Statistical Models for Assessing Agreement in Method Comparison Studies with Replicate Measurements

Bendix Carstensen*       Julie Simpson†
Lyle C. Gurrin‡
Statistical Models for Assessing Agreement in Method Comparison Studies with Replicate Measurements*

Bendix Carstensen, Julie Simpson, and Lyle C. Gurrin

Abstract

Method comparison studies are usually analyzed by computing limits of agreement. It is recommended that replicate measurements be taken by each method, but the resulting data are more cumbersome to analyze. We discuss the statistical model underlying the classical limits of agreement and extend it to the case with replicate measurements. As the required code to fit the models is non-trivial, we provide example computer code to fit the models, and show how to use the output to derive measures of repeatability and limits of agreement.

KEYWORDS: method comparison, Bland-Altman plot, mixed models

*We are grateful to Peter Dalgaard for (much needed) advice on the lme syntax.
1 Introduction

The problem of comparing two methods of measurement is still occasionally approached by computing correlation coefficients, despite the fact that this has been discouraged as irrelevant and misleading for more than 20 years [1, 2]. The preferred approach is to consider the differences between measurements by the two methods, and produce prediction limits for the difference between pairs of future measurements, known as the \textit{limits of agreement}.

When replicate measurements are taken with each method on each item (i.e. person or sample) measuring agreement becomes slightly more complicated. Bland and Altman [3] presented details of various approaches to adopt in this case, mainly based on calculations that can be performed “by hand”. Such tedious computations are unnecessary since the underlying concept of limits of agreement is merely a prediction from a statistical model that can be fitted with modern software for random effects models. The estimates of the variance components are given directly in the program output and can be used directly to generate limits of agreement and measures of repeatability of the methods.

This has the advantage of bypassing a lot of hand-calculations and makes it irrelevant whether the design is perfectly balanced or not.

Moreover, setting up a model focuses on the implications of the exchangeability properties of the replicate measurements, e.g. whether replicates are exchangeable \textit{within} each method by item stratum or only within items (paired or linked replicates).

2 Notation

In this paper we set up models for method comparison data with replicate measurements. The models that are needed are models where the residual variances differ by method, and this type of model is not very clearly presented in the manuals of any of the major software packages, so therefore we provide the code needed in \texttt{R}, \texttt{Stata} and \texttt{SAS}.

We assume the data are formatted as a dataset with four columns named:

\texttt{meth}, method of measurement, the number of methods being \( M \),
\texttt{item}, items (persons, samples) measured by each method, of which there are \( I \),
\texttt{repl}, replicate indicating repeated measurement of the same item by the same method, and
\texttt{y}, the measurement.
We denote the measurement by method $m$ on item $i$, replicate $r$ by $y_{mir}$.

When specifying mixed models we use Greek letters for fixed effects and Latin letters for random effects.

3 The classical approach

The classical setup for comparison of two measurement methods is one where one measurement by each method is taken on each item, that is, without replicates. In that case the recommendation is to compute the *limits of agreement*, a prediction interval for the difference between future measurements with the two methods on a new individual.

Underlying this approach is the two-way analysis of variance model:

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

The differences $y_{1i} - y_{2i}$ have variance $\sigma_1^2 + \sigma_2^2$, and the prediction interval for a difference between two new measurements is therefore:

$$\alpha_1 - \alpha_2 \pm 1.96 \sqrt{\sigma_1^2 + \sigma_2^2}$$

In practice, the term $\alpha_1 - \alpha_2$ is estimated by the mean difference, the last term is computed as the empirical standard deviation of the differences, and the 1.96 is replaced by 2 for convenience:

$$\bar{d} \pm 2 \text{s.d.}(d_i)$$

— this is what is commonly termed the limits of agreement.

This is formally incorrect as a prediction interval, since the errors in estimation of the parameters are not taken into account; formally the 95% prediction interval for the difference should be computed as:

$$\bar{d} \pm t_{0.975}(I - 1) \sqrt{1 + 1/I} \times \text{s.d.}(d_i)$$

where $I$ is the number if items. The term $t_{0.975}(I - 1) \sqrt{1 + 1/I}$ is 2.05 for $I = 30$ and less than 2 if $I > 61$, so the pragmatic method gives slight underestimates of the width of the limits of agreement for small studies. This is however based on a heavy exploitation of the normality assumption of the error terms ($e_{mi}$).

There are two rather more interesting assumptions in the model:

1. The variation of the differences is constant over the range of measurements.

2. The difference between the methods is constant over the range of measurements.
Figure 1: Measurements of subcutaneous fat (in mm) by two different observers. Data from the Steno Diabetes Center, 2006. The left panel is a Bland-Altman plot based on the means over replicates with limits of agreement based on these. The right panel is a Bland-Altman plot where the replicates are randomly matched, and (item \times repl) are used as independent items ignoring the exchangeability. The thick broken (gray) lines almost on top of the limits of agreement represent the correct limits of agreement computed from the variance component model in section 4.

These assumptions are checked by making a so-called Bland-Altman plot [2], where differences are plotted against averages of methods.

Figure 1 presents data from a comparison of measurements of subcutaneous fat by two observers at the Steno Diabetes Center. Measurements are in millimeters (mm). Each person is measured three times by each observer. The sequence of measurements is not considered to be of importance, so the replicate measurements are exchangeable within person (item) and observer (method).

The graph indicates that the underlying assumptions are reasonably well fulfilled. The limits of agreement in the first graph are based on the means of repeats within item and method. These limits of agreement can only be interpreted as prediction limits for the difference between means of three measurements by both methods, which is normally not relevant. Hence we must set up a framework that allows us to address the relevant prediction question based on single measurements.
4 Models for replicate measurements

To determine prediction limits for differences between single measurements we must resort to a more elaborate model for our data, where replicate measurements are explicitly modeled:

\[ y_{mir} = \alpha_m + \mu_i + c_{mi} + e_{mir}, \quad c_{mi} \sim N(0, \tau_m^2), \quad e_{mir} \sim N(0, \sigma_m^2) \] (1)

This is a model where the variation between items for method \( m \) is captured by \( \tau_m \) and the within item variation by \( \sigma_m \). The formulation of this model is general and refers to comparison of any number of methods — however, if only two methods are compared, separate values of \( \tau_1^2 \) and \( \tau_2^2 \) cannot be estimated, only their average, so in the case of only two methods we are forced to assume that \( \tau_1 = \tau_2 = \tau \).

Under this model the limits of agreement should be computed based on the standard deviation of the difference between a pair of measurements by the two methods on a new individual, \( j \), say:

\[ \text{var}(y_{1j} - y_{2j}) = 2\tau^2 + \sigma_1^2 + \sigma_2^2 \]

Therefore the limits of agreement are estimated by:

\[ \hat{\alpha}_1 - \hat{\alpha}_2 \pm 2 \times \sqrt{2\hat{\tau}^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2} \]

It therefore only remains to estimate the variance components in this linear mixed model, which can be done using standard software. Using the subcutaneous fat example, we present below the code and output for the statistical packages R, Stata and SAS.

