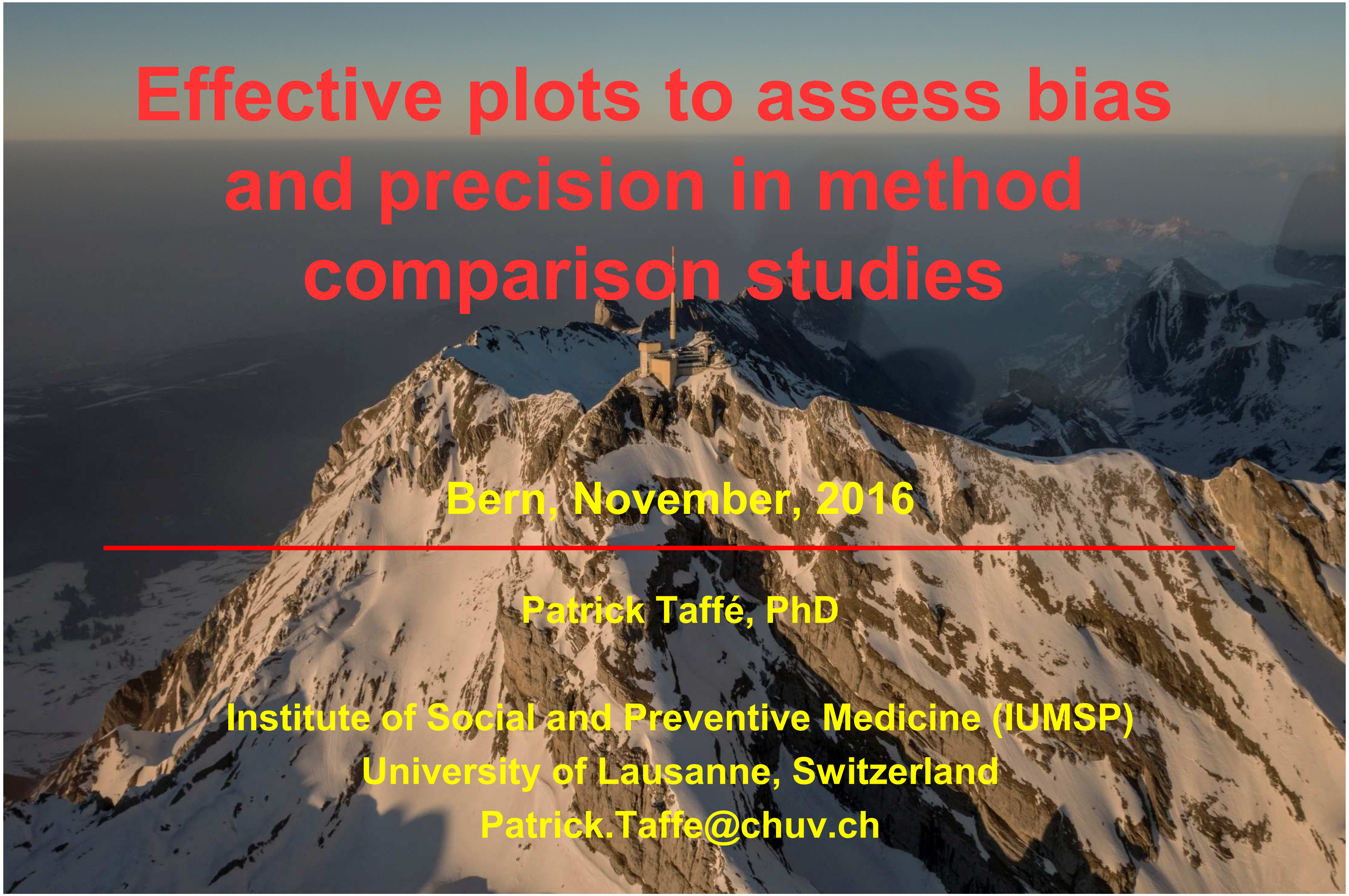


Effective plots to assess bias and precision in method comparison studies

Bern, November, 2016

Patrick Taffé, PhD

Institute of Social and Preventive Medicine (IUMSP)
University of Lausanne, Switzerland
Patrick.Taffe@chuv.ch

Outline

- Bland & Altman’s limits of agreement method (1986)
- Extension to proportional bias and heteroscedasticity (1999)
- A new methodology to quantify bias and precision
- Illustration with a simulated example

How to measure agreement between two measurement methods ?

Ex: blood pressure

STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT

J. Martin Bland, Douglas G. Altman
Department of Clinical Epidemiology and Social Medicine, St. George's Hospital Medical School, London SW17 ORE; and Division of Medical Statistics, MRC Clinical Research Centre, Northwick Park Hospital, Harrow, Middlesex

(*Lancet*, 1986; **i:** 307-310)

Statistical methods for assessing agreement between two ...
www.ncbi.nlm.nih.gov/pubmed/2868172

by JM Bland - 1986 - Cited by 35451 - Related articles
Lancet. 1986 Feb 8;1(8476):307-10. *Statistical methods for assessing agreement between two methods of clinical measurement*. Bland JM, Altman DG.

Bland & Altman (1986) : They wanted a measure of agreement which was easy to estimate and to interpret for a measurement on an individual patient.

An obvious starting point was a plot of the differences versus the mean of the measurements by the two methods :

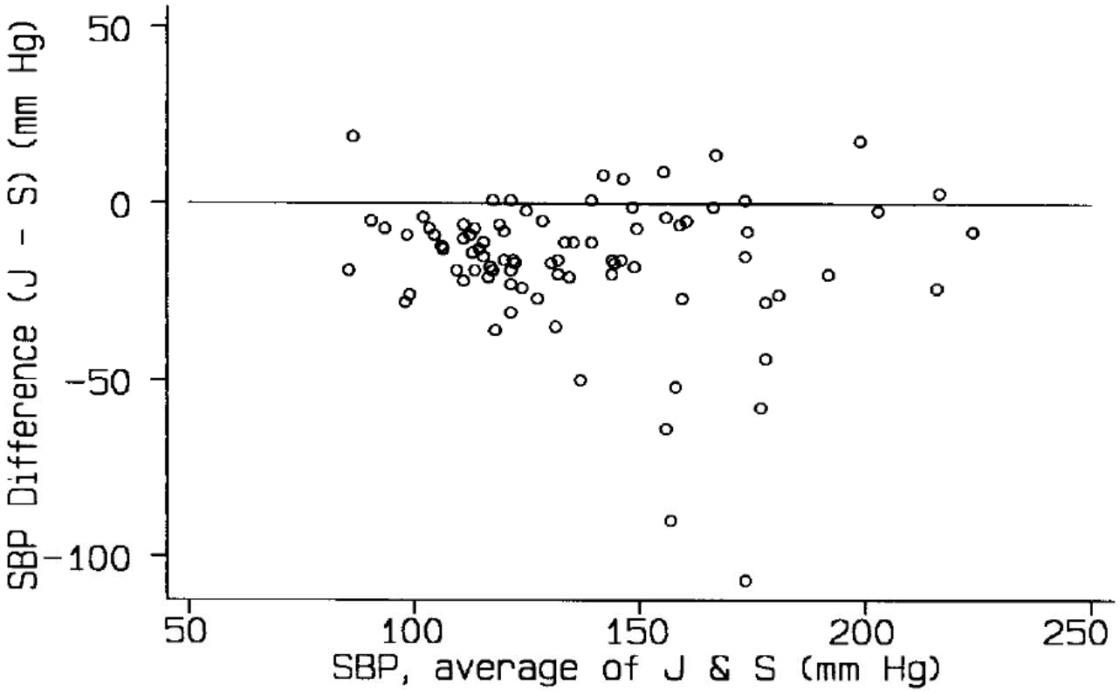


Figure 2 Systolic blood pressure: difference (J−S) versus average of values measured by observer J and machine S

The bias (differential bias) between the two measurement methods is estimated by the mean difference :

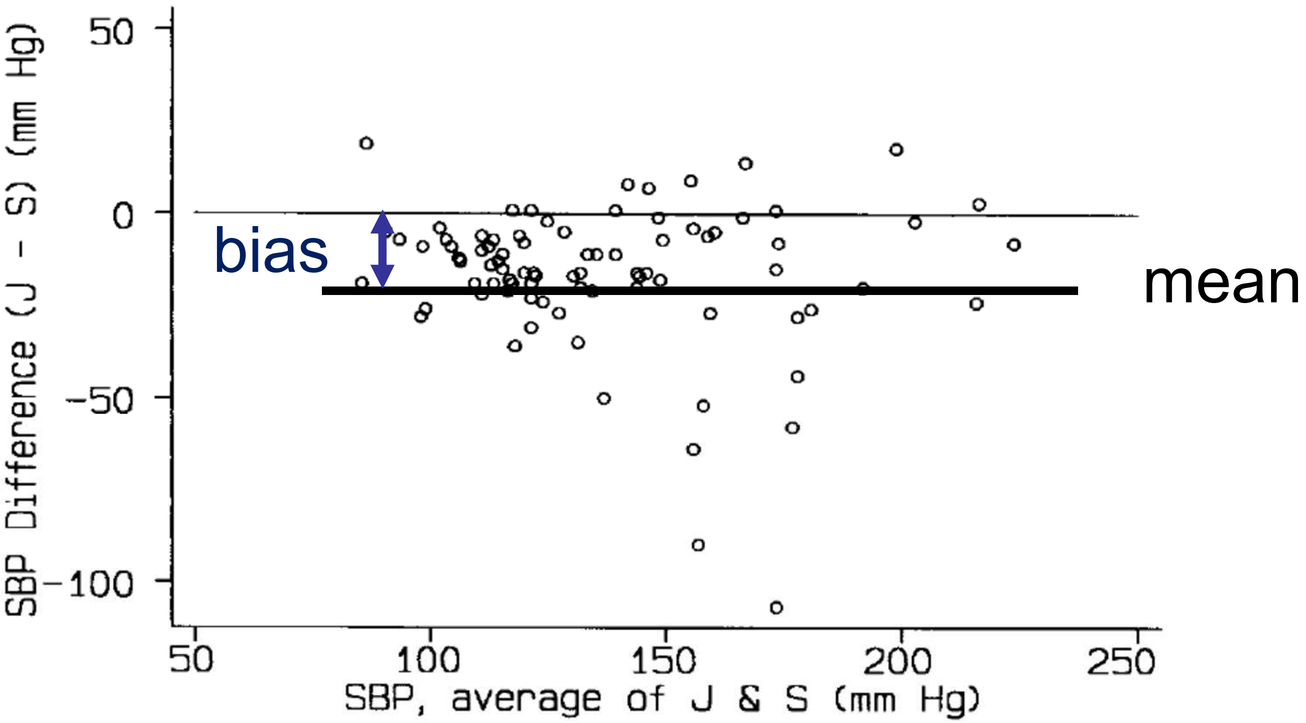


Figure 2 Systolic blood pressure: difference (J−S) versus average of values measured by observer J and machine S

