, H. 2008. Effects of reduced salinities on growth,
439 feed conversion efficiency and blood physiology of juvenile Atlantic halibut (*Hippoglossus*
440 *hippoglossus* L.). Aquaculture 274, 254-259.

441 Imsland, A.K., Gunnarsson, S., Roth, B., Foss, A., Le Deuff, S., Norberg, B., Thorarensen,
442 H., Helming, T. 2013. Long-term effect of photoperiod on growth, maturation and flesh
443 quality in turbot. Aquaculture 416-417; 152-160.

444 Imsland, A.K., 2001. Growth mechanisms in fishes and the use of stochastic models to
445 simulate growth and size-variation. Cand. scient. thesis in applied statistics, Dept. Math.,
446 University of Bergen, 105 pp.

447 Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from
448 restricted maximum likelihood. Biometrics 53, 983-997.

449 Kenward, M. G. and Roger, J. H. 2009. An improved approximation to the precision of fixed
450 effects from restricted maximum likelihood. Computational Statistics and Data Analysis, 53,
451 2583-2595.
452 Kozlov, M.V., Hurlbert, S.H. 2006. Pseudoreplication, chatter, and the international nature of
453 science: A response to D. V. Tatarnikov. Journal of Fundamental Biology (Moscow) 67, 145-
454 152.
455 Ling, E.N., Cotter, D., 2003. Statistical power in comparative aquaculture studies.
456 Aquaculture 224, 159–168.
457 Ling, E.N. 2007. Efficient analysis of growth trial data. Aquaculture research 38, 728-732.
458 Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2014). _nlme: Linear and
459 Nonlinear Mixed Effects Models. R package version 3.1-117, <URL: http://CRAN.R-
460 project.org/package=nlme>.
461 R Core Team (2014). R: A language and environment for statistical computing. R Foundation
462 for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
463 Ruhonen, K 1998. Individual measurements and nested designs in aquaculture experiments: a
464 simulation study. Aquaculture 165: 149–157.
465 Schram, E., Roques, J.A.C, Abbink, W, Yokohama, Y., Spanings, T., Vries, P., Bierman, S.,
466 Vis, H.v.d., Flik, G. 2012. The impact of elevated water nitrate concentration on physiology,
467 growth and feed intake of African catfish *Clarias gariepinus* (Burchell 1822). Aquaculture
468 Research, 1-13
469 Searcy-Bernal, R. 1994. Statistical power and aquaculture research. Aquaculture 127: 371-
470 388.
471 Shearer, K.D. 2000. Experimental design, statistical analysis and modelling of dietary nutrient
472 requirement studies of fish: a critical review. Aquac. Nutr. 6, 91-102.

473 Smart, T.S., Riley, J., Edwards, P. 1998. Statistical aspects of aquaculture research: sample
474 sizes for pond experiments. Aquac. Res. 29, 373-379.
475 Sokal, R.S., Rolf, F.J. 2012. Biometry. The principles and practices of statistics in biological
476 research. W.H.Freeman and Company, New York.
477 Ssebisubi, M., 2008. Effect of water temperature and feeding level on survival, feed
478 utilization, growth and whole body composition of *Oreochromis shiranus* (Trewavas, 1983)
479 at different body sizes. M.Sc. thesis, Dept. of Aquaculture and Fisheries Sci., University of
480 Malawi, 158pp.
481 Tang, P.C., 1938. The power function of the analysis of variance tests with tables and
482 illustrations of their use. Sta. Res. Memor. 2, 126-149.
483 Tobin, T., Kause, A., Mäntysaari, E.A., Martin, S.A.M, Houlihan, D.F., Dobly, A., Kiessling ,
484 A., Rungruangsak-Torrissen, K., Ritola, O., Ruohonen, K. 2006. Fat or lean? The quantitative
485 genetic basis for selection strategies of muscle and body composition traits in breeding
486 schemes of rainbow trout (*Oncorhynchus mykiss*)
487 Thorarensen, H., Gustavsson A., Mallya Y., Gunnarsson S., Arnason, J., Arnarson, I., Jonsson
488 F. A., Smaradottir, H., Zoega, T. G., Imsland, K. A., 2010. The effect of oxygen on the
489 growth and feed conversion of Atlantic halibut (*Hippoglossus hippoglossus* L.). Aquculture
490 309, 96-102.

