blandaltman: a command to create variants of Bland-Altman plots

Mark D. Chatfield
Faculty of Medicine
The University of Queensland
Brisbane, Australia
m.chatfield@uq.edu.au

Tim J. Cole
Great Ormond Street Institute of Child Health
University College London
London, UK

Henrica C. W. de Vet
Department of Epidemiology and Data Science
Amsterdam Medical Center
Amsterdam, The Netherlands

Louise Marquart-Wilson
School of Public Health
The University of Queensland
Brisbane, Australia

Daniel M. Farewell
School of Medicine
Cardiff University
Cardiff, UK

Abstract. Bland-Altman plots are useful in paired data settings such as measurement method comparison studies. A Bland-Altman plot has differences, percentage differences or ratios on the y-axis, and a mean of the data pairs on the x-axis. 95% limits of agreement are added to the plot, indicating the central 95% range of differences, percentage differences or ratios. This range can vary with the mean. We introduce the community-contributed **blandaltman** command, which uniquely in Stata can (1) create Bland-Altman plots featuring ratios in addition to differences and percentage differences, (2) allow the limits of agreement for ratios and percentage differences to vary as a function of the mean, and (3) add confidence intervals, prediction intervals and tolerance intervals to the plots.

Keywords: blandaltman, Bland-Altman plot, limits of agreement, agreement, baplot, batplot, concord, prediction, tolerance, interval, ratio, percentage difference

1 Introduction

When paired data arise from two different measurement techniques, e.g. a new method A and a conventional method B, the data can be plotted as in Figure 1a to visualise the correlation between the two methods. However, this plot is not the best for clearly showing the differences between the methods (Bland and Altman 1986). Bland and Altman (1986, 1999) introduced a plot for visualising agreement, which plots the difference between data pairs versus their arithmetic mean (Figure 1b). This is known as the Bland-Altman plot, and it can be used in other paired data settings such as measurement repeatability (Bland and Altman 1986, 1999) or longitudinal studies (Kirkwood and Sterne 2003). Variants of the Bland-Altman plot have ratios or percentage differences on the y-axis, and may have the geometric mean of data pairs on the x-axis (Dewitte 2002). Note that in geometric terms, the Bland-Altman plot rotates Figure 1a clockwise by 45˚ and linearly rescales the axes. Figure 1b shows the data inside the grey box on Figure 1a.

According to Bland and Altman (1999), 95% limits of agreement (LOA) provide an interval within which 95% of differences between measurements are expected to lie. If these limits are not too large (this is a contextual consideration in light of the intended use of the measurement method), then the methods can be considered interchangeable. Assuming differences are normally distributed, the LOA can be calculated using the mean and standard deviation (SD) of the paired differences (as mean ± 1.96 SD), and they are routinely added to a Bland-Altman plot as a pair of horizontal lines towards the top and bottom of the data cloud.

However, horizontal LOA are "meaningful only if we can assume the bias [*the mean difference*] and variability [*the SD of the difference*] are uniform throughout the range of measurement, assumptions which can be checked graphically" (Bland and Altman 1999, italics ours). In Figure 1b the mean difference changes little as the mean of data pairs varies, but the SD of the difference increases steeply with the mean of data pairs, so that the data cloud is shaped like a left-pointing arrowhead. Bland and Altman (1999) suggest this arrowhead pattern is the most common departure from the assumptions underlying horizontal LOA. In this instance, the LOA need to reflect the varying SD. The plot shows LOA calculated assuming that the SD increases linearly with the mean of data pairs, using the regression-based approach of Bland and Altman (1999) to adjust for non-constant means or SDs of differences.