If the differences are normally distributed, we would expect about 95% of the differences to lie between the mean +- 1.96*SD, the so called limits of agreement (LoA) (Bland & Altman, 1986):

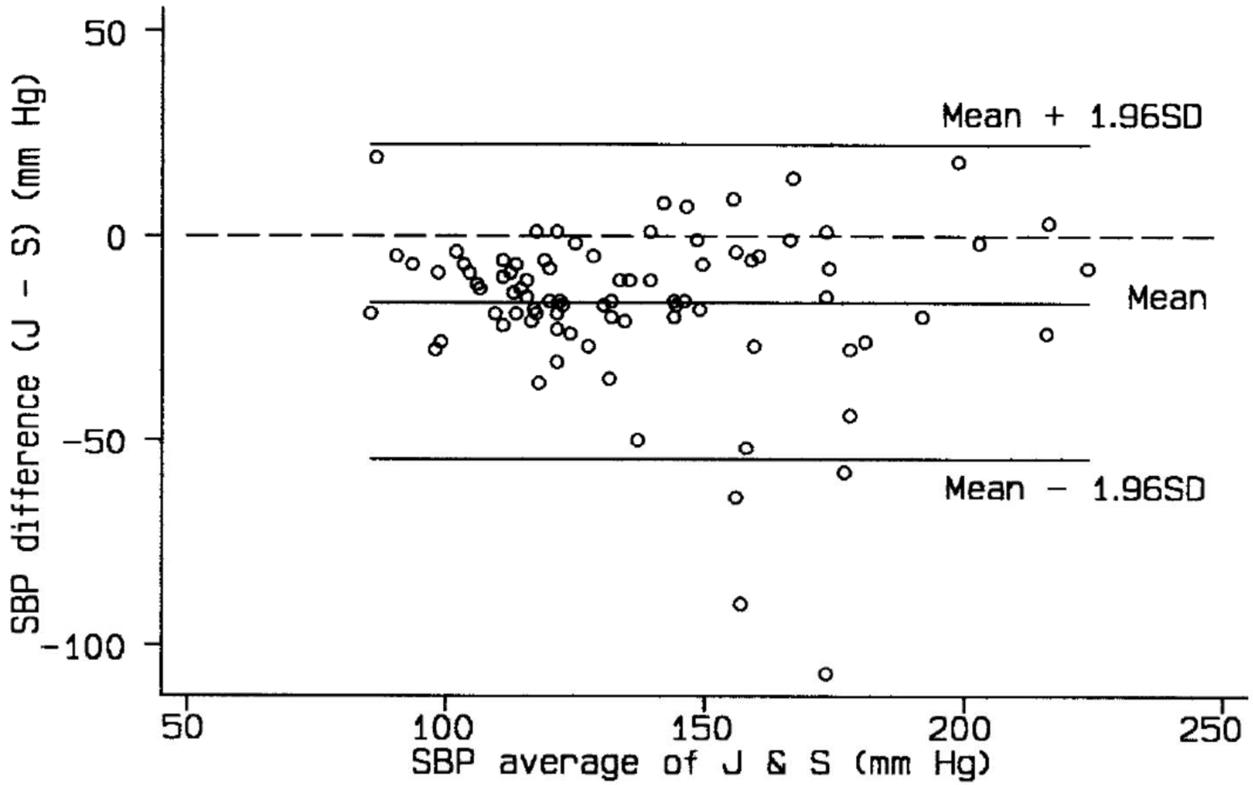


Figure 3 Systolic blood pressure: difference (J−S) versus average of values measured by observer J and machine S with 95% limits of agreement

The decision about what is acceptable agreement is a clinical one:

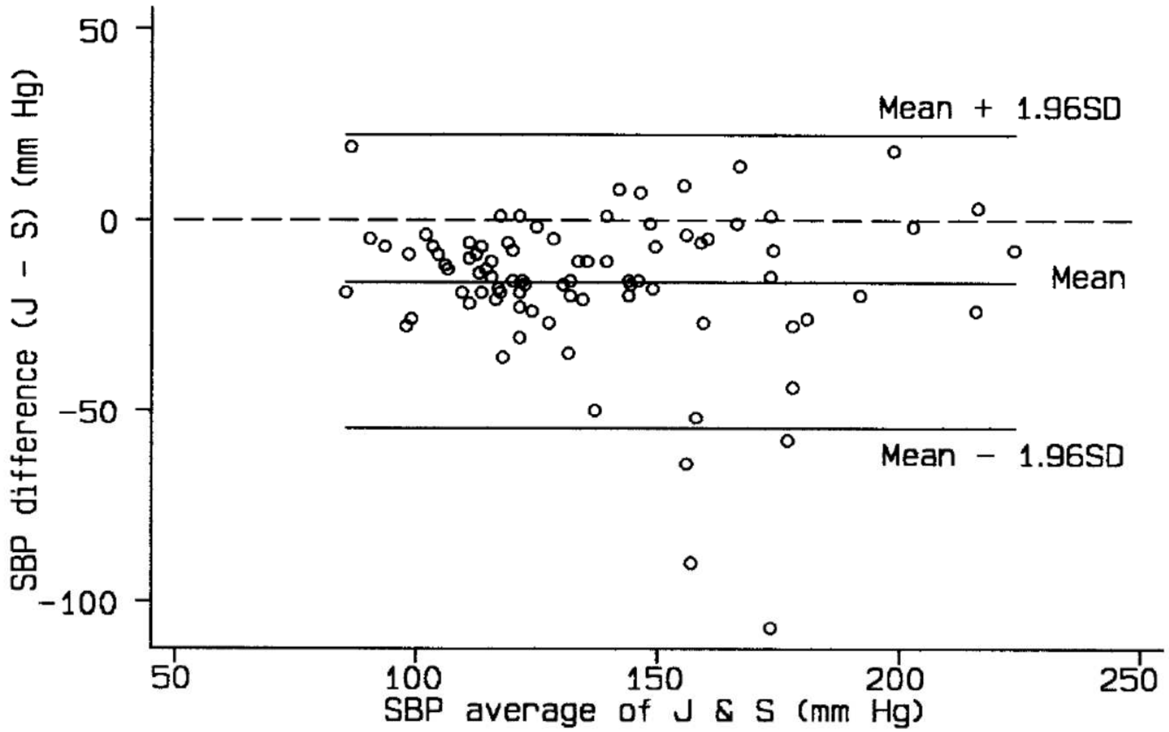


Figure 3 Systolic blood pressure: difference (J−S) versus average of values measured by observer J and machine S with 95% limits of agreement

We can see that the blood pressure machine (S) may give values between 55mmHg above the sphygmomanometer (J) reading to 22mmHg below it,

=> such differences would be unacceptable for clinical purposes

However, these estimates are meaningful only if we can assume bias and variability are uniform throughout the range of measurement, assumptions which can be checked graphically:

