Contents

Chapter 1

Formal tests

1.1 Formal Models and Tests

The Bland Altman plot is a simple tool for inspection of the data, but in itself it offers no formal testing procedure in this regard. To this end, the approach proposed by Altman and Bland (1983) is a formal test on the Pearson correlation coefficient of casewise differences and means (ρ_{AD}). According to the authors, this test is equivalent to a well established tests for equality of variances, known as the 'Pitman Morgan Test' (Pitman, 1939; Morgan, 1939).

1.1.1 Classical model for single measurements

The classical model is based on measurements y_{mi} by method $m = 1, 2$ on item $i = 1, 2 \ldots$

In the first instance, we require a simple model to describe a measurement by method m. We use the term *item* to denote an individual, subject or sample, to be measured, being randomly sampled from a population. Let y_{mi} be they_{mi} by method $m = 1, 2$ on item $i = 1, 2 \ldots$

$$y_{mi} = \alpha_m + \mu_i + e_{mi}$$

Here α_m is the fixed effect associated with method m, μ_i is the true value for subject i (fixed effect) and e_{mi} is a random effect term for errors with $e_{mi} \sim \mathcal{N}(0, \sigma^2_m)$.

Even though the separate variances can not be identified, their sum can be estimated by the empirical variance of the differences. Likewise the separate α can not be estimated, only their difference can be estimated as $\bar{d}$ (i.e. the inter-method bias). This model implies that the difference between the paired measurements can be expressed as

$$d_i = y_{1i} - y_{2i} \sim \mathcal{N}(\alpha_1 - \alpha_2, \sigma^2_1 - \sigma^2_2).$$

Importantly, this is independent of the item levels μ_i. As the case-wise differences are of interest, the parameters of interest are the fixed effects for methods α_m.

Using standard statistical theory, the variance of the case-wise difference, that allows for the calculation of the limits of agreement, can be calculated as

$$\mathrm{var}(d) = \omega^2_1 + \omega^2_2 - 2 \times \omega_1 2$$

1.1.2 Model Specification

The model underpinning the Bland-Altman approach can be presented as follows:

The case-wise differences $d_i = x_i - y_i$

$$\Sigma = \left(\begin{array}{cc} \sigma^2_1 + \sigma^2_2 & \frac{1}{2}(\sigma^2_1 - \sigma^2_2) \\ \frac{1}{2}(\sigma^2_1 - \sigma^2_2) & \sigma^2_b + \frac{1}{4}(\sigma^2_1 + \sigma^2_2) \end{array} \right)$$

1.1.3 Two Way ANOVA

Carstensen (2004) presents a model to describe the relationship between a value of measurement and its real value. The non-replicate case is considered first, as it is the context of the Bland Altman plots. This model assumes that inter-method bias is the only difference between the two methods.

A measurement y_{mi} by method m on individual i is formulated as follows;

$$y_{mi} = \alpha_m + \mu_i + e_{mi} \qquad e_{mi} \sim \mathcal{N}(0, \sigma_m^2) \tag{1.1}$$

The differences are expressed as $d_i = y_{1i} - y_{2i}$. For the replicate case, an interaction term c is added to the model, with an associated variance component. All the random effects are assumed independent, and that all replicate measurements are assumed to be exchangeable within each method.

Carstensen et al. (2008) develop their model from a standard two-way analysis of variance model, reformulated for the case of replicate measurements, with random effects terms specified as appropriate. Their model describing y_{mir}, again the rth replicate measurement on the ith item by the mth method ($m = 1, 2,\ i = 1, \ldots, N$, and $r = 1, \ldots, n$), can be written as

$$y_{mir} = \alpha_m + \mu_i + a_{ir} + c_{mi} + \epsilon_{mir}. \tag{1.2}$$

The fixed effects α_m and μ_i represent the intercept for method m and the 'true value' for item i respectively. The random-effect terms comprise an item-by-replicate interaction term $a_{ir} \sim \mathcal{N}(0, \varsigma^2)$, a method-by-item interaction term $c_{mi} \sim \mathcal{N}(0, \tau_m^2)$, and model error terms $\varepsilon \sim \mathcal{N}(0, \varphi_m^2)$. All random-effect terms are assumed to be independent. For the case when replicate measurements are assumed to be exchangeable for item i, a_{ir} can be removed.

The model expressed in (2) describes measurements by m methods, where $m = \{1, 2, 3 \ldots\}$. Based on the model expressed in (2), Carstensen et al. (2008) compute the limits of agreement as

$$\alpha_1 - \alpha_2 \pm 2\sqrt{\tau_1^2 + \tau_2^2 + \varphi_1^2 + \varphi_2^2}$$

Carstensen et al. (2008) notes that, for $m = 2$, separate estimates of τ_m^2 can not be obtained. To overcome this, the assumption of equality, i.e. $\tau_1^2 = \tau_2^2$ is required.

$$y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir}, \qquad e_{mi} \sim \mathcal{N}(0, \sigma_m^2), \quad c_{mi} \sim \mathcal{N}(0, \tau_m^2). \tag{1.3}$$

Of particular importance is terms of the model, a true value for item i (μ_i). The fixed effect of Roy's model comprise of an intercept term and fixed effect terms for both methods, with no reference to the true value of any individual item.

1.1.4 Formal Testing

While the Bland-Altman plot is useful for inspection of data, Kinsella (1986) notes the lack of formal testing offered by this methodology. Furthermore, Kinsella (1986) formulates a model for single measurement observations as a linear mixed effects model, i.e. a model that additively combines fixed effects and random effects:

$$Y_{ij} = \quad \mu + \beta_j + u_i + \epsilon_{ij} \qquad i = 1, \ldots, n \qquad j = 1, 2$$

The true value of the measurement is represented by μ while the fixed effect due to method j is β_j. For simplicity these terms can be combined into single terms; $\mu_1 = \mu + \beta_1$ and $\mu_2 = \mu + \beta_2$. The inter-method bias is the difference of the two fixed effect terms, $\beta_1 - \beta_2$. Each individual is assumed to give rise to a random error, represented by u_i. This random effects term is assumed to have mean zero and be normally distributed with variance σ^2. There is assumed to be an attendant error for each measurement on each individual, denoted ϵ_{ij}. This is also assumed to have mean zero. The variance of measurement error for both methods are not assumed to be identical for both methods variance, hence it is denoted σ^2_j. The set of observations (x_i, y_i) by methods X and Y are assumed to follow a bivariate normal distribution with expected values $E(x_i) = \mu_i$ and $E(y_i) = \tau_i$ respectively. The variance covariance of the observations $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 + \sigma^2_1 & \sigma^2 \\ \sigma^2 & \sigma^2 + \sigma^2_2 \end{bmatrix}$$