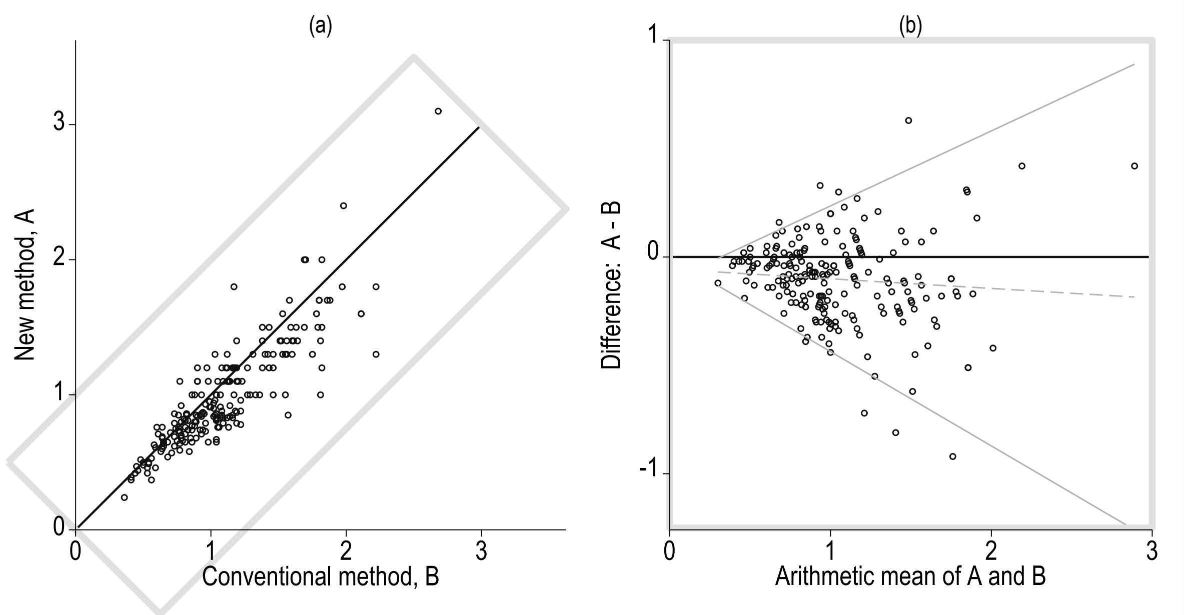

Figure 1. (a) Plot of two methods for measuring retinol-binding-protein-4 (μmol/L) from Brindle et al. (2017), with the line of equality. (b) Bland-Altman plot featuring differences with regression-based estimates of 95% limits of agreement (thin grey solid lines) and mean difference or "bias" (dashed line). The boxes show how rotating (a) clockwise by 45˚ and rescaling the axes leads to (b).

Log transformation often leads to the mean and SD of differences being constant, which in turn justifies using horizontal LOA. Figure 2a shows the same data as Figure 1a, but with the axes scaled logarithmically. Figure 2b is the corresponding Bland-Altman plot, where ratios (A/B) are plotted on the y-axis, and the geometric mean of the data pairs is plotted on the x-axis. Both axes are scaled logarithmically. Plotted this way the rotational symmetry between the plot of raw data and the Bland-Altman plot is preserved.

Bland-Altman plots can also have percentage differences on the y-axis, where the percentage difference is defined by dividing a difference by the arithmetic mean of the data pairs and multiplying by 100%. These percentage differences (which can range from -200% to +200%) are often plotted against the arithmetic mean of the data pairs (Dewitte 2002). Other ways of defining percentage differences are possible (Cole and Altman 2017). Dividing a difference by the logarithmic mean of the data pairs was recommended by economists (Tornqvist, Vartia, and Vartia 1985), and multiplying by 100% produces a percentage difference that can be calculated simply as 100(lnA – lnB) (Cole 2000). These percentage differences (which can range from -∞ to +∞) could be plotted instead of ratios on the y-axis of Figure 2b, and rotational symmetry preserved.

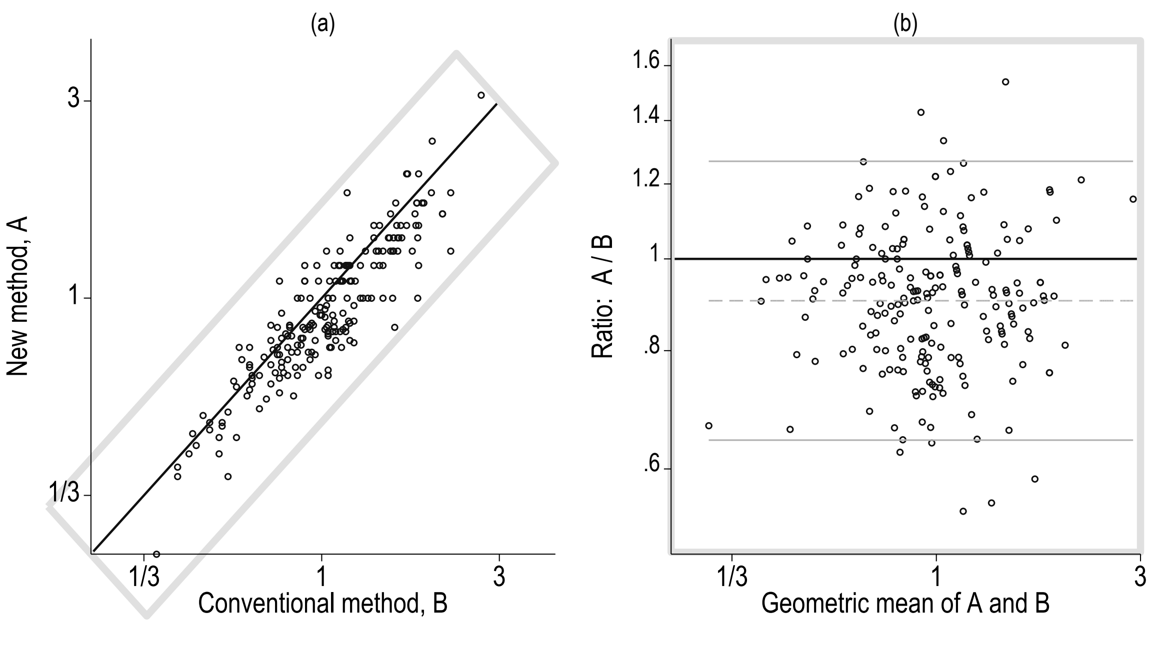

Figure 2. (a) Log-log plot of two methods for measuring retinol-binding-protein-4 (μmol/L) from Brindle et al. (2017), with the line of equality. (b) Bland-Altman plot featuring ratios with estimates of horizontal 95% limits of agreement (thin grey solid lines) and the geometric mean of ratios (dashed line). The boxes show how rotating (a) clockwise by 45˚ and rescaling the axes leads to (b).

Stata has no official Bland-Altman plot command, but there are several community-contributed commands. None of them create Bland-Altman plots with ratios, and only the **agree** command (Doménech 2021) gives percentage differences. For differences, **batplot** (Mander 2005) and **biasplot** (Taffé 2017) can draw regression-based LOA for datasets with one measurement per method per subject, and those with several measurements for the reference method per subject. The commands **concord** (Steichen and Cox 1998), **baplot** (Seed 2000), **kappaetc** (Klein 2018), **agree** (Doménech 2021) and **rmloa** (Linden 2021) present only horizontal LOA.