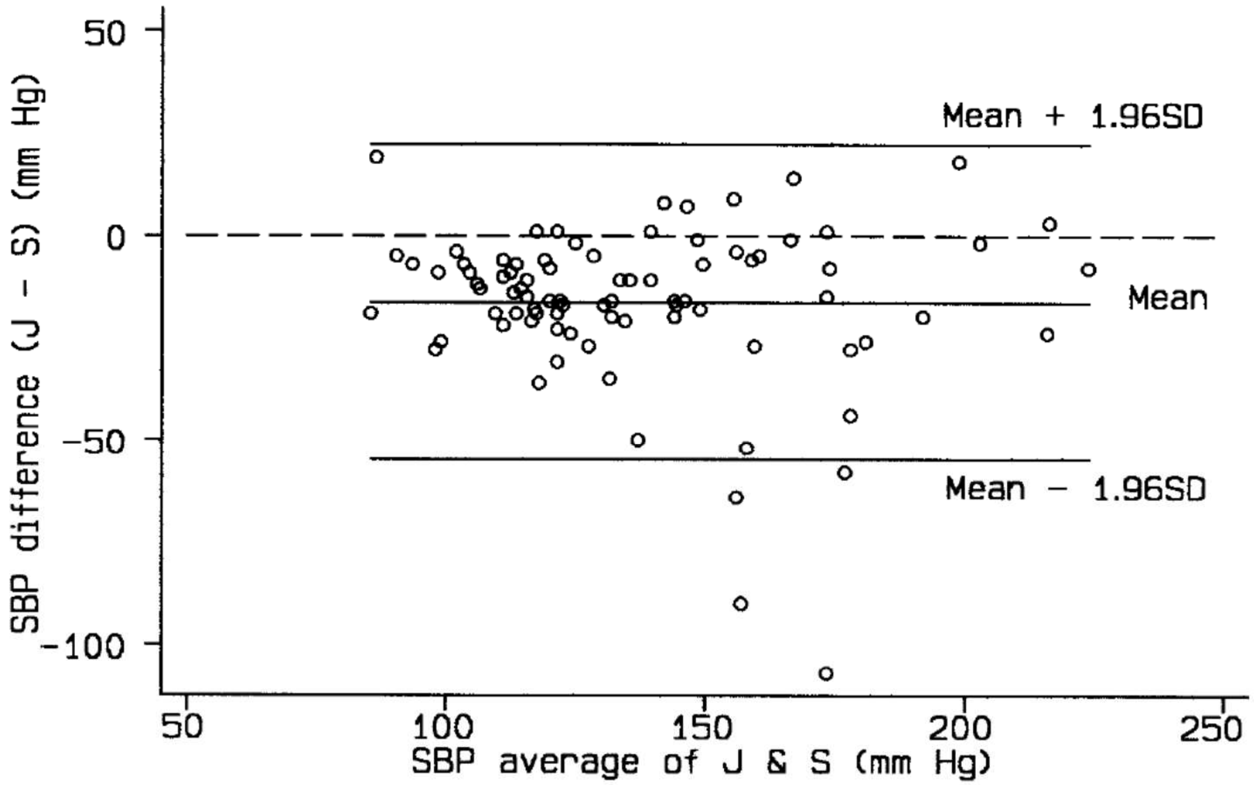
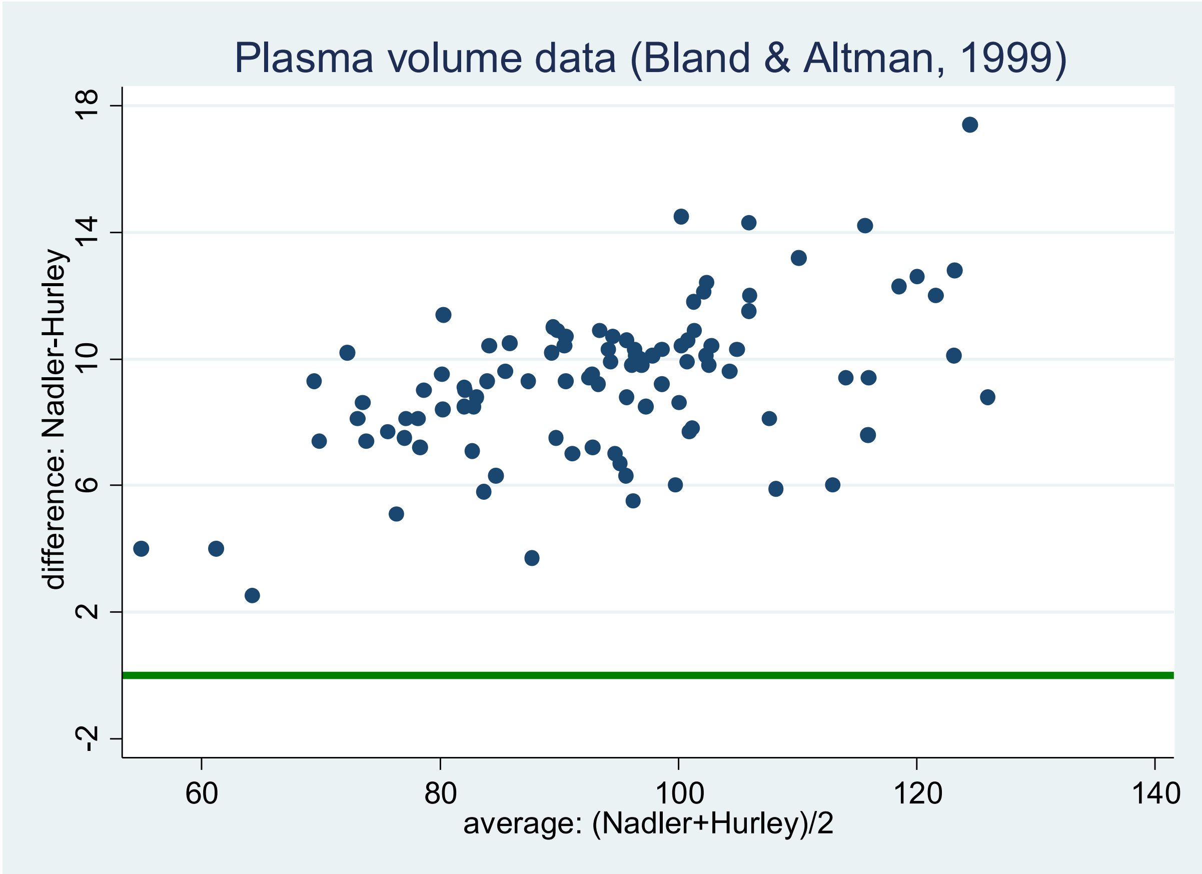


Figure 3 Systolic blood pressure: difference (J−S) versus average of values measured by observer J and machine S with 95% limits of agreement

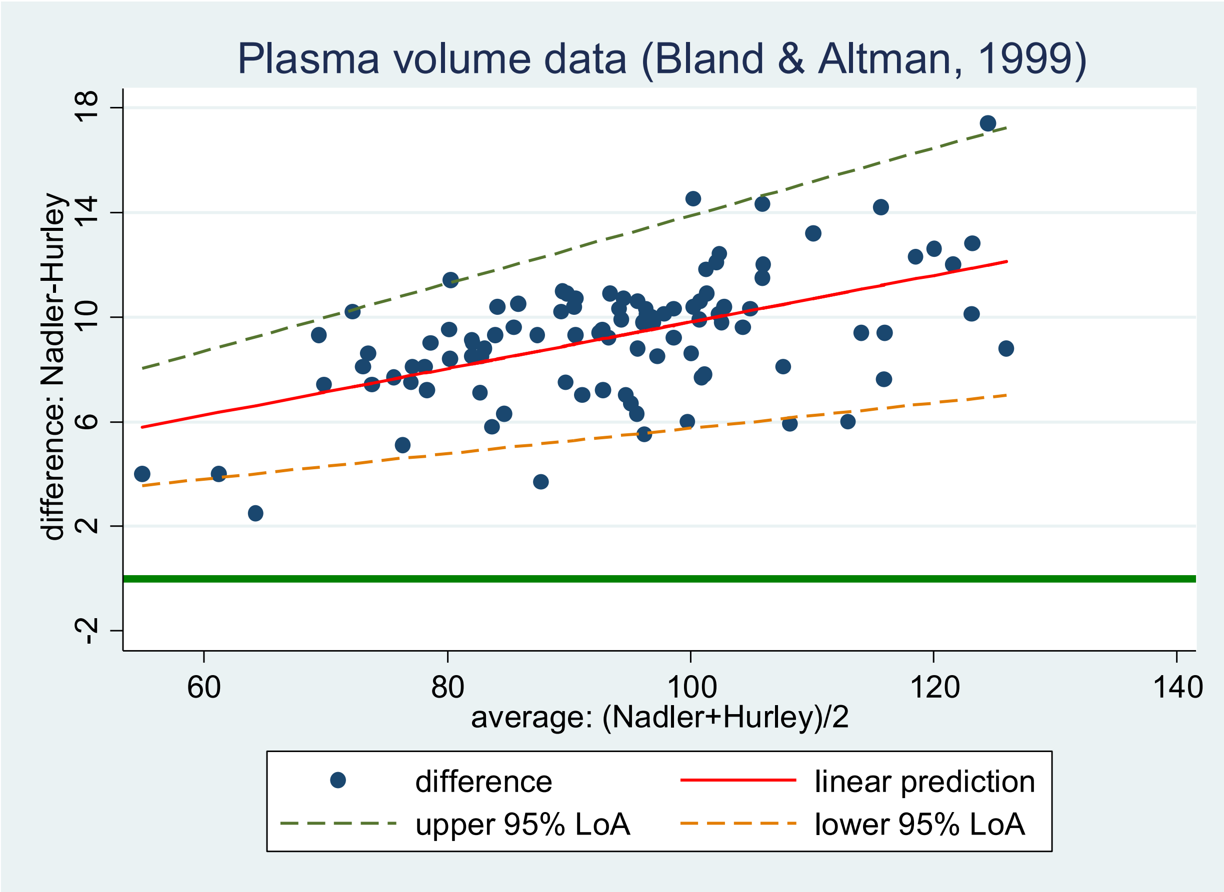
=> assumptions approximatively met

In some cases the variability of the measurements increases with the magnitude of the latent trait (heteroscedasticity), as well as with the mean difference (proportional bias):



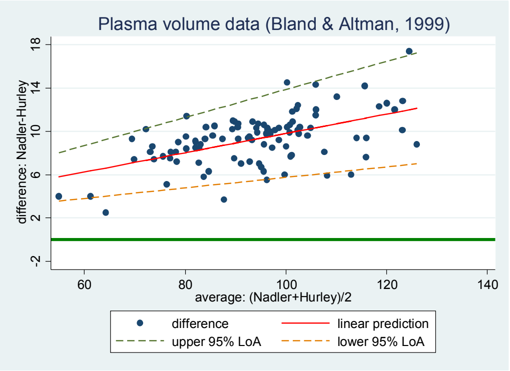
Plasma volume expressed in percentage of normal value: as measured by Nadler and Hurley

In this case, a linear regression of the differences on the averages can be estimated along with the LoA (Bland & Altman, 1999):