4.1 Practical estimation of the variance components

4.1.1 Data

For generality the dataset was set up with the variable names \text{meth}, \text{item}, \text{repl} and \text{y}. All three examples below are using this data set-up:

<table>
<thead>
<tr>
<th>meth</th>
<th>item</th>
<th>repl</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>KL</td>
<td>1</td>
<td>1</td>
<td>1.6</td>
</tr>
<tr>
<td>KL</td>
<td>1</td>
<td>2</td>
<td>1.7</td>
</tr>
<tr>
<td>KL</td>
<td>1</td>
<td>3</td>
<td>1.7</td>
</tr>
<tr>
<td>KL</td>
<td>3</td>
<td>1</td>
<td>2.8</td>
</tr>
<tr>
<td>KL</td>
<td>3</td>
<td>2</td>
<td>2.9</td>
</tr>
<tr>
<td>KL</td>
<td>3</td>
<td>3</td>
<td>2.8</td>
</tr>
<tr>
<td>KL</td>
<td>5</td>
<td>1</td>
<td>2.7</td>
</tr>
<tr>
<td>KL</td>
<td>5</td>
<td>2</td>
<td>2.8</td>
</tr>
</tbody>
</table>
The function to use in \( \mathcal{R} \) is \texttt{lme}, but the syntax is somewhat arcane, see e.g. [6]. If the \texttt{random} argument in \texttt{lme} is a list, and the name of the first element is the name of a variable in the dataset, all terms are nested in this variable. The example here requires that the variables \texttt{meth}, \texttt{item} and \texttt{repl} are factors.

\begin{verbatim}
> lme( y ~ meth + item,
+      random = list( item = pdIdent( ~ meth-1 ) ),
+      weights = varIdent( form = ~1 | meth ),
+      data=fat
+ )
Linear mixed-effects model fit by REML
  Data: fat
  Log-restricted-likelihood: 188.3488
  Fixed: y ~ meth + item
    (Intercept) methSL item2 item3
   1.6896001995 -0.0448837209 -0.8653286307 1.1326030428
  Random effects:
  Formula: ~meth - 1 | item
   Structure: Multiple of an Identity
       methKL methSL Residual
    StdDev: 0.059556 0.059556 0.07717392
  Variance function:
   Structure: Different standard deviations per stratum
   Formula: ~1 | meth
   Parameter estimates:
      KL     SL
   1.0000000 0.9383578
Number of Observations: 258
Number of Groups: 43
\end{verbatim}

\( \mathcal{R} \) gives the interaction s.d. and one of the residual s.d.s in the section named \texttt{Random effects}:, whereas the ratio of the residual standard deviations is found...
under the section Variance function. In this case the interaction s.d. is 0.059556, the residual s.d. for method KL is 0.077174 and for method SL it is $0.077174 \times 0.938358 = 0.072417$. The estimated difference in means between method 1 and 2 is 0.044837, so the limits of agreement are then given by:

$$0.044883 \pm 2 \times \sqrt{2 \times 0.059556^2 + 0.077174^2 + 0.072417^2} = (-0.23, 0.32)$$

### 4.1.3 Stata

The function to use in Stata is `xtmixed`, which is only available as of Stata version 9, [5, 7]. To calculate separate residual variances for each of the methods, `xtmixed` requires generation of new variables that has a unique code for each (method $\times$ item) and each (method $\times$ item $\times$ replicate) combination. Additionally, `xtmixed` parametrizes the residual variances, as the variance for the method with the smallest residual variance and the difference in residual variances between the two methods. Therefore we must take care to use the method with the smallest residual variance as the reference. Doing it the wrong way around produces some warning messages and estimates without standard errors.

Using the `var` option produces estimates of the variance parameters and not the s.d.s. The `nocons` option is required to exclude the usual residual variation term which is no longer required (output truncated to the right):

```stata
gen meth1 = (meth == 1)
gen MI = item + 100 * meth1
gen MIR = _n
xi: xtmixed y i.meth1 i.item || MI: || MIR:meth1, nocons var
```

```

|          | Coef. | Std. Err. | z   | P>|z| | [95% Co|
|----------|-------|-----------|-----|------|---------|
| _Imeth1_1 | .0448837 | .015868 | 2.83 | 0.005 | .013782 |
| _Item_2   | -.8653287 | .0735594 | -11.76 | 0.000 | -1.00950 |
| _Item_3   | 1.132603 | .0735594 | 15.40 | 0.000 | .988429 |

Random-effects Parameters | Estimate | Std. Err. | [95% Co|
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MI: Identity</td>
<td>var(_cons)</td>
<td>.0035469</td>
<td>.0011984</td>
</tr>
<tr>
<td>MIR: Identity</td>
<td>var(meth1)</td>
<td>.0007116</td>
<td>.0012102</td>
</tr>
<tr>
<td>var(Residual)</td>
<td>.0052442</td>
<td>.0007997</td>
<td>.003889</td>
</tr>
</tbody>
</table>
```

http://www.bepress.com/ijb/vol4/iss1/16
The residual variance for method 2 is 0.0052442 and for method 1 0.0052442 + 0.0007116 = 0.0059558, and the method by item interaction variance is 0.0035469. The estimated difference in means between method 1 and method 2 is 0.0448837, so the limits of agreement for the difference between method 1 and method 2 are:

\[ 0.0448837 \pm 2 \times \sqrt{2 \times 0.0035469 + 0.0052442 + 0.0059558} = (-0.23, 0.32) \]

### 4.1.4 SAS

The procedure to use is `proc mixed[4]`, and with the generic names of the variables we use the following code to fit the model (output truncated to the right):

```sas
proc mixed data = rdata ;
   class meth item ;
   model y = meth item / s;
   random meth * item ;
   repeated item / group = meth ;
run ;
```

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm Group Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>meth*item 0.003547</td>
</tr>
<tr>
<td>item meth 1 0.005956</td>
</tr>
<tr>
<td>item meth 2 0.005244</td>
</tr>
</tbody>
</table>

Solution for Fixed Effects

| Effect   meth item Estimate Error DF  t Value |
|----------|----------|---------|-----|--------|
| Intercept 1.6277 0.05259 42 30.95 |
| meth 1 0.04488 0.01587 42 2.83 |
| meth 2 0. 0 42 . . |
| item 1 0.01703 0.07356 42 0.23 |
| item 2 -0.8483 0.07356 42 -11.53 |
| item 3 1.1496 0.07356 42 15.63 |

SAS gives the desired variance components directly as in the model formulation and also the difference between means, so the limits of agreement are:

\[ 0.04488 \pm 2 \times \sqrt{2 \times 0.003547 + 0.005956 + 0.005244} = (-0.23, 0.32) \]

Note that SAS requires considerably less fidgeting with variables than do Stata, it has a syntax that is more in line with the way models are usually specified than that of R, and it gives estimates of the parameters used in the specification of the model. No wonder that `proc mixed` has become a de facto standard for fitting variance components models!
4.2 Limits of agreement

The limits of agreement based on the mixed model are shown in the right hand panel of figure 1. These correct limits are virtually indistinguishable from those based on a random pairing of replicates within item and using these item by replicate pairings as observations. We shall return to this point below.

5 Linked replicates

In the example above, we have assumed that the replicates were exchangeable within each method by item stratum. Sometimes, however, replicates are taken in parallel by each of the methods, which means that the values are linked by a common environment; typically time or sampling occasion.

5.1 The oximetry example

An example of this is the oximetry study, done at the Royal Children’s Hospital in Melbourne to examine the agreement between pulse oximetry and co-oximetry in small babies. Many were very sick and therefore had very low oxygen saturation levels — the normal range is between 95 and 100%. Each baby was measured three times by each method; performed at three different times for each infant.

There were 61 babies in the study, of these, four had only measurements on two occasions, and one on only one occasion.

Since replicates are linked across methods we need to incorporate this in the model by including an extra random effect common within each item by replicate stratum:

\[
y_{mir} = \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir},
\]

\[
a_{ir} \sim \mathcal{N}(0, \omega^2), \quad c_{mi} \sim \mathcal{N}(0, \tau_m^2), \quad e_{mir} \sim \mathcal{N}(0, \sigma_m^2)
\]

Recall that with only two methods we cannot estimate two separate, method-specific values of \(\tau\).

Note that the variance of the extra random effect \(a_{ir}\) cannot depend on method, but in principle it could depend on item-specific features, or some of it might be taken as a fixed effect, the latter could for example include an effect of time if replicates were taken at specific times.