Bland and Altman (1986, 1999) recommended calculating 95% confidence intervals for LOA. Yet surprisingly, none of the commands cited above displays confidence intervals on the plot[1]. Some authors have recommended prediction and tolerance intervals. (Ludbrook 2010, Vock 2016, Carkeet and Goh 2018, Francq, Berger, and Boachie 2020).

In this article we introduce the **blandaltman** command (Chatfield 2022), which uniquely in Stata can: (1) create Bland-Altman plots featuring ratios in addition to differences and percentage differences, (2) allow LOA for ratios and percentage differences to vary with the mean of data pairs, and (3) add confidence intervals, prediction intervals and tolerance intervals to the plots. We show how, by offering variants of Bland-Altman plots, the command can help decide how best to present LOA for measurement method comparison studies, based on how close to horizontal the regression-based LOA are.

The remainder of the article is organized as follows: section 2 presents the **blandaltman** command, syntax and options; section 3 shows the command in action, and section 4 has some conclusions.

2 The blandaltman command

2.1 Description

blandaltman produces Bland-Altman plots featuring differences, ratios or percentage differences on the y-axis. By default, regression-based estimates of bias and LOA appear on the plot to show how the distribution of differences, ratios or percentage differences, varies with the mean of data pairs. See Appendix for detail. Horizontal lines for bias and LOA can be produced instead.

Summary statistics are provided in the output. The distribution of differences and percentage differences are summarised by mean and SD, and these are used to calculate horizontal LOA. The distribution of ratios is summarised by geometric mean (GMean) and geometric standard deviation (GSD) (Limpert and Stahel 2011). These can be calculated by anti-logging the mean and SD of differences in log-transformed data. For ratios, all calculations are done using log-transformed data, before results such as LOA are anti-logged.

Confidence intervals for the bias and LOA can also be displayed, as well as a prediction interval and (up to three) tolerance intervals (see Appendix for detail), assuming the distribution of differences, ratios or percentage differences does not vary with the mean of data pairs.

2.2 Syntax

blandaltman *varA varB* [if *exp*] [in *range*], plot(*plot_type_list*) [<u>h</u>orizontal noregloa <u>noreg</u>bias hloa hbias <u>l</u>evel(#) <u>predint</u>erval <u>ticonf</u>idence(#) ticonfidence2(#) ticonfidence3(#) cibias ciloa cilevel(#) *minor_options*]

where *plot_type_list* is any combination of *plot_types*:

plot_type	Y-axis	X-axis
difference	A – B	(A+B)/2
ratio	A/B*	GMean(A,B)*
percentlmean	100(A – B)/LMean(A,B) = 100(lnA-lnB)	GMean(A,B)*

percentmean	100(A – B)/{(A+B)/2}	(A+B)/2*

*axis has a logarithmic scale

Multiple plots are created if several *plot_type*s are chosen.

GMean(A,B) = $(A\times B)^{1/2}$ is the geometric mean.

LMean(A,B) = (A – B)/(lnA – lnB) if A≠B, LMean(A,B) = A if A=B, is the logarithmic mean (Cole 2000).

Only positive-valued data are used with the options: ratio and percentlmean. With the exception of A=B=0, data pairs where A≥0 and B≥0 are used with the percentmean option.

2.3 Options

Main options

horizontal display *horizontal* rather than *regression-based* LOA and bias. Equivalent to specifying: noregloa noregbias hloa hbias.

noregloa prevents display of regression-based LOA.

noregbias prevents display of regression-based bias and LOA.

hloa display horizontal LOA.

hbias display horizontal bias. This option is assumed whenever horizontal LOA or a prediction interval or a tolerance interval is requested.

level(#) specifies the level, in percent, for #% LOA, #% prediction interval, #% tolerance interval with ticonf% confidence. The default is level(95).

predinterval display (horizontal) lines for a level% prediction interval.

ticonfidence(#) display (horizontal) lines for a level% tolerance interval with #% confidence.

ticonfidence2(#) display a second level% tolerance interval with #% confidence.

ticonfidence3(#) display a third level% tolerance interval with #% confidence.

ciloa display (exact) cilevel% confidence intervals for horizontal LOA. Requires horizontal or hloa to also be specified.

cibias display a cilevel% confidence interval for horizontal bias. Requires horizontal or hbias to also be specified.

cilevel(#) specifies the level, in percent, for confidence intervals for the bias and LOA. The default is cilevel(95).

Minor options

scopts(*scatter_options*) alter the display of the scatterplot.

regloaopts(*tw_function_options*) alter the display of the regression-based LOA.

regbiasopts(*tw_function_options*)   alter the display of the regression-based bias.

loaopts(*tw_function_options*)   alter the display of the horizontal LOA.

biasopts(*tw_function_options*)   alter the display of the horizontal bias line.

piopts(*tw_function_options*)   alter the display of the prediction interval.

tiopts(*tw_function_options*)   alter the display of the first tolerance interval.

tiopts2(*tw_function_options*)   alter the display of the second tolerance interval.

tiopts3(*tw_function_options*)   alter the display of the third tolerance interval.

ciloaopts(*tw_pcarrowi_options*)   alter the display of the confidence interval for the LOA.

cibiasopts(*tw_pcarrowi_options*)   alter the display of the confidence interval for the bias.

addplot(*plots*) add other plots to the Bland-Altman plot; see [G-3] addplot_option.

twoway_options any of the options for twoway graphs; see [G-3] twoway_options.

where

scatter_options are any of the options allowed with scatter; see [G-2] graph twoway scatter.

tw_function_options are any of the options allowed with twoway function; see [G-2] graph twoway function.

tw_pcarrowi _options are any of the options allowed with twoway pcarrowi; see [G-2] graph twoway pcarrowi.

