The Stata Journal (2017)
17, Number 1, pp. 208–221

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

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

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

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

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

Abstract. Bland and Altman's (1986, *Lancet* 327: 307–310) limits of agreement have been used in many clinical research settings to assess agreement between two methods of measuring a quantitative characteristic. However, when the variances of the measurement errors of the two methods differ, limits of agreement can be misleading. `biasplot` implements a new statistical methodology that Taffé (Forthcoming, *Statistical Methods in Medical Research*) recently developed to circumvent this issue and assess bias and precision of the two measurement methods (one is the reference standard, and the other is the new measurement method to be evaluated). `biasplot` produces three new plots introduced by Taffé: the "bias plot", "precision plot", and "comparison plot". These help the investigator visually evaluate the performance of the new measurement method. In this article, we introduce the user-written command `biasplot` and present worked examples using simulated data included with the package. Note that the Taffé method assumes there are several measurements from the reference standard and possibly as few as one measurement from the new method for each individual.

Keywords: gr0068, biasplot, limits of agreement, differential bias, proportional bias, Bland–Altman's plot, method comparison, measurement, empirical Bayes, BLUP

1 Introduction

Clinical researchers frequently use Bland and Altman's (1986) limits of agreement (LoA) to evaluate the agreement between two methods for measuring quantitative characteristics. Often this is motivated by a new, perhaps less expensive or easier, method of measurement against an established reference standard. To evaluate the comparability of the methods, the investigator collects measurements—one or several—from each method for a set of subjects. The investigator then computes Bland and Altman's LoA by adding and subtracting 1.96 times the estimated standard deviation to the mean differences. A scatterplot of the differences versus the means of the two variables with the LoA superimposed is then used to visually appraise the degree of agreement and quantify the magnitude. Further, the investigator adds a regression of the differences as a function of the means to the plot to indicate whether there is a bias and the direction of that bias (Bland and Altman 1999).

However, Bland and Altman's plot may be misleading when the variances of the measurement error for each method differ from one another. When this is the case, the regression line may show an upward or a downward trend when there is no bias or a zero slope when there is a bias. The literature has previously shown this problem. However, to the best of our knowledge, no simple-to-use and effective plots that evaluate bias and precision have been presented as an alternative (Hopkins 2004; Krouwer 2008; Carstensen, Simpson, and Gurrin 2008; Ludbrook 2010b,a; Carstensen 2010).

However, the purpose of this article is not a careful review of the literature. Interested readers should look at Nawarathna and Choudhary (2015) and the references therein for a recent review of measurement error. Rather, we will present the implementation of Taffé's (Forthcoming) method, which extends previously published methods to the setting of heteroskedastic measurement errors, particularly when heteroskedasticity is a function of the latent trait. We will perform the estimation in two steps, using an empirical Bayes approach to identify and quantify the amount of differential and proportional bias. Further, Taffé introduced two new plots, the "bias plot" and the "precision plot", to aid in assessing the new measurement method. These plots are not afflicted with the same issues as LoA plots but are still easily interpreted. The Taffé method requires that several measurements be made with the reference standard for each individual (usually more than five) and possibly only one measurement with the new method. The Taffé method allows each individual to have a different number of repeated measurements by each method and is applicable in all circumstances with or without differential or proportional bias and when the measurement errors are either homoskedastic or heteroskedastic.

2 The measurement error model

2.1 Formulation of the model

For a full presentation of the methodological theory, see Taffé (Forthcoming). Below we present an abridged version of the methods. Consider the measurement error model

$$
\begin{aligned}
y_{1ij} &= \alpha_1 + \beta_1 x_{ij} + \epsilon_{1ij} \qquad & \epsilon_{1ij}|x_{ij} \sim N\left\{0, \sigma^2_{\epsilon_1}\left(x_{ij};\boldsymbol{\theta}_1\right)\right\} \\
y_{2ij} &= \alpha_2 + \beta_2 x_{ij} + \epsilon_{2ij} \qquad & \epsilon_{2ij}|x_{ij} \sim N\left\{0, \sigma^2_{\epsilon_2}\left(x_{ij};\boldsymbol{\theta}_2\right)\right\} \\
x_{ij} &\sim f_x(\mu_x, \sigma^2_x)
\end{aligned}
$$