Kinsella (1986) demonstrates the estimation of the variance terms and relative precisions relevant to a method comparison study, with attendant confidence intervals for both. The measurement model introduced by Grubbs (1948, 1973) provides a formal procedure for estimating the variances σ^2, σ_1^2 and σ_2^2. Grubbs (1948) offers estimates, commonly known as Grubbs estimators, for the various variance components. These estimates are maximum likelihood estimates, which shall be revisited in due course.

$$\hat{\sigma^2} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} = Sxy$$
$$\hat{\sigma_1^2} = \sum \frac{(x_i - \bar{x})^2}{n-1} = S^2x - Sxy$$
$$\hat{\sigma_2^2} = \sum \frac{(y_i - \bar{y})^2}{n-1} = S^2y - Sxy$$

Thompson (1963) defines $\Delta_j = \sigma^2/\sigma_j^2, j = 1, 2$, to be a measure of the relative precision of the measurement methods, and demonstrates how to make statistical inferences about Δ_j. Based on the following identities,

$$\begin{aligned} C_x &= (n-1)S_x^2, \\ C_{xy} &= (n-1)S_{xy}, \\ C_y &= (n-1)S_y^2, \\ |A| &= C_x \times C_y - (C_{xy})^2, \end{aligned}$$

the confidence interval limits of Δ_1 are

$$\frac{C_{xy} - t(\frac{|A|}{n-2})^{\frac{1}{2}}}{C_x - C_{xy} + t(\frac{|A|}{n-2})^{\frac{1}{2}}} < \Delta_1 < \frac{C_{xy} + t(\frac{|A|}{n-2})^{\frac{1}{2}}}{C_x - C_{xy} - t(\frac{|A|}{n-1})^{\frac{1}{2}}}$$

The value t is the $100(1 - \alpha/2)\%$ upper quantile of Student's t distribution with $n - 2$ degrees of freedom (Kinsella, 1986). The confidence limits for Δ_2 are found by substituting C_y for C_x in (1.2). Negative lower limits are replaced by the value 0.

The case-wise differences and means are calculated as $d_i = x_i - y_i$ and $a_i = (x_i + y_i)/2$ respectively. Both d_i and a_i are assumed to follow a bivariate normal distribution with $E(d_i) = \mu_d = \mu_1 - \mu_2$ and $E(a_i) = \mu_a = (\mu_1 + \mu_2)/2$, and the variance matrix $\Sigma_{(a,d)}$ is

$$\Sigma_{(a,d)} = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2) \\ \frac{1}{2}(\sigma_1^2 - \sigma_2^2) & \sigma^2 + \frac{1}{4}(\sigma_1^2 + \sigma_2^2) \end{bmatrix}. \qquad (1.4)$$

1.1.5 Morgan Pitman Testing

An early contribution to formal testing in method comparison was made by both **?** and **?**, in separate contributions. The basis of this approach is that the distribution of the original measurements is bivariate normal. Morgan and Pitman noted that the correlation coefficient depends upon the difference $\sigma_1^2 - \sigma_2^2$, being zero if and only if $\sigma_1^2 = \sigma_2^2$.

The classical Pitman-Morgan test is a hypothesis test for equality of the variance of two data sets; $\sigma_1^2 = \sigma_2^2$, based on the correlation value $\rho_{a,d}$,and is evaluated as follows;

$$\rho(a,d) = \frac{\sigma_1^2 - \sigma_2^2}{\sqrt{(\sigma_1^2 + \sigma_2^2)(4\sigma_S^2 + \sigma_1^2 + \sigma_2^2)}} \tag{1.5}$$

The correlation constant takes the value zero if, and only if, the two variances are equal. Therefore a test of the hypothesis $H : \sigma_1^2 = \sigma_2^2$ is equivalent to a test of the hypothesis $H : \rho(D, A) = 0$. This corresponds to the well-known t test for a correlation coefficient with $n - 2$ degrees of freedom. Bartko (1994) describes the Morgan-Pitman test as identical to the test of the slope equal to zero in the regression of Y_{i1} on Y_{12}, a result that can be derived using straightforward algebra.

1.1.6 Pitman & Morgan Test -addins

This test assess the equality of population variances of Y_1 and Y_2 where there are bivariate normally distributed. $E(Y_1) = \mu_1$ and $E(Y_2) = \mu_2$, $\text{var}(Y_1) = \sigma_1^2$ and $\text{var}(Y_1)$ $= \sigma_2^2$. ($-1 \leq \rho \leq 1$).

$$\sigma_1^2 = \sigma_2^2$$

Pitman's test tests for zero correlation between the sums and products.

Correlation between differences and means is a test statistics for the null hypothesis of equal variances given bivariate normality.

1.1.7 The Pitman-Morgan Test-addins

The test of the hypothesis that the variances σ_1^2 and σ_2^2 are equal, which was devised concurrently by *Pitman* and *Morgan*, is based on the correlation of D with S, the coefficient being $\rho_{DS} = (\sigma_1^2 - \sigma_2^2)/(\sigma_D \sigma_S)$, which is zero if, and only if, $\sigma_1^2 = \sigma_2^2$. Consequently a test of H$''$: $\sigma_1^2 = \sigma_2^2$ is equivalent to a test of H$''$: $\rho_{DS} = 0$ and the test statistic is the familiar t-test for a correlation coefficient with $(n-2)$ degrees of freedom.

1.1.8 Paired sample *t* test

Bartko (1994) discusses the use of the well known paired sample t test to test for inter-method bias; $H : \mu_d = 0$. The test statistic is distributed a t random variable with $n-1$ degrees of freedom and is calculated as follows,

$$t^* = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}} \tag{1.6}$$

where $\bar{d}$ and s_d is the average of the differences of the n observations. Only if the two methods show comparable precision then the paired sample student t-test is appropriate for assessing the magnitude of the bias.