3 Examples

This section illustrates **blandaltman** in action. The first example shows how the estimated LOA on a Bland-Altman plot vary throughout the range of measurement. The second example shows how to add confidence intervals, a prediction interval, and tolerance intervals to a plot.

3.1 Laboratory measurements: exploring how LOA vary throughout the range of measurement

Brindle et al. (2017) described the simultaneous assessment of seven micronutrient and inflammation status biomarkers via a multiplex immunoassay method in a population of pregnant women. Results from their 7-Plex assay were compared with conventional immunoassay results on N=206 plasma samples. We focus on retinol-binding-protein-4 (Figure 1a), a surrogate biomarker for vitamin A deficiency, where low levels indicate deficiency. For simplicity we generate variables named A and B to represent measurements obtained using the new and conventional methods respectively.

```
. use http://fmwww.bc.edu/repec/bocode/l/labmeasures.dta, clear
. generate A = plexrbp4µmoll
. generate B = nimanurbp4µmoll
. blandaltman A B, plot(difference ratio percentlmean percentmean)
```

(*see output in Appendix*)

The above line of syntax produces the four Bland-Altman plots shown in Figure 3. As seen in Figure 3a, the estimated LOA for differences are far from horizontal. In contrast, the estimated LOA for ratios (Figure 3b) and percentage differences (Figure 3c and 3d) are close to horizontal, so horizontal LOA would be justified. Note that Figures 3b and 3c are equivalent (they differ only in their y-axis labelling). Figure 3d looks similar to Figure 3c, but plots a different definition of percentage difference against a different mean.

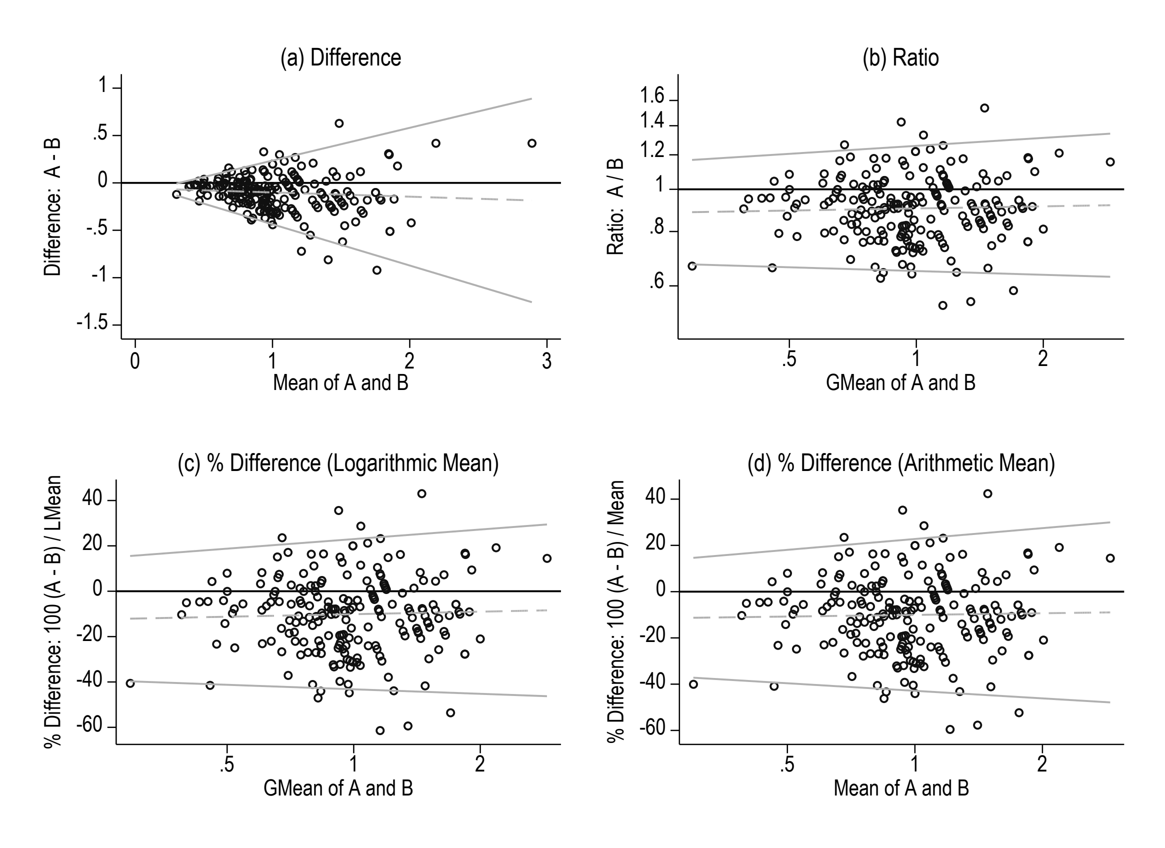

Figure 3. Bland-Altman plots featuring (a) differences, (b) ratios, (c) percentage differences using the logarithmic mean as denominator and (d) percentage differences using the arithmetic mean as denominator. Each plot shows regression-based estimates of LOA (grey solid lines) and bias (dashed line). (b) and (c) are equivalent plots, while (d) is different.