where y_{1ij} is the jth replicate measurement by method 1 on individual i, $j = 1, \dots, n_i$ and $i = 1, \dots, N$, 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 methods 1 and 2. We assume the variances of these errors, that is, $\sigma^2_{\epsilon_1}(x_{ij};\boldsymbol{\theta}_1)$ and $\sigma^2_{\epsilon_2}(x_{ij};\boldsymbol{\theta}_2)$, are heteroskedastic and increase with the level of the true latent trait, x_{ij}, in a way to be precisely specified later, depending on the vectors of unknown parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. For the reference method, $\alpha_2 = 0$ and $\beta_2 = 1$, whereas for method 1, we must estimate the differential α_1 and proportional β_1 biases from the data. The mean value of the latent variable x_{ij} is μ_x, and its variance is σ^2_x. When method 2 is the reference standard and method 1 the new method for evaluation, the model reduces to

$$
\begin{aligned}
y_{1ij} &= \alpha_1 + \beta_1 x_i + \epsilon_{1ij} \qquad & \epsilon_{1ij}|x_i \sim N\left\{0, \sigma^2_{\epsilon_1}\left(x_i;\boldsymbol{\theta}_1\right)\right\} \\
y_{2ij} &= x_i + \epsilon_{2ij} \qquad & \epsilon_{2ij}|x_i \sim N\left\{0, \sigma^2_{\epsilon_2}\left(x_i;\boldsymbol{\theta}_2\right)\right\} \\
x_i &\sim f_x(\mu_x, \sigma^2_x)
\end{aligned}
\tag{1}
$$

Note that this measurement error model is slightly different from the classical measurement error model; the heteroskedasticity depends on the latent trait and not on an observed average (Dunn 2004).

2.2 Estimation of the model

The estimation process has two steps:

Estimation step 1

Other methods treat x_i as a nuisance parameter and attempt to integrate it out from the joint likelihood function. The Taffé method fits the regression model for y_{2ij} using marginal maximum likelihood, allowing the variance of ϵ_{2ij} to be different for each decile of the empirical distribution of $\overline{y}_{2i}$ (that is, the mean of the individual repeated measurements $\overline{y}_{2i}$ is used as a rough approximation to x_i). Then, following an empirical Bayes approach, we predict x_i from the mean of its posterior distribution (that is, the mean of the conditional distribution of x_i given the vector $\mathbf{y}_{2i}$ of observations for individual i by method 2), which is the best linear unbiased prediction (BLUP) for x_i.

$$
\begin{aligned}
\widehat{x}_i &= E\left(x_i|\mathbf{y}_{2i}\right) \\
&= \int x_i \frac{f_{y_2}(\mathbf{y}_{2i}|x_i) f_x(x_i)}{\int f_{y_2}(\mathbf{y}_{2i}|x_i) f_x(x_i) dx_i} dx_i
\end{aligned}
\tag{2}
$$

For the sake of notational convenience, we have suppressed the dependence of the density functions f_{y_2} and f_x from their parameters, which have been estimated by maximum likelihood.

When f_x is the normal density, (2) is

$$\widehat{x}_i = \sigma_x^2 \boldsymbol{\iota}' \mathbf{V}_i^{-1} \left(\mathbf{y}_{2i} - \boldsymbol{\iota}\widehat{\mu}_x\right) + \widehat{\mu}_x$$

where $\boldsymbol{\iota}$ is an n_i vector of ones and $\mathbf{V}_i = \sigma_x^2 \boldsymbol{\iota}\boldsymbol{\iota}' + \text{diag}\{\sigma_{\epsilon_2}^2(x_i; \boldsymbol{\theta}_2)\}$ is the variance–covariance matrix of $\mathbf{y}_{2i}$.