When subtracting measurements by the two methods the effects \(a_{ir}\) cancel, so under this extended model we have the same expression for the variance of the differences as before:

\[
\text{var}(y_{1j} - y_{2j}) = 2\tau^2 + \sigma_1^2 + \sigma_2^2
\]
so the limits of agreement are again:

\[ \hat{\alpha}_1 - \hat{\alpha}_2 \pm 2 \times \sqrt{2\hat{\tau}^2 + \hat{\sigma}_1^2 + \hat{\sigma}_2^2} \]

Model (2) differs from the previous model (1) in the estimation of the variance components. The model where the replicates are non-exchangeable within method has some of the variation allocated to the item×replicate method.

It should be noted that the model with random effects of both method×item and item×replicate is a so-called “crossed” model and therefore usually will take longer time to fit.

### 5.2 Fitting the model

In the following we briefly indicate the code to fit the model with the crossed effects of meth×item and item×repl. The full code and the output generated is shown in the appendix.

#### 5.2.1 R

The convention in the lme syntax is that when the random option is a list and the first element has the name of a variable from the dataset all the effects are nested in this. In the example below, both meth and repl are nested in item, i.e. we have meth×item and item×repl as random effects.

The R-code for fitting the model is:

```r
lme( y ~ meth + item, 
    random=list( item = pdIdent( ~ meth-1 ), 
     repl = ~1 ), 
    weights = varIdent( form = ~1 | meth ), 
    data=ox )
```

#### 5.2.2 Stata

When using Stata we need to generate a few interaction variables prior to calling xtmixed:

```stata
. gen meth1 = (meth==1)
. gen meth2 = (meth==2)
. gen MI = item + 100*meth
. gen IR = item + 100*repl
. gen MIR = _n
. xi:xtmixed y i.meth i.item || _all:R.MI || _all:R.IR ///
   || MIR:meth2, nocons var
```
5.2.3 SAS

SAS has the absolutely simplest syntax — we just need to add the desired interaction:

```sas
proc mixed data = rdata ;
  class meth item repl;
  model y = meth item / s;
  random meth * item item * repl ;
  repeated item / group = meth ;
run ;
```

5.3 Results

For the oximetry data we have the following results for the variance components, when fitting the correct model as well as the model where we (wrongly) assume exchangeable replicates:

![Bland-Altman plots](image)

**Figure 2:** The oximetry data. Left panel: Bland-Altman plot for means over replicates (gray), and paired replicates (black). The individual replicates are connected with a gray line to the mean. Right panel: Bland-Altman plot for the individual replicates. Gray limits of agreement are based on estimates from a model assuming exchangeability of replicates within methods, black limits on the correct model for the linked replicates.
We see that failure to account for the \( i \times r \) interaction only slightly underestimates the total s.d.s, \( \Sigma_1 = \sqrt{\tau^2 + \omega^2 + \sigma_i^2} \) and \( \Sigma_2 = \sqrt{\tau^2 + \omega^2 + \sigma_r^2} \), but a substantial part of it is allocated to the wrong variance component, and so produces too wide limits of agreement.

Failure to take the replication structure into account results in over-estimation of the prediction interval for the difference between future measurements. This is illustrated in figure 2, where the left panel shows the limits obtained using classical methods, and the right panel shows the limits derived from mixed effects models. The difference between limits obtained by using the linked replicates as items, and fitting the correct model is very small in this case, whereas the effect of using means strongly underestimates the limits and failing to take account of the replication structure in the models strongly overestimates the limits.

### 6 Repeatability

The limits of agreement are not always the only issue of interest — the assessment of method specific repeatability and reproducibility are of interest in their own right. Repeatability can only be assessed when replicate measurements by each method are available.

The repeatability coefficient for a method is defined as the upper limits of a prediction interval for the absolute difference between two measurements by the same method on the same item under identical circumstances. If the standard deviation of a measurement is \( \sigma \) the repeatability coefficient is \( 2 \times \sqrt{\frac{2}{3}} \sigma = 2.83 \times \sigma \approx 2.8 \sigma \).

The repeatability of measurement methods is calculated differently under the two models; under the model assuming exchangeable replicates (1), the repeatability is based only on the residual standard deviation, i.e. \( 2.8 \sigma_m \); under the model for linked replicates (2) there are two possibilities depending on the circumstances.

If the variation between replicates within item can be considered a part of the repeatability it will be \( 2.8 \sqrt{\omega^2 + \sigma_m^2} \). However, if replicates are taken under substantially different circumstances, the variance component \( \omega^2 \) may be considered irrelevant in the repeatability and one would therefore base the repeatability on the measurement errors alone, i.e. use \( 2.8 \sigma_m \). In such cases one would presumably try to model the effects of differing replication circumstances by a systematic ef-
fect. Hence there is no subject-matter-free way of defining repeatability from the variance components in the models.

In the oximetry example the measurements were taken rather close in time and hence it would be natural to include the between replicate variation in the calculation of repeatability. For co-oximetry the repeatability is \(2.8 \times \sqrt{3.42^2 + 2.22^2} = 2.8 \times 4.08 = 11.4\%\) and for pulse oximetry it is \(2.8 \times \sqrt{3.42^2 + 3.99^2} = 2.8 \times 5.25 = 14.7\%\). Hence the upper 95\% limits for the absolute difference between two repeat measurements by the two methods is 11.4 and 14.7\% respectively, where as the limits of agreement (CO—pulse) are \((-9.9; 14.8)\%\). Thus the discrepancy between the two methods is largely attributable to the rather poor repeatability of both methods.

This conclusion would clearly not have been possible without taking replicate measurements by the two methods.

Had we deemed the between replicate variation to be irrelevant, the repeatabilities would have been only \(2.8 \times 2.22 = 6.2\%\) for CO and \(2.8 \times 3.99 = 11.2\%\) for pulse; substantially smaller, but still major contributors to the width of the limits of agreement.

7 Getting it wrong and getting it almost right

In a dataset with replicate measurements there are two ways to treat the data along the lines indicated by Bland & Altman [2] which covers the situation with only one measurement per method and item:

1. Take means over replicates within each method by item stratum.

2. Replicates within item are taken as items.

Suppose that we have the following model (model 2) for the measurements:

\[
y_{mir} = \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir},
\]

\[a_{ir} \sim \mathcal{N}(0, \omega^2), \quad c_{mi} \sim \mathcal{N}(0, \tau^2_m), \quad e_{mir} \sim \mathcal{N}(0, \sigma^2_m)\] (3)

Note that we are allowing the interaction between method and item to have separate variances for each method — with only two methods these cannot be estimated separately, but they can of course still be used in calculations. The random \(i \times r\) interaction term is only relevant if the replicates are linked across methods (paired replicates).

In the model the correct limits of agreement would be:

\[
\alpha_1 - \alpha_2 \pm 2 \sqrt{\tau^2_1 + \tau^2_2 + \sigma^2_1 + \sigma^2_2}
\]
7.1 Averaging over replicates

If we are using means of replicates to form the differences we have \((R_{mi} \text{ is the number of replicates by method } m \text{ on item } i)\):

\[
\bar{d}_i = \bar{y}_{1i} - \bar{y}_{2i} = \alpha_1 - \alpha_2 + \frac{\sum_r a_{ir}}{R_{1i}} - \frac{\sum_r a_{ir}}{R_{2i}} + c_{1i} - c_{2i} + \frac{\sum_r e_{1ir}}{R_{1i}} - \frac{\sum_r e_{2ir}}{R_{2i}}
\]

The terms with \(a_{ir}\) are only relevant for linked replicates in which case \(R_{1i} = R_{2i}\) and therefore the term vanishes. Thus:

\[
\text{var}(\bar{d}_i) = \tau_1^2 + \tau_2^2 + \sigma_1^2/R_{1i} + \sigma_2^2/R_{2i} < \tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2
\]

so the limits of agreement calculated based on the means are much too narrow as prediction limits for differences between future single measurements.