The values for horizontal LOA are displayed when the option **horizontal** is specified:

```
. blandaltman A B, plot(ratio percentlmean percentmean) horizontal

A: A
B: B

RATIOS...
Calculation                 N       GMean        GSD       Interval(s)
A/B                       206    .9033576   1.188854
                                95% limits of agreement:  .6435914  1.267971

PERCENTAGE DIFFERENCES (using Logarithmic Mean as denominator)...
Calculation                 N        Mean         SD       Interval(s)
100*(A-B)/LMean(A,B)      206   -10.16367   17.29902
                                95% limits of agreement: -44.06912  23.74177

PERCENTAGE DIFFERENCES (using Mean as denominator)...
Calculation                 N        Mean         SD       Interval(s)
100*(A-B)/[(A+B)/2]       206   -10.07663   17.10901
                                95% limits of agreement: -43.60967  23.45642
```

LOA from both definitions of percentage difference will often be very similar, as is the case here. Assuming both percentage differences are approximately normally distributed, and ratios are approximately lognormally distributed, there is little to choose between the 3 LOA above. Convention or personal preference may be the deciding factor in selecting one.

The previous code produces three Bland-Altman plots with horizontal LOA, one of which is shown in Figure 4. It features percentage differences (using arithmetic mean as denominator) on the y-axis and arithmetic mean on the x-axis, which is a popular choice in bioanalytical method validation studies (Dewitte 2002). By default, the x-axis is scaled logarithmically, which helps to space the data out more evenly. However, if a user wants a linear scale instead, they can specify the option **xscale(nolog)**. Assuming these percentage differences are approximately normally distributed, LOA are estimated to be -44% and +23%.

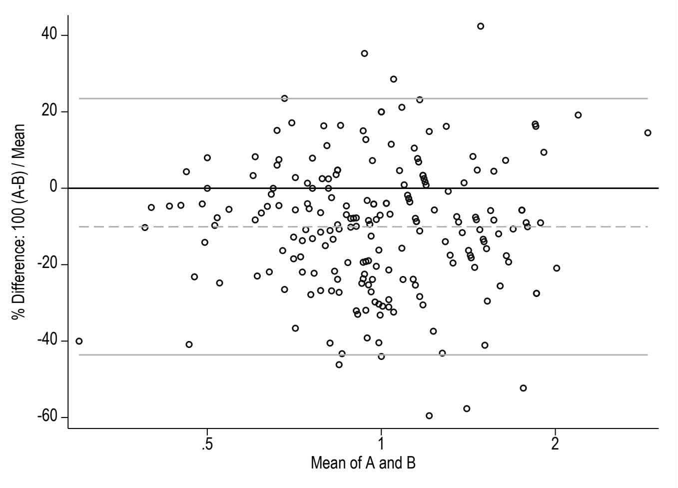

Figure 4. Bland-Altman plot featuring percentage differences (using arithmetic mean as denominator) assuming horizontal 95% LOA (grey solid lines) and bias (dashed line).

For control over the labelling of axes, **xlabel()** or **ylabel()** options can be specified. For example, in our plots featuring ratios (Figures 2b and 3b), we specified **ylabel(0.6 (0.2) 1.6)**. See Cox (2018, 2020) for other ways of labelling log scaled axes.

3.2 Peak Expiratory Flow Rate (PEFR) data: adding confidence intervals, prediction interval, and tolerance intervals

To demonstrate their method, Bland and Altman (1986) measured peak expiratory flow rate in 17 persons using both a Wright flow meter and a mini Wright flow meter. Like them, and others (Ludbrook 2010, Carkeet 2015, Vock 2016), we use the first measurement by each method. Figure 5 shows the data on an (overly busy) Bland-Altman plot. Bland and Altman saw no obvious relation between differences and means, and assumed differences were normally distributed. They estimated[2] 95% LOA to be -2.1 ± 2 × 38.8 l/min.

Figure 5 illustrates the various intervals that **blandaltman** can produce—see the Appendix and the references in this section for the meaning of prediction and tolerance intervals. The figure was created with the following syntax:

```
. use http://fmwww.bc.edu/repec/bocode/p/PEFR.dta, clear

. blandaltman Wright Mini, plot(difference) horizontal ///
```

```
  ciloa cibias predinterval ticonfidence(95) ///
  loaopts(lc(gs11) lp(solid)) ciloaopts(mc(gs11) lc(gs11) lp(solid)) ///
  piopts(lc(gs1) lp(dash_dot)) ///
  tiopts(lc(gs1) lp(dot)) ///
  legend(on order(2 "Bias (& 95% CI)" 4 "95% limits of agreement (& exact 95% CI)"
6 "95% prediction interval" 8 "95% tolerance interval with 95% confidence")
rowgap(*.7) col(1))

A: Wright                Wright peak expiratory flow rate (l/min)
B: Mini                  Mini Wright peak expiratory flow rate (l/min)

DIFFERENCES...
Calculation              N        Mean          SD        Interval(s)
A-B                     17   -2.117647    38.76513
                            95% limits of agreement: -78.09591   73.86061
                            95% prediction interval: -86.67853   82.44323
        95% tolerance interval with 95% confidence: -113.4634   109.2281
                                      95% CI (LLOA): -124.1608 -53.09493
                                      95% CI (ULOA):  48.85964   119.9255
                                95% CI (Mean diff.): -22.04884   17.81354
```