It is desirable to have a smooth estimate of the heteroskedasticity that does not depend on $\overline{y}_{2i}$ but rather on $\widehat{x}_i$, the BLUP for x_i. Therefore, Taffé suggests an approach similar to that of Bland and Altman (1999) by regressing the absolute values of the residuals $\widehat{\epsilon}_{2ij}$ from the linear regression model $y_{2ij} = \alpha_2^* + \beta_2^* \widehat{x}_i + \epsilon_{2ij}^*$ on $\widehat{x}_i$ by ordinary-least squares (OLS) to create a smooth estimate of the heterogeneous variance:

$$|\widehat{\epsilon}_{2ij}^*| = \theta_2^{(0)} + \theta_2^{(1)} \widehat{x}_i + v_{ij}$$

Under the normality assumption, $|\epsilon_{2ij}^*|$ follows a half-normal distribution with mean $E(|\epsilon_{2ij}^*|) = \sigma_{\epsilon_2}(\widehat{x}_i; \theta_2)\sqrt{2/\pi}$. Therefore, we obtain a smooth standard-deviation estimate as follows:

$$\widehat{\sigma}_{\epsilon_2}\left(\widehat{x}_i; \widehat{\theta}_2\right) = \widehat{E}\left(|\widehat{\epsilon}_{2ij}^*|\right)\sqrt{\pi/2} = \left(\widehat{\theta}_2^{(0)} + \widehat{\theta}_2^{(1)} \widehat{x}_i\right)\sqrt{\pi/2}$$

Note that Taffé suggests that the form of the heterogeneity need not be a straight line; we may consider other heterogeneity structures, and a graphical representation of $|\widehat{\epsilon}_{2ij}^*|$ versus $\widehat{x}_i$ provides a good start to visually check the plausibility of the straight-line model. It may be useful to assess the fit using a scatterplot of y_{2ij} versus $\widehat{x}_i$, with the estimated regression line and the 95% prediction limits computed as $\widehat{\alpha}_2^* + \widehat{\beta}_2^* \widehat{x}_i \pm 2\widehat{\sigma}_{\epsilon_2}(\widehat{x}_i; \boldsymbol{\theta}_2)$.

Estimation step 2

The second stage of the estimation process involves the estimation of the regression equation for y_{1ij} in (1) and estimation of the differential (α_1) and proportional (β_1) biases by OLS after substituting the BLUP for $\widehat{x}_i$ for the true unmeasured trait, x_i. We may then use the Wald test and 95% confidence intervals (CIs) for α_1 and β_1 to formally assess these biases. As before, we can obtain a smooth estimate of the variance by using OLS to fit the model $|\widehat{\epsilon}_{1ij}^*| = \theta_1^{(0)} + \theta_1^{(1)} \widehat{x}_i + \omega_{ij}$, where $|\widehat{\epsilon}_{1ij}^*|$ is the absolute value of the residuals $\widehat{\epsilon}_{1ij}^*$, from the linear regression model $y_{1ij} = \alpha_1^* + \beta_1^* \widehat{x}_i + \epsilon_{1ij}^*$. Then, based on the estimates $\widehat{\alpha}_1^*$ and $\widehat{\beta}_1^*$, the bias of the new method is estimated as

$$\text{bias}_i = \widehat{\alpha}_1^* + \widehat{x}_i \left(\widehat{\beta}_1^* - 1\right)$$

To visually assess the degree of bias, we obtain the "bias plot", after which the package is named, by graphing a scatterplot of y_{1ij} and y_{2ij} versus the BLUP for $\widehat{x}_i$, along with the two regression lines, while adding a second scale on the right showing the relationship between the estimated amount of bias and $\widehat{x}_i$.

Taffé shows, by simulation, that this methodology performs well and that the estimates of the differential α_1 and proportional β_1 biases are reasonably unbiased and consistent already for sample sizes of 100 persons, with 3 to 5 repeated measurements per individual from the reference method and only 1 measurement from the new method. However, to appropriately estimate the (heterogeneous) measurement error variances, one should have 10 to 15 repeated measurements per individual from the reference method and 1 or several measurements from the new method.

2.3 Recalibration of the new method

To remove the differential and proportional biases of the new method, we recalibrate it by computing $y^*_{1ij} = (y_{1ij} - \widehat{\alpha}^*_1)/\widehat{\beta}^*_1$. The "comparison plot" allows us to visualize the recalibration procedure.