7.2 Replicates as items

If replicates are taken as items, then the calculated differences are:

\[
d_{ir} = y_{1ir} - y_{2ir} = \alpha_1 - \alpha_2 + c_{1i} - c_{2i} + e_{1ir} - e_{2ir}
\]

which has variance \(\tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2\), and therefore using the empirical variance of the differences in principle gives the correct limits of agreement. However the differences are not independent:

\[
\text{cov}(d_{ir}, d_{is}) = \tau_1^2 + \tau_2^2, \quad \text{cor}(d_{ir}, d_{is}) = \frac{\tau_1^2 + \tau_2^2}{\tau_1^2 + \tau_2^2 + \sigma_1^2 + \sigma_2^2}
\]

This is negligible if the residual variances are very large compared to the interaction, so the estimate of the “correct” variance based on these differences is likely to be only slightly downwards biased.

If replicates are exchangeable within method by item strata it is not clear how to produce the differences — it can be done in a number of different ways since the replicates can be matched within item in several different ways. If replicates are paired at random, the variance will still be correct, assuming model (2) (without the \(i \times r\) interaction term)

\[
\text{var}(y_{1ir} - y_{2is}) = \tau_1^2 + \sigma_1^2 + \tau_2^2 + \sigma_2^2
\]

but again the differences will be positively correlated within item:

\[
\text{cov}(y_{1ir} - y_{2is}, y_{1it} - y_{2iu}) = \tau_1^2 + \tau_2^2
\]
so the estimate of $\tau_1^2 + \sigma_1^2 + \tau_2^2 + \sigma_2^2$ as the empirical variance of $y_{1ir} - y_{2is}$ for a random matching of replicates between methods will be an underestimate, albeit not a large one. In the fat dataset (with exchangeable replicates) the correct upper limit of agreement based on the model is 0.315, the upper limit based on the numbering in the dataset is 0.312, but the median upper limit over 1000 random matchings of replicates within items is 0.309.

8 Conclusion

Based on this, we offer the following general advice in the analysis of method comparison studies with replicate measurements:

- Do not use hand calculations — they are overly complicated and outdated in the computer age — software for mixed models was constructed for a reason.

- Set up the correct model, taking the exchangeability structure of the data into account: If replicates are linked across methods, include the item by replicate random effect, otherwise not.

- Fit the model and use the estimated parameters (and your subject-matter knowledge) to draw conclusions based on:
  - the limits of agreement between methods
  - repeatability of methods

- If you absolutely refuse to use modern statistical software, use (item $\times$ replicate) as items; if replicates are not linked, then make a random pairing. However, the correlations will bias the limits of agreement downward, and you will miss important information on the repeatability by not knowing the variance components. Your analysis will still be suboptimal, but not a totally wrong as it would be if you used averages over replicates.

Appendix: Programs

In this section we show the total results from fitting the models to the two datasets by the three packages.
The R-programs are completely self-contained since the two datasets used for illustration are part of the MethComp package. Currently (June 2008) the package is only available at www.biostat.ku.dk/bxc/MethComp.

Exchangeable replicates

```R
> library(MethComp)
> library(nlme)
> data(fat)
> fat <- data.frame( item=factor(fat$Id),
+ meth=fat$Obs,
+ repl=factor(fat$Rep),
+ y=fat$Sub )
> str(fat)
'data.frame': 258 obs. of 4 variables:
$ item: Factor w/ 43 levels "1","2","3","4",..: 1 1 1 3 3 3 5 5 5 11 ...
$ meth: Factor w/ 2 levels "KL","SL": 1 1 1 1 1 1 1 1 1 1 ...
$ repl: Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
$ y : num 1.6 1.7 1.7 2.8 2.9 2.8 2.7 2.8 2.9 3.9 ...
> lme( y ~ meth + item,
+ random = list( item = pdIdent(~meth-1) ),
+ weights = varIdent(form = ~1 | meth ),
+ data=fat)
Linear mixed-effects model fit by REML
Data: fat
Log-restricted-likelihood: 188.3488
Fixed: y ~ meth + item
 (Intercept) methKL methSL Residual
1.6896001995 -0.0448837209 -0.8653286307 1.1326030428 -1.0077856154
1.2014605811 -0.7673239282 -0.1844287691 -0.2510954358 0.6155712309
-0.5496348547 2.1282212996 -0.6750365145 1.2326030428 -0.9973239282
-0.3851590597 -0.0007302905 -0.0836984786 0.1815076070
-0.4347939144 0.2510954358 0.3163015214 0.0992697095 -1.1891236514
-0.0333333333 2.1163015214 0.3815076070 1.4666666667
-0.4666666667 -0.7991236514 0.8518257263 2.4148409403 -0.4666666667
-0.1170318119 0.2496348547 0.1481742737 -0.0170318119
Random effects:
Formula: ~meth - 1 | item
Structure: Multiple of an Identity
methKL methSL Residual
StdDev: 1.000000 0.9383578 Residual
Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | meth
Parameter estimates:
KL SL
1.000000 0.9383578
Number of Observations: 258
Number of Groups: 43
```
From the output (red entries) we get the following quantities:
\[
\alpha_{SL} - \alpha_{KL} = -0.0448837209 \\
\tau = 0.059559 \\
\sigma_{KL} = 0.07717392 \\
\sigma_{SL}/\sigma_{KL} = 0.9383578
\]

Linked replicates

```r
> library( MethComp )
Loading required package: R2WinBUGS
> library( nlme )

> data( ox )
> ox$item <- factor(ox$item)
> ox$repl <- factor(ox$repl)
> str( ox )
'data.frame': 354 obs. of 4 variables:
$ meth: Factor w/ 2 levels "CO","pulse": 1 1 1 1 1 1 1 1 1 1 ... 
$ item: Factor w/ 61 levels "1","2","3","4",..: 1 1 1 2 2 2 3 3 3 4 ... 
$ repl: Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ... 
$ y : num 78 76.4 77.2 68.7 67.6 68.3 82.9 80.1 80.7 62.3 ... 

> # The convention is that within a list in random, the termses subsequent to
> # item are nested within item
> # item are nested within item
> lme( y ~ meth + item,
+ random=list( item = pdIdent( ~ meth-1 ),
+ repl = ~1 ),
+ weights = varIdent( form = ~1 | meth ),
+ data=ox 
+
Linear mixed-effects model fit by REML
Data: ox
Log-restricted-likelihood: -911.7401
Fixed: y ~ meth + item
(Intercept) methpulse item2 item3 item4 item5
76.0428384 -2.4704462 -7.0216227 5.1497034 -10.7281860 -1.1137199
item6 item7 item8 item9 item10 item11
item12 item13 item14 item15 item16 item17
-47.3135668 3.3219575 -1.1293724 6.4653272 -0.4416475 13.9153464
item18 item19 item20 item21 item22 item23
1.5322522 -2.0861271 -1.0351969 6.2565251 -0.5367298 13.9153464
item24 item25 item26 item27 item28 item29
8.2772197 2.1049894 2.7779659 -10.3186089 -10.8197187 0.7833716
item30 item31 item32 item33 item34 item35
1.3644795 -29.2466418 5.6528703 6.8769614 8.7353406 0.9285974
item36 item37 item38 item39 item40 item41
3.0492155 3.6735649 7.5298316 2.7392939 -8.6159587 -0.1044011
item42 item43 item44 item45 item46 item47
item48 item49 item50 item51 item52 item53
-3.0309414 10.4662553 -24.8350417 -20.8508611 -0.3525354 -3.6222924
item54 item55 item56 item57 item58 item59
item60 item61
7.3963452 -1.7503731

Random effects:
  Formula: ~meth - 1 | item
  Structure: Multiple of an Identity
  methCO methpulse
StdDev: 2.928042 2.928042

  Formula: ~1 | repl %in% item
  (Intercept) Residual
StdDev: 3.415692 2.224868
```
Variance function:
Structure: Different standard deviations per stratum
Formula: ~1 | meth
Parameter estimates:
CO pulse
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CO pulse</td>
<td>1.000000</td>
</tr>
<tr>
<td></td>
<td>1.795365</td>
</tr>
</tbody>
</table>
Number of Observations: 354
Number of Groups:
  item repl %in% item
  61 177