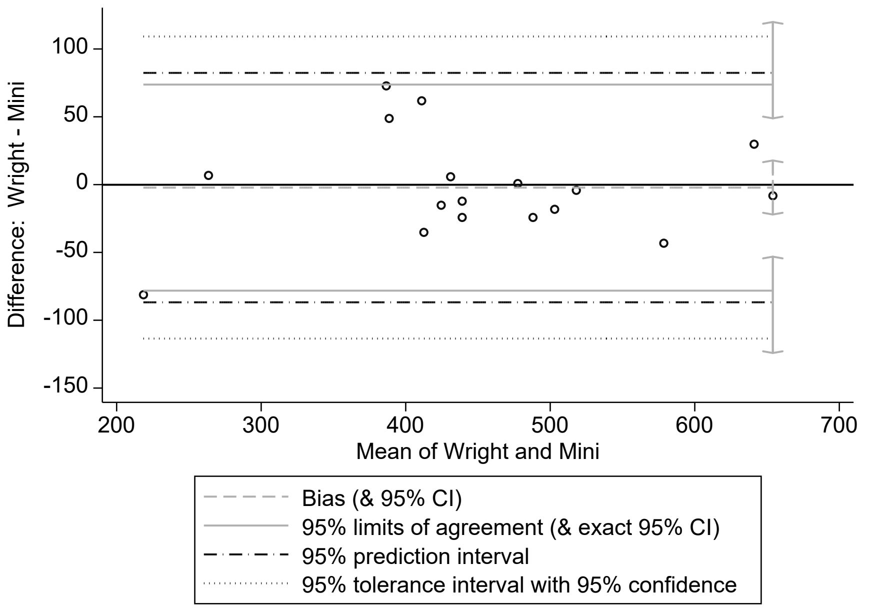

Figure 5. Bland-Altman plot of two methods measuring peak expiratory flow rate (l/min), illustrating the variety of intervals that **blandaltman** can produce. No one author has suggested using all of these intervals.

We now report which of these intervals different authors have recommended, and how they can be implemented with **blandaltman**. Bland and Altman recommended estimates of 95% LOA and 95% confidence intervals for LOA, and this advice features in reporting standards (Gerke 2020). Royston and Matthews (1991) considered methods to provide a best estimate of an interval containing the central 95% of a distribution. They considered the interval bounded by mean $\pm$ 1.96 SD (i.e. LOA) to be a good estimate, and they viewed a 95%-expectation tolerance interval (equivalent to a 95% prediction interval) to be of questionable value, as they did a 95% tolerance interval with ≥90% confidence.

A few authors prefer not to present estimates of LOA (Ludbrook 2010, Vock 2016, Carkeet and Goh 2018, Francq, Berger, and Boachie 2020). Some prefer a 95% prediction interval instead (Ludbrook 2010, Francq, Berger, and Boachie 2020), but not Vock (2016) who argued this is rarely appropriate. Others prefer a 95% tolerance interval with 50% confidence (Carkeet and Goh 2018).

Vock (2016) encouraged reporting a 95% tolerance interval with 95% confidence, as did Ludbrook (2010)[3]. Francq, Berger, and Boachie (2020) thought this interval may be too large (see Table A1 for how intervals depend on sample size), and suggested a 95% tolerance interval with 80% or 90% confidence might be presented if needed. Carkeet and Goh (2018) recommended reporting a 95% tolerance interval with 2.5% confidence as well as a 95% tolerance interval with 97.5% confidence.

The following syntax creates Bland-Altman plots with intervals recommended by the above-mentioned authors:

```
. blandaltman Wright Mini, plot(difference) name(Bland_Altman)  ///
horizontal ciloa ///
legend(on order(2 "Bias" 4 "95% limits of agreement (& exact 95% CI)") col(1))

. blandaltman Wright Mini, plot(difference) name(Ludbrook)  ///
noreg hbias predinterval ticonfidence(95) ///
legend(on order(2 "Bias" 4 "95% prediction interval" 6 "95% tolerance interval with
95% confidence") col(1))

. blandaltman Wright Mini, plot(difference) name(Francq_et_al) ///
noreg hbias predinterval ticonfidence(80) ///
legend(on order(2 "Bias" 4 "95% prediction interval" 6 "95% tolerance interval with
80% confidence") col(1))

. blandaltman Wright Mini, plot(difference) name(Vock) ///
noreg hbias ticonfidence(95) ///
legend(on order(2 "Bias" 4 "95% tolerance interval with 95% confidence") col(1))

. blandaltman Wright Mini, plot(difference) name(Carkeet_Goh) ///
noreg hbias ticonfidence(2.5) ticonfidence2(50) ticonfidence3(97.5) ///
legend(on order(2 "Bias" 4 "95% tolerance interval with 2.5% confidence" 6 "95%
tolerance interval with 50% confidence" 8 "95% tolerance interval with 97.5%
confidence") col(1))
```

4 Conclusion

The **blandaltman** command should help Stata users assessing agreement in measurement method comparison studies to follow the advice of Bland and Altman by visually assessing how estimated LOA vary throughout the range of measurement, and by reporting corresponding confidence intervals. The command is also flexible enough to allow users to follow recommendations of other authors involving the presentation of a prediction interval and/or a tolerance interval(s). More generally, in other paired data settings, the command could help users decide whether to summarise differences, ratios or percentage differences defined in one of two ways.