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. Because we would like to compare y_{2ij} with y^*_{1ij} (and not with y_{1ij}), we should recalculate a smooth estimate of the measurement errors variance of y^*_{1ij} by proceeding like before.

We can then compare the variances by making a scatterplot of the estimated standard deviations $\widehat{\sigma}_{\epsilon_1}(\widehat{x}_i;\boldsymbol{\theta}_1)$ and $\widehat{\sigma}_{\epsilon_2}(\widehat{x}_i;\boldsymbol{\theta}_2)$ versus $\widehat{x}_i$, which we call the "precision plot". It is possible that after recalibration, the new method will turn out to be more precise (locally or globally) than the reference standard.

3 The biasplot command

`biasplot` fits the measurement error model and provides estimates of the differential and proportional biases. It also allows the computation of the extended version of Bland and Altman's LoA when the variances of measurement errors are possibly heteroskedastic. We obtain the (extended) LoA, bias, precision, and comparison plots by specifying one of the options: `loa`, `bias`, `precision`, or `comp`.

3.1 Syntax

The syntax for using `biasplot` is

`biasplot` [*if*] [*in*], `idvar(`*varname*`)` `ynew(`*varname*`)` `yref(`*varname*`)` [`loa`
`bias precision comp pdfs`]

Note that you must choose at least one of the options `loa`, `bias`, `precision`, or `comp` for the program to run and save the corresponding graphs to the current directory. Also, the recalibrated values (`y1_corr`) of the new measurement method will be added to the dataset after computing the bias plot.

3.2 Options

`idvar(`*varname*`)` specifies the variable identifying the individual. `idvar()` is required.

`ynew(`*varname*`)` specifies the new measurement method's variable name. `ynew()` is required.

`yref(`*varname*`)` specifies the reference standard method's variable name. `yref()` is required.

`loa` computes and graphs the (extended) LoA.

`bias` graphs the bias plot.

`precision` graphs the precision plot.

`comp` graphs the comparison plot.

`pdfs` saves the graphs in `.pdf` format (instead of Stata's `.gph` format).

4 Examples

To illustrate the use of `biasplot`, we will consider three simulated datasets:

Simulated dataset 1

$$\begin{aligned} y_{1i} &= -4 + 1.2x_i + \epsilon_{1i} \qquad & \epsilon_{1i}|x_i \sim N\left\{0, (1+0.1x_i)^2\right\} \\ y_{2ij} &= x_i + \epsilon_{2ij} & \epsilon_{2ij}|x_i \sim N\left\{0, (2+0.2x_i)^2\right\} \\ x_i &\sim \text{Uniform}[10{-}40] & \end{aligned}$$

where $i = 1, \ldots, 100$ and the number of repeated measurements of individual i from the reference standard was $n_{1i} = 1$ and $n_{2i} \sim \text{Uniform}[10{-}15]$.

Simulated dataset 2

$$\begin{aligned} y_{1i} &= -4 + 1.2x_i + \epsilon_{1i} \qquad & \epsilon_{1i}|x_i \sim N\left\{0, (1+0.1x_i)^2\right\} \\ y_{2ij} &= x_i + \epsilon_{2ij} & \epsilon_{2ij}|x_i \sim N\left\{0, (2+0.2x_i)^2\right\} \\ x_i &\sim \text{Uniform}[10{-}40] & \end{aligned}$$

where $i = 1, \ldots, 100$ and the number of repeated measurements of individual i from the reference standard was $n_{1i} \sim \text{Uniform}[1{-}5]$ and $n_{2i} \sim \text{Uniform}[10{-}15]$.

Simulated dataset 3

$$y_{1i} = 3 + 0.9x_i + \epsilon_{1i} \qquad \epsilon_{1i}|x_i \sim N\left\{0, (2 + 0.06x_i)^2\right\}$$
$$y_{2ij} = x_i + \epsilon_{2ij} \qquad \epsilon_{2ij}|x_i \sim N\left\{0, (1 + 0.01x_i)^2\right\}$$
$$x_i \sim \text{Uniform}[10{-}40]$$

where $i = 1, \ldots, 100$ and the number of repeated measurements of individual i from the reference standard was $n_{1i} \sim \text{Uniform}[1{-}5]$ and $n_{2i} \sim \text{Uniform}[10{-}15]$.