From the output (red entries) we get the following quantities:

\[
\alpha_{\text{pulse}} - \alpha_{\text{CO}} = -2.4704462 \\
\tau = 2.928042 \\
\omega = 3.415692 \\
\sigma_{\text{CO}} = 2.224868 \\
\sigma_{\text{pulse}}/\sigma_{\text{CO}} = 1.796365
\]

**Stata**

**Exchangeable replicates**

. ** Indicator variable for methods
. ** (for the method with the largest residual variance)
  gen meth1 = ( meth == 1 )
. ** Interaction variable for method*item
  gen MI = item + 100 * meth1
. ** Generate a variable with a unique code for each method*item*replicate combination
  gen MIR = _n
. ** Linear mixed effects modelling
  xi: xtmixed y i.meth1 i.item || MI: || MIR: meth1, nocons var

Performing EM optimization:

Performing gradient-based optimization:

Iteration 0:  log restricted-likelihood = 185.333
Iteration 1:  log restricted-likelihood = 188.27598
Iteration 2:  log restricted-likelihood = 188.34852
Iteration 3:  log restricted-likelihood = 188.34884
Iteration 4:  log restricted-likelihood = 188.34884

Computing standard errors:

Mixed-effects REML regression  Number of obs = 258

<table>
<thead>
<tr>
<th>Group Variable</th>
<th>No. of Observations per Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Groups Minimum Average Maximum</td>
</tr>
<tr>
<td>MI</td>
<td>86 3 3.0 3</td>
</tr>
<tr>
<td>MIR</td>
<td>258 1 1.0 1</td>
</tr>
</tbody>
</table>

Wald chi2(43) = 11799.40
Log restricted-likelihood = 188.34884
Prob > chi2 = 0.0000
### Coefficients and Standard Errors

| Item     | Coef.   | Std. Err. | z       | P>|z|   | [95% Conf. Interval] |
|----------|---------|-----------|---------|-------|----------------------|
| _Iitem_1 | .0448837| .015868   | 2.83    | .005  | .0137829 .0759845    |
| _Iitem_2 | -.8653287| .0735594 | -11.76  | .000  | -1.009502 -.721154   |
| _Iitem_3 | 1.132603 | .0735594 | 15.40   | .000  | .9884293 1.276777    |
| _Iitem_4 | -1.007786| .0735594 | -13.70  | .000  | -1.151959 -.863611   |
| _Iitem_5 | 1.201461 | .0735594 | 16.33   | .000  | 1.057287 1.345634    |
| _Iitem_6 | -.7673239| .0735594 | -10.43  | .000  | -.914977  -.6231502  |
| _Iitem_7 | -.2510955| .0735594 | -3.41   | .001  | -.3952692 -.1069217  |
| _Iitem_8 | .6155712 | .0735594 | 8.37    | .000  | .4713975  .759745    |
| _Iitem_9 | -.5496349| .0735594 | -7.47   | .000  | -.6938086 -.4054611  |
| _Iitem_10| 2.128221 | .0735594 | 28.93   | .000  | 1.984048 2.272395    |
| _Iitem_11| -.6750365| .0735594 | -9.18   | .000  | -.8192103 -.5308628  |
| _Iitem_13| 1.232603 | .0735594 | 16.76   | .000  | 1.088429 1.376777    |
| _Iitem_15| -.9973239| .0735594 | -13.56  | .000  | -1.141498 -.8531502  |
| _Iitem_17| -.3851591| .0735594 | -10.43  | .000  | -.5293328 -.6231502  |
| _Iitem_18| .0496349 | .0735594 | 0.67    | .500  | -.0945389 .1938086  |
| _Iitem_20| .1815076 | .0735594 | 2.47    | .014  | .0373338 .3256814    |
| _Iitem_21| -.0007303| .0735594 | -0.01   | .992  | -.144904  .1434435   |
| _Iitem_22| -.0844288| .0735594 | -1.15   | .251  | -.2286025 .059745    |
| _Iitem_23| -.0836985| .0735594 | -1.14   | .255  | -.2278723 .059745    |
| _Iitem_24| .1815076 | .0735594 | 2.47    | .014  | .0373338 .3256814    |
| _Iitem_25| -.8170318| .0735594 | -11.11  | .000  | -.9612055 1.176047   |
| _Iitem_26| .3170317 | .0735594 | 4.31    | .000  | .172858  .4612055    |
| _Iitem_27| .0496349 | .0735594 | 0.67    | .500  | -.0945389 .1938086  |
| _Iitem_28| .3163015 | .0735594 | 4.30    | .000  | .172858  .4612055    |
| _Iitem_29| .9318727 | .0735594 | 12.67   | .000  | .787699  1.076047    |
| _Iitem_30| .3163015 | .0735594 | 4.30    | .000  | .172858  .4612055    |
| _Iitem_31| .0992697 | .0735594 | 1.35    | .177  | -.0449041 .2434435   |
| _Iitem_32| 2.116302 | .0735594 | 28.77   | .000  | 1.972128 2.260475    |
| _Iitem_33| .3815076 | .0735594 | 5.19    | .000  | .2373338 0.5256814   |
| _Iitem_37| 1.466667 | .0735594 | 19.94   | .000  | 1.322493 1.61084     |
| _Iitem_38| -.0170318| .0735594 | -0.23   | .817  | -.1612056 .1271419   |

### Random-effects Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>_Meth1_1</td>
<td>.0035469</td>
<td>.0011984</td>
<td>.0018291 .0068779</td>
</tr>
<tr>
<td>_MIR: _Meth1_1</td>
<td>.0007116</td>
<td>.0012102</td>
<td>.0000254 .0199439</td>
</tr>
<tr>
<td>_Residual</td>
<td>.0052442</td>
<td>.0007997</td>
<td>.0038893 .0070711</td>
</tr>
</tbody>
</table>

**Note:** LR test is conservative and provided only for reference.
From the output (red entries) we get the following quantities:

\[
\alpha_{KL} - \alpha_{SL} = 0.0448837 \\
\tau^2 = 0.0035469 \\
\sigma^2_{SL} = 0.0052442 \\
\sigma^2_{KL} - \sigma^2_{SL} = 0.0007116
\]

**Linked replicates**

- **Indicator variables for methods**
  - gen meth1 = (meth==1)
  - gen meth2 = (meth==2)

- **Interaction variables for method*item and item*replicate**
  - gen MI = item + 100*meth
  - gen IR = item + 100*rep1

- **Generate a variable with a unique code for each method*item*replicate combination**
  - gen MIR = _n

- **Model with random effects for method*item and replicate*item**
  - xi:xtmixed y i.meth i.item || _all:R.MI || _all:R.IR ///
    > | || MIR:meth2, nocons var
  - Performing EM optimization:
  - Performing gradient-based optimization:
  - Computing standard errors:

Mixed-effects REML regression

<table>
<thead>
<tr>
<th>Group Variable</th>
<th>No. of Groups</th>
<th>Minimum</th>
<th>Average</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>_all</td>
<td>1</td>
<td>354.0</td>
<td>354.0</td>
<td>354</td>
</tr>
<tr>
<td>MIR</td>
<td>354</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Log restricted-likelihood = -911.74012

