Design of Cost-Effective Cancer Biomarker Reproducibility Studies

by

Stephanie Carrieann Cooke

(Under the direction of Dr. Kevin Dobbin)

Abstract

The purpose of this study is to find an optimal experimental design for cancer biomarker reproducibility studies. A biomarker is defined as a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (CCR Focus March 9, 2010). Vital signs, such as blood pressure, can also be considered biomarkers. Experiments that measure biomarkers can be costly and time consuming. In response to these factors, this study has identified and developed an algorithm that determines optimal allocation of samples for the most effective experimental results. This biomarker reproducibility study estimates the variance of laboratory measurements by using the intraclass correlation coefficient (ICC), which, unlike the F test, uses no hypothesis testing of population means. The ICC evaluates the amount of overall variance relative to between-subject variability (von Eye et al. 2005). This study focuses on finding the most cost-efficient design for cancer biomarker studies assessing reproducibility.

Index words: cancer biomarker, experimental design, intraclass correlation coefficient (ICC)

Design of Cost-Effective Cancer Biomarker Reproducibility Studies

by

Stephanie Carrieann Cooke

A.A., Georgia Perimeter College, Physics 2004

B.S., Georgia Institute of Technology, Biology, minor Biomedical Engineering 2006

A Thesis Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Master of Science in Statistics

Athens, Georgia

2012

Design of Cost-Effective Cancer Biomarker Reproducibility Studies

by

Stephanie Carrieann Cooke

Approved:

Major Professor: Dr. Kevin Dobbin

Committee: Dr. William McCormick
Dr. Cheolwoo Park

Electronic Version Approved:

Dr. Maureen Grasso
Dean of the Graduate School
The University of Georgia
July 2012

Design of Cost-Effective Cancer Biomarker Reproducibility Studies

Stephanie Carrieann Cooke

July 23, 2012

Contents

1 Objective of Cancer Biomarker Designs **6**

2 Literature Review **8**

3 Background **15**

4 Results **22**

5 Conclusion **35**

6 Resources **37**

7 Appendix **39**

List of Figures

4.1 Between ICC 48 Sample Model A 30

4.2 Between ICC 48 Sample Model B 30

4.3 Between ICC 48 Sample Model C 30

4.4 48 Sample Bias Squared and Variance 31

4.5 48 Sample Mean 31

7.1 Between ICC 72 Sample Model A 45

7.2 Between ICC 72 Sample Model B 45

7.3 Between ICC 72 Sample Model C 45

7.4 72 Sample Bias Squared and Variance 46

7.5 72 Sample Mean 46

7.6 Between ICC 96 Sample Model A 57

7.7 Between ICC 96 Sample Model B 57

7.8 Between ICC 96 Sample Model C 57

7.9 96 Sample Bias Squared and Variance 58

7.10 96 Sample Mean 58

List of Tables

4.1 Balanced Data with 2 Replications for Sample Size of 48 25

4.2 Balanced Data with 1 Replication for Sample Size of 48 26

4.3 Unbalanced Data for Sample Size of 48 27

4.4 Balanced LM Data with 2 Replications for Sample Size of 48 28

4.5 Balanced LM Data with 1 Replication for Sample Size of 48 29

4.6 LMER 2 Replication Design with Interaction 32

4.7 LMER 1 Replication Design with Interaction 32

4.8 LMER 2 Replication Design with no Interaction 32

4.9 LMER 1 Replication Design with no Interaction 32

4.10 Linear Model 2 Replication Design with no Interaction 33

4.11 Linear Model 1 Replication Design with no Interaction 33

4.12 Model A 48 Sample Confidence Interval 33

4.13 Model B 48 Sample Confidence Interval 33

4.14 Model C 48 Sample Confidence Interval 33

4.15 48 Sample 2 Replicate Design with 6 Samples and 4 Laboratories 34

4.16 48 Sample 1 Replicate Design with 12 Samples and 4 Laboratories 34

4.17 48 Sample Unbalanced Design with 8 Samples and 4 Laboratories 34

7.1 Balanced LMER Data with 2 Replications for Sample Size of 72 for LMER model 40

7.2 Balanced LMER Data with 1 Replication for Sample Size of 72 41
7.3 Unbalanced LMER Data for Sample Size of 72 42
7.4 Balanced LM Data with 2 Replications for Sample Size of 72 43
7.5 Balanced LM Data with 1 Replication for Sample Size of 72 44
7.6 2 Replication Design with Interaction 47
7.7 1 Replication Design with Interaction 47
7.8 2 Replication Design with no Interaction 47
7.9 1 Replication Design with no Interaction 47
7.10 Linear Model 2 Replication Design with no Interaction 48
7.11 Linear Model 1 Replication Design with no Interaction 48
7.12 Model A 72 Sample Confidence Interval 48
7.13 Model B 72 Sample Confidence Interval 48
7.14 Model C 72 Sample Confidence Interval 48
7.15 72 Sample 2 Replicate Design with 9 Samples and 4 Laboratories 49
7.16 72 Sample 1 Replicate Design with 18 Samples and 4 Laboratories 49
7.17 72 Sample Unbalanced Design with 12 Samples and 4 Laboratories 50
7.18 LMER Balanced Data with 2 Replications for Sample Size of 96 52
7.19 LMER Balanced Data with 1 Replication for Sample Size of 96 53
7.20 LMER Unbalanced Data for Sample Size of 96 54
7.21 Balanced LM Data with 2 Replications for Sample Size of 96 55
7.22 Balanced LM Data with 1 Replication for Sample Size of 96 56
7.23 LMER 2 Replication Design with Interaction 59
7.24 LMER 1 Replication Design with Interaction 59
7.25 LMER 2 Replication Design with no Interaction 59
7.26 LMER 1 Replication Design with no Interaction 59
7.27 Linear Model 2 Replication Design with no Interaction 60

7.28 Linear Model 1 Replication Design with no Interaction 60
7.29 Model A 96 Sample Confidence Interval 60
7.30 Model B 96 Sample Confidence Interval 60
7.31 Model C 96 Sample Confidence Interval 60
7.32 96 Sample 2 Replicate Design with 12 Samples and 4 Laboratories 61
7.33 96 Sample 1 Replicate Design with 24 Samples and 4 Laboratories 61
7.34 96 Sample Unbalanced Design with 16 Samples and 4 Laboratories 62

Chapter 1

Objective of Cancer Biomarker Designs

In reproducibility studies, the same experiment is performed several times, and an attempt is made to determine whether the same result is produced. An example of a reproducibility study was done by Dobbin et al. (2005) to measure gene expression values from microarrays. It used the between laboratory ICC as the measure of reproducibility with multiple laboratories. Another study that used reproducibility was the McShane et al. study (2000). This study compared the reproducibility of p53 measurements in different tumors. The laboratory reproducibility between laboratories is compared. Another example is the study done by Jessup et al. (2009) to determine the expression of 18q LOH. The chromosome 18q LOH (long arm of the chromosome) is a prognostic marker. It was shown that stage II colon cancer patients with 18q LOH had a smaller rate of survival. This study measured whether the loss of heterozygosity on the long arm of chromosome 18 was reproducible in different laboratories. The reproducibility must be established before it can be used in a large clinical trial. Reproducibility studies are important because they will test whether the results of an event in one situation can be reproduced in another situation. In many cases

the results from different laboratories will be biased, and therefore the decisions made by doctors for the patients will be based on biased information. Reproducibility ensures the laboratory results from different laboratories are the same and not biased. The intraclass correlation coefficient (ICC) is one measurement of reproducibility. The variability of the ICC is determined by measuring the mean square error (MSE), and designs that result in smaller MSE are preferred. When the MSE is small, if the experiment is repeated, then the same ICC is expected. The closer the ICC is to 1, the greater the likelihood of experimental reproducibility. When the ICC is 1, the results in one laboratory should be reproduced in another laboratory. The ICC is a better measurement than the regular Pearson correlation, because the ICC considers experimental design. These experiments can be expensive, and in some cases, measurement of the sample just once can be costly. Therefore, the focus of this study is to construct a design that provides the most cost-effective information about an experiment. In a traditional statistical setting (such as an F-test), groups are arranged to detect statistical differences. A reproducibility study estimates the similarly between the groups. In this study, the ICC was used as a measure of reproducibility to test whether the laboratories are the same.

Chapter 2

Literature Review

Saito et al. (2006) examined designs in which multiple raters evaluated multiple subjects to determine the inter-rater reliability of rating scales. A two-way analysis of variance (ANOVA) model without interaction was used, which identified subjects and raters as random effects. A study was performed on the number of subjects and the number of raters that minimized the variance of the ICC with a fixed total number of ratings. The relative ratio is the ratio of rater variance to error variance. A large number of raters improved the accuracy of the ICC when the relative ratio was greater. The Saito et al. (2006) included 144 observations, but since the study utilized 12 raters to obtain these 144 observations, this approach becomes an expensive proposition in terms of reproducibility. Their investigation was a rater study, and it cost nothing to rate a single item repeatedly, but it can be expensive to repeat an experiment even once across 12 laboratories. Also, the logistics of employing more than a small number of laboratories would not prove feasible. For the same rater to rate the same item is inexpensive, but having the same laboratory measure the same sample several times is expensive. Each laboratory experiment can also be complex. Additionally, as more steps exist in a protocol, the likelihood that people will not follow that exact protocol increases. Duplication of protocol is imperative to the validity of a study.