4.1 Dataset 1

In dataset 1, there are between 10 and 15 repeated measurements by the reference standard and only 1 by the new measurement method for each individual. The differential and proportional biases are -4 and 1.2, respectively. The standard deviation of the measurement errors is heteroskedastic for both measurement methods and increases with the level of the underlying true latent trait. However, the dispersion of the reference standard is twice that of the new measurement method.

```
. use sample1.dta, clear
. biasplot, idvar(id) ynew(y1) yref(y2) loa
(0 observations deleted)
Bias and Precision Plots
Variables - Please check -
id Variable:  id
New Method Y Variable:  y1
Reference Method Y Variable:  y2
Running ...
Generating Bland and Altman LoA Plot
Bland and Altman LoA Plot saved to current working directory
Please wait ...
diff_bias=-3.211, 95%CI=[-5.1218891;-1.300695]
prop_bias=1.189, 95%CI=[1.096019;1.2812098]
Bias Plot Omitted
Comparison Plot Omitted
Precision Plot Omitted
End of Commands
. biasplot, idvar(id) ynew(y1) yref(y2) bias
  (output omitted)
. biasplot, idvar(id) ynew(y1) yref(y2) comp
  (output omitted)
. biasplot, idvar(id) ynew(y1) yref(y2) precision
  (output omitted)
```

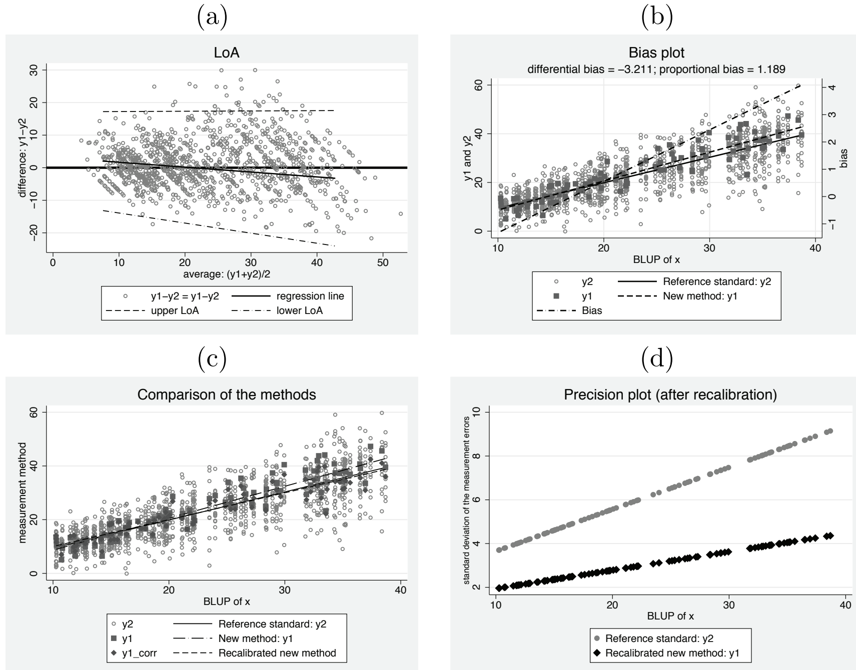

Figure 1. (a) Bland and Altman's LoA plot; (b) bias plot showing the amount of bias of the new measurement method; (c) scatterplot illustrating that recalibration of the new measurement method (that is, `y1_corr`) was effective; and (d) precision plot showing the precision (that is, standard deviation of the measurement error) of each measurement method.