Plasma volume expressed in percentage of normal value: as measured by Nadler and Hurley

In that case, the LoA are more difficult to interpret (width not constant),



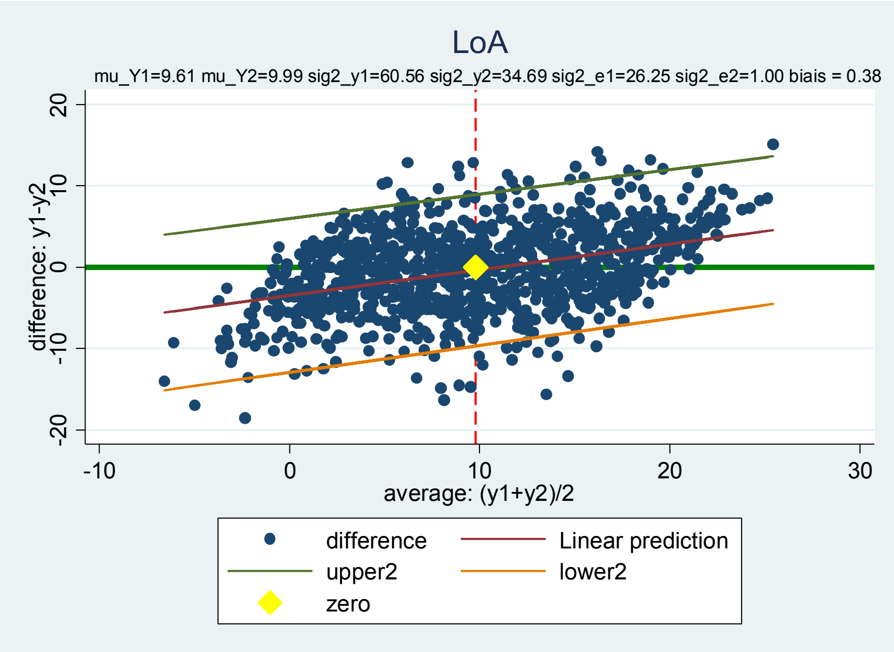
and more importantly,

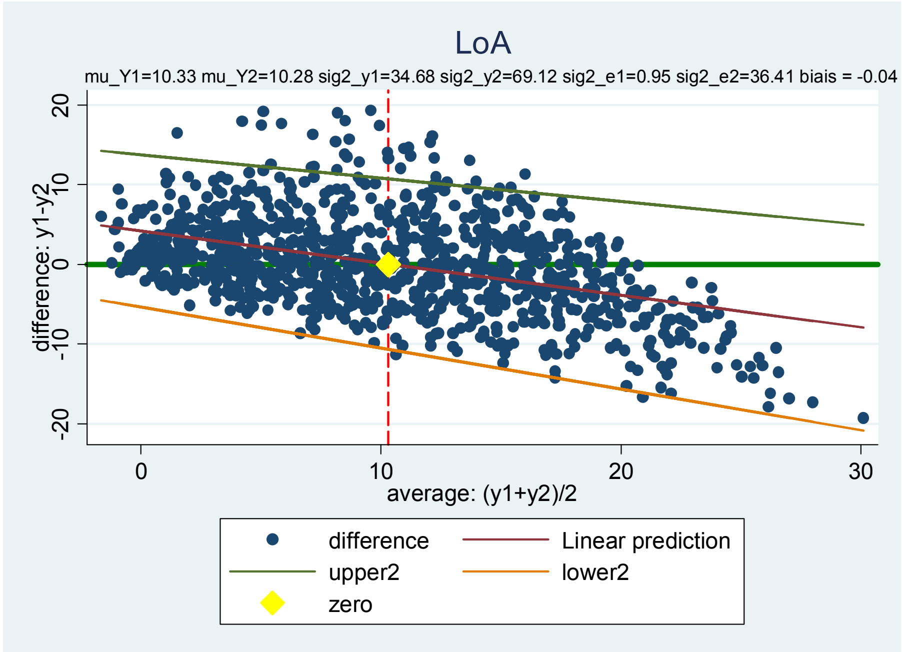
there are settings where Bland & Altman's plots are misleading !

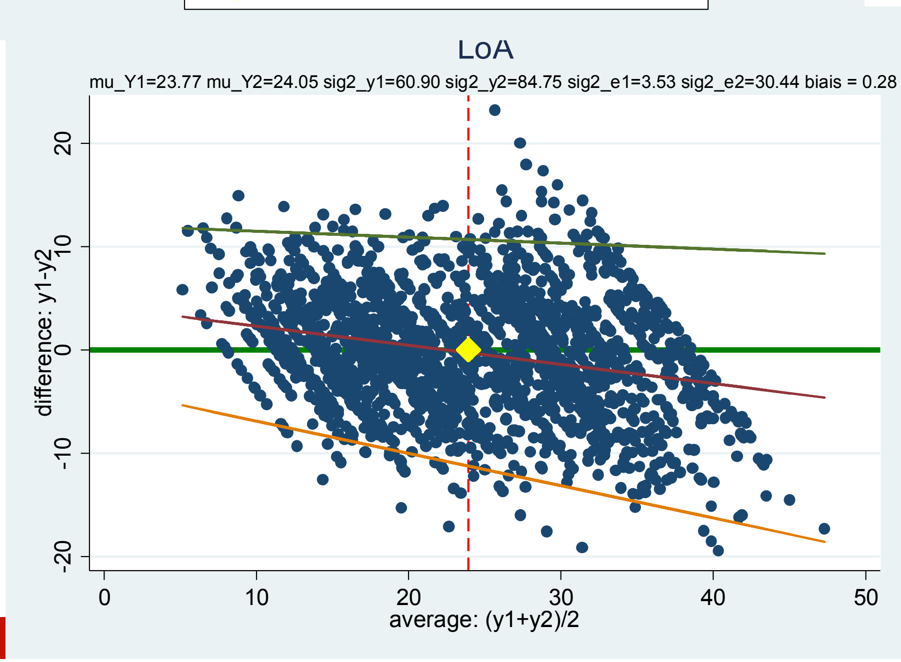
Indeed, we will show that

when variances of the measurement errors of the two methods are different, Bland and Altman's plots may be misleading…

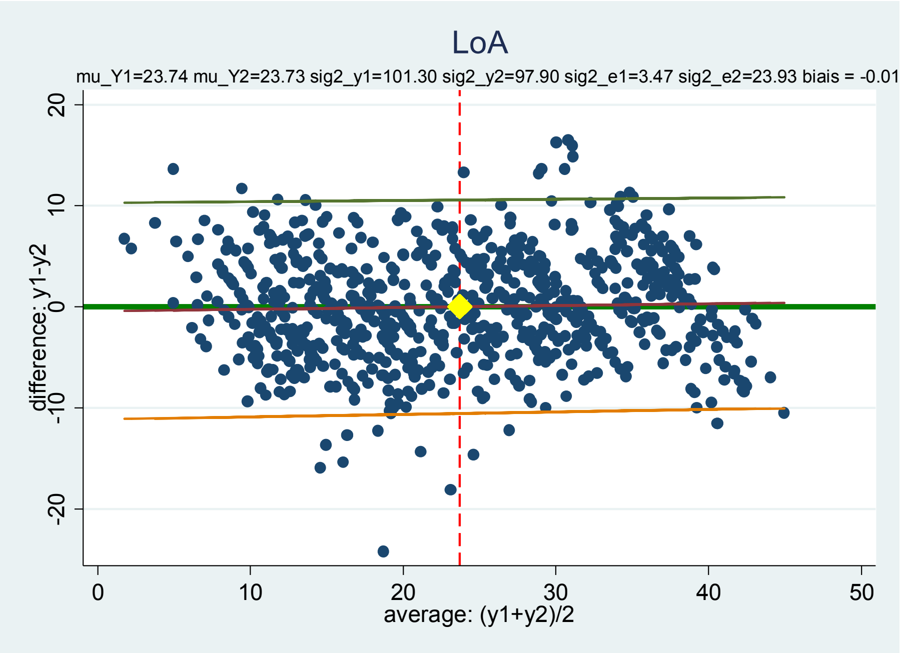
Simulated examples where the regression line shows an upward or a downward trend but there is no bias…

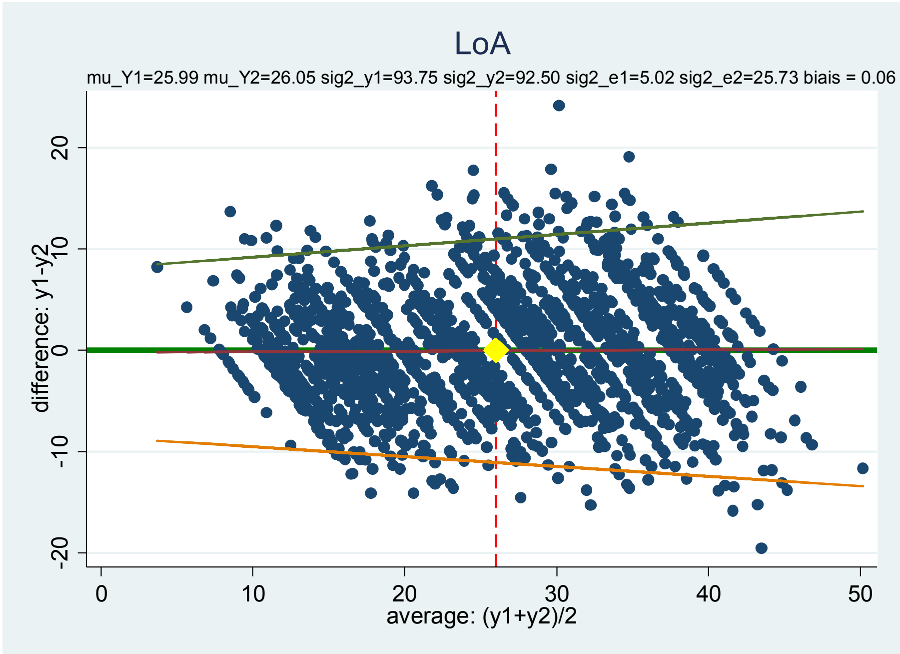






or a zero slope and there is a bias…





Therefore, the goal of my presentation is to introduce a new methodology for the evaluation of the **agreement** between two methods of measurement, where the first is the *reference standard* and the other the *new method* to be evaluated:

Effective plots to assess bias and precision in method comparison studies

Patrick Taffé

Institute for Social and Preventive Medicine, University of Lausanne, Switzerland
Patrick.Taffe@chuv.ch

Accepted September 2015, online October 2016

IUMSP
Institut universitaire de médecine sociale et préventive, Lausanne

More specifically, the objectives of this new methodology are to

- identify and quantify the amounts of differential and proportional biases,
- develop a method of recalibration in order to correct the bias of the new measurement method,
- and compare its precision with that of the reference standard.

