Diagnostic measures for the analysis of method comparison studies using the linear mixed effects model

Transfer Report Presentation

Kevin O'Brien

kevin.obrien@ul.ie

Dept. of Mathematics & Statistics,
University *of* Limerick

November 2010

Method comparison studies:

- Method comparison studies are used to assess the relative agreement between two methods that measure the same variable.
- Two methods of measurement giving consistently similar results, with the same level of precision are considered to be in agreement.
- Commonly a new method is compared to one currently in use to see whether their measurements are indeed comparable, and whether these two methods can be used interchangeably.
- Inter-method bias can be described as the tendency for one method to give a measurement higher than the counterpart method. The presence of inter-method bias indicates lack of agreement.
- Importance of this area is reflected by the 18,860 citations of Bland and Altman's 1986 paper in the Lancet.

Bland-Altman plots

- (Bland and Altman, 1986 and 1999) noted the inappropriate use of correlation for method comparison, and developed a simple graphical approach , known as a Bland-Altman plot, to compare two measurements methods.
- The approach requires the calculation and plotting the case-wise average and case-wise differences.
- The mean of the case-wise differences serves an estimate for the inter-method bias. A horizontal line to represent this is added to the plot.
- Bland and Altman also advise the use of scatterplots as a preliminary analysis.

Limits of Agreement

- Bland and Altman (1986) introduces a further element to their methodology, the limits of agreement.
- The standard deviation of the case-wise differences is determined.
- The 95% limits of agreement are calculated as the inter-method bias plus and minus 1.96 standard deviation of casewise differences
- Again, horizontal line are added to the Bland-Altman plot.
- These limits are expected to contain the difference between measurements by the two methods for 95% of pairs of future measurements on similar individuals.

Repeatability

- Repeatability is relevant to the study of method comparison because the repeatabilities of the two methods of measurement limit the amount of agreement which is possible.
- If one method has poor repeatability i.e. there is considerable variation in repeated measurements on the same subject, the agreement between the two methods is bound to be poor too.
- Bland and Altman (1986) complemented their methodology with the coefficient of repeatability. The coefficient of repeatability is a precision measure which represents the value below which the absolute difference between two repeated measurements can be expected to lie with 95% probability .

Example: PEFR Data

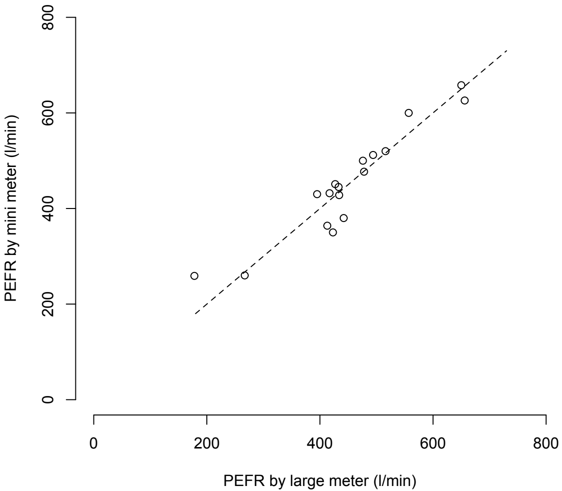
- This example uses data from Bland and Altman (1986). Two measurements of peak expiratory flow rate (PEFR) are compared. One of these measurements uses a “Large” meter and the other a “Mini” meter.
- The purpose of the procedure is to assess the agreement between both methods.