The laboratory effect is analogous to the rater effect. Their rater-to-rater variation applies to our laboratory-to-laboratory variation. One rater may rate a little high just as one laboratory may score a little higher than another. The Saito et al. (2006) study is similar to our research, but the key differences are costs and context. The Saito et al. (2006) study context included multiple raters who performed assessments, whereas this study included multiple laboratories which performed assays. These cost-free raters, in the Saito et al. (2006) study, can be compared to laboratories in this study, which in theory are costly. The Saito et al. (2006) study has many raters, and in contrast, our study involved only four laboratories.

The McShane et al. study (2000) was performed to assess the reproducibility of immunohistochemistry for measuring the p53 expression in bladder tumors. The McShane study, which compared laboratories, did not answer the question under the examination of this study. The objective of this study is to design a reproducibility study to give the best or smallest root MSE with a fixed total number of observations. McShane et al. (2000) used a different design than the design used in this study, which examined laboratory-to-laboratory reproducibility. They used one design, while this study compared multiple designs to determine the best design for the study. Their experiment implicated fifty paraffin blocks that were selected randomly from invasive primary bladder tumors. After each block was stained, intralaboratory and interlaboratory studies were observed with consideration for staining and scoring to establish whether the stains were positive or negative. McShane et al. (2000) used a design with two replications within the laboratory per sample and compared laboratories on a scoring method, which produced integers from 0 to 4 with the result that the data were nominal. This study assumes continuous data from a normal distribution.

Bland et al. (2010) stated that in order to determine whether a new measurement technique should replace an established one, the clinical measurement contrast of the new technique to the established one can be necessary to establish whether the two techniques agree. However,

such a comparison can be performed incorrectly by using correlation coefficients, because, sometimes, the meaning of statistical correlation is ambiguous. The correlation coefficient "r" measures the strength of a linear relation between two variables, not the agreement between them. Bland et al. (2010) pointed out that data which seem to be in poor agreement can generate strong correlations. An example was reported by Serfontein and Jaroszewicz (1978) that compared two methods for the measurement of gestational age. Babies with a gestational age of 35 weeks by one method had gestations between 34 and 39.5 by the other method; but, the comparison produced a high r value (0.85).

Another observation from Bland et al. (2010) is that a change in the scale of measurement affects the agreement but not the correlation. For example, subcutaneous fat can be measured by skinfold calipers. The calipers will measure two thicknesses of fat. If caliper measurements were plotted against half-caliper measurements a straight line with a slope of 2.0 should be obtained. The correlation would be 1.0, but the measurements would not agree, because the fat thickness was obtained by two methods. In addition, another example described by Bland et al. (2010) is when the range of the true quantity sample is wide, the correlation will be larger than if the range is small. When a subject has a peak expiratory flow meter (PEFR) that is less than 500 l/min, r is 0.88; while for those with greater PEFRs, r is 0.90. These are less than the total correlation of 0.94, but according to Bland et al. (2010) it is not reasonable to argue that agreement is worse below 500 l/min and worse above 500 l/min than it is for everybody. Investigators usually try to compare two methods over the whole range of values encountered, so a high correlation can result.

There may be a "Gold Standard" established measurement method that works. Sometimes, researchers try to find a cheaper version of the Gold Standard. The assays must be approved by ethical committees in order for them to be used. We assume there is no Gold Standard when using the ICC to access reproducibility.

For a set of samples, two different measurements are performed on each sample resulting in

two times the samples. Each of the samples is shown on the Bland Altman plot by the mean of the two measurements as the abscissa value and the difference between the two values as the ordinate value.

The ICC has been used extensively and, therefore, is applicable to this study. However, there are other correlation measurement methods, such as the method used in the 1989 Lin et al. study, which developed a new reproducibility index. This index is the correlation involving two readings that should ideally fall on a 45 degree line through the origin. One reading comes from the x-axis, and the other reading comes from the y-axis. This index, known as the Lin concordance correlation coefficient (CCC), measures how well a new reading of observations replicates an original set. Lin's CCC is an alternative to the ordinary Pearson correlation coefficient, which measures the linear relationship between two sets of measurements. The slope of the line relating the two readings should be 1 for the second reading to reproduce the first reading. The statistical properties of Lin's estimate are adequately assessed using an inverse hyperbolic tangent transformation (also known as Fisher transformation). Correlation measures can miss systematic shifts and do not capture the deviation from the 45 degree line. The Lin Concordance Correlation Coefficient measures the deviation from the 45 degree line. Calculating the Lin Concordance Correlation Coefficient for experimental designs would be nontrivial for our study, because several plots would be needed; therefore, the Coefficient is not used in our study. Our model has two sources of variation, including the biological sample and laboratory, so more scatter plots would need to be formed.

The coefficient ρ_c is derived to allow the researcher to determine how well the relationship between the measurements is represented by a line through the origin at a 45-degree angle. If the two measurements have identical results, the measurements are represented by a line

through the origin at an angle of 45 degrees. The equation for ρ_c is as follows:

$$\rho_c = 1 - \frac{E[(Y_1 - Y_2)^2]}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} \tag{2.1}$$

The Pearson Correlation Coefficient R is defined as:

$$R = \frac{\Sigma(Y_1 - \mu_1)(Y_2 - \mu_2)}{[\Sigma(Y_1 - \mu_1)^2 \Sigma(Y_2 - \mu_2)^2]^{\frac{1}{2}}} \tag{2.2}$$

(Kutner et al. 2005)

The expected squared perpendicular deviation from the line when the measurements are correlated is $E[(Y_1 - Y_2)^2]$, and $\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2$ is the expected squared perpendicular deviation from the line when the measurements are uncorrelated.

ρ_c can also be written as

$$\rho_c = R * C_b. \tag{2.3}$$

R is the Pearson correlation coefficient. C_b, which is the second measurement component, is a bias factor that is calculated as follows:

$$C_b = \frac{2}{\frac{v+1}{v+u^2}} \tag{2.4}$$

where C_b measures "how far the best fit line deviates from the 45 degree line,"

$$v = \frac{s_1}{s_2} \tag{2.5}$$

and

$$u = \frac{\mu_1 - \mu_2}{\sqrt{s_1 * s_2}} \tag{2.6}$$

The means of the first and the second set of measurements are μ_1 and μ_2, respectively.

The standard deviations of the first and the second set of measurements are s_1 and s_2 , respectively.

If $\beta_1 = \frac{s_1}{s_2}$ and $\beta_0 = \mu_1$ - β_1 μ_2 are the slope and intercept from the conditional distribution of Y_1 given Y_2, then

$$\rho_c = \frac{2\beta_1 {s_2}^2}{({s_1}^2 + {s_2}^2)[(\beta_0 - 0) + (\beta_1 - 1)\mu_2]^2}. \tag{2.7}$$

The inverse hyperbolic transformation or the Fisher Z transformation is

$$Z = log(\frac{1 + \rho c}{1 - \rho c}) \tag{2.8}$$

(Lin 1989 and Versaudaraan, 2009).

In addition, this study could have been done such that an F test was completed to compare the sample means. The null hypothesis of this F test would be that the means of the laboratories are equal, while the alternative is that the means of the laboratories are not equal. Therefore,

$H_0 : \mu_{Laboratory1} = \mu_{Laboratory2} = \mu_{Laboratory3} = \mu_{Laboratory4}$

H_A: not all population means of laboratories are equal

and

$$F = \frac{MSL}{MSE} \tag{2.9}$$

such that MSL is the mean square of the laboratories and MSE is the mean square of the error. Please see, for example, Table 4.8 without interaction.

This test could have been done in this study to compare whether the laboratories are getting the same average result. In contrast, this ICC laboratory study does not reject the null hypothesis when there are no laboratory effects but instead the ICC determines how well the laboratories agree with each other.

The Dobbin et al. (2005) study determines whether 4 laboratories are reproducible using unbalanced designs. This study used the ICC and hierarchical cluster analysis to determine that complete tumor microarray analysis can be performed for a single study at multiple independent laboratories. From this study, it was shown that there is a possibility to develop a standardized assay that has sufficient reproducibility for clinical use.

Chapter 3

Background

The sample size is an integer multiple of the number of biological samples and laboratories for a design when every cell has the same number of observations (known as a balanced design). This is not true for an experiment, where each cell does not have the same number of observations, known as an unbalanced design. By replication, it is possible to estimate the experimental error variance and measure the variability between the laboratories and the biological samples. A repeated biological sample or laboratory effect reveals any difference in the response compared to a prior response resulting from experimental error. When the experimental error variance is low, the response is extremely reproducible. With a large variance, the response is not very reproducible (Kutner et al. 2005).

The ICC is a variance decomposition method to evaluate the portion of general variance attributable to between-subject variability. Theoretically, the ICC formula for a population is expressed as

$$ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}, \tag{3.1}$$

where σ_B^2, is the variance between cases and σ_W^2 indicates the variance within cases. The variance between cases and the variance within cases cover the total variance of the ratings. Both components are estimated using Analysis of Variance (ANOVA) methods (von Eye et

al. 2005).

A data vector y is a random sample from an unknown population. $f(y|\theta)$ is the probability density function (PDF) that determines the probability of observing the data vector y given the parameter θ. For a set of parameter values, the PDF will have some data that has a higher probability than other data. One method of estimation is maximum likelihood (ML). Given the data and a model, among all probability densities, the goal of maximum likelihood is to find the model that is most likely to have produced the data.

In order to reach this goal, we determine the most likely PDF by setting $L(\theta|y)$=$f(y|\theta)$, where $L(\theta|y)$ is the likelihood of the parameter θ given y. When y is the data vector and θ is the vector of parameters in the distribution function of y, this function can be written as $f(y|\theta)$. Therefore, for some, the given value of θ, $f(y|\theta)$ is the density function of y. In conclusion, the maximum likelihood finds the θ that maximizes that likelihood, given the data.

