Comparing Clinical Methods of Measurement

or:

Statistical Analysis of Method Comparison Studies

Course at Centre for Clinical Research, Haukeland University Hospital, Bergen, by

Bendix Carstensen

Senior Statistician, Steno Diabetes Center
Niels Steensens Vej 2, DK-2820 Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bxc@steno.dk http://BendixCarstensen.com/MethComp/Courses/Bergen.2014

Attendance

You should attend this course if you ever have come across a clinician that asked you which of two methods are the better for measuring blood-pressure (say). Or a researcher trying to convince you (and the rest of the world) that a correlation of 0.93 is compelling evidence that two methods are in agreement.

You will learn how to provide them with what they need rather than what they want. You will work through a number of realistic examples on your own computer.

What it's about

The comparison of two methods of measurement is normally done by measuring a number of samples by two different methods. The data will therefore be pairs of measurements, one pair for each sample. This type of data is most frequently analyzed using the so-called "Bland-Altman" procedure of plotting the difference against the mean for each pair of observations. This has become the *de facto* standard for analysis of method comparison studies without replicates [1, 2]. But still some papers appear that rely on correlations in some form or another.

The Bland-Altman procedure produces limits of agreement, that is prediction limits for the difference between a measurement by one method and a measurement by the other. The Bland-Altman approach has been expanded with practical methods for the analysis of situations with replicate measurements for multiple methods [5].

The course will give an introduction to the standard procedures, and put these in a proper modeling framework. More elaborate designs of method comparison studies, in particular studies with replicate measurements by each method will be covered. Studies with more than two methods and with non-constant, and even non-linear relationships between methods will be covered too.

As a bonus you will also learn why you should only compute the coefficient of variation from log-transformed data.

I will illustrate the methods by examples from the literature, and demonstrate the use of the `MethComp` package for R, and you will be subjected to hands-on exercises using the `MethComp` package on your own computer.

If desired and feasible, analysis of your own data would be possible too.

Details of the subject

If the the traditional Bland-Altman plot is rotated 45 degrees we get a plot of measurements by one method versus those of the other, with a line that allows conversion from measurement by one method to what would have been the measurement by the other, with a prediction interval. And this plot can be used both ways [4, 3].

The core of the approach I will present is a proper statistical model for the observations, which makes it easier to adequately address the parameters of interest (and identify those *not* of interest), as well as to assess the assumptions underlying the model.

If the comparison experiment has been properly conducted with replicate measurements, the standard approach by Bland and Altman is a special case of standard random effects models. The model to use depends on the structure of the replication scheme, *i.e.* whether the replicate measurements are exchangeable or linked.

The general model for measurements y_{mir} by method m on individual i, replication r is one where we assume a "true" value μ_i for the ith individual and a method-specific bias α_m and three components of variance ω, τ_m and σ_m:

$$\begin{aligned} y_{mir} &= \alpha_m + \mu_i + a_{ir} + c_{mi} + e_{mir} \\ \text{s.d.}(a_{ir}) &= \omega, \quad \text{s.d.}(c_{mi}) = \tau_m, \quad \text{s.d.}(e_{mir}) = \sigma_m \end{aligned}$$

I will show how this model can be analyzed and reported in the form of a plot converting from one method to another. Basically, the parameters of interest are the variance components τ and σ, whereas the variance component ω and the individual means μ_i are nuisance parameters, and depending on the design of the experiment, alien to the comparison of the methods.

If basic assumptions such as constant difference between methods or homoschedastic errors are not met, transformation of measurements may remedy this. However, only the log-transform provides results which are interpretable on the transformed scale.

However for any transformation of measurements it is possible to do the correct analysis on the transformed scale, and subsequently back-transform resulting limits of agreement etc. to the original measurement scale, providing a direct conversion between methods enabling conversion with relevant prediction limits both ways.

Even if transformation can provide a scale with constant variance it may still be so that the difference between the measurement methods is not constant but linear:

$$y_{mir} = \alpha_m + \beta_m(\mu_i + a_{ir} + c_{mi}) + e_{mir}$$

It is possible to estimate in this model allowing for linear relationships between methods, but it is no longer a standard mixed-effects model, and requires either an extra layer of iteration or brute force such as MCMC estimation.

I shall outline how the latter is implemented in the `MethComp` package via a link to either `OpenBUGS`[6] or `JAGS`[7]. Participants will be given instructions on what to install and how to check the installation before the course.

References

[1] JM Bland and DG Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, i:307–310, 1986.

[2] JM Bland and DG Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8:136–160, 1999.

[3] B. Carstensen. *Comparing Clinical Measurement Methods: A practical guide*. Wiley, 2010.

[4] B. Carstensen. Comparing methods of measurement: Extending the LoA by regression. *Stat Med*, 29:401–410, Feb 2010.

[5] B Carstensen, J Simpson, and LC Gurrin. Statistical models for assessing agreement in method comparison studies with replicate measurements. *International Journal of Biostatistics*, 4(1):Article 16, 2008.

[6] D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Stat Med*, 28(25):3049–3067, Nov 2009.

[7] Martyn Plummer. Jags: A program for analysis of bayesian graphical models using gibbs sampling.