PEFR Data

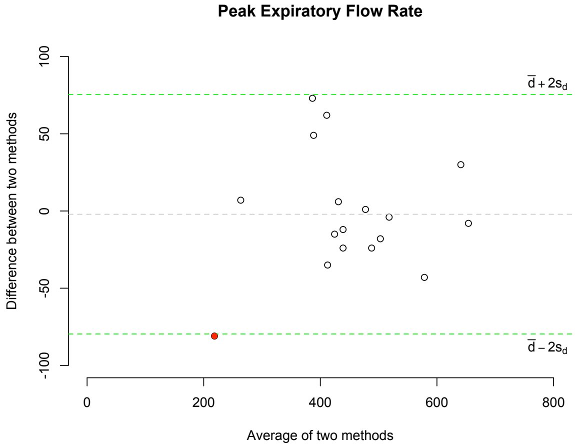
	Wright (1st)	Wright (2nd)	Mini (1st)	Mini (2nd)
Subject	(l/min)	(l/mi)	(l/min)	(l/min)
1	494	490	512	525
2	395	397	430	415
3	516	512	520	508
4	434	401	428	444
⋮	⋮	⋮	⋮	⋮

The table above tabulates the observed PEFR measured with both Wright peak flow and Mini peak flow meters, on 17 individuals. Each individual is measure twice with each meter. Bland and Altman only use the first measurements of each meter.

Scatter plot for PEFR data



Bland-Altman plot for PEFR data



Bland-Altman Plot : Multiple measurements

- Method comparison studies will commonly use replicate measurements on each individual.
- The Bland-Altman plot was specifically constructed for a single pair of measurements.
- Bland and Altman (1999) extend the limits of agreement approach to data with repeated measurements.
- The proposed approaches entail either treating measurements as a single measurement, or by considering the average of a set of repeated measures.
- Carstensen et al (2008) disapprove of this approach, and recommend using linear mixed effects models for assessing agreement in the presence of replicate measurements.

Linear Mixed Effects model

- A linear mixed effects (LME) model is a statistical model containing both fixed effects and random effects.
- LME models are a generalization of the classical linear model, which contain fixed effects only.
- LME models are commonly formulated in the notation described in Laird and Ware (1982)

$$y = X\beta + Zb + \varepsilon$$

- Of particular interest is the variance estimates associated with the random effects parameters

$$\mathrm{var}\begin{pmatrix} b \\ \varepsilon \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & \Sigma \end{pmatrix}$$

LME models in method comparison

- Hamlett et al (2003) describe a formulation of the LME model to describe all responses of the $i-$th subject $(\boldsymbol{y}_i)$, in the presence of two predictor variables.

$$\boldsymbol{y}_i = \boldsymbol{X}_i\beta + \boldsymbol{Z}_i Z\boldsymbol{b}_i + \varepsilon_i$$

- Roy (2009) uses this approach to build an LME methodology for assessing agreement between two methods. The measurements are the responses, and both methods of measurement are the predictor variables.
- Roy's approach allows for specifying the structure of the between-subject variance $\boldsymbol{D}$ and the within-subject variance Σ of the measurement methods.

Roy's agreement criteria

Roy (2009) 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 no significant difference in the within-subject variabilities.
- There should be no difference in the the overall variability of both methods.
- Roy demonstrates 3 formal tests for comparing the variabilities.
- These tests are realized by specifying different structures for the variance-covariance matrices.

Variability tests

The three variability tests are as follows

- Testing whether both methods have equal between-subject variances
- Testing whether both methods have equal within-subject variances
- Testing whether both methods have equal overall variability

JSR Data

Roy includes a table from Bland and Altman's 1999 paper which shows a set of systolic blood pressure data from a study in which simultaneous measurements were made by each of two experienced observers (denoted J and R) using a sphygmomanometer and by a semi-automatic blood pressure monitor (denoted S). Three sets of readings were made in quick succession. Her demonstrates her procedure with both comparisons.

Roy's Methodology R implementation

Roy's methodology requires the construction of four very similar LME models. The first of Roy's candidate model (MCS1) can be implemented using the following code;

```
MCS1 = lme(BP ~ method-1, data = dat,
random = list(subject=pdSymm(~ method-1)),
weights = varIdent(form=~1|method),
correlation = corSymm(form=~1 | subject/obs), method="ML")
```

The other three candidate models are constructed by using the same code, but exchanging the phrase "CompSymm" for "Symm" when appropriate.