A method related to ML is restricted maximum likelihood (REML). This applies ML to linear functions of y. An example of a linear function is $K^T y$ for which K^T causes $K^T y$ to contain none of the fixed effects that are part of the model for y. The variance components of REML are estimated without being affected by fixed effects.

Exact mathematical results are not feasible for this research, because closed form expressions cannot be obtained for all of the experimental designs considered. Therefore, simulations are used to compare the designs.

The following Random Effects Model will be considered

$$Y_{BLR} = \mu + b_B + l_L + \epsilon_{BLR} \tag{3.2}$$

where $B = 1, 2, ..., B_0$, B_0=number of biological samples, and b_B is the effect of biological sample B, $L = 1, 2, ..., L_0$, L_0=number of laboratories, and l_L is the effect of laboratory L,

$R = 1, 2, ..., R_0$, R_0=number of replications. In addition, μ is constant, b_B are independent $N(0, \sigma_b^2)$, l_L are independent $N(0, \sigma_l^2)$, and ϵ_{BLR} are independent $N(0, \sigma_e^2)$.

The model used has no interaction term, yet if an interaction term was used the model would have the form

$$Y_{BLR} = \mu + b_B + l_L + (bl)_{BL} + \epsilon_{BLR}. \tag{3.3}$$

There is a method by Scheffe (1959) in the book *The Analysis of Variance* that tests for an interaction in a 2 way layout with 1 replication per cell, yet our models assumes that there is no interaction term.

In order to use the method of maximum likelihood, it is assumed that Y_{BLR} are jointly normally distributed. The density function of the multivariate normal distribution is given as

$$f(Y) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp[-\frac{1}{2}(Y-\mu)'\Sigma^{-1}(Y-\mu)]. \tag{3.4}$$

The joint density function for Y_{BLR} is shown in Equation 3.5: this applies to other places below. The mean vector μ in this equation is Zu. $\sigma^2 Y$ is equivalent to the variance-covariance matrix Σ below:

$$f(Y) = \frac{1}{(2\pi)^{n_T/2}|\Sigma|^{1/2}} exp[-\frac{1}{2}(Y-X\beta)'\Sigma^{-1}(Y-X\beta)] \tag{3.5}$$

In order to find the maximum likelihood estimates of the parameters, the log of Equation 3.5 is obtained and is shown below:

$$log_e L = -\frac{n_T}{2} log_e(2\pi) - \frac{1}{2} log_e|\Sigma| - \frac{1}{2}(Y-X\beta)'\Sigma^{-1}(Y-X\beta) \tag{3.6}$$

(Kutner et al. 2005)

Maximum likelihood estimates are then found for σ_b^2, σ_l^2, and σ_e^2.

This is a specific example for three biological samples, two laboratories, and two replications. The observation matrix is a 12×1 matrix, and this is denoted as Y. Other examples can be provided as well for different numbers of biological samples, laboratories, and replications. The ICC is calculated from Equation 3.2.

12X1 Observation Matrix, Y is shown in Equation 3.7:

$$\begin{bmatrix} Y_{B1L1R1} \\ Y_{B1L1R2} \\ Y_{B1L2R1} \\ Y_{B1L2R2} \\ Y_{B2L1R1} \\ Y_{B2L1R2} \\ Y_{B2L2R1} \\ Y_{B2L2R2} \\ Y_{B3L1R1} \\ Y_{B3L1R2} \\ Y_{B3L2R1} \\ Y_{B3L2R2} \end{bmatrix} \tag{3.7}$$

The covariance matrix is defined as follows where $\otimes$ is the Kronecker Product:

$$I_3 \otimes \begin{bmatrix} \sigma_b^2 + \sigma_l^2 + \sigma_e^2 & \sigma_b^2 + \sigma_l^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 + \sigma_l^2 & \sigma_b^2 + \sigma_l^2 + \sigma_e^2 & \sigma_b^2 & \sigma_b^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_l^2 + \sigma_e^2 & \sigma_b^2 + \sigma_l^2 \\ \sigma_b^2 & \sigma_b^2 & \sigma_b^2 + \sigma_l^2 & \sigma_b^2 + \sigma_l^2 + \sigma_e^2 \end{bmatrix} + (J_{3,3} - I_3) \otimes \begin{bmatrix} \sigma_l^2 & \sigma_l^2 & 0 & 0 \\ \sigma_l^2 & \sigma_l^2 & 0 & 0 \\ 0 & 0 & \sigma_l^2 & \sigma_l^2 \\ 0 & 0 & \sigma_l^2 & \sigma_l^2 \end{bmatrix} \tag{3.8}$$

such that $J_{3,3}=$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{3.9}$$

and $I_3=$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.10}$$

For between laboratories, the estimate is

$$ICC_B = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_l^2 + \hat{\sigma}_e^2}. \tag{3.11}$$

The within laboratory estimate is

$$ICC_W = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_e^2}. \tag{3.12}$$

Within laboratory experiments are experiments that are done in the same laboratory. Between laboratory experiments are experiments from different laboratories that are compared. For example, 12 samples, three laboratories, and two replicates per laboratory equal 72 observations. σ_b^2 is driven by biological variation. In laboratory reproduction studies, the

researcher is often more interested in between laboratory agreement than within laboratory agreement. The real independent replications are the samples. Complete within laboratory replicates show how well laboratories agree with themselves. Ideally, the researcher would like to use most of the replications on between laboratory replications and not within laboratory replications, because the ICC_B is driven by biological variation. Multiple within laboratory replications give good within laboratory estimates but poor between laboratory estimates. The analysis of variance depends, in many situations, on whether every cell of data has the same number of observations. When every cell has the same number of observations, the data are balanced. When the data are unbalanced, some cells have different numbers of replicates. Unbalanced laboratory designs can give more between laboratory replications.

It is not obvious whether the same marker will give the same result on the same person. The intention is to measure whether, if tested twice in the same laboratory, the same answer will be found. Then, a measurement will be taken to determine whether two different laboratories give the same answer. It is not ethical to test this on people if it is not reproducible in the same laboratory. The logistics and reproducibility determine whether it can be performed in multiple laboratories compared to just one laboratory, as well as in a different laboratory. According to Burdick et al. (2005) (page 80) a confidence interval is constructed as:

$$\frac{\sigma_o^2}{\gamma_Y} \tag{3.13}$$

where the numerator and denominator are defined by

$$\sigma_o^2 = max[0, \frac{(B_0 - 1)MSB}{L_0 R_0 W_2} - \frac{(L_0 B_0 R_0 - L_0 - B_0 + 1)MSE}{lrW_3}] \tag{3.14}$$

and

$$\gamma_Y = \frac{(l-1)MSL}{brW_1} + \frac{(b-1)MSB}{brW_2} + \frac{(lbr - l - b + 1)MSE}{lbrW_3} \tag{3.15}$$

where W_1, W_2, and W_3 are equal to chi square random variables with degrees of freedom L_0-1, B_0-1 and L_0*B_0*R_0-L_0-B_0+1, respectively. The values found at the 2.5 and 97.5 percentiles are the left and right ends of the confidence interval, respectively. σ_o^2 and γ_Y are generalized pivotal quantities as define in Burdick et al. (2005). The mean squares of the sample are chi square distributed so the W's are chi squared. 100,000 values are distributed because it is hard to estimate the percentiles in the tails. The confidence interval should contain the true values of the parameters 95 percent of the time.

This study is interested in measuring between laboratory reproducibility as well as possible. The between laboratories are hardest to estimate because there are a limited number of laboratories. If one can estimate between laboratory reproducibility, then one can also estimate within laboratory reproducibility. The within laboratory study can be done with only one laboratory. Intuitively the 1 replication design is better than the other 2 designs because the 1 replication design uses the most biological samples. The more biological sample replicates there are the better the estimates. A large sample size gives the best estimation of biological variance to minimize the variance. Some believed an unbalanced design was the best design because the one replication design has no within laboratory replications. Also, it seemed with a 1 replication design that experimental error could not be estimated well.

Chapter 4

Results

The data used in this study was simulated in R and is not laboratory data. We generated data from different models using R and the effect of errors were generated as normal random variables. The parameters were fixed at different settings. An algorithm was developed based upon the equations described in the background section, and it was applied to samples of size 48, 72 and 96 to study the optimal experimental design. The 48 sample results are shown in this section, while the 72 and 96 samples are in the Appendix due to similarities. In general, it was found that the 1 replication design was the optimal design, because it has the smallest root MSE of any of the samples. If the root MSE is smaller, then this implies a better design.

The 2 replicate linear design has the highest root MSE; and therefore, it is the worst design. Since the 2 replicate design has more replicates, it uses fewer biological samples because the population or total number of observations is fixed; and this causes the design to have a poor root MSE. 2 biological samples in each cell means fewer biological samples. The 48 sample designs have higher root MSE values than the 72 sample designs and the 96 sample designs due to the sample size. The 96 sample designs have smaller root MSE designs than the 48 sample designs and the 72 sample designs.

48 Sample

The balanced 2 replicate design is shown in Table 4.15. Each cell has two observations indicated with "x's". The 2 replicate design has 6 biological samples and 4 laboratories. The balanced and unbalanced 1 replicate designs are shown in Table 4.16 and Table 4.17, respectively. The 1 replicate design has 12 biological samples and 4 laboratories, and the unbalanced design has 8 biological samples with 4 laboratories. Each of these Tables has biological samples and laboratories that equal 48 observations.