The LoA plot indicates a slight positive bias of the new measurement method for low values of the estimated latent trait level (that is, BLUP of x) and a negative bias for high values. On the contrary, the bias plot illustrates that the bias is negative for low values and positive for high values. The estimated differential bias is -3.21 95% CI $= [-5.36, -1.06]$, and the estimated proportional bias is 1.19 95% CI $= [1.10, 1.28]$. These values are close, and the CIs cover the true values. The precision plot shows that after recalibration, the new measurement method is about twice as precise as the reference standard and that both measurement methods are more precise for lower values than for higher values of the true latent trait. One can see on the comparison plot that recalibration of the new measurement method was effective in removing bias.

4.2 Dataset 2

Dataset 2 is similar to dataset 1, except there are between one to five repeated measurements (instead of just one) by the new measurement method for each individual:

```
. use sample2.dta, clear
. biasplot, idvar(id) ynew(y1) yref(y2) loa
(0 observations deleted)
Bias and Precision Plots
Variables - Please check -
id Variable:  id
New Method Y Variable:  y1
Reference Method Y Variable:  y2
Running ...
Generating Bland and Altman LoA Plot
Bland and Altman LoA Plot saved to current working directory
Please wait ...
diff_bias=-3.527, 95%CI=[-5.4047941;-1.6485342]
prop_bias=1.181, 95%CI=[1.1038746;1.2575773]
Bias Plot Omitted
Comparison Plot Omitted
Precision Plot Omitted
End of Commands
. biasplot, idvar(id) ynew(y1) yref(y2) bias
  (output omitted)
. biasplot, idvar(id) ynew(y1) yref(y2) comp
  (output omitted)
. biasplot, idvar(id) ynew(y1) yref(y2) precision
  (output omitted)
```

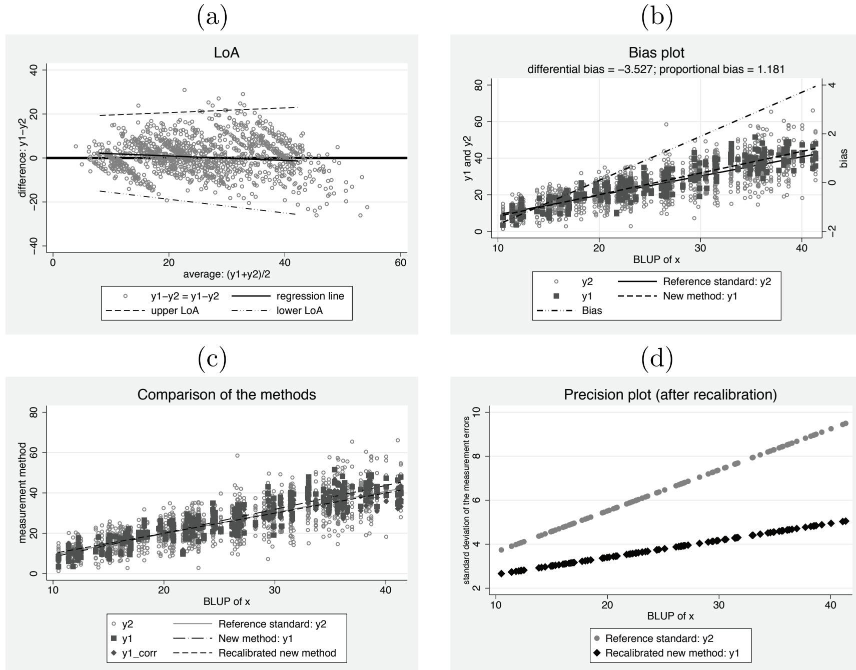

Figure 2. (a) Bland and Altman's LoA plot; (b) bias plot showing the amount of bias of the new measurement method; (c) scatterplot illustrating that the recalibration of the new measurement method (that is, `y1_corr`) was effective; and (d) precision plot showing the precision (that is, standard deviation of the measurement error) of each measurement method.

Consistently with the results for dataset 1, the LoA plot indicates a slight positive bias of the new measurement method for low values of the estimated latent trait level (that is, BLUP of x) and a negative bias for high values. In contrast, the bias plot (correctly) illustrates that the bias is negative for low values and positive for high values. The estimated differential bias is -3.53 95% CI $= [-5.13, -1.92]$, and the estimated proportional bias is 1.18 95% CI $= [1.12, 1.24]$. Note that these CIs are, as expected, more narrow than when we have only one measurement by the new method (that is, dataset 1). The conclusions for the precision and comparison plots are the same as for dataset 1.