1.1.9 Bland-Altman correlation test

The approach proposed by Altman and Bland (1983) is a formal test on the Pearson correlation coefficient of case-wise differences and means (ρ_{AD}). According to the authors, this test is equivalent to the 'Pitman Morgan Test'. For the Grubbs data, the correlation coefficient estimate (r_{AD}) is 0.2625, with a 95% confidence interval of (-0.366, 0.726) estimated by Fishers 'r to z' transformation (Cohen, Cohen, West, and Aiken, Cohen et al.). The null hypothesis (ρ_{AD} =0) fail to be rejected. Consequently the null hypothesis of equal variances of each method would also fail to be rejected. There has no been no further mention of this particular test in Bland and Altman

(1986), although Bland and Altman (1999) refers to Spearman's rank correlation coefficient. Bland and Altman (1999) comments 'we do not see a place for methods of analysis based on hypothesis testing'. Bland and Altman (1999) also states that consider structural equation models to be inappropriate.

1.1.10 Regression Methods

Conventional regression models are estimated using the ordinary least squares (OLS) technique, and are referred to as 'Model I regression' (Cornbleet and Cochrane, 1979; Ludbrook, 1997). A key feature of Model I models is that the independent variable is assumed to be measured without error. As often pointed out in several papers (Altman and Bland, 1983; Ludbrook, 1997), this assumption invalidates simple linear regression for use in method comparison studies, as both methods must be assumed to be measured with error.

The use of regression models that assumes the presence of error in both variables X and Y have been proposed for use instead (Cornbleet and Cochrane, 1979; Ludbrook, 1997). These methodologies are collectively known as 'Model II regression'. They differ in the method used to estimate the parameters of the regression.

Regression estimates depend on formulation of the model. A formulation with one method considered as the X variable will yield different estimates for a formulation where it is the Y variable. With Model I regression, the models fitted in both cases will entirely different and inconsistent. However with Model II regression, they will be consistent and complementary.

Regression approaches are useful for a making a detailed examination of the biases across the range of measurements, allowing bias to be decomposed into fixed bias and proportional bias. Fixed bias describes the case where one method gives values that are consistently different to the other across the whole range. Proportional bias describes the difference in measurements getting progressively greater, or smaller, across the range of measurements. A measurement method may have either an attendant fixed bias or proportional bias, or both. (?). Determination of these biases shall be discussed in due course.

1.1.11 Bradley-Blackwood Method

Bradley and Blackwood (1989) offers a formal simultaneous hypothesis test for the mean and variance of two paired data sets. Using simple linear regression of the differences of each pair against the sums, a line is fitted to the model, with estimates for intercept and slope ($\hat{\beta}_0$ and $\hat{\beta}_1$). The null hypothesis of this test is that the mean (μ) and variance (σ^2) of both data sets are equal if the slope and intercept estimates are equal to zero(i.e $\sigma_1^2 = \sigma_2^2$ and $\mu_1 = \mu_2$ if and only if $\beta_0 = \beta_1 = 0$)

A test statistic is then calculated from the regression analysis of variance values (Bradley and Blackwood, 1989) and is distributed as 'F' random variable. The degrees of freedom are $\nu_1 = 2$ and $\nu_1 = n - 2$ (where n is the number of pairs). The critical value is chosen for α% significance with those same degrees of freedom. Bartko (1994) amends this methodology for use in method comparison studies, using the averages of the pairs, as opposed to the sums, and their differences. This approach can facilitate simultaneous usage of test with the Bland-Altman methodology. Bartko's test statistic take the form:

$$F.test = \frac{(\Sigma d^2) - SSReg}{2MSReg}$$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Averages	1	0.04	0.04	0.74	0.4097
Residuals	10	0.60	0.06		

Table 1.1.1: Regression ANOVA of case-wise differences and averages for Grubbs Data

For the Grubbs data, $\Sigma d^2 = 5.09$, $SSReg = 0.60$ and $MSreg = 0.06$ Therefore the test statistic is 37.42, with a critical value of 4.10. Hence the means and variance of the Fotobalk and Counter chronometers are assumed to be simultaneously equal.

Importantly, this methodology determines whether there is both inter-method bias and precision present, or alternatively if there is neither present. It has previously been demonstrated that there is a inter-method bias present, but as this procedure does not allow for separate testing, no conclusion can be drawn on the comparative precision of both methods.

1.1.12 The Problem of Identifiability

Dunn (2002) highlights an important issue regarding using models such as these, the identifiability problem. This comes as a result of there being too many parameters to be estimated. Therefore assumptions about some parameters, or estimators used, must be made so that others can be estimated.

For example α may take the value of the inter-method bias estimate from Bland - Altman methodology.

For example in literature the variance ratio $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$ must often be assumed to be equal to 1 (Linnet, 1998).Dunn (2002) considers methodologies based on two methods with single measurements on each subject as inadequate for a serious study on the measurement characteristics of the methods. This is because there would not be enough data to allow for a meaningful analysis. There is, however, a contrary argument that in many practical settings it is very difficult to get replicate observations when the measurement method requires invasive medical procedure.

1.1.13 Identifiability

In many models, nave assumptions are required to overcome issues of identifiabilty. Precision is defined by the reciprocal of the variance of the random errors. Also it is assumed that the error variance is independent of the amount of material being measured. However, in practice, this is often not the case. Variability increases over the scale of measurements over many cases. Estimators of scale parameters are estimable only if the analyst is prepared to make nave, if not unacceptable, assumptions.

Equation 4 ψ and ε are statistically independent of each other. Contamination effect that arises from non-specificity / specimen specific bias. Random error is measured by . Homogenity of variances is assumed. If there are no replicate measures, both variances are completely confounded, and there is no way of telling them apart. Scaling of new measurements is measured by .