The methodology requires several measurements by the reference standard and possibly only one by the new method for each individual.

It is applicable in all circumstances with or without differential and/or proportional biases and when the measurement errors are either homoscedastic or heteroscedastic.

Get ready !

2 The measurement error model

2.1 Formulation of the model

Consider the measurement error model:

$$y_{1ij} = \alpha_1 + \beta_1 x_{ij} + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_{ij};\boldsymbol{\theta}_1))$$
$$y_{2ij} = \alpha_2 + \beta_2 x_{ij} + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_{ij};\boldsymbol{\theta}_2))$$
$$x_{ij} \sim f_x(\mu_x, \sigma^2_x)$$

where y_{1ij} be the jth replicate measurement by method 1 on individual i, $j = 1, \ldots, n_i$ and $i = 1, \ldots, N$, whereas y_{2ij} is obtained by method 2, x_{ij} is a latent variable with density f_x representing the true unknown trait, and ε_{1ij} and ε_{2ij} represent measurement errors by method 1 and 2.

$$y_{1ij} = \alpha_1 + \beta_1 x_{ij} + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_{ij}; \boldsymbol{\theta}_1))$$
$$y_{2ij} = \alpha_2 + \beta_2 x_{ij} + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_{ij}; \boldsymbol{\theta}_2))$$

$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

It is assumed that the variances of these errors, i.e. $\sigma^2_{\varepsilon_1}(x_{ij}; \boldsymbol{\theta}_1)$ and $\sigma^2_{\varepsilon_2}(x_{ij}; \boldsymbol{\theta}_2)$, are heteroscedastic and depend on the level of the true unknown variable x_{ij}, as well as on the vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ of unknown parameters.

For the reference method, for instance method 2, $\alpha_2 = 0$ and $\beta_2 = 1$, whereas for method 1 the differential α_1 and proportional β_1 biases have to be estimated from the data.

The mean value of the latent variable x_{ij} is μ_x and its variance σ_x^2.

It is assumed that the latent variable is constant for individual *i*, i.e. $x_{ij} \equiv x_i$ (this assumption may be relaxed).

When method 2 is the reference standard and method 1 the new method to be evaluated, the model reduces to:

$$
\begin{aligned}
&y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma_{\varepsilon_1}^2(x_i; \boldsymbol{\theta}_1)) \\
&y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma_{\varepsilon_2}^2(x_i; \boldsymbol{\theta}_2)) \\
&x_i \sim f_x(\mu_x, \sigma_x^2)
\end{aligned}
$$

$$y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1))$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2))$$
$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

We have considered a simple linear relationship between y_{1ij} and x_i to identify the differential and proportional biases.

It is possible, however, in our framework to consider instead a non-linear function of x_i but in that case the bias no longer decomposes into two components with clear interpretations.

$$
\begin{aligned}
&y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1)) \\
&y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2)) \\
&x_i \sim f_x(\mu_x, \sigma_x^2)
\end{aligned}
$$

Nawarathna and Choudhary (Stat in Med, 2015) estimate the parameters of this model by bivariate maximum likelihood.

Their approach is complicated by the evaluation of the integrals in the marginal likelihood function and requires special numerical methods such as Laplace approximation or Gauss-Hermite quadrature.

We have developed another more simple way to estimate this model by a two-stage procedure, which performs effectively as demonstrated by the simulation study.

$$y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1))$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2))$$
$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

2.2 Estimation of the model

In the first stage, we estimate the regression model for y_{2ij}, by marginal maximum likelihood accounting non-parametrically for the heteroscedasticity.

$$y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1))$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2))$$
$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

Then, we adopt an empirical Bayes approach to predict x_i by the mean of its posterior distribution, which is the best linear unbiased prediction (BLUP) for x_i:

$$\hat{x}_i = E(x_i \mid \mathbf{y}_{2i})$$
$$= \int x_i \frac{f_{y_2}(\mathbf{y}_{2i} \mid x_i) f_x(x_i)}{\int f_{y_2}(\mathbf{y}_{2i} \mid x_i) f_x(x_i)\, dx_i} dx_i$$

$$y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1))$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2))$$
$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

In the second stage, we proceed to the estimation of the regression equation for y_{1ij} and of the differential α_1 and proportional β_1 biases simply by OLS after having substituted the BLUP $\hat{x}_i$ for the true unmeasured trait x_i.

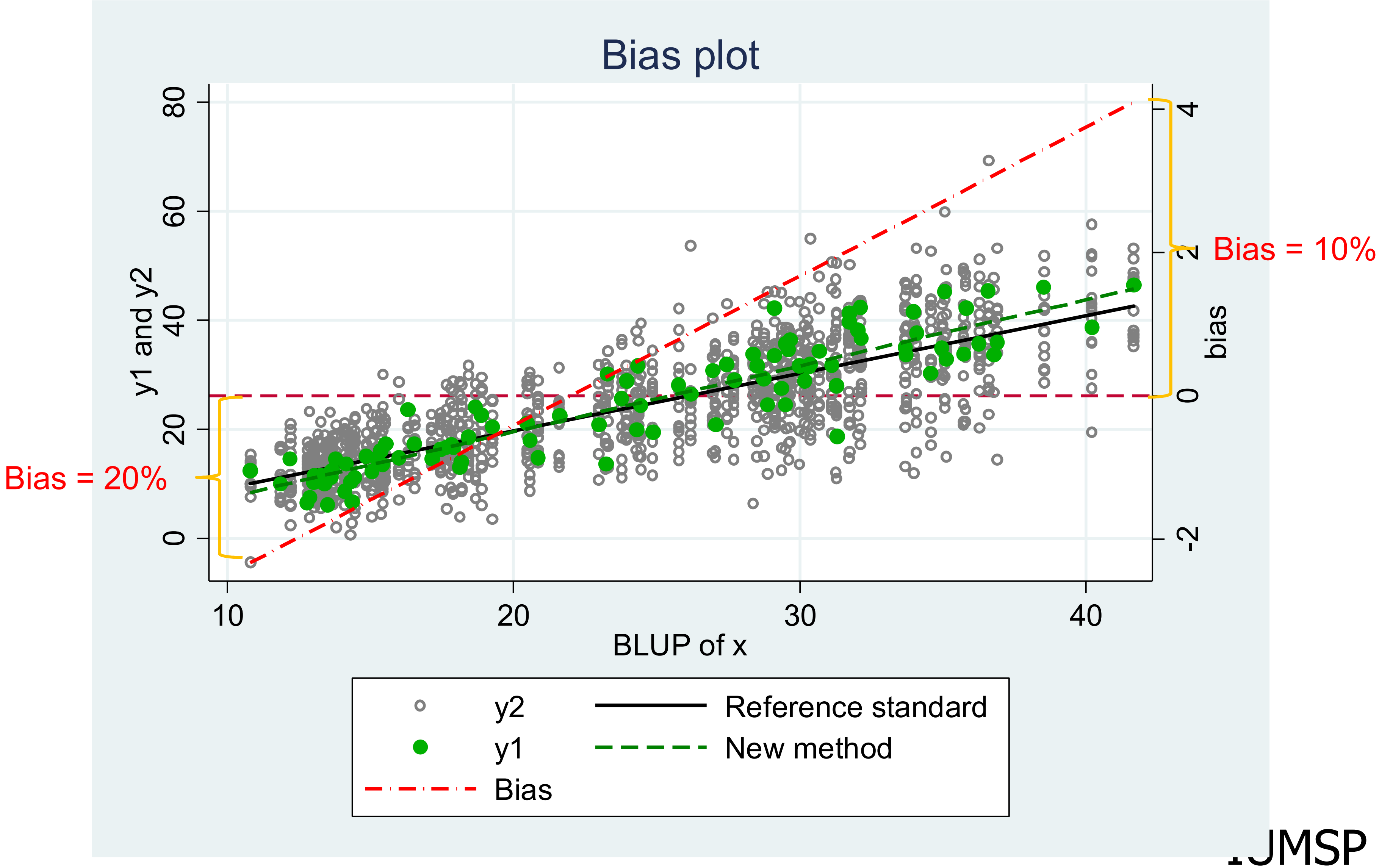
IUMSP
Institut universitaire de médecine sociale et préventive, Lausanne

$$y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1))$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2))$$
$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

Based on the estimates $\hat{\alpha}_1^*$ and $\hat{\beta}_1^*$ of the differential and proportional biases one can compute an estimate of the bias of the new method:

$$bias_i = \hat{\alpha}_1^* + \hat{x}_i(\hat{\beta}_1^* - 1)$$

A very useful figure to visualize the bias of the new method (i.e. method 1) is the "bias plot".