4.3 Dataset 3

In dataset 3, there are between 10 to 15 repeated measurements by the reference standard and 1 to 5 repeated measurements by the new measurement method for each individual. The differential bias amounts to 3, and the proportional bias amounts to 0.9. The standard deviation of the measurement errors is heteroskedastic for both measurement methods and increases with the level of the underlying true latent trait. However, the dispersion of the reference standard is much lower than that of the new measurement method.

```
. use sample3.dta, clear
. biasplot, idvar(id) ynew(y1) yref(y2) loa
(0 observations deleted)
Bias and Precision Plots
Variables - Please check -
id Variable:  id
New Method Y Variable:  y1
Reference Method Y Variable:  y2
Running ...
Generating Bland and Altman LoA Plot
Bland and Altman LoA Plot saved to current working directory
Please wait ...
diff_bias=2.714, 95%CI=[1.3859422;4.0417949]
prop_bias=.902, 95%CI=[.84553455;.95922297]
Bias Plot Omitted
Comparison Plot Omitted
Precision Plot Omitted
End of Commands
. biasplot, idvar(id) ynew(y1) yref(y2) bias
  (output omitted)
. biasplot, idvar(id) ynew(y1) yref(y2) comp
  (output omitted)
. biasplot, idvar(id) ynew(y1) yref(y2) precision
  (output omitted)
```

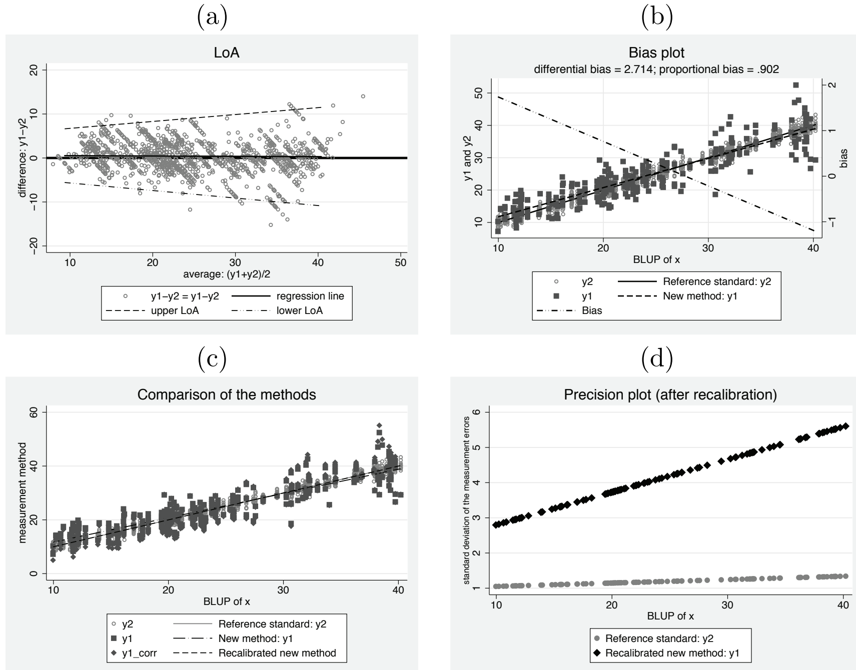

Figure 3. (a) Bland and Altman's LoA plot; (b) bias plot showing the amount of bias of the new measurement method; (c) scatterplot illustrating that the recalibration of the new measurement method (that is, `y1_corr`) was effective; and (d) precision plot showing the precision (that is, standard deviation of the measurement error) of each measurement method.

The LoA plot does not indicate any bias from the new measurement method, whereas the bias plot illustrates that the bias is positive for low values and negative for high values. The estimated differential bias is 2.71 95% CI $= [1.37, 4.06]$, and the estimated proportional bias is 0.9 95% CI $= [0.85, 0.96]$. The precision plot shows that after recalibration, the new measurement method is clearly less precise than the reference standard. Note that the dispersion of measurement errors of the reference standard is almost constant throughout the whole range of the latent trait, whereas that of the new method is, in comparison, sharply increasing. Again, the recalibration of the new measurement method was very effective in removing bias, as illustrated by the comparison plot.