1.2 Measurement Error Models

Dunn (2002) proposes a measurement error model for use in method comparison studies. Consider n pairs of measurements X_i and Y_i for $i = 1, 2, ...n$.

$$X_i = \tau_i + \delta_i \tag{1.7}$$

$$Y_i = \alpha + \beta\tau_i + \epsilon_i$$

In the above formulation is in the form of a linear structural relationship, with τ_i and $\beta\tau_i$ as the true values , and δ_i and ϵ_i as the corresponding measurement errors. In the case where the units of measurement are the same, then $\beta = 1$.

$$E(X_i) = \tau_i \tag{1.8}$$

$$E(Y_i) = \alpha + \beta\tau_i$$

$$E(\delta_i) = E(\epsilon_i) = 0$$

The value α is the inter-method bias between the two methods.

$$z_0 = d = 0 \tag{1.9}$$

$$z_{n+1} = z_n^2 + c \tag{1.10}$$

Chapter 2

SEMS

2.1 Structural Equation modelling

Structural Equation modelling is a statistical technique used for testing and estimating causal relationships using a combination of statistical data and qualitative causal assumptions. This technique was proposed by Lewis et al. (1991) and G. Kelly as a method of assessing the reliability of a new measurement technique. It can indicate the presence of bias. However Bland and Altman (1987) have criticized it on the basis that it offers no insights into the variability about the line of equality.

In this paper, the SEM method is used to assess the linear relationship between the new method and the standard method.

Structural analysis is a generalization of regression analysis.

In Hopkins' papers, a critique of the Bland-Altman plot he makes the following remark:

What's needed for a comparison of two or more measures is a generic approach more powerful even than regression to model the relationship and error structure of each measure with a latent variable representing the true value.

Hopkins also adds that he himself is collaborating in research utilising SEM and Mixed Effects modelling. This is a methodology proposed by Kelly (1985).

2.2 Structural Equations

The structural equation model is a regression approach that allows to estimate a linear regression when independent variables are measured with error (Carrasco, 2004).

Kelly (1985) proposed the Structural Equation method

The Structural equations model is used to estimate the linear relationship between new and standards method. The Delta method is used to find the variance of the estimated parameters.

Altman and Bland [1987]criticize it for a reason that should come as no surprise: Knowing the data are consistent with a structural equation with a slope of 1 says something about the absence of bias but nothing about the variability about Y = X (the difference between the measurements), which, as has already been stated, is all that really matters.

2.2.1 Bland and Altmans Critique

Does not assess alternative statistical approaches It is unnecessary to perform elaborate statistical analysis There are two aspects of agreement that must be considered

Bias describing the average agreement Individual variability

Dr Kelly has considered only the first of these. It is , however, of no clinical value to know that two methods agree on average if one has no idea of the between subject variability.

The Structural equations approach has merit in that it avoids the biased estimation of the slope and intercept that occurs in ordinary least square regression.

however, it is quite wrong to argue solely from a lack of bias that two methods can be regarded as comparable

It is, however, the variability around Y=X that is of major interest

Chapter 3

Error In Variable Models

3.1 Conclusions about Existing Methodologies

Scatterplots are recommended by Altman and Bland (1983) for an initial examination of the data, facilitating an initial judgement and helping to identify potential outliers. They are not useful for a thorough examination of the data. O'Brien et al. (1990) notes that data points will tend to cluster around the line of equality, obscuring interpretation.

The Bland Altman methodology is well noted for its ease of use, and can be easily implemented with most software packages. Also it doesn't require the practitioner to have more than basic statistical training. The plot is quite informative about the variability of the differences over the range of measurements. For example, an inspection of the plot will indicate the 'fan effect'. They also can be used to detect the presence of an outlier.

Ludbrook (1997, 2002)criticizes these plots on the basis that they presents no information on effect of constant bias or proportional bias. These plots are only practicable when both methods measure in the same units. Hence they are totally unsuitable for conversion problems. The limits of agreement are somewhat arbitrarily constructed. They may or may not be suitable for the data in question. It has been found that the limits given are too wide to be acceptable. There is no guidance on how to deal with outliers. Bland and Altman recognize effect they would have on the limits of agreeement, but offer no guidance on how to correct for those effects.

There is no formal testing procedure provided. Rather, it is upon the practitioner opinion to judge the outcome of the methodology.

3.2 Background

In method comparison studies, it is of importance to assure that the presence of a difference of medical importance is detected. For a given difference, the necessary number of samples depends on the range of values and the analytical standard deviations of the methods involved. For typical examples, the present study evaluates the statistical power of least-squares and Deming regression analyses applied to the method comparison data.

3.3 Constant and Proportional Bias

Linear Regression is a commonly used technique for comparing paired assays. The Intercept and Slope can provide estimates for the constant bias and proportional bias occurring between both methods. If the basic assumptions underlying linear regression are not met, the regression equation, and consequently the estimations of bias are undermined. Outliers are a source of error in regression estimates.

Constant or proportional bias in method comparison studies using linear regression can be detected by an individual test on the intercept or the slope of the line regressed from the results of the two methods to be compared.

- Model I regression
- Model II regression

Regression approaches are useful for a making a detailed examination of the biases across the range of measurements, allowing bias to be decomposed into fixed bias and proportional bias. Fixed bias describes the case where one method gives values that are consistently different to the other across the whole range. Proportional bias describes the difference in measurements getting progressively greater, or smaller, across the range of measurements. A measurement method may have either an attendant fixed bias or proportional bias, or both. (**?**). Determination of these biases shall be discussed in due course.

3.4 Model I and II Regression

On account of the fact that one set of measurements are linearly related to another, one could surmise that Linear Regression is the most suitable approach to analyzing comparisons. This approach is unsuitable on two counts. Firstly one of the assumptions of Regression analysis is that the independent variable values are without error.

In method comparison studies one must assume the opposite; that there is error present in the measurements. Secondly a regression of X on Y would yield and entirely different result from Y on X.

Model I regression is unsuitable for method comparison studies. Even in the case where one method is a gold standard , it is disputed as to whether it is a valid approach. Model II regression is suitable for method comparison studies, but it is more difficult to execute. Both Model I and II regression models are unduly influenced by outliers. Regression Models can not easily be used to analyze repeated measurements

Cornbleet and Cochrane (1979) argue for the use of methods that based on the assumption that both methods are imprecisely measured ,and that yield a fitting that is consistent with both 'X on Y' and 'Y on X' formulations. These methods uses alternatives to the OLS approach to determine the slope and intercept.

They describe three such alternative methods of regression; Deming, Mandel, and Bartlett regression. Collectively the authors refer to these approaches as Model II regression techniques.

- The authors make the distinction between model I and model II regression types.
- Model II regression is the appropriate type when the predictor variable x is measured with imprecision. Cornbleet and Cochrane remark that clinical laboratory measurements usually increase in absolute imprecision when larger values are measured.
- Model II regression is the appropriate type when the predictor variable x is measured with imprecision. Cornbleet and Cochrane remark that clinical laboratory measurements usually increase in absolute imprecision when larger values are measured.

