0.1   Roy's Approach

For the purposes of comparing two methods of measurement, Roy (2009b) presents a methodology utilizing linear mixed effects model. This methodology provides for the formal testing of inter-method bias, between-subject variability and within-subject variability of two methods.

Roy (2009b) uses an approach based on linear mixed effects (LME) models for the purpose of comparing the agreement between two methods of measurement, where replicate measurements on items (often individuals) by both methods are available. Three tests of hypothesis appropriate are provided for evaluating the agreement between the two methods of measurement under this sampling scheme.

Importantly Roy (2009b) further proposes a series of three tests on the variance components of an LME model, which allow decisions on the second and third of Roy's criteria. For these tests, four candidate LME models are constructed.

Roy (2009b) proposes the use of LME models to perform a test on two methods of agreement to comparing the agreement between two methods of measurement, where replicate measurements on items (often individuals) by both methods are available, determining whether they can be used interchangeably. The methodology proposed by Roy (2009b) is largely based on Hamlett et al. (2004), which in turn follows on from Lam et al. (1999).

Roy (2009b) proposes a novel method using the LME model with Kronecker product covariance structure in a doubly multivariate set-up to assess the agreement between a new method and an established method with unbalanced data and with unequal replications for different subjects. By doubly multivariate set up, Roy means that the information on each patient or item is multivariate at two levels, the number of methods and number of replicated measurements. Further to Lam et al. (1999), it is

assumed that the replicates are linked over time. However it is easy to modify to the unlinked case.

The well-known "Limits of Agreement", as developed by Bland and Altman (1986) are not referred to directly, but are easily computable using the framework proposed by Roy (2009b). Further discussion will be provided in due course.

Further to this, Roy (2009b) demonstrates an suite of tests that can be used to determine how well two methods of measurement, in the presence of repeated measures, agree with each other.

- No Significant inter-method bias

- No difference in the between-subject variabilities of the two methods

- No difference in the within-subject variabilities of the two methods

The formulation presented above usefully facilitates a series of significance tests that advise as to how well the two methods agree. These tests are as follows:

- A formal test for the equality of between-item variances,

- A formal test for the equality of within-item variances,

- A formal test for the equality of overall variances.

These tests are complemented by the ability to consider the inter-method bias and the overall correlation coefficient. Two methods can be considered to be in agreement if criteria based upon these methodologies are met. Additionally Roy makes reference to the overall correlation coefficient of the two methods, which is determinable from variance estimates. Other important aspects of the method comparison study are consequent. The limits of agreement are computed using the results of the reference model.

Differences in between-subject variabilities of the two methods arise when one method is yielding average response levels for individuals than are more variable than the average response levels for the same sample of individuals taken by the other method. Differences in within-subject variabilities of the two methods arise when one method is yielding responses for an individual than are more variable than the responses for this same individual taken by the other method. The two methods of measurement can be considered to agree, and subsequently can be used interchangeably, if all three null hypotheses are true.

0.1.1 LME Model Specification

Let y_{mir} denote the rth replicate measurement on the ith item by the mth method, where $m = 1, 2$, $i = 1, \ldots, N$, and $r = 1, \ldots, n_i$. When the design is balanced and there is no ambiguity we can set $n_i = n$. The LME model underpinning Roy's approach can be written

$$y_{mir} = \beta_0 + \beta_m + b_{mi} + \epsilon_{mir}. \tag{1}$$

Here β_0 and β_m are fixed-effect terms representing, respectively, a model intercept and an overall effect for method m.

The b_{1i} and b_{2i} terms represent random effect parameters corresponding to the two methods, having $\mathrm{E}(b_{mi}) = 0$ with $\mathrm{Var}(b_{mi}) = g_m^2$ and $\mathrm{Cov}(b_{mi}, b_{m'i}) = d_{12}$. The random error term for each response is denoted ϵ_{mir} having $\mathrm{E}(\epsilon_{mir}) = 0$, $\mathrm{Var}(\epsilon_{mir}) = \sigma_m^2$, $\mathrm{Cov}(b_{mir}, b_{m'ir}) = \sigma_{12}$, $\mathrm{Cov}(\epsilon_{mir}, \epsilon_{mir'}) = 0$ and $\mathrm{Cov}(\epsilon_{mir}, \epsilon_{m'ir'}) = 0$.