491 Table 1. Variance and power in 24 independent growth studies on fish.

Study	Species	Treatment levels[1]	No. of tanks[2]	N[3]	Average final body mass (g)[4]	d (% of grand mean)[5]	CV_ε[6]	CV_β[7]	Observed power[8]	Minimum detectable difference at 80% power[9]
1	Halibut	5	3	47	122	24	0.32	0.00	99	11
2	Turbot	3	3	36	330.3	30	0.28	0.09	44	36
3	Tilapia	3	6	16	11.3	56	0.37	0.04	100	33
4	Arctic charr	7	4	50	4.7	30	0.25	0.07	100	22
5	Arctic charr	7	4	39	10.9	17	0.28	0.08	49	28
6	Arctic charr	6	4	50	90	12	0.21	0.09	23	32
7	Arctic charr	6	3	35	230.8	11	0.24	0.04	34	21
8	Arctic charr	6	3	132	672.8	4	0.15	0.02	40	8

9	Arctic charr	6	3	64	1067.9	4	0.18	0.00	20	9
10	Arctic charr	6	3	60	1437.5	10	0.17	0.00	98	15
11	Arctic charr	6	3	96	886.7	17	0.39	0.06	55	27
12	Arctic charr	16	3	30	2.3	37	0.26	0.06	100	33
13	Arctic charr	6	3	90	1082.9	6	0.16	0.03	23	12
14	Arctic charr	16	4	151	4.7	19	0.26	0.06	97	23
15	Atlantic cod	5	3	13	800	18	0.36	0.00	41	31
16	Atlantic cod	5	3	12	1497.3	13	0.33	0.00	60	6
17	Atlantic cod	5	3	46	248.7	7	0.32	0.05	12	24
18	Atlantic cod	6	3	15	791.8	20	0.35	0.00	46	32
19	Atlantic cod	6	3	32	105.2	37	0.32	0.12	37	55
20	Atlantic cod	3	6	56	1.9	16	0.36	0.07	38	17
21	Atlantic cod	2	9	105	1.8	17	0.39	0.10	92	14

22	Atlantic cod	2	5	31	0.23	13	0.48	0.11	29	28
23	Atlantic cod	2	5	35	0.52	8	0.36	0.00	44	12
24	Atlantic cod	2	5	14	0.08	13	0.56	0.00	13	31

[1]Number of treatments tested in the experiment.

[2]Number of tanks tested for each treatment.

[3]Number of fish in each tank.

[4]Mean body-mass of fish (g) in a study.

[5]Maximum difference between treatments means ((% of grand mean).

[6]Error coefficient of variation (CV_{ε}).

[7]Coefficient of variation for tanks within treatment (CV_{β}).

[8]Retrospective power (%) at the end of studies.

[9]Effect size (% of grand mean) at 80% power.

Data from: 1-Thorarensen *et al.* (2010); 2-Le Deuff *et al.*(2010); 3-Ssebisubi, (2008);.4–14-Sigurgeirsson *et al.*, unpublished ; 15-Sigurgeirsson and Árnason, unpublished; 16-21-Árnason *et al.*, unpublished; 22–24-Edelsparre and Pálsson, unpublished.

Table 2. Summary of analyses from simulation studies on data sampled from artificial populations, two with graded responses (Res11% and Res45%) and one population with no response to treatment (Res0%). Randomized normally distributed data with equal variances was generated based on the population responses assuming that CV_{ε} was 30.6 and CV_{β} was 4.5. The treatment level required to give a maximum response was 100% for all artificial populations and the maximum response was 100%.