3.4.1 Model I regression [Criterion v Test]

Cornbleet and Cochrane (1979) define this analysis as the case in which the independent variable, X, is measured without error, with y as the dependent variable.

Simple Linear Regression is well known statistical technique, wherein estimates for slope and intercept of the line of best fit are derived according to the Ordinary Least Square (OLS) principle. This method is known to Cornbleet and Cochrane (1979) as Model I regression.

Simple linear regression is defined as such with the name 'Model I regression' by Cornbleet and Cochrane (1979), in contrast to 'Model II regression'.

3.4.2 Model II regression [Test V Test]

In this type of analysis, both of the measurement methods are test methods, with both expected to be subject to error. Deming regression is an approach to model II regression.

In Model I regression, the independent variable is assumed to be measured without error. For method comparison studies, both sets of measurement must be assumed to be measured with imprecision and neither case can be taken to be a reference method. Arbitrarily selecting either method as the reference will yield two conflicting outcomes. A fitting based on 'X on Y' will give inconsistent results with a fitting based on 'Y on X'. Consequently model I regression is inappropriate for such cases.

Simple linear regression calculates a line of best fit for two sets of data, n which the independent variable, X, is measured without error, with y as the dependent variable.

SLR (Model I) regression is considered by many Altman and Bland (1983); Cornbleet and Cochrane (1979); Ludbrook (1997) to be wholly unsuitable for method comparison studies, although recommended for use in calibration studies [Corncoch]. Even in the case where one method is a gold standard , it is disputed as to whether it is a valid approach. Model II regression is more suitable for method comparison studies, but it is more difficult to execute. Both Model I and II regression models are unduly influenced by outliers. Regression Models can not be used to analyze repeated measurements

Conversely, Cornbleet and Cochrane (1979) state that when the independent variable X is a precisely measured reference method, Model I regression may be considered suitable. They qualify this statement by referring the X as *the correct value*, tacitly implying that there must still be some measurement error present. The validity of this approach has been disputed elsewhere.

This regression method also calculates a line of best fit for two sets of data. It differs from Model I regression in that it is derived in a way that factors in for error in the x-axis, as well as the y-axis. Cornbleet and Cochrane (1979) refer to it as 'Model II regression'.

Model II regression method also calculates a line of best fit for two sets of data. It differs from Model I regression in that it is derived in a way that factors in for error in the x-axis, as well as the y-axis. Cornbleet and Cochrane (1979) refer to it as 'Model II regression'.

- Model II regression [Test V Test] In this type of analysis, both of the measurement methods are test methods, with both expected to be subject to error. Deming regression is an approach to model II regression.

- Model I regression [Criterion v Test] Cornbleet and Cochrane (1979) define this analysis as the case in which the independent variable, X, is measured without error, with Y as the dependent variable.

- In method comparison studies, the X variable is a precisely measured reference method. In the Cornbleet and Cochrane (1979) paper It is argued that criterion may be regarded as the correct value. Other papers dispute this.

3.4.3 Contention

Several papers have commented that this approach is undermined when the basic assumptions underlying linear regression are not met, the regression equation, and consequently the estimations of bias are undermined. Outliers are a source of error in regression estimates.

In method comparison studies, the X variable is a precisely measured reference method. In the Cornbleet and Cochrane (1979) paper. It is argued that criterion may be regarded as the correct value. Other papers dispute this.

3.4.4 Comparison of Various Model II regressions

Cornbleet and Cochrane (1979) comparing the three methods, citing studies by other authors, concluding that Deming regression is the most useful of these methods. They found the Bartlett method to be flawed in determining slopes.

However the author point out that *clinical laboratory measurements usually increase in absolute imprecision when larger values are measured.*However one of the assumptions that underline Deming and Mandel regression is constancy of the measurement errors throughout the range of values.

3.5 Deming Regression

The appropriate estimates were derived by Kummel (1879), but were popularized in the context of medical statistics and clinical chemistry by Deming (1943).

3.6 Error-In-Variable Models for Regression

Errors-in-variables models or measurement errors models are regression models that account for measurement errors in the independent variables, as well as the dependent variable.

The Bland Altman Plot is uninformative about the comparative influence of proportional bias and fixed bias. Model II approaches, such as Deming regression, can provide independent tests for both types of bias.

As with conventional regression methodologies, Deming regression calculates an estimate for both the slope and intercept for the fitted line, and standard errors thereof. Therefore there is sufficient information to carry out hypothesis tests on both estimates, that are informative about presence of fixed and proportional bias.

The most commonly known Model II methodology is known as Deming's Regression, (also known an Ordinary Least Product regression). Deming regression is recommended by Cornbleet and Cochrane (1979) as the preferred Model II regression for use in method comparison studies.

Informative analysis for the purposes of method comparison, Deming Regression is a regression technique taking into account uncertainty in both the independent and dependent variables.

Demings method always results in one regression fit, regardless of which variable takes the place of the predictor variables.

To compute the slope by Demings formula, normally distributed error of both variables is assumed, as well as a constant level of imprecision throughout the range of measurements.

3.6.1 Compuational Aspects of Deming Regression

Deming approaches the matter by simultaneously minimizing the sum of the square of the residuals of both variables. This derivation results in the best fit to minimize the sum of the squares of the perpendicular distances from the data points.

Deming regression method also calculates a line of best fit for two sets of data. It differs from simple linear regression in that it is derived in a way that factors in for error in the x-axis, as well as the y-axis.

Deming regression is a regression fitting approach that assumes error in both variables. Deming regression is recommended by Cornbleet and Cochrane (1979) as the preferred Model II regression for use in method comparison studies.

Deming regression is a type of error-in-variable regression approach that assumes that the ratio $\lambda = \sigma^2_\epsilon/\sigma^2_\eta$ is known.

This approach would be appropriate when errors in y and x are both caused by measurements, and the accuracy of measuring devices or procedures are known. The case when $\lambda = 1$ is also known as the ***orthogonal regression***.

As stated previously, the fundamental flaw of simple linear regression is that it allows for measurement error in one variable only. This causes a downward biased slope estimate.