Tables 4.1, 4.2 and 4.3 show the simulation results from each design and different variance parameter settings. All models were fit using REML. Each row represents 1,000 simulations. Model A has the estimates of σ_e^2=1.0, σ_l^2=0.5 and σ_b^2=3.5. Model B has the estimates of σ_e^2=1.0, σ_l^2=1.0 and σ_b^2=4.67. Model C has the estimates of σ_e^2=1.0, σ_l^2=1.0 and σ_b^2=2.33. Table 4.1 shows balanced data with 2 replications for a sample of size 48. Table 4.2 has balanced data with 1 replication for a sample size of 48. Table 4.3 has unbalanced data for a sample size of 48. Model B has the smallest root MSE for the within ICC, and model A has the smallest root MSE for the between ICC for a sample of size 48. The 1 replication design has the smallest root MSE, and the 2 replication design has the largest root MSE for a sample of size 48. Since REML is used some of the variance values will result in zero estimates for variance parameters. For the balanced 2 replication data, Model B has the least number of zeros for the laboratory variance. For the balanced 1 replication data and the unbalanced data, Model C has the least number of zeros for the laboratory variance. For each model, there are no zeros for the biological variance.

Tables 4.4 and 4.5 show simulation results, but models were instead fit using maximum likelihood. Table 4.4 shows balanced Linear Model data with 2 replications for a sample of size 48. These results are different from the Table 4.1 REML 2 replicate balanced design data. Table 4.5 shows balanced data with 1 replication for a sample of size 48. The 1 replicate design results for REML are the same as the 1 replicate ML design results.

Figures 4.1, 4.2 and 4.3 show a graphical comparison of the designs for 48 total observations. Figure 4.4 and 4.5 show the bias and variance of the between-laboratory intraclass correlation coefficient as a function of the true between ICC. It is shown on the bias squared plot that the center of the graph is noisier than the tails of the graph. This is because there is less variability in the tails of the graph and more variability in the center of the graph. On the variance plot, the 1 replicate design has the smallest standard error and is therefore the best design. The biggest advantage is around a correlation of 0.5, because there is the greatest difference between the designs as seen in the right of Figure 4.4. These figures show that the root MSE is minimized for the 1 replicate design.

ANOVA tables are presented in Tables 4.6-4.11. Note that when an interaction term is added to the model (Table 4.7), there are no degrees of freedom for error. The assumption of no interaction is key to our comparisons. If there is an interaction, this design will not necessarily be optimal. For a model with an interaction term, one would have to do more work to determine the best design.

Tables 4.12-4.14 compare results of confidence intervals from the 2 replicate and 1 replicate designs for Model A, Model B and Model C. Although the 1 replicate intervals are narrower, the coverage is not maintained at 95 percent. This shows a drawback of the one replicate design.

As discussed previously, results for 72 and 96 observations were similar and are presented in the Appendix. The 96 sample size had the smallest root MSE values, and the 48 sample size had the largest root MSE values.

Table 4.1: Balanced Data with 2 Replications for Sample Size of 48 (6 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.473943	0.520506	1.007539	0.777778	0.714861	0.157000	0.169065	0.000951
B	4.635215	1.040408	1.008789	0.823633	0.765159	0.142249	0.153733	0.000865
C	2.312374	1.040416	1.008744	0.699700	0.633618	0.174608	0.186613	0.001050

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.473943	0.520506	1.007539	0.700000	0.640196	0.173142	0.183098	0.001030
B	4.635215	1.040408	1.008789	0.700150	0.646354	0.176411	0.184347	0.001037
C	2.312374	1.040416	1.008744	0.538106	0.497335	0.188491	0.192758	0.001085

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	31
B	0	9
C	0	12

Table 4.2: Balanced Data with 1 Replication for Sample Size of 48 (4 Laboratories and 12 Samples)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.397129	0.504965	1.006476	0.777778	0.742996	0.108038	0.113447	0.000638
B	4.533890	1.009069	1.007877	0.823633	0.792627	0.092482	0.097497	0.000549
C	2.259984	1.009188	1.007759	0.699700	0.660621	0.128883	0.134616	0.000757

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.397129	0.504965	1.006476	0.700000	0.667439	0.128558	0.132555	0.000746
B	4.533890	1.009069	1.007877	0.700150	0.674013	0.134852	0.137295	0.000772
C	2.259984	1.009188	1.007759	0.538106	0.519189	0.154488	0.155565	0.000875

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	30
B	0	15
C	0	12

Table 4.3: Unbalanced Data for Sample Size of 48 (8 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.541710	0.510118	1.008353	0.777778	0.731882	0.137597	0.144984	0.000816
B	4.722069	1.016031	1.009583	0.823633	0.781079	0.122363	0.129494	0.000729
C	2.359783	1.015923	1.009675	0.699700	0.651317	0.156671	0.163896	0.000922

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.541710	0.510118	1.008353	0.700000	0.658108	0.156758	0.162183	0.000912
B	4.722069	1.016031	1.009583	0.700150	0.664467	0.161063	0.164890	0.000928
C	2.359783	1.015923	1.009675	0.538106	0.514734	0.176350	0.177805	0.001000

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	27
B	0	16
C	0	11

Table 4.4: Balanced LM Data with 2 Replications for Sample Size of 48 (6 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	6.246232	0.552390	3.760658	0.777778	0.326396	0.420337	0.616645	0.003469
B	8.307065	1.402464	4.864537	0.823633	0.332660	0.422147	0.647368	0.003642
C	4.048780	1.677400	3.214922	0.699700	0.275827	0.410964	0.590247	0.003321

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	6.246232	0.552390	3.760658	0.700000	0.313359	0.409814	0.563268	0.003169
B	8.307065	1.402464	4.864537	0.700150	0.311175	0.400835	0.558399	0.003142
C	4.048780	1.677400	3.214922	0.538106	0.248576	0.373767	0.472641	0.002659

	Number of σ_b^2 less than 0	Number of σ_l^2 less than 0
A	289	523
B	288	431
C	319	369

Table 4.5: Balanced LM Data with 1 Replication for Sample Size of 48 (4 Laboratories and 12 Samples)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.396536	0.502595	1.008847	0.777778	0.742555	0.108120	0.113662	0.000639
B	4.533647	1.008100	1.008847	0.823633	0.792470	0.092535	0.097597	0.000549
C	2.259711	1.008100	1.008847	0.699700	0.660367	0.128933	0.134737	0.000758

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.396536	0.502595	1.008847	0.700000	0.667399	0.128551	0.132558	0.000746
B	4.533647	1.008100	1.008847	0.700150	0.674002	0.134847	0.137292	0.000772
C	2.259711	1.008100	1.008847	0.538106	0.519156	0.154472	0.155554	0.000875

	Number of σ_b^2 less than 0	Number of σ_l^2 less than 0
A	0	58
B	0	26
C	0	26

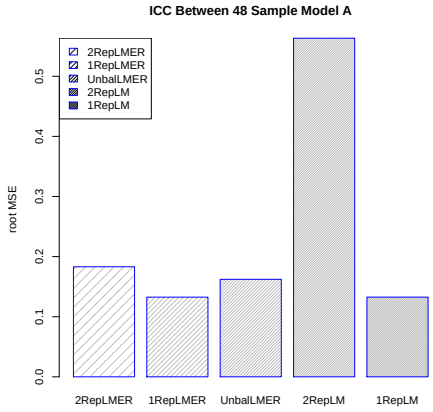

Figure 4.1: ICC between values of a 48 sample for model A.

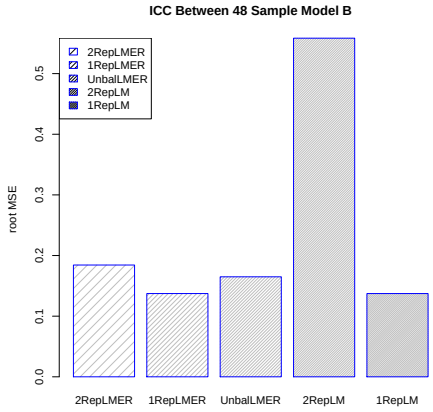

Figure 4.2: ICC between values of a 48 sample for model B.

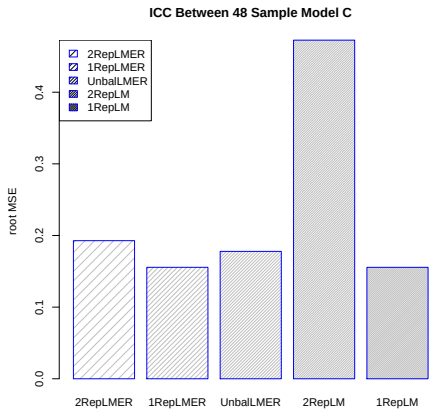

Figure 4.3: ICC between values of a 48 sample for model C.

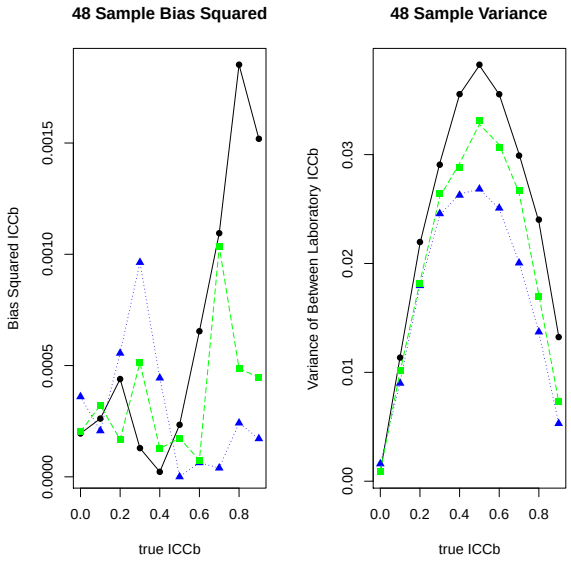

Figure 4.4: Bias squared and variance of 1 replicate, 2 replicate, and unbalanced designs. The circle symbol is the 2 replicate design, the 1 replicate design is the triangle symbol, and the unbalanced design is the square symbol.