Notes

<Footnotes to go here instead of at the end of this document>

5 Acknowledgement

blandaltman uses some syntax from the command **niceloglabels** (Cox, 2018, 2020) to provide publication-ready plots when axes are scaled logarithmically. We thank the editor and a referee for help structuring the article, and Professor Annette Dobson for helpful feedback.

6 Programs and supplemental materials

To install the latest version of software files, type

`. ssc install blandaltman` (to install program files)

`. net get blandaltman` (to install ancillary files)

7 References

Bland, J. M., and D. G. Altman. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 327: 307–10.

Bland, J. M., and D. G. Altman. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8: 135–160.

Brindle, E., L. Lillis, R. Barney, S. Y. Hess, K.R. Wessells, C. T. OueÂdraogo, S. Stinca, M. Kalnoky, R. Peck, A. Tyler, C. Lyman, and D. S. Boyle. 2017. Simultaneous assessment of iodine, iron, vitamin A, malarial antigenemia, and inflammation status biomarkers via a multiplex immunoassay method on a population of pregnant women from Niger. *PLoS ONE* 12: e0185868.

Carkeet, A. 2015. Exact parametric confidence intervals for Bland–Altman limits of agreement. *Optometry and Vision Science* 92: e71–e80.

Carkeet, A., and Y. T. Goh. 2018. Confidence and coverage for Bland–Altman limits of agreement and their approximate confidence intervals. *Statistical Methods in Medical Research* 27: 1559–1574.

Chatfield, M. 2021. tolerance: Stata module to calculate tolerance intervals (normal distribution). Statistical Software Components S459009, Boston College Department of Economics.

Chatfield, M. 2022. blandaltman: Stata module to create Bland-Altman plots featuring differences, percentage differences or ratios, with options to add a variety of lines and intervals. Statistical Software Components S459040, Boston College Department of Economics.

Cole, T. J. 2000. Sympercents: symmetric percentage differences on the 100 log(e) scale simplify the presentation of log transformed data. *Statistics in Medicine* 19: 3109–25.

Cole, T. J., and D. G. Altman. 2017. Statistics Notes: What is a percentage difference? *BMJ* 358: j3663.

Cox, N. J. 2018. Speaking Stata: Logarithmic binning and labelling. *Stata Journal* 18: 262–286.

Cox, N. J. 2020. Software update for niceloglabels. *Stata Journal* 20: 1028–1030.

Dewitte, K. 2002. Application of the Bland–Altman Plot for interpretation of method-comparison studies: a critical investigation of its practice. *Clinical Chemistry* 48: 799–801.

Doménech J. M. 2021. Passing-Bablok & Bland-Altman methods: User-written command agree for Stata [computer program]. V1.2.2. Bellaterra: Universitat Autònoma de Barcelona.

Francq, B. G., M. Berger, and C. Boachie. 2020. To tolerate or to agree: A tutorial on tolerance intervals in method comparison studies with BivRegBLS R Package. *Statistics in Medicine* 39: 4334–4349.

Gerke, O. 2020. Reporting standards for a Bland-Altman agreement analysis: a review of methodological reviews. *Diagnostics* 10: 334.

Howe, W. G. 1969. Two-sided tolerance limits for normal populations - some improvements. *Journal of the American Statistical Association* 64: 610–620.

Kirkwood, B. R., and J. A. C. Sterne. 2003. *Essential Medical Statistics.* Second edition. John Wiley & Sons, Inc.

Klein, D. 2018. Implementing a general framework for assessing interrater agreement in Stata. *Stata Journal* 18: 871–901.

Limpert, E., and W. A. Stahel. 2011. Problems with using the normal distribution – and ways to improve quality and efficiency of data analysis. *PLoS ONE* 6: e21403.

Linden, A. 2021. rmloa: Stata module to compute limits of agreement for data with repeated measures. Statistical Software Components S458980, Boston College Department of Economics.

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

Mander, A. 2005. batplot: Stata module to produce Bland-Altman plots accounting for trend. Statistical Software Components S448703, Boston College Department of Economics.

Meeker, W. Q., G. J. Hahn, and L. A. Escobar. 2017. *Statistical Intervals: A guide for practitioners and researchers.* Second edition. John Wiley & Sons, Inc.

Royston, P., and J. N. S. Matthews. 1991. Estimation of reference ranges from normal samples. *Statistics in Medicine* 10: 691–695.

Seed, P. 2000. sbe33: Comparing several methods of measuring the same quantity. *Stata Technical Bulletin* 54: 23–25.

Shieh, G. 2018. The appropriateness of Bland-Altman's approximate confidence intervals for limits of agreement. *BMC Medical Research Methodology* 18:45.

Steichen, T. J., and N. J. Cox. 1998. sg84: Concordance correlation coefficient. *Stata Technical Bulletin* 54: 23–25.

Taffé P., M. Peng, V. Stagg, and T. Williamson. 2017. biasplot: A package to effective plots to assess bias and precision in method comparison studies. *Stata Journal* 17: 208–221.

Tornqvist, L., P. Vartia, and Y.O. Vartia. 1985. How should relative changes be measured? *The American Statistician* 39:43–46.