3.6.2 Kummel's Estimates

For a given λ, Kummel (1879) derived the following estimate that would later be used for the Deming regression slope parameter. The intercept estimate α is simply estimated in the same way as in conventional linear regression, by using the identity $\bar{Y} - \hat{\beta}\bar{X}$;

$$\hat{\beta} = \frac{S_{yy} - \lambda S_{xx} + [(S_{yy} - \lambda S_{xx})^2 + 4\lambda S^2_{xy}]^{1/2}}{2S_{xy}} \tag{3.1}$$

, with λ as the variance ratio. As stated previously λ is often unknown, and therefore must be assumed to equal one.

Carroll and Ruppert (1996) states that Deming regression is acceptable only when the precision ratio (λ,in their paper as η) is correctly specified, but in practice this is often not the case, with the λ being underestimated. Several candidate models, with varying variance ratios may be fitted, and estimates of the slope and intercept are produced. However no model selection information is available to determine the best fitting model.

As with conventional regression methodologies, Deming regression calculates an estimate for both the slope and intercept for the fitted line, and standard errors thereof. Therefore there is sufficient information to carry out hypothesis tests on both estimates, that are informative about presence of fixed and proportional bias.

```
Use of deming regression in method comparison studies.
Henk Konings
```

Accuracy is closeness to the true value, or alternatively, having a low measurement error.

The determination of a true value for a biological specimen is difficult and sometimes impossible.

Precision is expressed in terms of standard deviation, coefficient of variance or variance.

In Deming regression, the errors between methods are assigned to both methods in proportion to the variances of the methods.

3.6.3 Variance Ratio

The measurement error (lambda or λ) is specified with measurement error variance related as

$$\lambda = \sigma^2_y/\sigma^2_x$$

(where σ^2_x and σ^2_y is the measurement error variance of the x and y variables, respectively).

In the case where λ is equal to one, (i.e. equal error variances), the methodology is equivalent to ***orthogonal regression***.

The sum of squared distances from measured sets of values to the regression line is minimized at an angles specified by the ratio λ of the residual variance of both variables.

The variance of the ratio,λ, specifies the angle. When λ is one, the angle is 45 degrees. In ordinary linear regression, the distances are minimized in the vertical directions (Linnet, 1999).

The Deming regression line is estimated by minimizing the sums of squared deviations in both the x and y directions at an angle determined by the ratio of the analytical standard deviations for the two methods.

In cases involving only single measurements by each method, λ may be unknown and is therefore assumes a value of one. While this will bias the estimates, it is less biased than ordinary linear regression.

This ratio can be estimated if multiple measurements were taken with each method, but if only one measurement was taken with each method, it can be assumed to be equal to one.

Orthogonal regression describes the special case where the variance ratio is equal to one.

3.6.4 Estimating the Variance ratio

In cases involving only single measurements by each method, λ may be unknown and is therefore assumes a value of one. While this will bias the estimates, it is less biased than ordinary linear regression.

$$x_i = \mu + \beta_0 + \epsilon_{xi}$$
$$y_i = \mu + \beta_1 + \epsilon_{yi}$$

The inter-method bias is the difference of these biases. In order to determine an estimate for the residual variances, one of the method biases must be assumed to be zero, i.e. $\beta_0 = 0$. The inter-method bias is now represented by β_1.

$$\begin{aligned} x_i &= \mu + \epsilon_{xi} \\ y_i &= \mu + \beta_1 + \epsilon_{yi} \end{aligned}$$

The residuals can be expressed as

$$\begin{aligned} \epsilon_{xi} &= x_i - \mu \\ \epsilon_{yi} &= y_i - (\mu + \beta_1) \end{aligned}$$

The variance of the residuals are equivalent to the variance of the corresponding observations, $\sigma^2_{\epsilon x} = \sigma^2_x$ and $\sigma^2_{\epsilon y} = \sigma^2_y$.

$$\lambda = \frac{\sigma^2_{yx}}{\sigma^2_y}. \tag{3.2}$$

Assuming constant standard deviations, and given duplicate measurements, the analytical standard deviations are given by

$$\begin{aligned} SD^2_{ax} &= \frac{1}{2n} \sum (x_{2i} - x_{1i})^2 \\ SD^2_{ay} &= \frac{1}{2n} \sum (y_{2i} - y_{1i})^2 \end{aligned}$$

Using duplicate measurements, one can estimate the analytical standard deviations and compute their ratio. This ratio is then used for computing the slope by the Deming method.[Linnet]

3.6.5 Inferences for Deming Regression

The Intercept and Slope are calculated according to Combleet & Gochman, 1979. The standard errors and confidence intervals are estimated using the jackknife method (Armitage et al., 2002).

The 95% confidence interval for the Intercept can be used to test the hypothesis that A=0. This hypothesis is accepted if the confidence interval for A contains the value 0. If the hypothesis is rejected, then it is concluded that A is significantly different from 0 and both methods differ at least by a constant amount.

The 95% confidence interval for the Slope can be used to test the hypothesis that B=1. This hypothesis is accepted if the confidence interval for B contains the value 1. If the hypothesis is rejected, then it is concluded that B is significantly different from 1 and there is at least a proportional difference between the two methods.

3.6.6 Expaning the use of Deming Regression for MCS

As noted before, Deming regression is an important and informative methodology in method comparison studies. For single measurement method comparisons, Deming regression offers a useful complement to LME models.

3.7 Drawbacks of Deming Regression

As stated previously, the fundamental flaw of simple linear regression is that it allows for measurement error in one variable only. This causes a downward biased slope estimate.

Deming's Regression suffers from some crucial drawback. Firstly it is computationally complex, and it requires specific software packages to perform calculations. Secondly it is uninformative about the comparative precision of two methods of measurement. Most importantly Carroll and Ruppert (1996) states that

Deming's regression is acceptable only when the precision ratio (λ, in their paper as η) is correctly specified, but in practice this is often not the case, with the λ being underestimated. This underestimation leads to an overcorrection for attenuation.

3.8 Diagnostics for Deming Regression

Model selection and diagnostic technique are well developed for classical linear regression methods. Typically an implementation of a linear model fit will be accompanied by additional information, such as the coefficient of determination and likelihood and information criterions, and a regression ANOVA table. Such additional information has not, as yet, been implemented for Deming regression.

3.9 Performance in the presence of outliers

All least square estimation methods are sensitive to outliers. In common with all regression methods, Deming regression is vulnerable to outliers.