<p>| y | Coef.  | Std. Err. | z     | P&gt;|z|  | [95% Conf. Interval] |
|---|--------|-----------|------|------|-----------------|
| Item_2 | -1.021622 | 4.422289 | -0.23 | 0.813 | -9.781245 7.553808 |
| Item_3 | 5.149703  | 4.422289 | 1.16  | 0.244 | 3.517823 13.81723 |
| Item_4 | 10.72819  | 4.422289 | 2.43  | 0.015 | 19.39571 2.60659 |
| Item_5 | -1.113719 | 4.422289 | -0.25 | 0.801 | 9.781245 7.553808 |
| Item_6 | 3.164994  | 4.422289 | 0.72  | 0.474 | 5.502532 11.83252 |
| Item_7 | 9.706565  | 4.422289 | 2.19  | 0.029 | 10.039039 18.37409 |
| Item_8 | 3.056866  | 4.422289 | 0.80  | 0.421 | 5.110666 12.22439 |
| Item_9 | -4.182137 | 4.422289 | -0.95 | 0.344 | 12.84966 4.48539 |
| Item_10| -14.42224 | 4.422289 | -3.26 | 0.001 | 23.08977 -5.754717 |
| Item_11| 12.75037  | 4.422289 | 2.88  | 0.004 | 4.082848 21.4179  |</p>
<table>
<thead>
<tr>
<th>Item</th>
<th>Estimate</th>
<th>Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>_Item_12</td>
<td>-47.31357</td>
<td>4.42229</td>
<td>(-55.98109, -38.64604)</td>
</tr>
<tr>
<td>_Item_13</td>
<td>-1.129371</td>
<td>4.42229</td>
<td>(-9.79698, 7.538155)</td>
</tr>
<tr>
<td>_Item_14</td>
<td>-1.906526</td>
<td>4.42229</td>
<td>(-9.204257, 5.338155)</td>
</tr>
<tr>
<td>_Item_15</td>
<td>-5.536731</td>
<td>4.42229</td>
<td>(-9.204257, 8.130795)</td>
</tr>
<tr>
<td>_Item_16</td>
<td>13.91535</td>
<td>4.715682</td>
<td>(7.538155, 20.292405)</td>
</tr>
<tr>
<td>_Item_17</td>
<td>4.582033</td>
<td>4.42229</td>
<td>(1.906526, 7.25754)</td>
</tr>
<tr>
<td>_Item_18</td>
<td>-2.086126</td>
<td>4.42229</td>
<td>(-4.532629, 0.359377)</td>
</tr>
<tr>
<td>_Item_19</td>
<td>-1.035196</td>
<td>4.715682</td>
<td>(-5.22378, 3.153488)</td>
</tr>
<tr>
<td>_Item_20</td>
<td>6.465328</td>
<td>4.42229</td>
<td>(1.906526, 10.92405)</td>
</tr>
<tr>
<td>_Item_21</td>
<td>-0.536731</td>
<td>4.42229</td>
<td>(-4.532629, 3.458155)</td>
</tr>
<tr>
<td>_Item_22</td>
<td>13.91535</td>
<td>4.715682</td>
<td>(7.538155, 20.292405)</td>
</tr>
<tr>
<td>_Item_23</td>
<td>1.532253</td>
<td>4.42229</td>
<td>(0.359377, 2.705126)</td>
</tr>
<tr>
<td>_Item_24</td>
<td>-2.086126</td>
<td>4.42229</td>
<td>(-4.532629, 0.359377)</td>
</tr>
<tr>
<td>_Item_25</td>
<td>6.465328</td>
<td>4.42229</td>
<td>(1.906526, 10.92405)</td>
</tr>
<tr>
<td>_Item_26</td>
<td>-0.536731</td>
<td>4.42229</td>
<td>(-4.532629, 3.458155)</td>
</tr>
<tr>
<td>_Item_27</td>
<td>4.582033</td>
<td>4.42229</td>
<td>(0.359377, 8.744686)</td>
</tr>
<tr>
<td>_Item_28</td>
<td>8.27722</td>
<td>4.42229</td>
<td>(3.705126, 12.84438)</td>
</tr>
<tr>
<td>_Item_29</td>
<td>2.104989</td>
<td>4.715682</td>
<td>(0.450578, 3.7594)</td>
</tr>
<tr>
<td>_Item_30</td>
<td>2.777965</td>
<td>4.42229</td>
<td>(0.6368, 4.909126)</td>
</tr>
<tr>
<td>_Item_31</td>
<td>-10.31861</td>
<td>4.42229</td>
<td>(-18.98613, -1.651082)</td>
</tr>
<tr>
<td>_Item_32</td>
<td>-10.81972</td>
<td>4.42229</td>
<td>(-19.48724, -2.152191)</td>
</tr>
<tr>
<td>_Item_33</td>
<td>0.7833705</td>
<td>4.42229</td>
<td>(-7.884156, 9.450897)</td>
</tr>
<tr>
<td>_Item_34</td>
<td>2.64448</td>
<td>4.42229</td>
<td>(0.60, 4.68856)</td>
</tr>
<tr>
<td>_Item_35</td>
<td>-29.24664</td>
<td>4.42229</td>
<td>(-37.91417, -20.57911)</td>
</tr>
<tr>
<td>_Item_36</td>
<td>10.46626</td>
<td>4.42229</td>
<td>(2.37, 18.56456)</td>
</tr>
<tr>
<td>_Item_37</td>
<td>-24.83504</td>
<td>4.715682</td>
<td>(-34.07761, -15.59248)</td>
</tr>
<tr>
<td>_Item_38</td>
<td>-20.85086</td>
<td>4.42229</td>
<td>(-29.51839, -12.18333)</td>
</tr>
<tr>
<td>_Item_39</td>
<td>-0.3525351</td>
<td>4.42229</td>
<td>(-9.020062, 8.314991)</td>
</tr>
<tr>
<td>_Item_40</td>
<td>-3.622292</td>
<td>4.42229</td>
<td>(-11.69847, 5.63586)</td>
</tr>
<tr>
<td>_Item_41</td>
<td>1.42991</td>
<td>4.42229</td>
<td>(-7.237617, 10.09744)</td>
</tr>
<tr>
<td>_Item_42</td>
<td>12.83856</td>
<td>4.42229</td>
<td>(2.90, 22.76456)</td>
</tr>
<tr>
<td>_Item_43</td>
<td>9.797168</td>
<td>4.42229</td>
<td>(2.22, 17.3524)</td>
</tr>
<tr>
<td>_Item_44</td>
<td>13.35012</td>
<td>4.42229</td>
<td>(3.02, 23.66024)</td>
</tr>
<tr>
<td>_Item_45</td>
<td>13.35012</td>
<td>4.42229</td>
<td>(3.02, 23.66024)</td>
</tr>
<tr>
<td>_Item_46</td>
<td>15.66574</td>
<td>4.42229</td>
<td>(3.54, 27.78674)</td>
</tr>
<tr>
<td>_Item_47</td>
<td>7.396344</td>
<td>4.42229</td>
<td>(1.67, 13.10212)</td>
</tr>
<tr>
<td>_Item_48</td>
<td>-1.750373</td>
<td>4.42229</td>
<td>(-4.0, 0.4925)</td>
</tr>
<tr>
<td>_cons</td>
<td>76.04284</td>
<td>3.138534</td>
<td>(24.23, 128.292405)</td>
</tr>
</tbody>
</table>