When two methods of measurement are in agreement, there is no significant differences between β_1 and β_2, g_1^2 and g_2^2, and σ_1^2 and σ_2^2. Here β_0 and β_m are fixed-effect terms representing, respectively, a model intercept and an overall effect for method m.

The model can be reparameterized by gathering the β terms together into (fixed

effect) intercept terms $\alpha_m = \beta_0 + \beta_m$. The b_{1i} and b_{2i} terms are correlated random effect parameters having $E(b_{mi}) = 0$ with $Var(b_{mi}) = g_m^2$ and $Cov(b_{1i}, b_{2i}) = d_{12}$.

The random error term for each response is denoted ϵ_{mir} having $E(\epsilon_{mir}) = 0$, $Var(\epsilon_{mir}) = \sigma_m^2$, $Cov(\epsilon_{1ir}, \epsilon_{2ir}) = \sigma_{12}$, $Cov(\epsilon_{mir}, \epsilon_{mir'}) = 0$ and $Cov(\epsilon_{1ir}, \epsilon_{2ir'}) = 0$. Two methods of measurement are in complete agreement if the null hypotheses $H_1 \colon \alpha_1 = \alpha_2$ and $H_2 \colon \sigma_1^2 = \sigma_2^2$ and $H_3 \colon g_1^2 = g_2^2$ hold simultaneously. Roy (2009b) uses a Bonferroni correction to control the familywise error rate for tests of $\{H_1, H_2, H_3\}$ and account for difficulties arising due to multiple testing.

0.1.2 Variance Covariance Matrices

Under Roy's model, random effects are defined using a bivariate normal distribution. Consequently, the variance-covariance structures can be described using 2×2 matrices. A discussion of the various structures a variance-covariance matrix can be specified under is required before progressing. The following structures are relevant: the identity structure, the compound symmetric structure and the symmetric structure.

The differences in the models are specifically in how the the D and Λ matrices are constructed, using either an unstructured form or a compound symmetry form. To illustrate these differences, consider a generic matrix A,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

A symmetric matrix allows the diagonal terms a_{11} and a_{22} to differ. The compound symmetry structure requires that both of these terms be equal, i.e $a_{11} = a_{22}$.

The identity structure is simply an abstraction of the identity matrix. The compound symmetric structure and symmetric structure can be described with reference to the following matrix (here in the context of the overall covariance Block-$\boldsymbol{\Omega}_i$, but

4

equally applicable to the component variabilities $\boldsymbol{D}$ and $\boldsymbol{\Sigma}$);

$$\begin{pmatrix} \omega_1^2 & \omega_{12} \\ \omega_{12} & \omega_2^2 \end{pmatrix}$$

Symmetric structure requires the equality of all the diagonal terms, hence $\omega_1^2 = \omega_2^2$. Conversely compound symmetry make no such constraint on the diagonal elements. Under the identity structure, $\omega_{12} = 0$. A comparison of a model fitted using symmetric structure with that of a model fitted using the compound symmetric structure is equivalent to a test of the equality of variance.

Independence

As though analyzed using between subjects analysis.

$$\begin{pmatrix} \psi^2 & 0 & 0 \\ 0 & \psi^2 & 0 \\ 0 & 0 & \psi^2 \end{pmatrix}$$

Compound Symmetry

Assumes that the variance-covariance structure has a single variance (represented by ψ^2) for all 3 of the time points and a single covariance (represented by ψ_{ij}) for each of the pairs of trials.

$$\begin{pmatrix} \psi^2 & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi^2 & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi^2 \end{pmatrix}$$

0.1.3 Model Terms (Roy 2009)

- Let y_{mir} be the response of method m on the ith subject at the $r-$th replicate.

- Let $\boldsymbol{y}_{ir}$ be the 2×1 vector of measurements corresponding to the $i-$th subject at the $r-$th replicate.

- Let $\boldsymbol{y}_i$ be the $R_i \times 1$ vector of measurements corresponding to the $i-$th subject, where R_i is number of replicate measurements taken on item i.

- Let $\alpha_m i$ be the fixed effect parameter for method for subject i.

- Formally ARoy2009 uses a separate fixed effect parameter to describe the true value μ_i, but later combines it with the other fixed effects when implementing the model.