Bland Altman's 1986 paper contains a data set, measurement of mean velocity of circumferential fibre shortening (VCF) by the long axis and short axis in M-mode echocardiography. Evident in this data set are outliers. Choosing the most noticeable, we shall use the deming regression method on this data set, both with and without this outlier, to assess its influence.

- In the presence of the outlier, the intercept and slope are estimated to be -0.0297027 and 1.0172959 respectively.

- Without the outlier the intercept and slope are estimated to be -0.11482220 and 1.09263112 respectively.

- We therefore conclude that Deming Regression is adversely affected by outliers , in the same way model I regression is.

3.10 Inference Procedures

A 95% confidence interval for the intercept estimate can be used to test the intercept, and hence fixed bias, is equal to zero. This hypothesis is accepted if the confidence interval for the estimate contains the value 0 in its range. Should this be, it can be concluded that fixed bias is not present. Conversely, if the hypothesis is rejected, then it is concluded that the intercept is non zero, and that fixed bias is present.

Testing for proportional bias is a very similar procedure. The 95% confidence interval for the slope estimate can be used to test the hypothesis that the slope is equal to 1. This hypothesis is accepted if the confidence interval for the estimate contains the value 1 in its range. If the hypothesis is rejected, then it is concluded that the slope is significant different from 1 and that a proportional bias exists.

3.11 Using LME models to estimate the ratio (BXC)

$$y_{mi} = \mu + \beta_m + b_i + \epsilon_{mi}$$

with β_m is a fixed effect for the method m and b_i is a random effect associated with patient i, and ϵ_{mi} as the measurement error. This is a simple single level LME model. Pinheiro and Bates (1994) provides for the implementation of fitting a model. The variance ratio of the residual variances is immediately determinable from the output. This variance ratio can be use to fit a Deming regression, as described in chapter 1.

3.12 Zhange Example

For convenience, a new data set shall be introduced to demonstrate Deming regression. Measurements of transmitral volumetric flow (MF) by doppler echocardiography, and left ventricular stroke volume (SV) by cross sectional echocardiography in 21 patients with aortic valve disease are tabulated in Zhang et al. (1986). This data set features in the discussion of method comparison studies in Altman (1991, p.398) .

Patient	MF (cm^3)	SV (cm^3)	Patient	MF (cm^3)	SV (cm^3)	Patient	MF (cm^3)	SV (cm^3)
1	47	43	8	75	72	15	90	82
2	66	70	9	79	92	16	100	100
3	68	72	10	81	76	17	104	94
4	69	81	11	85	85	18	105	98
5	70	60	12	87	82	19	112	108
6	70	67	13	87	90	20	120	131
7	73	72	14	87	96	21	132	131

Table 3.12.1: Transmitral volumetric flow(MF) and left ventricular stroke volume (SV) in 21 patients. (Zhang et al 1986)

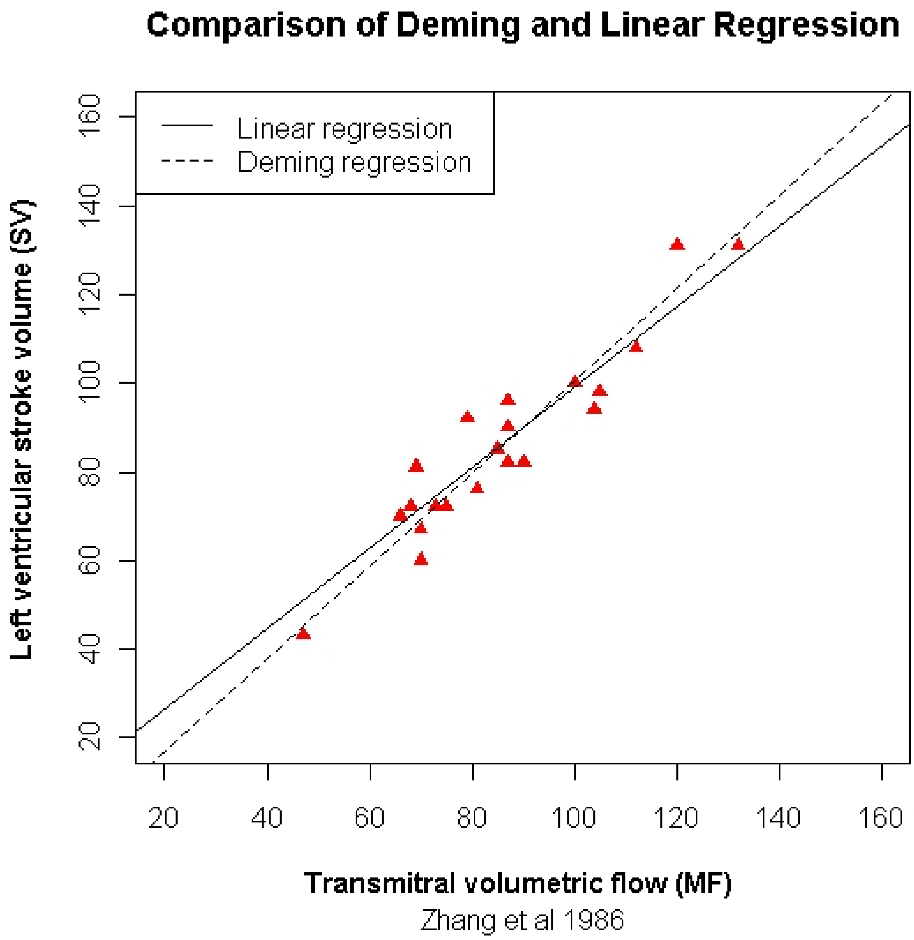

Figure 3.12.1: Deming Regression For Zhang's Data

3.13 Implementations

Thus far, one of the few R implementations of Deming regression is contained in the 'MethComp' package. (Carstensen et al., 2008).

Unless specified otherwise, the variance ratio λ has a default value of one. A means of computing likelihood functions would potentially allow for an algorithm for estimating the true variance ratio.

3.14 Methods

Theoretical calculations and simulations were used to consider the statistical power for detection of slope deviations from unity and intercept deviations from zero. For situations with proportional analytical standard deviations, weighted forms of regression analysis were evaluated.