Roy's likelihood ratio tests

Comparing Models:
A form test may be implemented using a likelihood ration test, implemented in `R` with the `anova()` command.

```
> anova(MCS1,MCS2)
      Model df  AIC    BIC    logLik  Test    L.Ratio  p-value
MCS1  1     8   4077.5 4111.3 -2030.7
MCS2  2     7   4075.6 4105.3 -2030.8 1 vs 2  0.15291  0.6958
>
```

Further to this test, we fail to reject the model MCS1 in favour of MCS2. Therefore we conclude both methods have the same between-subject variance.

Roy's fixed effects estimates

Fixed Effects estimates:
The fixed effects estimates of any model can be used to determine the Inter-method bias. To compute the inter-method bias simple subtract one estimate from the other.
However , to obtain explicit parameter estimates, such as p-value, a variation of the code, specifying a intercept, can be used;

```
MCS1 = lme(BP ~ method, data = dat, ......
```

Limits of Agreement

General LME models:
Bendix Carstensen et al demonstrate a method of computing the limits of agreement using simple LME models. However, if the observations are assumed to be linked, an extra interaction term must be added to Carstensen LME model.

Roy's LME models:
Roy's LME models can also be used to compute the limits of agreement. The formulation of the model already accounts for linkage, so no additional terms are necessary. Limits of agreement are not given explicitly in the code output, but a simple R function can be used to construct them.

LME Diagnostics for LME models

- Cook and Weisberg (1982) developed case deletion model diagnostics, such as Cook's distance for linear models.
- Christensen, Pearson and Johnson (1992) examined case deletion diagnostics, in particular the LME equivalent of Cooks distance, for assessing influential observations in LME models.
- Haslett and Hayes (1998) presented a general theory for residuals in the general linear model (GLM) framework, in which marginal and conditional residuals are described.

Henderson Equation

Henderson's Equation for estimating LME parameters can be formulated as

$$\begin{pmatrix} X'\Sigma^{-1}X & X'\Sigma^{-1}Z \\ Z'\Sigma^{-1}X & X'\Sigma^{-1}X + D^{-1} \end{pmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} = \begin{pmatrix} X'\Sigma^{-1}y \\ Z'\Sigma^{-1}y \end{pmatrix}.$$

Henderson's equations can be rewritten as a GLM model in form $(T'W^{-1}T)\delta = T'W^{-1}y_a$ using

$$\delta = \begin{pmatrix} \beta \\ b \end{pmatrix}, \; y_a = \begin{pmatrix} y \\ \psi \end{pmatrix}, \; T = \begin{pmatrix} X & Z \\ 0 & I \end{pmatrix}, \text{ and } W = \begin{pmatrix} \Sigma & 0 \\ 0 & D \end{pmatrix},$$

(ψ represents "quasi-data", a construct used to analyze GLMs.)

LMEs as Augmented GLMs

- This augmented GLM formulation allows the analyses developed by Haslett and Hayes (1998) to be used on LME models.
- It is proposed to investigate diagnostic measures, such as residuals Mahalanobis distances, for validating method comparison studies models.

Bibliography

- D.G Altman and J.M. Bland (1983) - Measuring agreement in method comparison studies. (*Statistical Methods in Medical Research* ; **8**:135–160.)
- J.M. Bland and D.G Altman (1986) - Measuring agreement in method comparison studies. (*Statistical Methods in Medical Research* ; **8**:135–160.)
- J.M. Bland and D.G Altman (1999) - Measuring agreement in method comparison studies. (*Statistical Methods in Medical Research* ; **8**:135–160.)
- N.M. Laird and J.H Ware (1982) - Random-Effects Models for Longitudinal Data, (*Biometrics*, **38**: 963974.)

Bibliography (contd)

- Hamlett et al (1983) - Measuring agreement in method comparison studies. (*Statistical Methods in Medical Research* ; **8**:135–160.)
- Carstensen (2008) - Measuring agreement in method comparison studies. (*Statistical Methods in Medical Research* ; **8**:135–160.)
- Anuradha Roy (2009) - Random-Effects Models for Longitudinal Data, (*Biometrics*, **38**: 963974.)