- Let u_{1i} and u_{2i} be the random effects corresponding to methods for item i.

- $\boldsymbol{\epsilon}_i$ is a n_i-dimensional vector comprised of residual components. For the blood pressure data $n_i = 85$.

- $\boldsymbol{\beta}$ is the solutions of the means of the two methods. In the LME output, the bias ad corresponding t-value and p-values are presented. This is relevant to ARoy2009's first test.

0.2   Using LME for method comparison

Due to the prevalence of modern statistical software, Carstensen et al. (2008) advocates the adoption of computer based approaches, such as LME models, to method comparison studies. Carstensen et al. (2008) remarks upon 'by-hand' approaches advocated in Bland and Altman (1999) discouragingly, describing them as tedious, unnecessary and 'outdated'. Rather than using the 'by hand' methods, estimates for required LME parameters can be read directly from program output. Furthermore, using computer

approaches removes constraints associated with 'by-hand' approaches, such as the need for the design to be perfectly balanced.

0.2.1 Roy's Approach

For the purposes of comparing two methods of measurement, Roy (2009a) presents a framework that utilizes linear mixed effects model. This methodology provides for the formal testing of inter-method bias, between-subject variability and within-subject variability of two methods. The formulation contains a Kronecker product covariance structure in a doubly multivariate setup. By doubly multivariate set up, Roy means that the information on each patient or item is multivariate at two levels, the number of methods and number of replicated measurements. Further to Lam et al. (1999), it is assumed that the replicates are linked over time. However it is easy to modify to the unlinked case.

Roy (2009a) sets out three criteria for two methods to be considered in agreement. Firstly that there be no significant bias. Second that there is no difference in the between-subject variabilities, and lastly that there is no significant difference in the within-subject variabilities. Roy further proposes examination of the the overall variability by considering the second and third criteria be examined jointly. Should both the second and third criteria be fulfilled, then the overall variabilities of both methods would be equal.

A formal test for inter-method bias can be implemented by examining the fixed effects of the model. This is common to well known classical linear model methodologies. The null hypotheses, that both methods have the same mean, which is tested against the alternative hypothesis, that both methods have different means. The inter-method bias and necessary $t-$value and $p-$value are presented in computer output. A decision on whether the first of Roy's criteria is fulfilled can be based on these values.

Importantly Roy (2009a) further proposes a series of three tests on the variance components of an LME model, which allow decisions on the second and third of Roy's criteria. For these tests, four candidate LME models are constructed. The differences in the models are specifically in how the the D and Λ matrices are constructed, using either an unstructured form or a compound symmetry form. To illustrate these differences, consider a generic matrix A,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

A symmetric matrix allows the diagonal terms a_{11} and a_{22} to differ. The compound symmetry structure requires that both of these terms be equal, i.e $a_{11} = a_{22}$.

The first model acts as an alternative hypothesis to be compared against each of three other models, acting as null hypothesis models, successively. The models are compared using the likelihood ratio test. Likelihood ratio tests are a class of tests based on the comparison of the values of the likelihood functions of two candidate models.

0.2.2   Correlation

In addition to the variability tests, Roy advises that it is preferable that a correlation of greater than 0.82 exist for two methods to be considered interchangeable. However if two methods fulfil all the other conditions for agreement, failure to comply with this one can be overlooked. Indeed Roy demonstrates that placing undue importance to it can lead to incorrect conclusions. Roy (2009a) remarks that current computer implementations only gives overall correlation coefficients, but not their variances. Consequently it is not possible to carry out inferences based on all overall correlation coefficients.

0.2.3 Variability test 1

The first test determines whether or not both methods A and B have the same between-subject variability, further to the second of Roy's criteria.

$$H_0: \ d_A = d_B$$

$$H_A: \ d_A \neq d_B$$

This test is facilitated by constructing a model specifying a symmetric form for D (i.e. the alternative model) and comparing it with a model that has compound symmetric form for D (i.e. the null model). For this test $\hat{\mathbf{\Lambda}}$ has a symmetric form for both models, and will be the same for both.