3.15 Weighted Deming Regression

Weighted linear regression allows for non-constancy of the standard deviation of the y variable. However it is assumed that X is without measurement error. Weighted Deming regression takes into account the non-constant proportional measurement errors in both variables.Despite the non-constancy, it is necessary to retain the constant value of λ.

In **both forms** of Deming regression, λ is assumed to be constant through out the range of measurements. For WDR weights w_i are used to compute the sums of squares and cross products. The weights are inversely proportional to the squared analytical variance at any given value.

3.16 Linnet - References

The statistical procedures are described in: Linnet K. Necessary sample size for method comparison studies based on regression analysis. Clin Chem 1999; 45: 882-94. Linnet K. Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. Clin Chem 1998; 44: 1024-1031. Linnet K. Evaluation of regression procedures for methods comparison studies. Clin Chem 1993; 39. 424-432. Linnet K. Estimation of the linear relationship between measurements of two methods with proportional errors. Stat Med 1990; 9: 1463-1473.

3.17 Bootstap Techniques

Use of Bootstap Techniques to obtain Confidence Interval estimates

Carpenter, J., Bithell, J. (2000) Bootstrap condence intervals: when, which, what? A practical guide for medical statisticians. Stat Med, 19 (9), 11411164.

3.18 Ordinary Least Product Regression

Ludbrook (1997) states that the grouping structure can be straightforward, but there are more complex data sets that have a hierarchical(nested) model.

Observations between groups are independent, but observations within each groups are dependent because they belong to the same subpopulation. Therefore there are two sources of variation: between-group and within-group variance.

3.19 Least Products Regression

Used as an alternative to Bland-Altman Analysis, this method is also known as 'Geometric Mean Regression' and 'Reduced Major Axis Regression'. This regression model minimizes the areas of the right triangles formed by the data points' vertical and horizontal deviations from the fitted line and the fitted line.

Model II regression analysis caters for cases in which random error is attached to both dependent and independent variables. Comparing methods of measurement is just such a case.(Ludbrook)

Least products regression is the reviewer's preferred technique for analysing the Model II case. In this, the sum of the products of the vertical and horizontal deviations of the x,y values from the line is minimized.

Least products regression analysis is suitable for calibrating one method against another. It is also a sensitive technique for detecting and distinguishing fixed and proportional bias between methods.

Least-products regression can lead to inflated SEEs and estimates that do not tend to their true values an N approaches infinity (Draper and Smith, 1998).

Bibliography

Altman, D. and J. Bland (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society. Series D (The Statistician) 32*(3), 307–317.

Altman, D. G. (1991). *Practical Statistics for Medical Research*. Chapman and Hall.

Bartko, J. (1994). Measures of agreement: A single procedure. *Statistics in Medicine 13*, 737–745.

Bland, J. and D. Altman (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet i*, 307–310.

Bland, J. and D. Altman (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research 8*(2), 135–160.

Bradley, E. L. and L. G. Blackwood (1989). Comparing paired data: A simultaneous test for means and variances. *The American Statistician 43*(4), 234–235.

Carrasco, J. L. (2004). Structural equation model. In *Encyclopedia of Biopharmaceutical Statistics, Second Edition*, pp. 1–7. Taylor & Francis.

Carroll, R. and D. Ruppert (1996). The use and misuse of orthogonal regression in linear errors-in-variables models. *The American Statistician 50*(1), 1–6.

Carstensen, B. (2004). Comparing and predicting between several methods of measurement. *Biostatistics 5*(3), 399–413.

Carstensen, B., J. Simpson, and L. C. Gurrin (2008). Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics 4*(1).

Cohen, J., P. Cohen, S. West, and L. Aiken. *Applied multiple regression / correlation analysis for the behavioral sciences* (Third ed.). Laurence Erlbaum Associates.

Cornbleet, P. J. and D. Cochrane (1979). Regression methods for assessing agreement between two methods of clinical measurement. *Journal of Clinical Chemistry 24*(2), 342–345.

Dunn, G. (2002). *Statistical Evaluation of Measurement Error* (Second ed.). Stanford: American Mathematical Society and Digital Press.

Grubbs, F. (1948). On estimating precision of measuring instruments and product variability. *Journal of the American Statistical Association 43*, 243–264.

Grubbs, F. (1973). Errors of measurement, precision, accuracy and the statistical comparison of measuring instruments. *Technometrics 15*(1), 53–66.

Kelly, G. E. (1985). Use of the structural equations model in assessing the reliability of a new measurement technique. *Applied Statistics*, 258–263.

Kinsella, A. (1986). Estimating method precision. *The Statistician 35*, 421–427.

Kummel, C. (1879). Reduction of observation equations which contain more than one observed quantity. *The Analyst 6*, 97–105.

Lewis, P., P. Jones, J. Polak, and H. Tillitson (1991). The problem of conversion in method comparison studies. *Applied Statistics 40*, 105–112.

Linnet, K. (1998). Performance of deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clinical Chemistry 44*, 1024–1031.

Linnet, K. (1999). Necessary sample size for method comparison studies based on regression analysis. *Clinical Chemistry 45*(6), 882–894.

Ludbrook, J. (1997). Comparing methods of measurement. *Clinical and Experimental Pharmacology and Physiology 24*, 193–203.

Ludbrook, J. (2002). Statistical techniques for comparing measurers and methods of measurement: a critcal review. *Clinical and Experimental Pharmacology and Physiology 29*, 527–536.

Morgan, W. A. (1939). A test for the signicance of the difference between two variances in a sample from a normal bivariate population. *Biometrika 31*, 13–19.

O'Brien, E., J. Petrie, W. Littler, M. de Swiet, P. L. Padfield, D. Altman, M. Bland, A. Coats, and N. Atkins (1990). The British Hypertension Society protocol for the evaluation of blood pressure measuring devices. *Journal of Hypertension 8*, 607–619.

Pinheiro, J. and D. Bates (1994). *Mixed Effects Models in S and S plus* (2nd ed.). Reading, Massachusetts: Springer.

Pitman, E. J. G. (1939). A note on normal correlation. *Biometrika 31*, 9–12.

Thompson, W. (1963). Precision of simultaneous measurement procedures. *Journal of American Statistical Association 58*, 474–479.

Zhang, Y., S. Nitter-Hauge, H. Ihlen, K. Rootwelt, and E. Myhre (1986). Measurement of aortic regurgitation by doppler echocardiography. *British Heart Journal 55*, 32–38.