Random-effects Parameters | Estimate | Std. Err. | [95% Conf. Interval] |
---------------------------|----------|----------|----------------------|
_all: Identity | var(R.MI) | 8.573426 | 2.25398 | (5.12119, 14.35284) |
_all: Identity | var(R.IR) | 11.66695 | 2.263471 | (7.976607, 17.06462) |
MIR: Identity | var(meth2) | 11.00559 | 3.624397 | (5.771552, 20.9862) |
MIR: Identity | var(Residual) | 4.950042 | 1.784803 | (2.441726, 10.03508) |

 cons | 76.04284 | 3.138534 | (24.23, 128.292405) |
LR test vs. linear regression: $\chi^2(3) = 55.24$  Prob > $\chi^2 = 0.0000$

Note: LR test is conservative and provided only for reference

From the output (red entries) we get the following quantities:

\[
\begin{align*}
\alpha_{\text{pulse}} - \alpha_{\text{CO}} &= -2.470446 \\
\tau^2 &= 8.573426 \\
\omega^2 &= 11.66695 \\
\sigma^2_{\text{CO}} &= 4.950042 \\
\sigma^2_{\text{pulse}} - \sigma^2_{\text{CO}} &= 11.00559
\end{align*}
\]

SAS

Exchangeable replicates

```
proc mixed data = rdata ;
class meth item ;
model y = meth item / s;
random meth * item ;
repeated item / group = meth ;
run ;
```

NOTE: Convergence criteria met.
NOTE: The PROCEDURE MIXED printed pages 1-2.
NOTE: PROCEDURE MIXED used (Total process time):
real time 3.75 seconds
cpu time 1.52 seconds

The Mixed Procedure

Model Information

Data Set WORK.RDATA
Dependent Variable y
Covariance Structure Variance Components
Group Effect meth
Estimation Method REML
Residual Variance Method None
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Containment

Class Level Information

Class    Levels    Values
meth     2         KL SL
item     43        1 2 3 4 5 6 7 8 9 10 11 13 14
          15 16 17 18 19 20 21 22 24 25
          27 28 29 30 31 32 33 34 35 36
          37 38 39 40 41 42 43 44 45 46

Dimensions

Covariance Parameters 3
Columns in X 46
Columns in Z 86
Subjects 1
Max Obs Per Subject 258

Number of Observations
Number of Observations Read 258  
Number of Observations Used 258  
Number of Observations Not Used 0  

Iteration History  
Iteration  Evaluations  -2 Res Log Like  Criterion  
0  1  -353.24387418  
1  1  -376.69765836  0.00000000  

Convergence criteria met.  

Covariance Parameter Estimates  
Cov Parm Group Estimate  
meth*item 0.003547  
meth KL meth 0.005956  
meth SL meth 0.005244  

Fit Statistics  
-2 Res Log Likelihood -376.7  
AIC (smaller is better) -370.7  
AICC (smaller is better) -370.6  
BIC (smaller is better) -363.3  

Solution for Fixed Effects  
Effect meth item Estimate Standard Error DF t Value Pr > |t|  
Intercept 1.6277 0.05259 42 30.95 <.0001  
meth KL 0.04488 0.01587 42 2.83 0.0071  
meth SL 0 0 22  
item 1 0.01703 0.07356 42 0.23 0.8180  
item 2 -0.8481 0.07356 42 -11.53 <.0001  
item 3 1.1496 0.07356 42 15.63 <.0001  
item 4 -0.9908 0.07356 42 -13.47 <.0001  
item 5 1.2185 0.07356 42 16.56 <.0001  
item 6 -0.7503 0.07356 42 -10.20 <.0001  
item 7 -0.1674 0.07356 42 -2.28 0.0280  
item 8 -0.2341 0.07356 42 -3.18 0.0028  
item 9 0.6326 0.07356 42 8.60 <.0001  
item 10 -0.5326 0.07356 42 -7.24 <.0001  
item 11 2.1453 0.07356 42 29.16 <.0001  
item 12 -0.6580 0.07356 42 -8.95 <.0001  
item 13 1.2496 0.07356 42 16.99 <.0001  
item 14 -0.9803 0.07356 42 -13.33 <.0001  
item 15 -0.3681 0.07356 42 -5.00 <.0001  
item 16 0.01630 0.07356 42 0.22 0.8257  
item 17 -0.06740 0.07356 42 -0.92 0.3648  
item 18 0.1985 0.07356 42 2.70 0.0100  
item 19 -0.06667 0.07356 42 -0.91 0.3699  
item 20 0.2681 0.07356 42 3.65 0.0007  
item 21 0.3341 0.07356 42 4.54 <.0001  
item 22 0.3333 0.07356 42 4.53 <.0001  
item 23 -0.4333 0.07356 42 -5.89 <.0001  
item 24 0.01630 0.07356 42 0.22 0.8257  
item 25 2.1333 0.07356 42 29.00 <.0001  
item 26 -1.1721 0.07356 42 -15.93 <.0001  
item 27 -0.01630 0.07356 42 -0.22 0.8257  
item 28 2.1333 0.07356 42 29.00 <.0001  
item 29 0.8341 0.07356 42 11.34 <.0001  
item 30 0.3985 0.07356 42 5.42 <.0001  
item 31 1.4837 0.07356 42 20.17 <.0001  
item 32 -0.4496 0.07356 42 -6.11 <.0001
From the output (red entries) we get the following quantities:

\[ \alpha_{KL} - \alpha_{SL} = 0.04488 \]

\[ \tau^2 = 0.003547 \]

\[ \sigma^2_{KL} = 0.005956 \]

\[ \sigma^2_{SL} = 0.005244 \]

**Linked replicates**

```plaintext
20 proc mixed data = rdata ;
21 class meth item repl ;
22   model y = meth item / s ;
23   random meth *item item *repl ;
24   repeated item / group = meth ;
25 run ;
```

NOTE: Convergence criteria met.
NOTE: The PROCEDURE MIXED printed pages 1-2.
NOTE: PROCEDURE MIXED used (Total process time):
      real time 3:22.36
      cpu time 2:51.92

The Mixed Procedure

Model Information

Data Set WORK.RDATA
Dependent Variable y
Covariance Structure Variance Components
Group Effect meth
Estimation Method REML
Residual Variance Method None
Fixed Effects SE Method Model-Based
Degrees of Freedom Method Containment

Class Level Information

Class Levels Values
meth 2 CO pu
item 61 1 2 3 4 5 6 7 8 9 10 11 12 13
   14 15 16 17 18 19 20 21 22 23
   24 25 26 27 28 29 30 31 32 33
   34 35 36 37 38 39 40 41 42 43
   44 45 46 47 48 49 50 51 52 53
   54 55 56 57 58 59 60 61
repl 3 1 2 3

Dimensions

Covariance Parameters 4
Columns in X 64
Columns in Z 299
Subjects 1

Published by The Berkeley Electronic Press, 2008
Max Obs Per Subject 354

Number of Observations

| Number of Observations Read | 354 |
| Number of Observations Used | 354 |
| Number of Observations Not Used | 0 |

Iteration History

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Evaluations</th>
<th>-2 Res Log Like</th>
<th>Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1878.72378376</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1823.48059503</td>
<td>0.00000054</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1823.48033506</td>
<td>0.00000014</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1823.48031459</td>
<td>0.00000010</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1823.48024763</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>

Convergence criteria met.

Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Cov Parm</th>
<th>Group</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>meth*item</td>
<td></td>
<td>8.5734</td>
</tr>
<tr>
<td>item*repl</td>
<td></td>
<td>11.6670</td>
</tr>
<tr>
<td>item</td>
<td>meth CO</td>
<td>4.9500</td>
</tr>
<tr>
<td>item</td>
<td>meth pu</td>
<td>15.9556</td>
</tr>
</tbody>
</table>

Fit Statistics

-2 Res Log Likelihood 1823.5
AIC (smaller is better) 1831.5
AICC (smaller is better) 1831.6
BIC (smaller is better) 1842.7

Solution for Fixed Effects

<p>| Effect | meth | item | Estimate | Standard Error | DF | t Value | Pr &gt; |t| |
|--------|------|------|----------|----------------|----|---------|-------|
| Intercept | CO |     | 71.8220  | 3.1482         | 60  | 22.81   | &lt;.0001 |
| meth |     | CO |         | 2.4704        | 60  | 3.90    | 0.0002 |
| item | 1   |     | 1.7504  | 4.4223         | 60  | 0.40    | 0.6937 |
| item | 2   |     | -5.2713 | 4.4223         | 60  | -1.19   | 0.2380 |
| item | 3   |     | 6.9001  | 4.4223         | 60  | 1.56    | 0.1239 |
| item | 4   |     | -8.9778 | 4.4223         | 60  | -2.03   | 0.0468 |
| item | 5   |     | 0.6367  | 4.4223         | 60  | 0.14    | 0.8860 |
| item | 6   |     | 4.9154  | 4.4223         | 60  | 1.11    | 0.2708 |
| item | 7   |     | 11.4569 | 4.4223         | 60  | 2.59    | 0.0120 |
| item | 8   |     | 5.3072  | 4.4223         | 60  | 1.20    | 0.2348 |
| item | 9   |     | -2.4318 | 4.4223         | 60  | -0.55   | 0.5844 |
| item | 10  |     | -12.6719| 4.4223         | 60  | -2.87   | 0.0057 |
| item | 11  |     | 14.5007 | 4.4223         | 60  | 3.28    | 0.0017 |
| item | 12  |     | -45.5632| 4.4223         | 60  | -10.30  | &lt;.0001 |
| item | 13  |     | 5.0723  | 4.4223         | 60  | 1.15    | 0.2559 |
| item | 14  |     | 0.6210  | 4.4223         | 60  | 0.14    | 0.8888 |
| item | 15  |     | 8.0069  | 4.4223         | 60  | 1.81    | 0.0752 |
| item | 16  |     | 1.2136  | 4.4223         | 60  | 0.27    | 0.7847 |
| item | 17  |     | 15.6657 | 4.7157         | 60  | 3.32    | 0.0015 |
| item | 18  |     | 3.2826  | 4.4223         | 60  | 0.74    | 0.4608 |
| item | 19  |     | -0.3358 | 4.4223         | 60  | -0.08   | 0.9397 |
| item | 20  |     | 0.7152  | 4.7157         | 60  | 0.15    | 0.8800 |
| item | 21  |     | 8.2157  | 4.4223         | 60  | 1.86    | 0.0681 |
| item | 22  |     | 1.3087  | 4.4223         | 60  | 0.30    | 0.7683 |
| item | 23  |     | 6.3324  | 4.4223         | 60  | 1.43    | 0.1574 |
| item | 24  |     | 10.0276 | 4.4223         | 60  | 2.27    | 0.0270 |
| item | 25  |     | 3.8554  | 4.7157         | 60  | 0.82    | 0.4168 |
| item | 26  |     | 4.5283  | 4.4223         | 60  | 1.02    | 0.3100 |</p>
<table>
<thead>
<tr>
<th>item</th>
<th>value1</th>
<th>value2</th>
<th>value3</th>
<th>value4</th>
<th>value5</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>-8.5682</td>
<td>4.4223</td>
<td>60</td>
<td>-1.94</td>
<td>0.0574</td>
</tr>
<tr>
<td>28</td>
<td>-9.0693</td>
<td>4.4223</td>
<td>60</td>
<td>-2.05</td>
<td>0.0447</td>
</tr>
<tr>
<td>29</td>
<td>2.5337</td>
<td>4.4223</td>
<td>60</td>
<td>0.57</td>
<td>0.5688</td>
</tr>
<tr>
<td>30</td>
<td>4.3949</td>
<td>4.4223</td>
<td>60</td>
<td>0.99</td>
<td>0.3243</td>
</tr>
<tr>
<td>31</td>
<td>-27.4963</td>
<td>4.4223</td>
<td>60</td>
<td>-6.22</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>32</td>
<td>-1.4032</td>
<td>4.4223</td>
<td>60</td>
<td>1.67</td>
<td>0.0993</td>
</tr>
<tr>
<td>33</td>
<td>8.6273</td>
<td>4.4223</td>
<td>60</td>
<td>1.09</td>
<td>0.2821</td>
</tr>
<tr>
<td>34</td>
<td>4.7996</td>
<td>4.4223</td>
<td>60</td>
<td>0.61</td>
<td>0.5469</td>
</tr>
<tr>
<td>35</td>
<td>9.2802</td>
<td>4.4223</td>
<td>60</td>
<td>0.99</td>
<td>0.3243</td>
</tr>
<tr>
<td>36</td>
<td>1.6460</td>
<td>4.4223</td>
<td>60</td>
<td>0.37</td>
<td>0.7111</td>
</tr>
<tr>
<td>37</td>
<td>-18.9965</td>
<td>4.4223</td>
<td>60</td>
<td>-4.30</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>38</td>
<td>-14.5440</td>
<td>4.4223</td>
<td>60</td>
<td>-3.29</td>
<td>0.0017</td>
</tr>
<tr>
<td>39</td>
<td>3.7834</td>
<td>4.4223</td>
<td>60</td>
<td>0.86</td>
<td>0.3957</td>
</tr>
<tr>
<td>40</td>
<td>6.2634</td>
<td>4.4223</td>
<td>60</td>
<td>1.42</td>
<td>0.1618</td>
</tr>
<tr>
<td>41</td>
<td>5.1758</td>
<td>4.4223</td>
<td>60</td>
<td>1.17</td>
<td>0.2465</td>
</tr>
<tr>
<td>42</td>
<td>-1.2806</td>
<td>4.4223</td>
<td>60</td>
<td>-0.29</td>
<td>0.7731</td>
</tr>
<tr>
<td>43</td>
<td>-23.0847</td>
<td>4.4223</td>
<td>60</td>
<td>-4.90</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>44</td>
<td>-19.1005</td>
<td>4.4232</td>
<td>60</td>
<td>-4.32</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>45</td>
<td>1.3978</td>
<td>4.4223</td>
<td>60</td>
<td>0.32</td>
<td>0.7530</td>
</tr>
<tr>
<td>46</td>
<td>-1.9719</td>
<td>4.4223</td>
<td>60</td>
<td>-0.42</td>
<td>0.6736</td>
</tr>
<tr>
<td>47</td>
<td>3.1803</td>
<td>4.4223</td>
<td>60</td>
<td>0.72</td>
<td>0.4748</td>
</tr>
<tr>
<td>48</td>
<td>14.5889</td>
<td>4.4223</td>
<td>60</td>
<td>3.30</td>
<td>0.0016</td>
</tr>
<tr>
<td>49</td>
<td>11.5475</td>
<td>4.4223</td>
<td>60</td>
<td>2.61</td>
<td>0.0114</td>
</tr>
<tr>
<td>50</td>
<td>15.1005</td>
<td>4.4223</td>
<td>60</td>
<td>3.41</td>
<td>0.0012</td>
</tr>
<tr>
<td>51</td>
<td>15.2457</td>
<td>4.4223</td>
<td>60</td>
<td>3.45</td>
<td>0.0010</td>
</tr>
<tr>
<td>52</td>
<td>17.4161</td>
<td>4.4223</td>
<td>60</td>
<td>3.94</td>
<td>0.0002</td>
</tr>
<tr>
<td>53</td>
<td>9.1467</td>
<td>4.4223</td>
<td>60</td>
<td>2.07</td>
<td>0.0429</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

From the output (red entries) we get the following quantities:

\[
\alpha_{CO} - \alpha_{pulse} = 2.4704 \\
\tau^2 = 8.5734 \\
\omega^2 = 11.6670 \\
\sigma^2_{CO} = 4.9500 \\
\sigma^2_{pulse} = 15.9556
\]

References