	ANOVA			Second order polynomial			Three parameter logistic regression		
	Res45%	Res11%	Res0%	Res45%	Resp11%	Resp0%	Resp45%	Resp11%	Resp0%
Mean critical treatment (±95% range)[1]	99.7 (96-104)	95.0 (90-103)	92.3 (90-101)	110.7 (107-113)	108.8 (102-113)	96.5 (85-108)	101.5 (97-107)	97.0 (88-128)	-
Median critical treatment (%)	100	95	92	111	108	98	101	92	-
Mean maximum response (95% range)[2]	100 (95-105)	101 (97-106)	100 (94-105)	103 (99-106)	102 (99-106)	105 (101-109)	96.8 (93-101)	97 (93-112)	-

Mean effect size as % of grand mean (95% range)	18.3 (10.2-26.9)	13.1 (8.8-17.5)	9.0 (5.8-12.0)	-	-	-	-	-	-
Mean square residual deviation[3]	8.4	12.0	23.4	9.6	10.2	16.5	8.9	3.8	-
Proportion of analyses showing a significant effect of treatment	100%	36%	1%	100%	51%	5%	-	-	-
Analysis producing an error message	-	-	-	-	-	-	0.1%	20%	67%

[1]The treatment effect required to give maximum response

[2]Estimated maximum effect.

[3]The mean square residual deviation between predicted responses and population responses.

Table 3. A summary of variability of final body mass and experimental design in 29 growth studies of 24 species of fish published in 2013 and 2014 in Aquaculture. The CV_β were estimated based on reported standard errors and levels of replication in studies where simple ANOVA was used for statistical analysis.

	Mean	Range	Mean factorial increase[2]
CV_ε (%)[1]	27.9	23-36	1.78
CV_β (%)	4.9	0-49	1.75
Level of replication (rearing units / treatment)	3	2-6	
Number of fish in each rearing unit	25.7	4-100	

[1]Information on CV_ε was only available in 4 studies.

[2]Final divided by the initial CV_ε and CV_β.

522 Figure captions

523 Figure 1. The three populations used in the model simulations: Res45% where the minimum
524 treatment gave a response that was 45% less than the maximum; Res11% where the minimum
525 treatment gave a response that was 11% less than the maximum; and Res0% where treatment
526 had no effect on response. The units for treatment and response are shown as percentages. For
527 Res11% and Res45%, a treatment level of 100% will produce a 100% response.

528 Figure 2: Development of CV_{ε} with increasing body mass in experiments on (a) tilapia,
529 Atlantic halibut and turbot, (b) Atlantic cod and (c) Arctic charr. (The different lines
530 represent separate studies). The increase in body mass is shown as factorial increase (mean
531 body-mass / mean initial body-mass).

532 Figure 3: Development of CV_{β} with increasing body mass in experiments on (a) tilapia,
533 Atlantic halibut, and turbot, (b) Atlantic cod and (c) Arctic charr. (The different lines
534 represent separate studies). The increase in body mass is shown as factorial increase (mean
535 body-mass / mean initial body-mass).

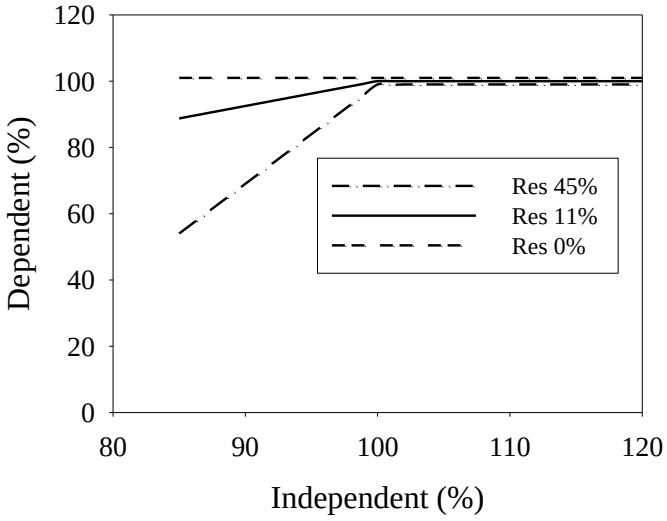
536 Figure 4: Coefficients of variation in growth studies of Atlantic cod and Arctic charr at
537 different final mean body mass. a) CV_{ε} and mean final body-mass. The intercepts for the two
538 species were significantly different ($p<0.0001$) while the slopes of the regression lines for the
539 two species were not significantly different. The regression lines (interrupted for the Atlantic
540 cod, continuous for Arctic charr) with a common slope was CV_{β} = Intercept - 0.006 × body
541 mass with the intercepts being 25.7 and 40.6 for the Arctic charr and Atlantic cod
542 respectively. b) CV_{β} and mean final body-mass. Neither slopes nor intercepts were
543 significantly different. The common regression line was: CV_{β} = 6.67 - 0.005 × body-mass (R^2:
544 0.38).