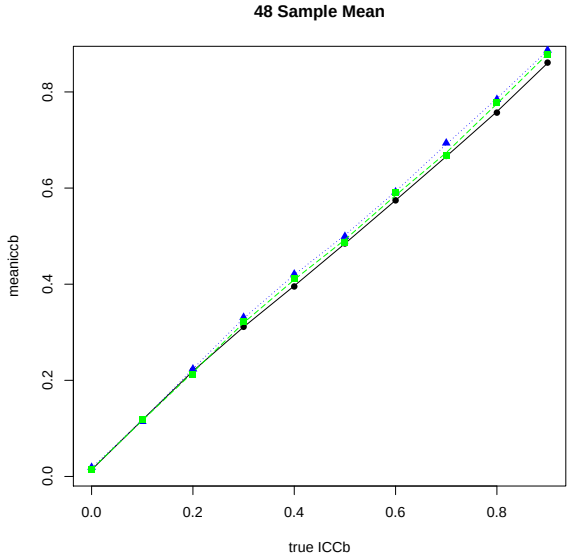

Figure 4.5: Mean of 1 replicate, 2 replicate, and unbalanced designs. The circle symbol is the 2 replicate design, the 1 replicate design is the triangle symbol, and the unbalanced design is the square symbol.

Table 4.6: LMER 2 Replication Design with Interaction

	df	df actual	E(MS)
Biological Sample	b-1	5	$\sigma^2_\epsilon + 2\text{L}\sigma^2_\alpha + 2\sigma^2_{\alpha\beta}$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + 2\text{B}\sigma^2_\beta + 2\sigma^2_{\alpha\beta}$
Interaction	(b-1)(l-1)	15	$\sigma^2_\epsilon + 2\sigma^2_{\alpha\beta}$
Error	(r-1)bl	24	σ^2_ϵ
Total	rb$l$-1	47	

Table 4.7: LMER 1 Replication Design with Interaction

	df	df actual	E(MS)
Biological Sample	b-1	11	$\sigma^2_\epsilon + \text{L}\sigma^2_\alpha + \sigma^2_{\alpha\beta}$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + \text{B}\sigma^2_\beta + \sigma^2_{\alpha\beta}$
Interaction	(b-1)(l-1)	33	$\sigma^2_\epsilon + \sigma^2_{\alpha\beta}$
Error	(r-1)bl	0	σ^2_ϵ
Total	rb$l$-1	47	

Table 4.8: LMER 2 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	5	$\sigma^2_\epsilon + 2\text{L}\sigma^2_\beta$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + 2\text{B}\sigma^2_\alpha$
Error	rb$l$-b-l+1	39	σ^2_ϵ
Total	rb$l$-1	47	

Table 4.9: LMER 1 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	11	$\sigma^2_\epsilon + \text{L}\sigma^2_\beta$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + \text{B}\sigma^2_\alpha$
Error	(b-1)(l-1)	33	σ^2_ϵ
Total	bl-1	47	

Table 4.10: Linear Model 2 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	5	$\hat{\sigma}_b^2 = \frac{MS(B)-MS(E)}{L}$
Laboratory Effect	l-1	3	$\hat{\sigma}_l^2 = \frac{MS(L)-MS(E)}{B}$
Error	rb$l$-b-l+1	39	$\hat{\sigma}_e^2 = MS(E)$
Total	rb$l$-1	47	

Table 4.11: Linear Model 1 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	11	$\hat{\sigma}_b^2 = \frac{MS(B)-MS(E)}{L}$
Laboratory Effect	l-1	3	$\hat{\sigma}_l^2 = \frac{MS(L)-MS(E)}{B}$
Error	(b-1)(l-1)	33	$\hat{\sigma}_e^2 = MS(E)$
Total	bl-1	47	

Table 4.12: Model A 48 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.606335	0.084901	0.922
2 Replicate	0.635270	0.117607	0.963

Table 4.13: Model B 48 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.6166412	0.09128614	0.944
2 Replicate	0.6606639	0.1222401	0.963

Table 4.14: Model C 48 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.6260818	0.07160277	0.929
2 Replicate	0.6916461	0.0878267	0.961

Table 4.15: 48 Sample 2 Replicate Design with 6 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	xx	xx	xx	xx
	2	xx	xx	xx	xx
	3	xx	xx	xx	xx
	4	xx	xx	xx	xx
	5	xx	xx	xx	xx
	6	xx	xx	xx	xx

Table 4.16: 48 Sample 1 Replicate Design with 12 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	x	x	x	x
	2	x	x	x	x
	3	x	x	x	x
	4	x	x	x	x
	5	x	x	x	x
	6	x	x	x	x
	7	x	x	x	x
	8	x	x	x	x
	9	x	x	x	x
	10	x	x	x	x
	11	x	x	x	x
	12	x	x	x	x

Table 4.17: 48 Sample Unbalanced Design with 8 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	xx	xx	x	x
	2	xx	xx	x	x
	3	xx	xx	x	x
	4	xx	xx	x	x
	5	x	x	xx	xx
	6	x	x	xx	xx
	7	x	x	xx	xx
	8	x	x	xx	xx

Chapter 5

Conclusion

The purpose of this thesis was to find an optimal experimental design for cancer biomarker reproducibility studies. It focused on identifying the most cost-effective design for cancer biomarker investigations assessing reproducibility. An algorithm was developed to determine optimal allocations of samples for the most effective experimental results. This algorithm was applied to sample sizes of 48, 72 and 96.

The results from this thesis demonstrate that the 1 replicate design gave the lowest root MSE for the between ICC. This is somewhat surprising, because there are no within-laboratory replicates; yet, this makes intuitive sense, because the 1 replicate designs use most of the biological samples. These results rely critically on two assumptions: that the effects are normally distributed and that there is no laboratory by sample interaction as seen in Equation 3.2.

There are two commonly used criteria for comparing designs, which are the variance and the confidence interval. Usually, in most cases, the best design has the smallest variance and most narrow confidence interval width. From the designs that were utilized, it was found that the 1 replicate design has the smallest variance and the most narrow confidence interval. In this thesis, it was found that the one replication design was the best design. In studies

such as the Dobbin et al. (2005) study, it was thought that the unbalanced design was the best design. This thesis demonstrates that the one replication design is the best design.

Chapter 6

Resources

Bland, J., Altman, D. (2010). Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement. *International Journal of Nursing Studies,47,* 931-936.

Burdick, R., Borror, C., Montgomery, D. *Design and Analysis of Gauge R and R Studies: Making Decisions with Confidence Intervals in Random and Mixed ANOVA Models*, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA, 2005.

Dobbin, K., Beer, D., Meyerson, M., Yeatman, T., Gerald, W., Jacobson, J., Conley, B., Buetow, K., Heiskanen, M., Simon, R., Minna, J., Girard, L., Misek, D., Taylor, J., Hanash, S., Naoki, K., Hayes, D., Ladd-Acosta, C., Enkemann, S., Viale, A., Giordano, T. (2005, January 15). Interlaboratory Comparability Study of Cancer Gene Expression Analysis using Oligonucleotide Microarrays. *Clinical Cancer Research, 11* 565-572.

Jessup, J., Dobbin, K., Hamilton, S., Thibodeau, S., Redston, M., Taube, S., Wang, Z., Benedetti, J. and the Program for the Assessment of Clinical Cancer Tests (PACCT) 18qLOH Team. (2009) Interlaboratory assay reproducibility study for loss of heterozygosity on chromosome 18 (18qLOH) in colon cancer. Journal of Clinical Oncology, 27: Supplement, Meet-

ing Abstract 4052.

Kutner, M., Nachtsheim, C., Neter, J., Li, W. *Applied Linear Statistical Models*, McGraw-Hill Irwin New York, NY, 2005. Lin, L. (1989, March). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics,45,*255-268.

McShane, L. (2000, May 1). Reproducibility of p53 Immunohistochemistry in Bladder Tumors. *Clinical Cancer Research, 6,* 1854-1864.

Saito, Y., Sozu, T., Hamada, C., Yoshimura, I. (2006). Effective Number of Subjects and Number of Raters for Inter-Rater Reliability Studies. *Statistics in Medicine,25,* 1547-1560.

Scheffe, H., *The Analysis of Variance*, John Wiley and Sons, Inc. Canada, 1959.

Serfontein, G., Jaroszewicz, A. (1978). Estimation of Gestational Age at Birth: Comparison of Two Methods. *Arch Dis. Child, 53,*509-511.

Versaudaraan, B. S. (2009, February 20). Lin's Concordance Correlation Coefficient (Correlation Coefficient with Small Sample). [web blog comment]. Retrieved from http://purabuana.

von Eye, A., Mun, E. (2005). *Analyzing Rater Agreement: Manifest Variable Methods* Lawrence Erlbaum Associates, Publishers. wordpress.com/2009/02/20/lins-concordance-correlation-coefficient-correlation-coefficient-with-small-sample/

Chapter 7

Appendix

72 Sample

In the 72 sample design, it is seen that the 1 replicate linear model and the 1 replicate LMER model have the smallest root MSE. As stated previously, the linear model has no interaction term, and, therefore, extra degrees of freedom cause the linear model to have a good fit. Figures 7.1 through 7.3 show the ICC between root MSE values for model A, model B and model C respectively. These results are seen in Tables 7.1 through 7.5. The squared bias and variance are shown in Figure 7.4. It is apparent in Figure 7.4 that the squared bias is small compared to the variance. The mean of each design is seen in Figure 7.5. In Table 7.15, the 2 replicate design is shown where each cell has two "x's" to indicate two observations. The balanced 1 replicate design is shown in Table 7.16. Table 7.17 shows the unbalanced design. ANOVA tables are show in Tables 7.6-7.11. When an interaction term is added to the model, there are no degrees of freedom for error. The results for confidence intervals of 1 replicate and 2 replicate designs are shown in Tables 7.12-7.14 for Model A, Model B and Model C.