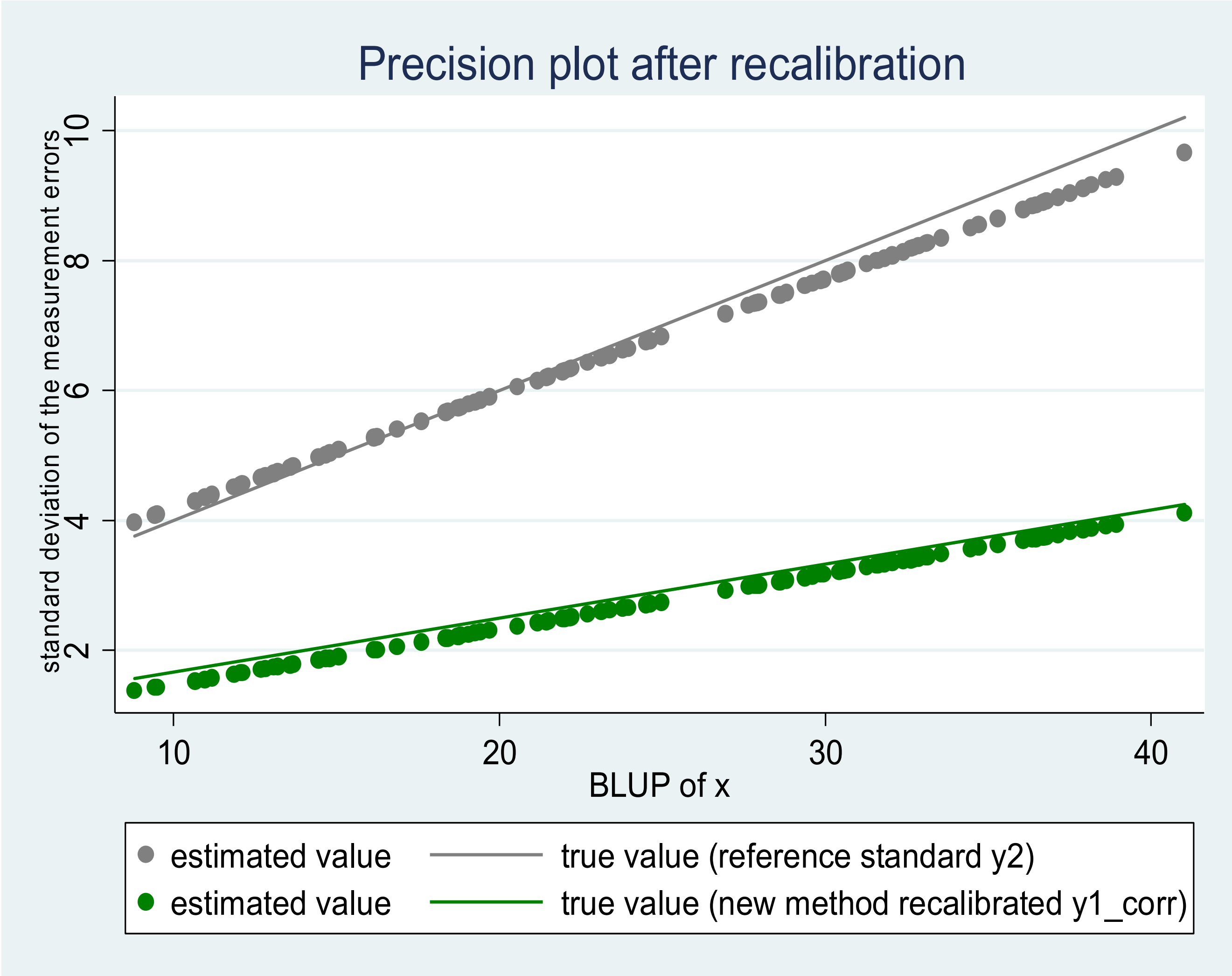
2.3 Recalibration of the new method

To remove the differential and proportional biases of the new method we proceed to its recalibration by computing:

$$y_{1ij}^{*} = (y_{1ij} - \hat{\alpha}_1^{*}) / \hat{\beta}_1^{*}$$

Now that y_{2ij} and y_{1ij}^{*} are on the same scale we can compare the variances of the measurement errors to determine which method is more precise.

We proceed to the comparison of the variances by making a scatter plot of the estimated standard deviations $\hat{\sigma}_{\varepsilon_1}(\hat{x}_i;\boldsymbol{\theta}_1)$ and $\hat{\sigma}_{\varepsilon_2}(\hat{x}_i;\boldsymbol{\theta}_2)$ versus $\hat{x}_i$, which we call “precision plot” :



IUMSP
Institut universitaire de médecine sociale et préventive, Lausanne

2.4 Why Bland and Altman's plot may be misleading

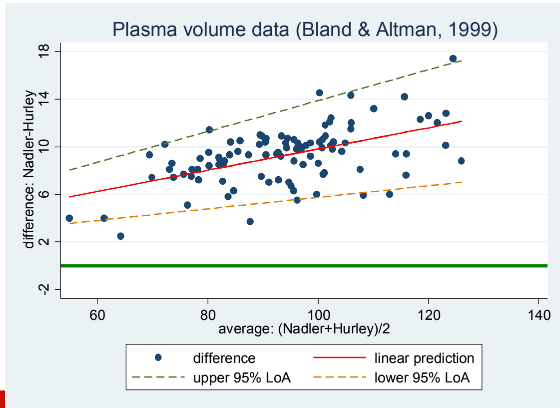
Bland and Altman have suggested to plot the differences $D_{ij} = y_{1ij} - y_{2ij}$ versus the averages $A_{ij} = (y_{1ij} + y_{2ij})/2$, and add to the plot the regression line of the relationship between D_{ij} and A_{ij} in addition to the LoA.

The problem is that the regression line may show a positive or negative slope when there is no bias or have a zero slope in the presence of a bias.

The reason is related to the fact that in the regression of D_{ij} on A_{ij}:

$$D_{ij} = \alpha + \beta A_{ij} + \varepsilon_{ij}$$

A_{ij} cannot be considered as being exogenous it is, rather, endogenous.



$$
\begin{aligned}
&y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1)) \\
&y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2)) \\
&x_i \sim f_x(\mu_x, \sigma^2_x)
\end{aligned}
$$

OLS estimation provides unbiased estimates only when:

$$
\operatorname{cov}(A_{ij}, \varepsilon_{ij}) = 0 \quad \Leftrightarrow \quad \frac{\sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1)}{\sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2)} = \frac{\beta_1}{\beta_2}
$$

i.e. there is no bias whenever the variances of the measurement errors are strictly equal to the proportional bias,

a special condition that has little chance to truly hold in practice…

3 A simulation study

Extensive simulations demonstrated that our methodology to assess biases, recalibrate the new method, and compare the precision of the two measurement methods performed very well

- for sample sizes of 100 individuals
- and between 10 to 15 measurements per individual by the reference standard
- and as few as only 1 by the new method.

For our simulations we considered the following data generating process:

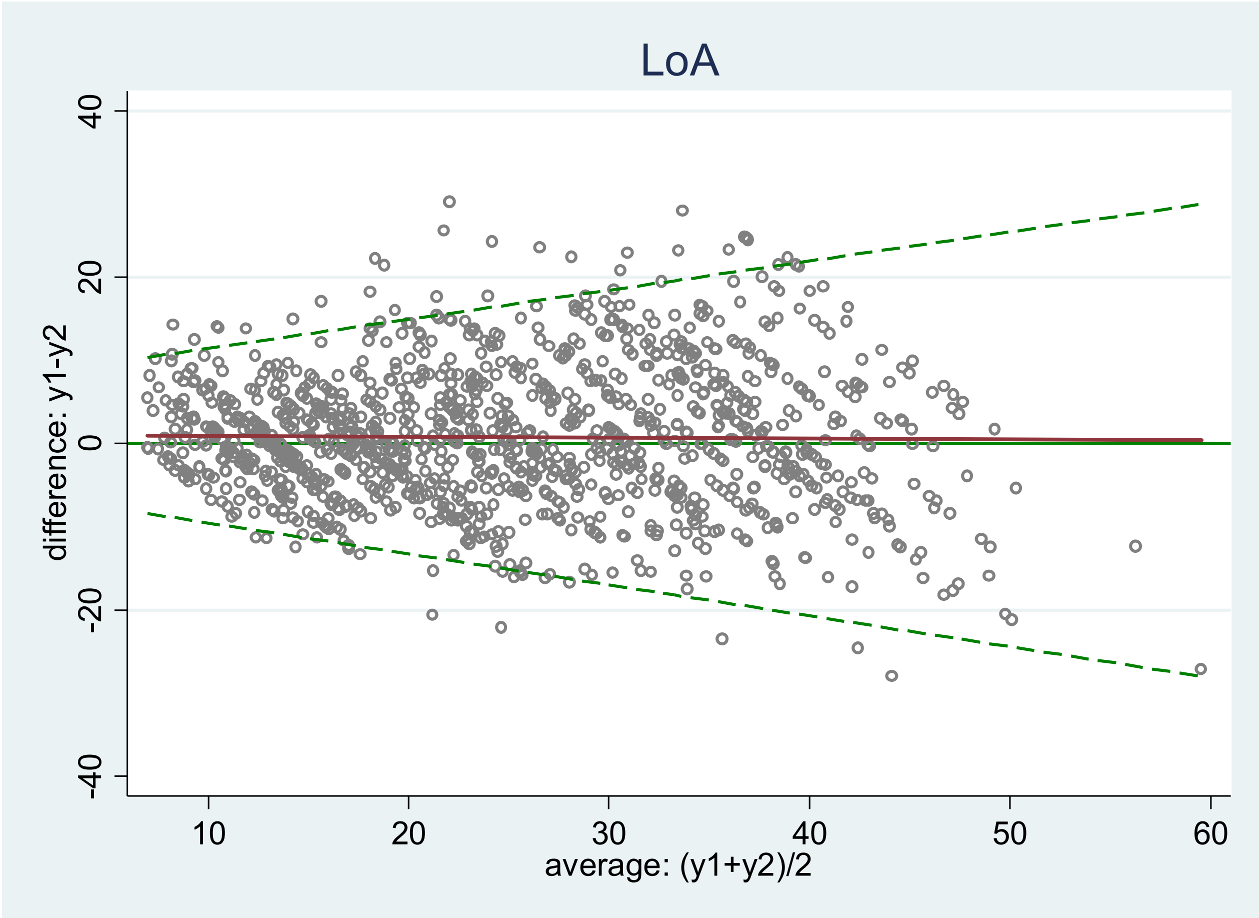
$$y_{1i} = -4 + 1.2\, x_i + \varepsilon_{1i}, \quad \varepsilon_{1i} \sim N(0, (1+0.1\, x_i)^2)$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, (2+0.2\, x_i)^2)$$
$$x_i \sim Uniform[10-40]$$