545 Figure 5: Minimum detectable difference (MDD), shown as % of the grand mean in growth
546 studies with five treatments levels when statistical power is 80%. a) Mean CV_{ε} and mean
547 CV_{β}. b) High CV_{ε} and high CV_{β}. c) Low CV_{ε} and low CV_{β}.

548

549

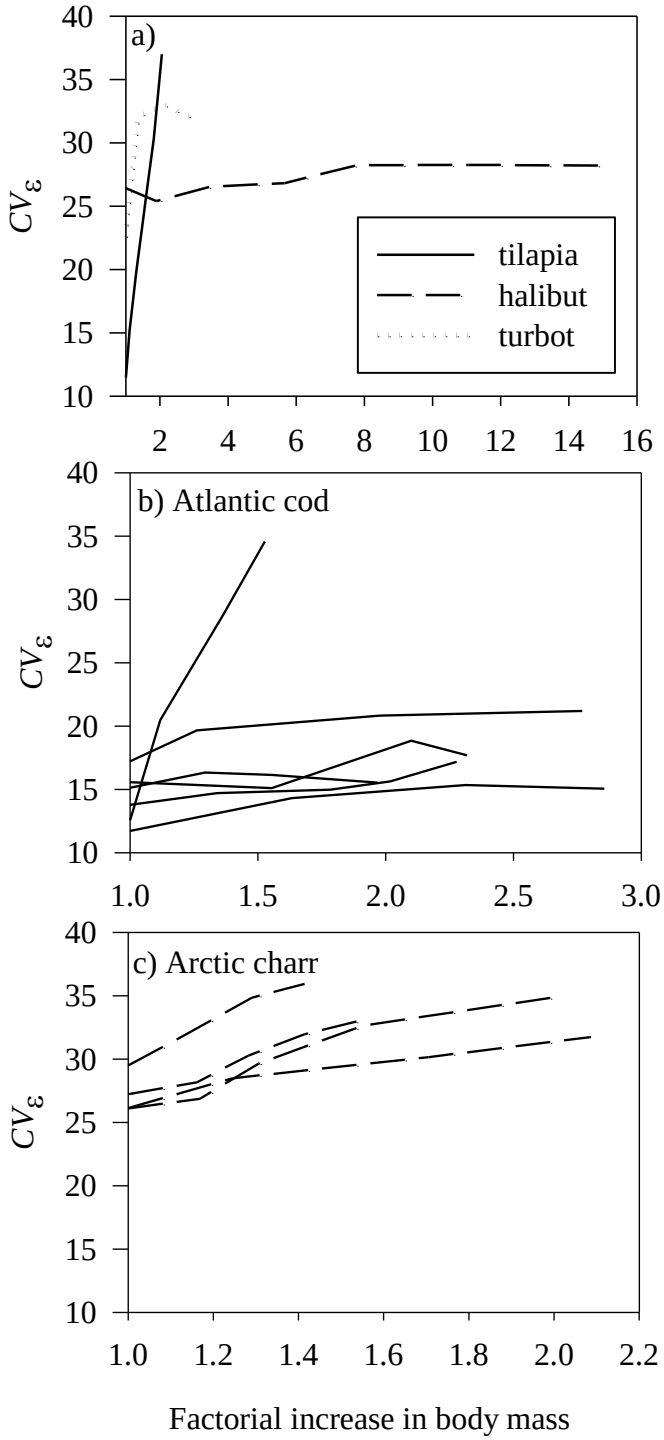
550 Figure 1.