0.2.4 Variability test 2

This test determines whether or not both methods A and B have the same within-subject variability, thus enabling a decision on the third of Roy's criteria.

$$H_0: \ \lambda_A = \lambda_B$$

$$H_A: \ \lambda_A = \lambda_B$$

This model is performed in the same manner as the first test, only reversing the roles of $\hat{\mathbf{D}}$ and $\hat{\mathbf{\Lambda}}$. The null model is constructed a symmetric form for $\hat{\mathbf{\Lambda}}$ while the alternative model uses a compound symmetry form. This time $\hat{\mathbf{D}}$ has a symmetric form for both models, and will be the same for both.

As the within-subject variabilities are fundamental to the coefficient of repeatability, this variability test likelihood ratio test is equivalent to testing the equality of two coefficients of repeatability of two methods. In presenting the results of this test, Roy (2009a) includes the coefficients of repeatability for both methods.

0.2.5   Variability test 3

The last of the variability test examines whether or not methods A and B have the same overall variability. This enables the joint consideration of second and third criteria.

$$H_0 : \quad \sigma_A = \sigma_B$$

$$H_A : \quad \sigma_A = \sigma_B$$

The null model is constructed a symmetric form for both $\hat{\boldsymbol{D}}$ and $\hat{\boldsymbol{\Lambda}}$ while the alternative model uses a compound symmetry form for both.

0.2.6   Model Terms (Roy 2009)

It is important to note the following characteristics of this model.

Let the number of replicate measurements on each item i for both methods be n_i, hence $2 \times n_i$ responses. However, it is assumed that there may be a different number of replicates made for different items. Let the maximum number of replicates be p. An item will have up to $2p$ measurements, i.e. $\max(n_i) = 2p$.

Later on $\boldsymbol{X}_i$ will be reduced to a 2×1 matrix, to allow estimation of terms. This is due to a shortage of rank. The fixed effects vector can be modified accordingly.

$\boldsymbol{Z}_i$ is the $2n_i \times 2$ model matrix for the random effects for measurement methods on item i.

$\boldsymbol{b}_i$ is the 2×1 vector of random-effect coefficients on item i, one for each method.

$\boldsymbol{\epsilon}$ is the $2n_i \times 1$ vector of residuals for measurements on item i.

$\boldsymbol{G}$ is the 2×2 covariance matrix for the random effects.

$\boldsymbol{R}_i$ is the $2n_i \times 2n_i$ covariance matrix for the residuals on item i.

The expected value is given as $\mathrm{E}(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$. (Hamlett et al., 2004)

The variance of the response vector is given by $\mathrm{Var}(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{G}\boldsymbol{Z}'_i + \boldsymbol{R}_i$ (Hamlett et al., 2004).

$\boldsymbol{b}_i$ is a $m-$dimensional vector comprised of the random effects.

$$\boldsymbol{b}_i = \begin{pmatrix} b_{1i} \\ b_{21} \end{pmatrix} \tag{2}$$

$\boldsymbol{V}$ represents the correlation matrix of the replicated measurements on a given method. $\boldsymbol{\Sigma}$ is the within-subject VC matrix.

$\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ are positive definite matrices. The dimensions of $\boldsymbol{V}$ and $\boldsymbol{\Sigma}$ are $3 \times 3 (= p \times p)$ and $2 \times 2 (= k \times k)$.

It is assumed that $\boldsymbol{V}$ is the same for both methods and $\boldsymbol{\Sigma}$ is the same for all replications.

$\boldsymbol{V} \bigotimes \boldsymbol{\Sigma}$ creates a $6 \times 6 (= kp \times kp)$ matrix. $\boldsymbol{R}_i$ is a sub-matrix of this.

Bibliography

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.

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).

Hamlett, A., L. Ryan, and R. Wolfinger (2004). On the use of PROC MIXED to estimate correlation in the presence of repeated measures. *Proceedings of the Statistics and Data Analysis Section, SAS Users Group International 198-229*, 1–7.

Lam, M., K. Webb, and D. O'Donnell (1999). Correlation between two variables in repeated measurements. *American Statistical Association, Proceedings of the Biometric Session*, 213–218.

Roy, A. (2009a). An application of linear mixed effects model to assess the agreement between two methods with replicated observations. *Journal of Biopharmaceutical Statistics 19*, 150–173.

Roy, A. (2009b). An application of the linear mixed effects model to ass the agreement

between two methods with replicated observations. *Journal of Biopharmaceutical Statistics* *19*, 150–173.