where $i =,1,...,100$ and the number of repeated measurements of individual i from the reference standard was $n_i \sim Uniform[10-15]$.

The new method has differential bias of -4 and a proportional bias of 1.2 .

The variance of the measurement errors from method 1 is smaller than that of the reference method 2.

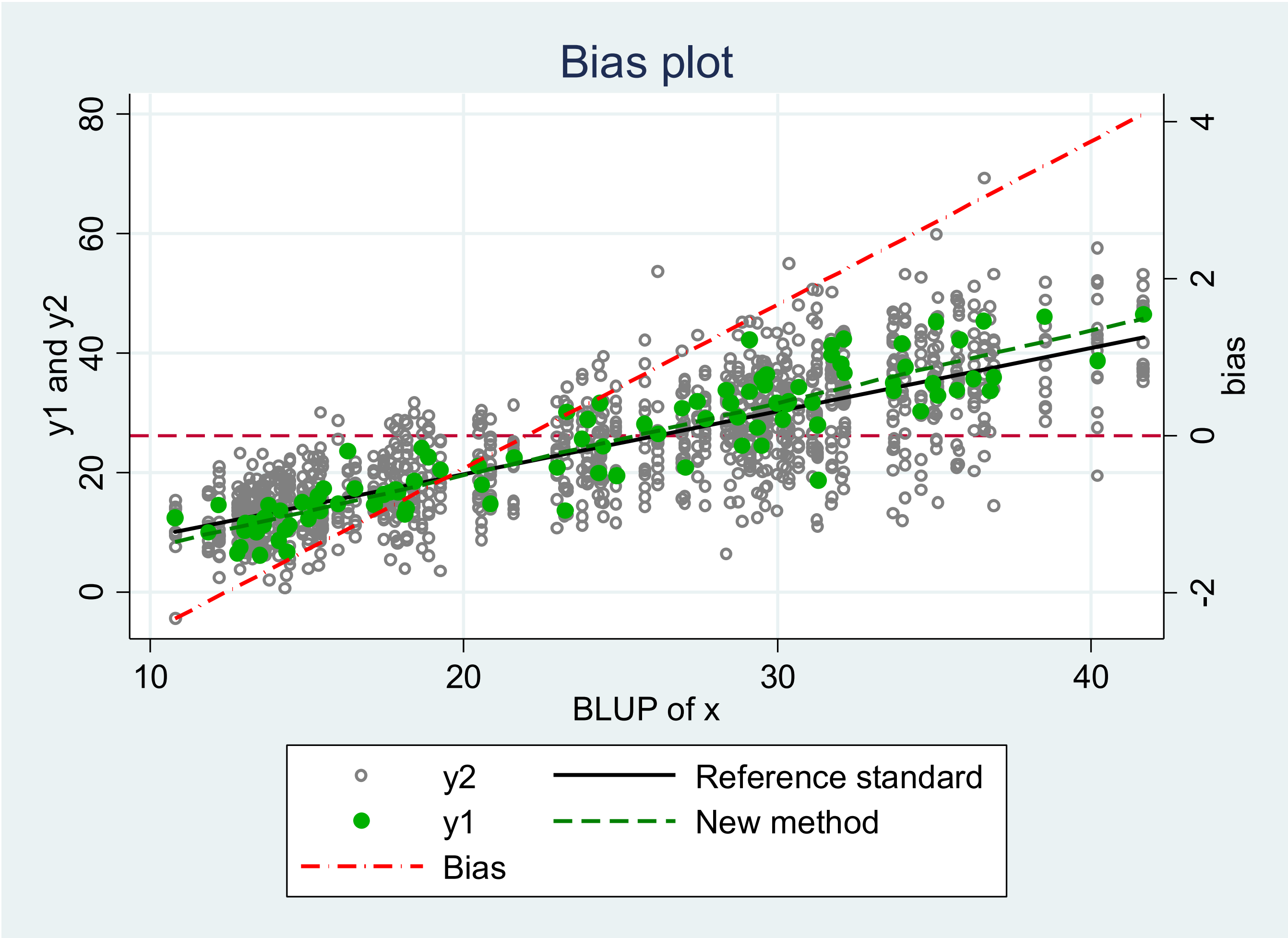
The Bland and Altman' LoA plot extended to the setting where there is heteroscedasticity of the measurement errors does not seem to indicate any bias:



IUMSP

Institut universitaire de médecine sociale et préventive, Lausanne

whereas the bias plot illustrates that the new method underestimates the trait up to 22 and then overestimates it, thereby clearly illustrating the occurrence of differential and proportional biases:



IUMSP
Institut universitaire de médecine sociale et préventive, Lausanne

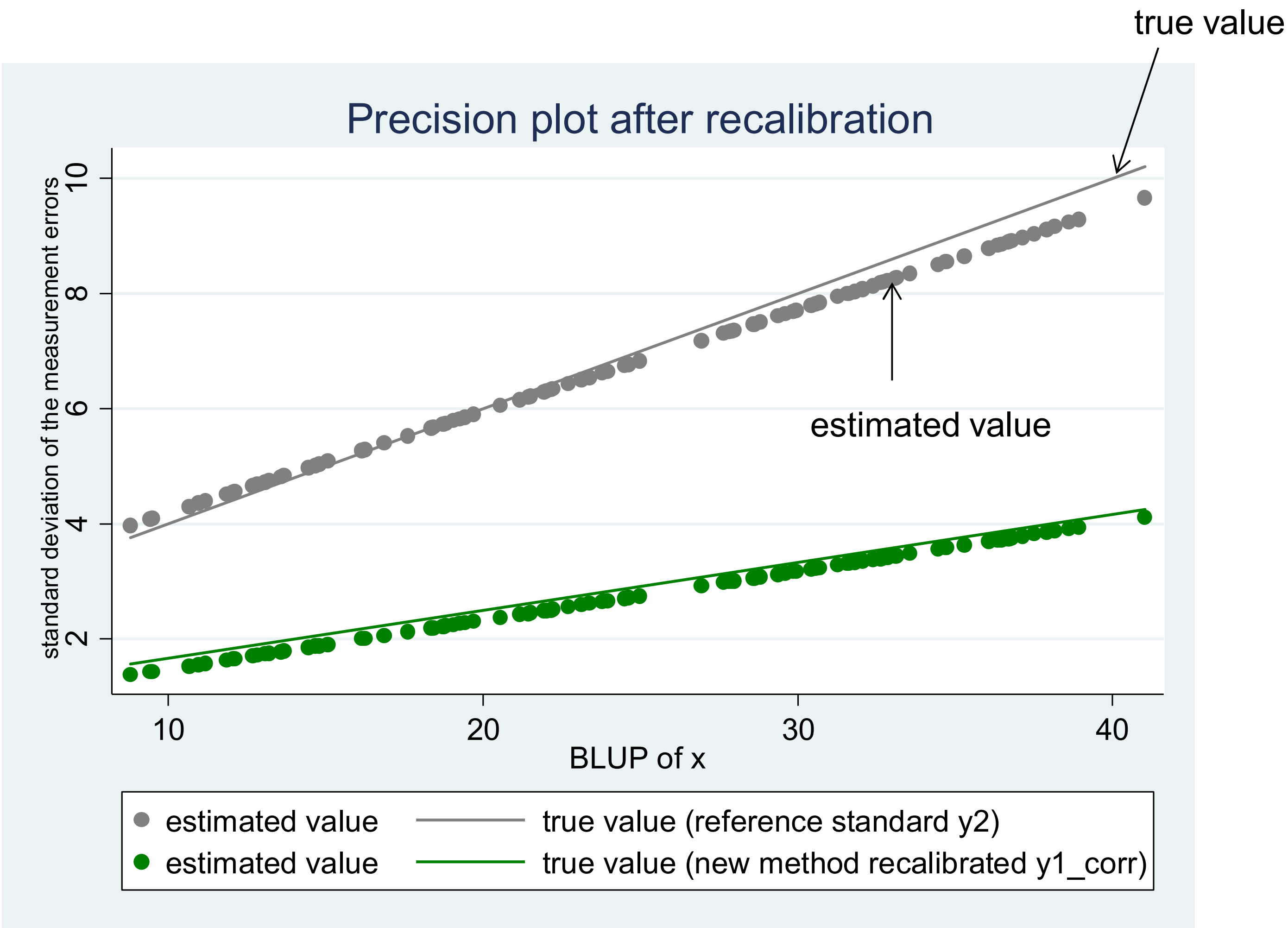
$$y_{1ij} = \alpha_1 + \beta_1 x_i + \varepsilon_{1ij}, \quad \varepsilon_{1ij} \sim N(0, \sigma^2_{\varepsilon_1}(x_i; \boldsymbol{\theta}_1))$$
$$y_{2ij} = x_i + \varepsilon_{2ij}, \quad \varepsilon_{2ij} \sim N(0, \sigma^2_{\varepsilon_2}(x_i; \boldsymbol{\theta}_2))$$
$$x_i \sim f_x(\mu_x, \sigma^2_x)$$