551

552

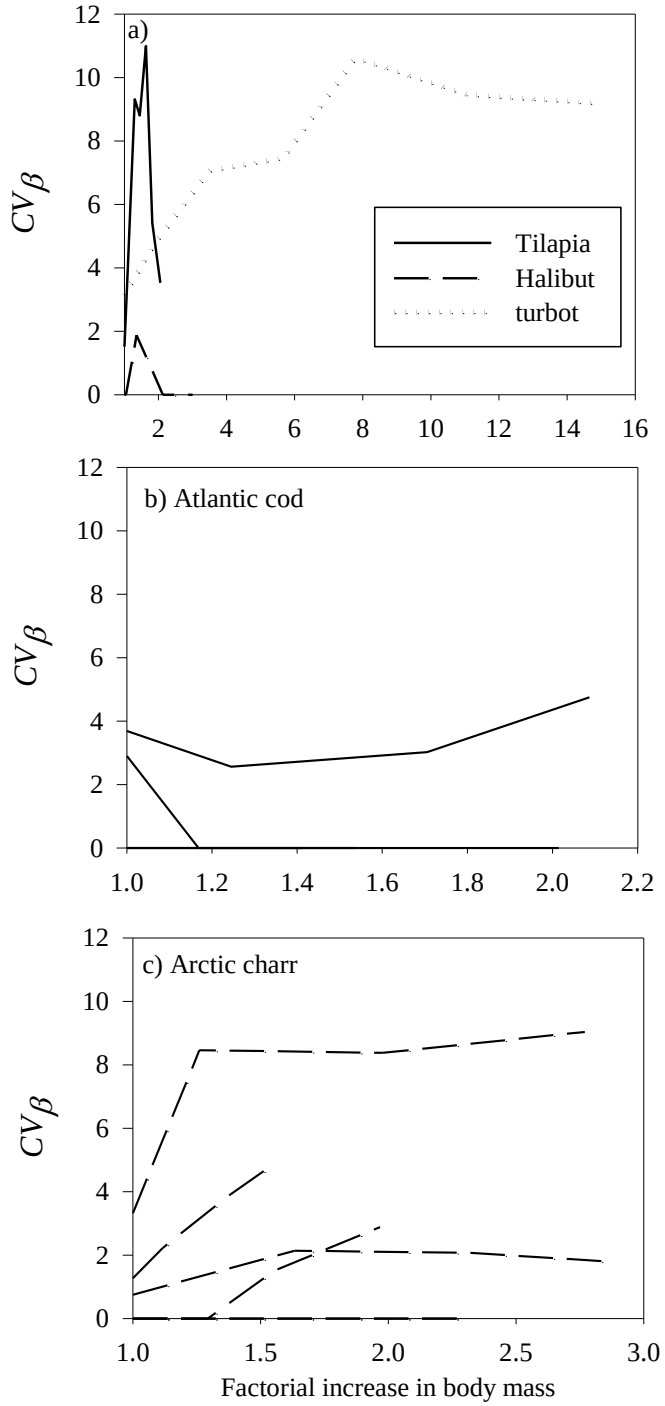
553 Figure 2.



554

555

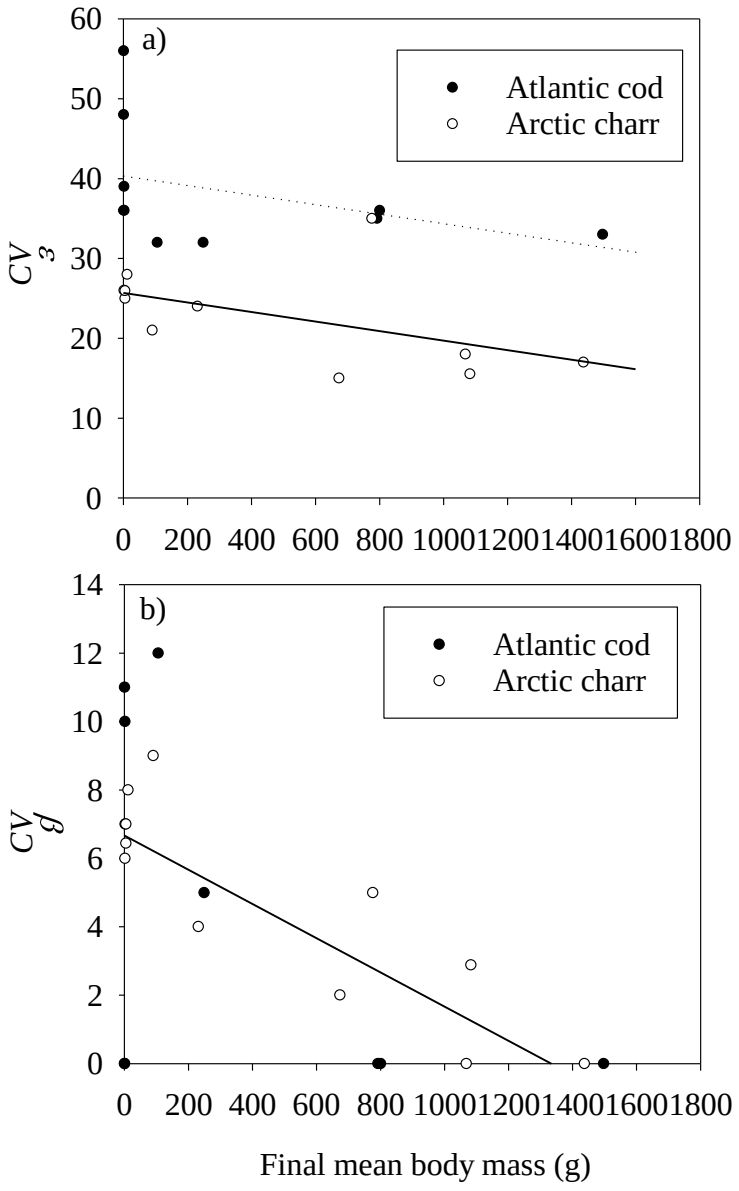
556 Figure 3.