Table 7.1:
Balanced LMER Data with 2 Replications for Sample Size of 72 for LMER model(9 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.512658	0.495996	1.008481	0.777778	0.736984	0.123299	0.129814	0.000730
B	4.689649	0.987329	1.008830	0.823633	0.786306	0.108900	0.115068	0.000647
C	2.335742	0.987329	1.008830	0.699700	0.656311	0.140870	0.147333	0.000829

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.512658	0.495996	1.008481	0.700000	0.664454	0.142154	0.146462	0.000824
B	4.689649	0.987329	1.008830	0.700150	0.672166	0.147789	0.150343	0.000846
C	2.335742	0.987329	1.008830	0.538106	0.520789	0.162882	0.163719	0.000921

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	19
B	0	9
C	0	7

Table 7.2: Balanced LMER Data with 1 Replication for Sample Size of 72 (4 Laboratories and 18 Samples)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.532900	0.473642	1.004642	0.777778	0.761682	0.079506	0.081080	0.000456
B	4.714364	0.954569	1.005213	0.823633	0.809220	0.067274	0.068768	0.000387
C	2.351171	0.954572	1.005210	0.699700	0.681787	0.096726	0.098323	0.000553

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.532900	0.473642	1.004642	0.700000	0.691702	0.101019	0.101309	0.000570
B	4.714364	0.954569	1.005213	0.700150	0.698693	0.110611	0.110566	0.000622
C	2.351171	0.954572	1.005210	0.538106	0.545119	0.131078	0.131200	0.000738

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	17
B	0	12
C	0	6

Table 7.3: Unbalanced LMER Data for Sample Size of 72 (12 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.455155	0.505918	1.000857	0.777778	0.747202	0.098851	0.103424	0.000582
B	4.607147	1.009222	1.001313	0.823633	0.795947	0.085459	0.089791	0.000505
C	2.303155	1.009436	1.001325	0.699700	0.666630	0.116484	0.121031	0.000681

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.455155	0.505918	1.000857	0.700000	0.672078	0.121356	0.124468	0.000700
B	4.607147	1.009222	1.001313	0.700150	0.677845	0.129541	0.131384	0.000739
C	2.303155	1.009436	1.001325	0.538106	0.524695	0.145577	0.146121	0.000822

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	18
B	0	5
C	0	6

Table 7.4: Balanced LM Data with 2 Replications for Sample Size of 72 (9 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	6.338824	0.613537	4.117168	0.777778	0.331262	0.412350	0.607650	0.003419
B	8.423143	1.436159	5.330453	0.823633	0.336714	0.414134	0.639081	0.003596
C	4.125490	1.647999	3.423890	0.699700	0.287196	0.402658	0.576308	0.003242

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	6.338824	0.613537	4.117168	0.700000	0.319897	0.397409	0.549776	0.003093
B	8.423143	1.436159	5.330453	0.700150	0.317833	0.389632	0.545736	0.003070
C	4.125490	1.647999	3.423890	0.538106	0.261549	0.362906	0.456128	0.002566

	Number of σ_b^2 less than 0	Number of σ_l^2 less than 0
A	279	459
B	277	394
C	313	312

Table 7.5: Balanced LM Data with 1 Replication for Sample Size of 72 (4 Laboratories and 18 Samples)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.532640	0.472601	1.005684	0.777778	0.761493	0.079542	0.081152	0.000457
B	4.714247	0.954099	1.005684	0.823633	0.809150	0.067300	0.068807	0.000387
C	2.351052	0.954099	1.005684	0.699700	0.681685	0.096753	0.098369	0.000553

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.532640	0.472601	1.005684	0.700000	0.691687	0.101014	0.101305	0.000570
B	4.714247	0.954099	1.005684	0.700150	0.698689	0.110609	0.110563	0.000622
C	2.351052	0.954099	1.005684	0.538106	0.545107	0.131069	0.131191	0.000738

	Number of σ_b^2 less than 0	Number of σ_l^2 less than 0
A	0	44
B	0	16
C	0	16

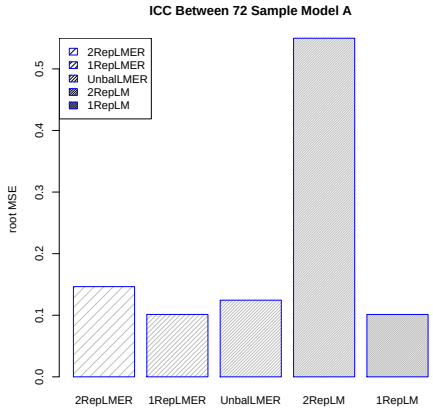

Figure 7.1: ICC between values of a 72 sample for model A.

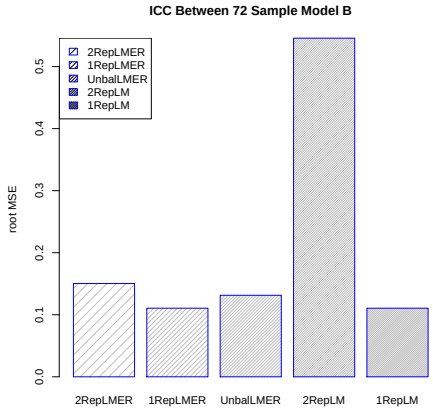

Figure 7.2: ICC between values of a 72 sample for model B.

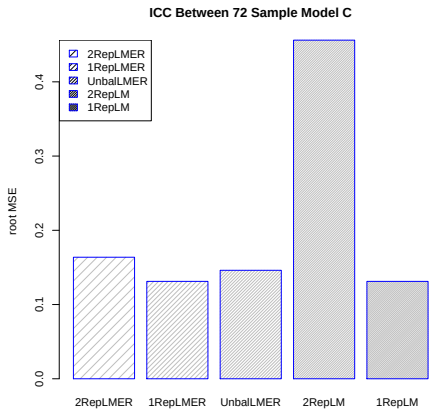

Figure 7.3: ICC between values of a 72 sample for model C.

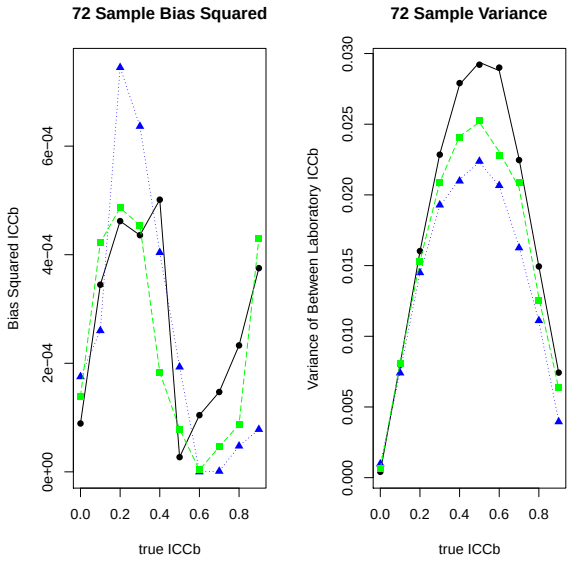

Figure 7.4: Bias squared and variance of 1 replicate, 2 replicate, and unbalanced designs. The circle symbol is the 2 replicate design, the 1 replicate design is the triangle symbol, and the unbalanced design is the square symbol.

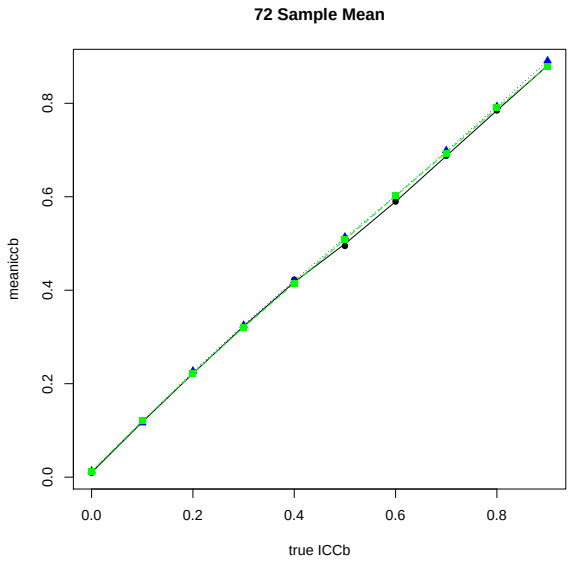

Figure 7.5: Mean of 1 replicate, 2 replicate, and unbalanced designs. The circle symbol is the 2 replicate design, the 1 replicate design is the triangle symbol, and the unbalanced design is the square symbol.