Actually, estimation of the measurement error model allowed us to identify

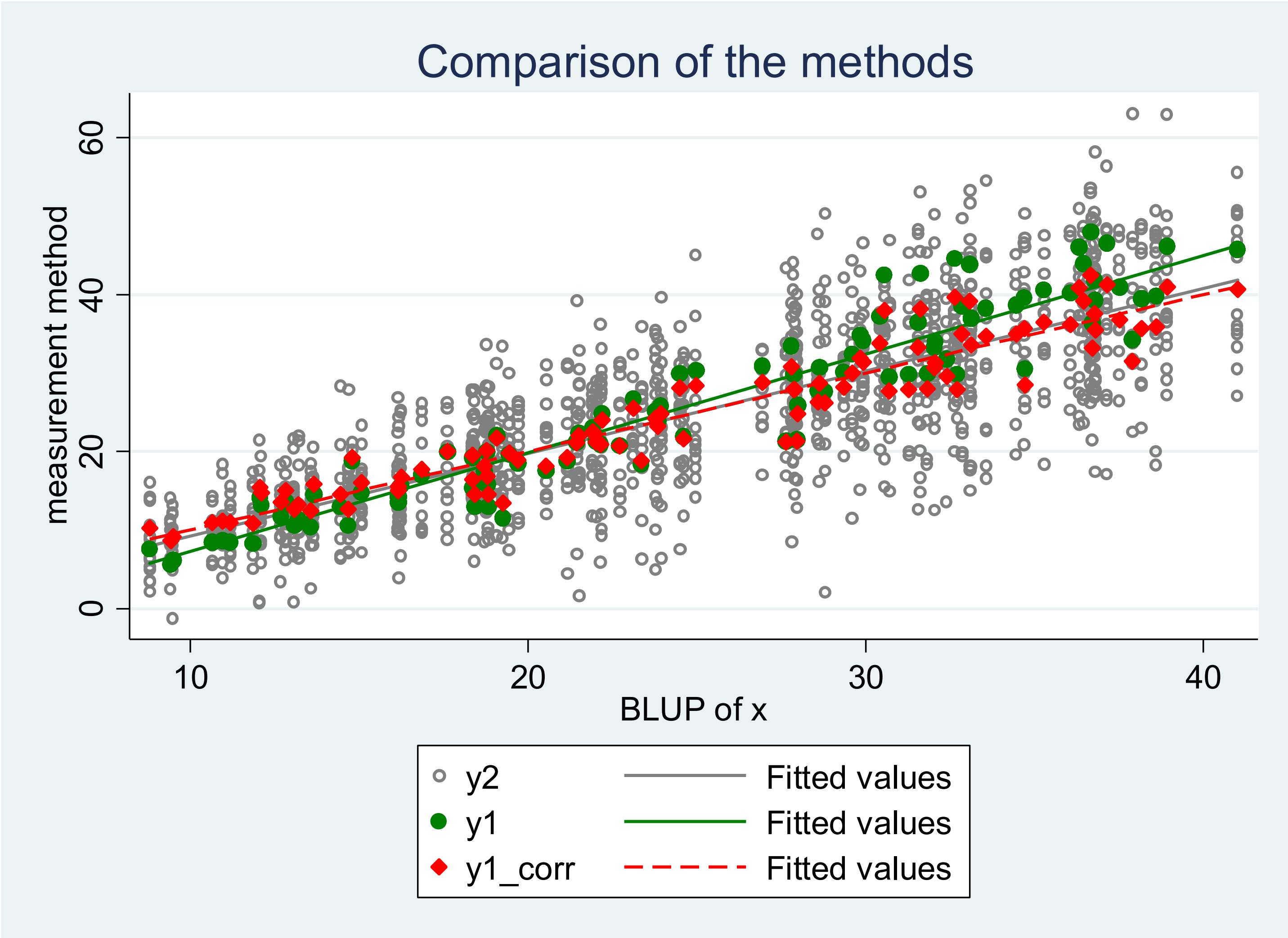
a differential bias of -3.85 95%CI= [-6.81; -0.88] (true value is -4)

and a proportional bias of 1.19 95%CI = [1.08; 1.29] (true value is 1.2).

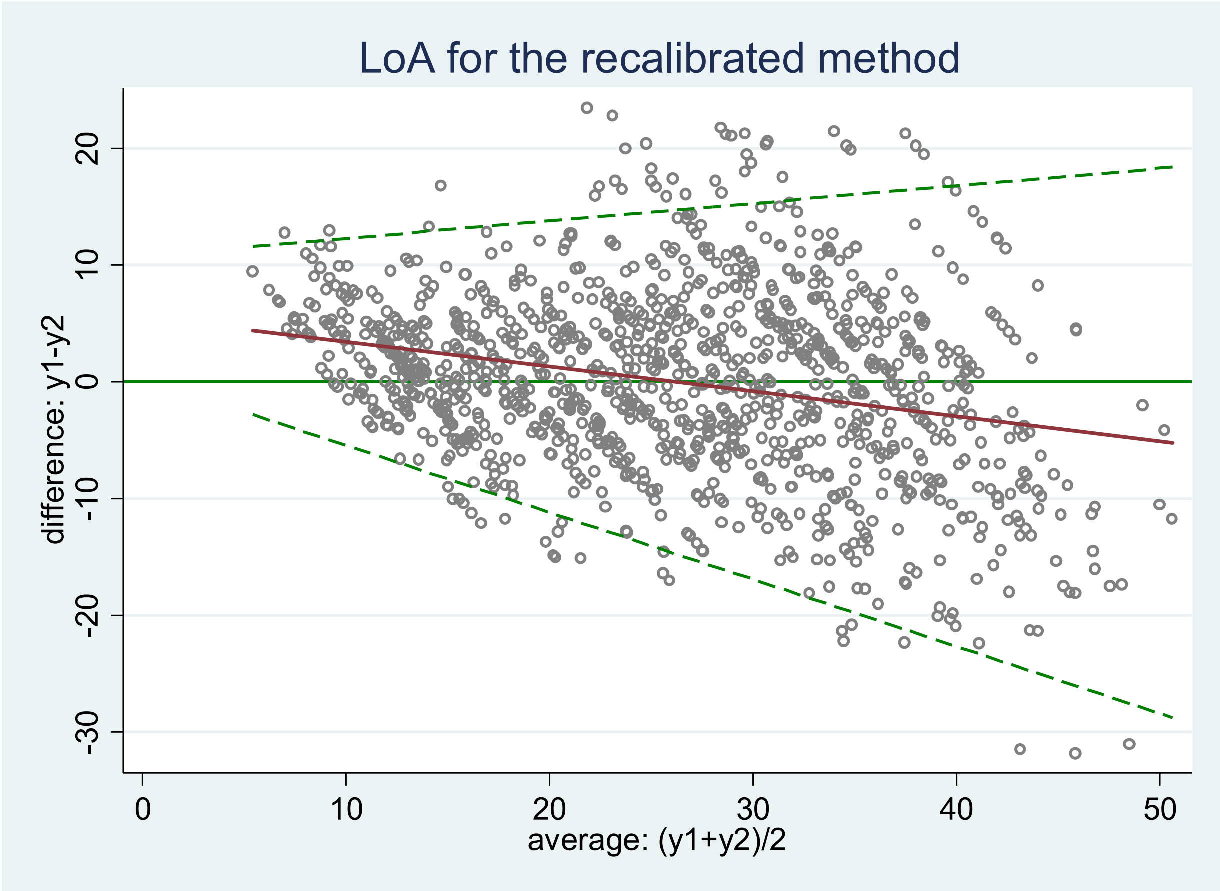
The variances of the measurement errors can already be well estimated with 10~15 measurements by the reference standard and only 1 by the new method:



Finally, the comparison plot allows us to visualize the performance of our recalibration procedure:



We computed Bland and Altman' LoA plot for the recalibrated method to illustrate that in the absence of bias the figure may mislead the reader into believing that there is a bias:



In summary,

We have developed a new methodology to assess the bias and precision of a new measurement method relative to the reference standard,

which does not have the shortcomings of Bland and Altman’s LoA methodology.

It is, however, in spirit of the original paper in the sense that new graphical representations of the bias and of the performance of the method to be evaluated are proposed.

In addition, we have shown a very simple way to recalibrate the new method to be able to use it in place of the more complex and costly reference standard.

biasplot: A Stata package to effective plots to assess bias and precision in method comparison studies

Patrick Taffé
Institute for Social and Preventive Medicine, University of Lausanne, Switzerland
Patrick.Taffe@chuv.ch

Mingkai Peng
Department of Community Health Sciences, University of Calgary, Canada
Mingkai.peng@ucalgary.ca

Vicki Stagg
Calgary Statistical Support, Canada
Vicki@calgarystatisticalsupport.com

Tyler Williamson
Department of Community Health Sciences, University of Calgary, Canada
Tyler.williamson@ucalgary.ca

Will appear soon in the Stata Journal ☺

IUMSP

Institut universitaire de médecine sociale et préventive, Lausanne