557

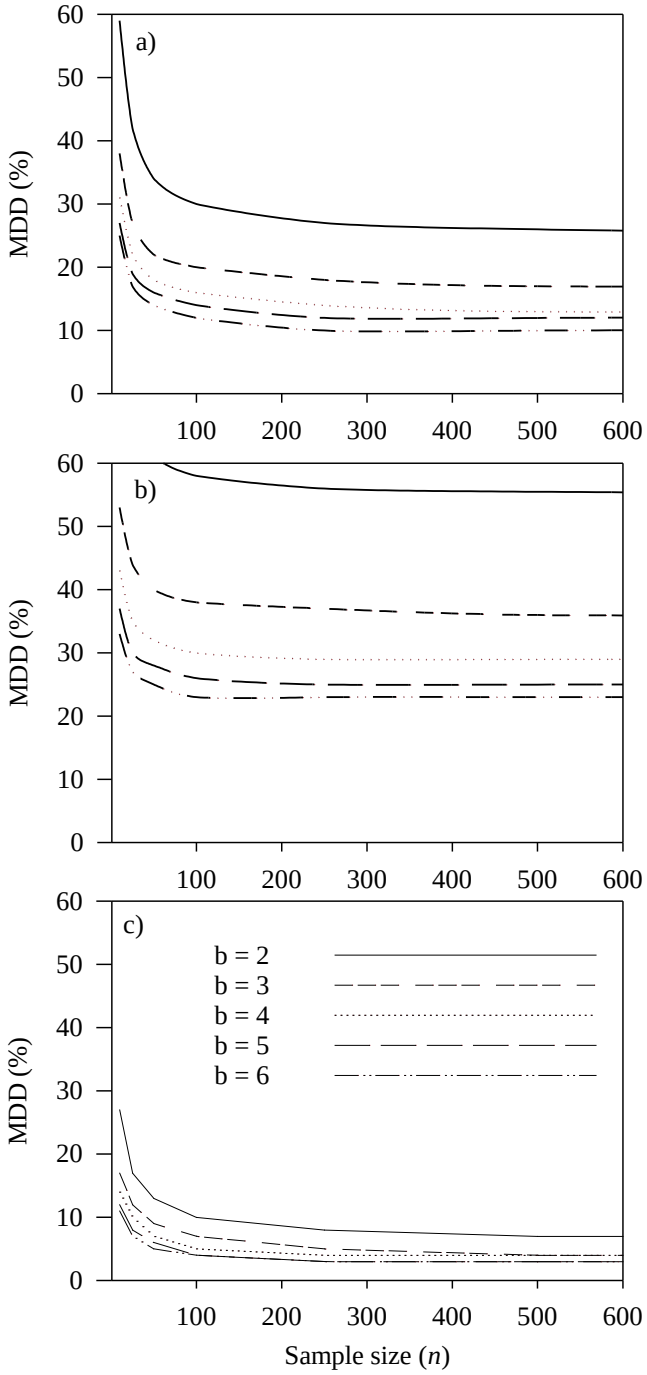
558

559 Figure 4.



560

561 Figure 5.



562

563