Table 7.6: LMER 2 Replication Design with Interaction

	df	df actual	E(MS)
Biological Sample	b-1	8	$\sigma^2_\epsilon + 2\text{L}\sigma^2_\alpha + 2\sigma^2_{\alpha\beta}$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + 2\text{B}\sigma^2_\beta + 2\sigma^2_{\alpha\beta}$
Interaction	(b-1)(l-1)	24	$\sigma^2_\epsilon + 2\sigma^2_{\alpha\beta}$
Error	(r-1)bl	36	σ^2_ϵ
Total	rb$l$-1	71	

Table 7.7: LMER 1 Replication Design with Interaction

	df	df actual	E(MS)
Biological Sample	b-1	17	$\sigma^2_\epsilon + \text{L}\sigma^2_\alpha + \sigma^2_{\alpha\beta}$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + \text{B}\sigma^2_\beta + \sigma^2_{\alpha\beta}$
Interaction	(b-1)(l-1)	51	$\sigma^2_\epsilon + \sigma^2_{\alpha\beta}$
Error	(r-1)bl	1	σ^2_ϵ
Total	rb$l$-1	71	

Table 7.8: LMER 2 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	8	$\sigma^2_\epsilon + 2\text{L}\sigma^2_\beta$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + 2\text{B}\sigma^2_\alpha$
Error	rb$l$-b-l+1	60	σ^2_ϵ
Total	rb$l$-1	71	

Table 7.9: LMER 1 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	17	$\sigma^2_\epsilon + \text{L}\sigma^2_\beta$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + \text{B}\sigma^2_\alpha$
Error	(b-1)(l-1)	51	σ^2_ϵ
Total	bl-1	71	

Table 7.10: Linear Model 2 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	8	$\hat{\sigma}_b^2 = \frac{MS(B)-MS(E)}{L}$
Laboratory Effect	l-1	3	$\hat{\sigma}_l^2 = \frac{MS(L)-MS(E)}{B}$
Error	rb$l$-b-l+1	60	$\hat{\sigma}_e^2 = MS(E)$
Total	rb$l$-1	71	

Table 7.11: Linear Model 1 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	17	$\hat{\sigma}_b^2 = \frac{MS(B)-MS(E)}{L}$
Laboratory Effect	l-1	3	$\hat{\sigma}_l^2 = \frac{MS(L)-MS(E)}{B}$
Error	(b-1)(l-1)	51	$\hat{\sigma}_e^2 = MS(E)$
Total	bl-1	71	

Table 7.12: Model A 72 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.536804	0.081189	0.902
2 Replicate	0.573068	0.113866	0.959

Table 7.13: Model B 72 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.5627318	0.09452231	0.926
2 Replicate	0.609554	0.1239345	0.946

Table 7.14: Model C 72 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.565991	0.05781505	0.92
2 Replicate	0.6298454	0.08614601	0.948

Table 7.15: 72 Sample 2 Replicate Design with 9 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	xx	xx	xx	xx
	2	xx	xx	xx	xx
	3	xx	xx	xx	xx
	4	xx	xx	xx	xx
	5	xx	xx	xx	xx
	6	xx	xx	xx	xx
	7	xx	xx	xx	xx
	8	xx	xx	xx	xx
	9	xx	xx	xx	xx

Table 7.16: 72 Sample 1 Replicate Design with 18 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	x	x	x	x
	2	x	x	x	x
	3	x	x	x	x
	4	x	x	x	x
	5	x	x	x	x
	6	x	x	x	x
	7	x	x	x	x
	8	x	x	x	x
	9	x	x	x	x
	10	x	x	x	x
	11	x	x	x	x
	12	x	x	x	x
	13	x	x	x	x
	14	x	x	x	x
	15	x	x	x	x
	16	x	x	x	x
	17	x	x	x	x
	18	x	x	x	x

Table 7.17: 72 Sample Unbalanced Design with 12 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	xx	xx	x	x
	2	xx	xx	x	x
	3	xx	xx	x	x
	4	xx	xx	x	x
	5	xx	xx	x	x
	6	xx	xx	x	x
	7	x	x	xx	xx
	8	x	x	xx	xx
	9	x	x	xx	xx
	10	x	x	xx	xx
	11	x	x	xx	xx
	12	x	x	xx	xx

96 Sample

The results from the 96 sample design demonstrate that the 1 replicate linear model and the 1 replicate LMER model have the smallest root MSE. Since the linear model has no interaction term, the extra degrees of freedom cause the linear model to have a good fit. Tables 7.18 through 7.22 show these results. Figures 7.6 through 7.8 have the ICC between root MSE values for model A, model B and model C, respectively. Figure 7.9 has the squared bias and the variance. It is demonstrated in Figure 7.9 that the squared bias is small compared to the variance. Figure 7.10 shows the mean of each design. The balanced 2 replicate design is shown in Table 7.32, where each cell has two "x's" to indicate two observations. The balanced 1 replicate design is shown in Table 7.33, and the unbalanced design is shown in Table 7.34. Tables 7.23-7.28 show ANOVA tables. There are no degrees of freedom for error when an interaction term is added to the model. Confidence intervals for 1 replicate and 2 replicate designs are shown in Tables 7.29-7.31 for Model A, Model B and Model C.

Table 7.18: LMER Balanced Data with 2 Replications for Sample Size of 96 (12 Samples and 4 Laboratories) A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.528249	0.518816	0.991930	0.777778	0.755995	0.088650	0.091244	0.000513
B	4.705081	1.035895	0.992085	0.823633	0.803833	0.076004	0.078504	0.000442
C	2.351100	1.035908	0.992074	0.699700	0.676521	0.105390	0.107857	0.000607

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.528249	0.518816	0.991930	0.700000	0.679625	0.115280	0.117010	0.000658
B	4.705081	1.035895	0.992085	0.700150	0.684432	0.126331	0.127242	0.000716
C	2.351100	1.035908	0.992074	0.538106	0.532067	0.143091	0.143147	0.000805

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	11
B	0	4
C	0	2

Table 7.19: LMER Balanced Data with 1 Replication for Sample Size of 96 (4 Laboratories and 24 Samples) A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.572809	0.512805	0.991626	0.777778	0.770734	0.064704	0.065054	0.000366
B	4.762290	1.021237	0.991889	0.823633	0.816993	0.054291	0.054669	0.000308
C	2.382567	1.021238	0.991889	0.699700	0.692670	0.079647	0.079917	0.000450

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.572809	0.512805	0.991626	0.700000	0.695209	0.092067	0.092145	0.000518
B	4.762290	1.021237	0.991889	0.700150	0.699641	0.105727	0.105676	0.000595
C	2.382567	1.021238	0.991889	0.538106	0.546662	0.122739	0.122976	0.000692

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	13
B	0	4
C	0	6

Table 7.20: LMER Unbalanced Data for Sample Size of 96 (16 Samples and 4 Laboratories) A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.491497	0.523546	0.996940	0.777778	0.758476	0.079487	0.081758	0.000460
B	4.658418	1.042889	0.997153	0.823633	0.806281	0.067853	0.070004	0.000394
C	2.324347	1.042824	0.997164	0.699700	0.678670	0.095108	0.097359	0.000548

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.491497	0.523546	0.996940	0.700000	0.681603	0.103970	0.105534	0.000594
B	4.658418	1.042889	0.997153	0.700150	0.686388	0.115342	0.116103	0.000653
C	2.324347	1.042824	0.997164	0.538106	0.532202	0.132729	0.132794	0.000747

	Number of σ_b^2 Equal to 0	Number of σ_l^2 Equal to 0
A	0	16
B	0	6
C	0	3

Table 7.21: Balanced LM Data with 2 Replications for Sample Size of 96 (12 Samples and 4 Laboratories)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	6.361417	0.776383	4.287492	0.777778	0.328845	0.404461	0.604124	0.003399
B	8.436554	1.718036	5.563335	0.823633	0.334482	0.405117	0.634999	0.003573
C	4.152323	1.886488	3.541916	0.699700	0.284315	0.398156	0.575252	0.003236

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	6.361417	0.776383	4.287492	0.700000	0.312150	0.386271	0.547251	0.003079
B	8.436554	1.718036	5.563335	0.700150	0.309243	0.377950	0.543609	0.003058
C	4.152323	1.886488	3.541916	0.538106	0.251311	0.355204	0.456394	0.002568

	Number of σ_b^2 less than 0	Number of σ_l^2 less than 0
A	274	451
B	271	369
C	305	300

Table 7.22: Balanced LM Data with 1 Replication for Sample Size of 96 (4 Laboratories and 24 Samples)
A : σ_e^2=1.0 σ_l^2=0.5 σ_b^2=3.5 B : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=4.67 C : σ_e^2=1.0 σ_l^2=1.0 σ_b^2=2.33

	σ_b^2	σ_l^2	σ_e^2	TrueICC_W	$\hat{ICC}_W$	EstStdDev$\hat{ICC}_W$	$rMSE_W$	Std Err $rMSE_W$
A	3.572701	0.512372	0.992059	0.777778	0.770649	0.064767	0.065126	0.000366
B	4.762248	1.021067	0.992059	0.823633	0.816963	0.054328	0.054709	0.000308
C	2.382524	1.021067	0.992059	0.699700	0.692628	0.079698	0.079971	0.000450

	σ_b^2	σ_l^2	σ_e^2	TrueICC_B	$\hat{ICC}_B$	EstStdDev$\hat{ICC}_B$	$rMSE_B$	Std Err $rMSE_B$
A	3.572701	0.512372	0.992059	0.700000	0.695202	0.092066	0.092145	0.000518
B	4.762248	1.021067	0.992059	0.700150	0.699639	0.105726	0.105675	0.000595
C	2.382524	1.021067	0.992059	0.538106	0.546656	0.122737	0.122973	0.000692

	Number of σ_b^2 less than 0	Number of σ_l^2 less than 0
A	0	28
B	0	8
C	0	8

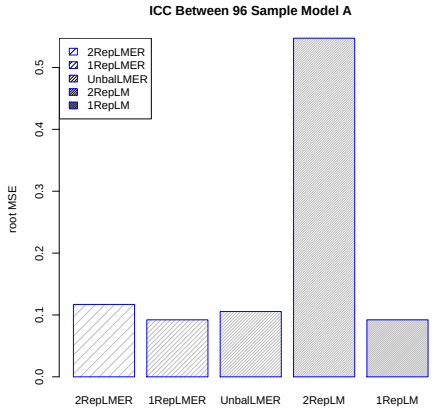

Figure 7.6: ICC between values of a 96 sample for model A.