5 Discussion

Using simulated data where the relationship between the true latent trait and the two measurement methods is known, we have illustrated that `biasplot` was effective in removing existing bias of the new measurement method and in assessing the precision of the two measurement methods after recalibration. We have also illustrated that there are settings where Bland and Altman's LoA methodology is misleading, whereas `biasplot` allows one to properly identify, quantify, and correct for any biases.

`biasplot` is widely applicable when one has repeated measurements from the reference standard or possibly as few as one measurement per individual from the new method to be evaluated. Also, `biasplot` is useful when measurement errors are either homoskedastic or heteroskedastic. When one's focus is mainly on identifying and correcting for the bias, as few as three to five repeated measurements from the reference standard and only one from the new measurement method may be large enough to provide good point estimates of the proportional and differential biases and CIs with appropriate coverage rates. However, when one compares the precision of the two measurement methods, it is better to have at least 10 to 15 repeated measurements by the reference standard (and possibly only 1 from the new measurement method) to appropriately estimate the (heteroskedastic) standard deviations of the 2 measurement methods. Actually, it is important to have repeated measurements from the reference standard because our methodology relies essentially on the BLUP of x_i, whereas repeated measurements from the new method will increase precision of the estimated heteroskedastic relationship.

In summary, we have implemented in `biasplot` a new estimation procedure to assess bias and precision of a quantitative measurement method relative to the reference standard, which performs very well in many settings, particularly when several measurements from the reference standard and possibly only one from the new measurement method are available. This method enables measurement errors to be either homoskedastic or heteroskedastic and provides bias, precision, and comparison plots to allow the investigator to visually and clinically appraise the performance of the new method. These plots do not have the shortcomings of Bland and Altman's LoA and are still in the spirit of the original article.

6 References

Bland, J. M., and D. G. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327: 307–310.

———. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8: 135–160.

Carstensen, B. 2010. Comparing methods of measurement: Extending the LoA by regression. *Statistics in Medicine* 29: 401–410.

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

Dunn, G. 2004. *Statistical Evaluation of Measurement Errors: Design and Analysis of Reliability Studies*. 2nd ed. London: Arnold.

Hopkins, W. G. 2004. Bias in Bland–Altman but not regression validity analyses. *Sportscience* 8: 42–46.

Krouwer, J. S. 2008. Why Bland–Altman plots should use X, not (Y+X)/2 when X is a reference method. *Statistics in Medicine* 27: 778–780.

Ludbrook, J. 2010a. Confidence in Altman–Bland plots: A critical review of the method of differences. *Clinical and Experimental Pharmacology and Physiology* 37: 143–149.

———. 2010b. Linear regression analysis for comparing two measurers or methods of measurement: But which regression? *Clinical and Experimental Pharmacology and Physiology* 37: 692–699.

Nawarathna, L. S., and P. K. Choudhary. 2015. A heteroscedastic measurement error model for method comparison data with replicate measurements. *Statistics in Medicine* 34: 1242–1258.

Taffé, P. Forthcoming. Effective plots to assess bias and precision in method comparison studies. *Statistical Methods in Medical Research*.

About the authors

Patrick Taffé is a biostatistician in the Institute for Social and Preventive Medicine at the University of Lausanne and a lecturer in the Faculty of Biology and Medicine at the University of Lausanne, Switzerland.

Mingkai Peng is a statistical associate in the Department of Community Health Sciences at the University of Calgary and a member of the O'Brien Institute for Public Health.

Vicki Stagg is the Senior Statistical Analyst with Calgary Statistical Support (http://calgarystatisticalsupport.com/).

Tyler Williamson is an assistant professor of biostatistics in the Department of Community Health Sciences at the University of Calgary and a member of the Alberta Children's Hospital Research Institute and the O'Brien Institute for Public Health.