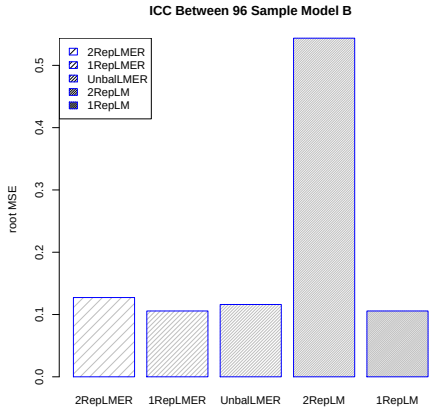

Figure 7.7: ICC between values of a 96 sample for model B.

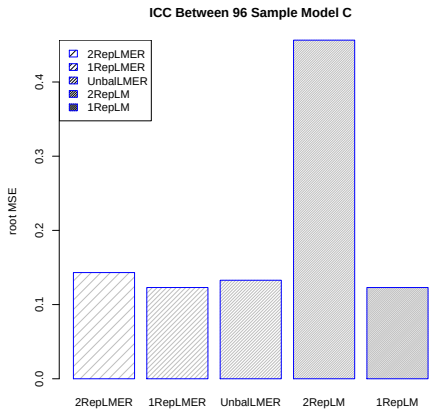

Figure 7.8: ICC between values of a 96 sample for model C.

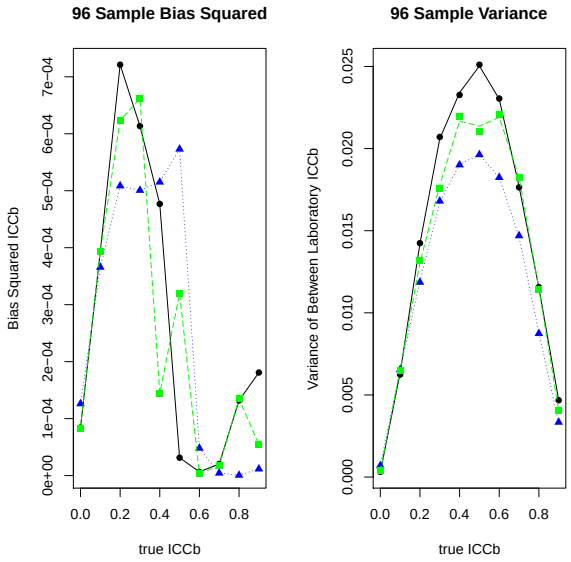

Figure 7.9: Bias squared and variance of 1 replicate, 2 replicate, and unbalanced designs. The circle symbol is the 2 replicate design, the 1 replicate design is the triangle symbol, and the unbalanced design is the square symbol.

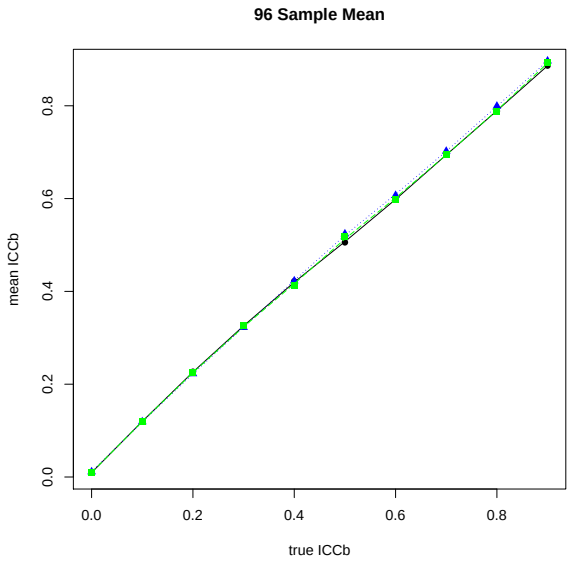

Figure 7.10: Mean of 1 replicate, 2 replicate, and unbalanced designs. The circle symbol is the 2 replicate design, the 1 replicate design is the triangle symbol, and the unbalanced design is the square symbol.

Table 7.23: LMER 2 Replication Design with Interaction

	df	df actual	E(MS)
Biological Sample	b-1	11	$\sigma^2_\epsilon + 2\text{L}\sigma^2_\alpha + 2\sigma^2_{\alpha\beta}$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + 2\text{B}\sigma^2_\beta + 2\sigma^2_{\alpha\beta}$
Interaction	(b-1)(l-1)	33	$\sigma^2_\epsilon + 2\sigma^2_{\alpha\beta}$
Error	(r-1)bl	48	σ^2_ϵ
Total	rbl-1	95	

Table 7.24: LMER 1 Replication Design with Interaction

	df	df actual	E(MS)
Biological Sample	b-1	23	$\sigma^2_\epsilon + \text{L}\sigma^2_\alpha + \sigma^2_{\alpha\beta}$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + \text{B}\sigma^2_\beta + \sigma^2_{\alpha\beta}$
Interaction	(b-1)(l-1)	68	$\sigma^2_\epsilon + \sigma^2_{\alpha\beta}$
Error	(r-1)bl	1	σ^2_ϵ
Total	rb$l$-1	95	

Table 7.25: LMER 2 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	11	$\sigma^2_\epsilon + 2\text{L}\sigma^2_\beta$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + 2\text{B}\sigma^2_\alpha$
Error	rb$l$-b-l+1	81	σ^2_ϵ
Total	rb$l$-1	95	

Table 7.26: LMER 1 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	23	$\sigma^2_\epsilon + \text{L}\sigma^2_\beta$
Laboratory Effect	l-1	3	$\sigma^2_\epsilon + \text{B}\sigma^2_\alpha$
Error	(b-1)(l-1)	69	σ^2_ϵ
Total	bl-1	95	

Table 7.27: Linear Model 2 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	11	$\hat{\sigma}_b^2 = \frac{MS(B)-MS(E)}{L}$
Laboratory Effect	l-1	3	$\hat{\sigma}_l^2 = \frac{MS(L)-MS(E)}{B}$
Error	rb$l$-b-l+1	81	$\hat{\sigma}_e^2 = MS(E)$
Total	rb$l$-1	95	

Table 7.28: Linear Model 1 Replication Design with no Interaction

	df	df actual	E(MS)
Biological Sample	b-1	23	$\hat{\sigma}_b^2 = \frac{MS(B)-MS(E)}{L}$
Laboratory Effect	l-1	3	$\hat{\sigma}_l^2 = \frac{MS(L)-MS(E)}{B}$
Error	(b-1)(l-1)	69	$\hat{\sigma}_e^2 = MS(E)$
Total	bl-1	95	

Table 7.29: Model A 96 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.505002	0.083529	0.868
2 Replicate	0.539424	0.112881	0.944

Table 7.30: Model B 96 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.5422297	0.09719413	0.907
2 Replicate	0.584088	0.1182391	0.943

Table 7.31: Model C 96 Sample Confidence Interval

	Mean Width	Standard Deviation Width	Coverage Probability
1 Replicate	0.5313649	0.0578626	0.89
2 Replicate	0.5948094	0.0802446	0.946

Table 7.32: 96 Sample 2 Replicate Design with 12 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	xx	xx	xx	xx
	2	xx	xx	xx	xx
	3	xx	xx	xx	xx
	4	xx	xx	xx	xx
	5	xx	xx	xx	xx
	6	xx	xx	xx	xx
	7	xx	xx	xx	xx
	8	xx	xx	xx	xx
	9	xx	xx	xx	xx
	10	xx	xx	xx	xx
	11	xx	xx	xx	xx
	12	xx	xx	xx	xx

Table 7.33: 96 Sample 1 Replicate Design with 24 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	x	x	x	x
	2	x	x	x	x
	3	x	x	x	x
	4	x	x	x	x
	5	x	x	x	x
	6	x	x	x	x
	7	x	x	x	x
	8	x	x	x	x
	9	x	x	x	x
	10	x	x	x	x
	11	x	x	x	x
	12	x	x	x	x
	13	x	x	x	x
	14	x	x	x	x
	15	x	x	x	x
	16	x	x	x	x
	17	x	x	x	x
	18	x	x	x	x
	19	x	x	x	x
	20	x	x	x	x
	21	x	x	x	x
	22	x	x	x	x
	23	x	x	x	x
	24	x	x	x	x

Table 7.34: 96 Sample Unbalanced Design with 16 Samples and 4 Laboratories

		Laboratory			
		1	2	3	4
Biological Sample	1	xx	xx	x	x
	2	xx	xx	x	x
	3	xx	xx	x	x
	4	xx	xx	x	x
	5	xx	xx	x	x
	6	xx	xx	x	x
	7	xx	xx	x	x
	8	xx	xx	x	x
	9	x	x	xx	xx
	10	x	x	xx	xx
	11	x	x	xx	xx
	12	x	x	xx	xx
	13	x	x	xx	xx
	14	x	x	xx	xx
	15	x	x	xx	xx
	16	x	x	xx	xx