Vangel, M. G. 2005. Tolerance Interval. In *The Encyclopedia of Biostatistics*, 2nd Edition, edited by P. Armitage and T. Colton. John Wiley & Sons, Ltd.

Vardeman, S. B. 1992. What about the Other Intervals? *The American Statistician* 46: 193–197.

Vock, M. 2016. Intervals for the assessment of measurement agreement: Similarities, differences, and consequences of incorrect interpretations. *Biometrical Journal* 58: 489–501.

About the authors

Mark Chatfield is a biostatistician, and has enjoyed using Stata for 20 years. He is currently undertaking a PhD, amongst other things.

Tim Cole is a professor in medical statistics. He has longstanding interests in the statistics of human growth, where logarithms are useful for modelling purposes.

Henrica de Vet is a professor in clinimetrics. A clinical epidemiologist by training, her focus is on methodological research on the development and improvement of measurement instruments.

Louise Marquart-Wilson is a biostatistician with interests in generalised linear mixed models.

Daniel Farewell is a professor in statistics. He is interested in the development of novel statistical theory and methodology, especially for causal inference and missing data.

A Appendix

A.1 Calculation and reporting of regression-based estimates for bias and 95% LOA

The approach described in Bland and Altman (1999, sec 3.2), for differences, can be written as follows:

A linear regression is used to estimate how the mean of differences (Mean_Y) varies linearly with the mean of the data pairs. A second linear regression is used to estimate how the SD of differences (SD_Y) varies linearly with the mean of the data pairs. These two relationships are then combined to estimate LOA (95%LOA_Y) that vary linearly with an average of the data pairs.

Equivalently, for a Bland-Altman plot of differences against means, use Y = A – B and X = (A+B)/2 in the more general approach we outline below.

Mean_Y (or Bias):

Fit a linear regression of Y on X

```
. regress Y X
```

The resulting regression equation estimates the mean of Y as a linear function of X:

Mean_Y = $b_0 + b_1$ X

SD_Y

First, obtain the residuals from the above regression.

. predict resid, resid

Second, calculate the absolute values of the residual, and adjust by multiplying by $\sqrt{\frac{\pi}{2}}$. (Given a value of X, it is assumed that residuals are normally distributed with standard deviation σ, and therefore the mean of absolute residuals is $\sigma\sqrt{\frac{2}{\pi}}$.)

. generate adj_abs_resid = abs(resid) * sqrt(_pi/2)

Third, fit a linear regression of adj_abs_resid on X

. regress adj_abs_resid X

The resulting regression equation estimates the standard deviation of Y as a linear function of X:

SD_Y = $b_2 + b_3$ X

Estimated 95% LOA for Y

95%LOA_Y = Mean_Y ± 1.96 SD_Y

= $b_0 + b_1$ X ± 1.96 ($b_2 + b_3$ X)

i.e.

Lower limit (LLOA) = $b_0 - 1.96\ b_2 + (b_1 - 1.96\ b_3)$X

Upper limit (LLOA) = $b_0 + 1.96\ b_2 + (b_1 + 1.96\ b_3)$X

For a Bland-Altman plot of ratios against geometric means, we apply the approach to log-transformed data but express relationships in terms of ratios and geometric means:

we use Y = lnA – lnB = ln(A/B) = ln(Ratio) and X = (lnA + lnB)/2 = ln(GMean(A,B))

For a Bland-Altman plot of percentage differences 100(lnA – lnB)%, against geometric means:

we use Y = 100(lnA – lnB) and X = (lnA + lnB)/2 = ln(GMean(A,B))

For a Bland-Altman plot of percentage differences 100(A – B)/[(A+B)/2]%, against arithmetic means:

we use Y = 100(A – B)/[(A+B)/2] and X = ln((A+B)/2) = ln(Mean(A,B))

The following output relates to section 3.1:

```
. blandaltman plexrbp4µmoll nimanurbp4µmoll, ///
plot(difference percentmean percentlmean ratio)

A: plexrbp4µmoll          7-Plex RBP4 (µmol/L)
B: nimanurbp4µmoll        NiMaNu RBP4 (µmol/L)

DIFFERENCES...
Calculation                 N        Mean         SD        Interval(s)
A-B                       206   -.1022816   .2013955
```

```
. regress difference mean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      1.51
       Model |  .061142301         1  .061142301   Prob > F        =    0.2204
    Residual |  8.25368545       204  .040459242   R-squared       =    0.0074
-------------+----------------------------------   Adj R-squared   =    0.0025
       Total |  8.31482775       205  .040560135   Root MSE        =    .20114

------------------------------------------------------------------------------
    __000002 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __000003 |  -.0445447   .0362355    -1.23   0.220    -.1159888    .0268994
       _cons |  -.0559129   .0402386    -1.39   0.166    -.1352497     .023424
------------------------------------------------------------------------------
              -> regression-based bias:      -.0559129 + -.0445447 × Mean(A,B)

. regress adj_abs_resid mean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =     48.97
       Model |  1.22288054         1  1.22288054   Prob > F        =    0.0000
    Residual |  5.09388346       204  .024970017   R-squared       =    0.1936
-------------+----------------------------------   Adj R-squared   =    0.1896
       Total |  6.31676401       205  .030813483   Root MSE        =    .15802

------------------------------------------------------------------------------
    __00000C | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __000003 |   .1992128   .0284665     7.00   0.000     .1430865    .2553392
       _cons |   -.027725   .0316114    -0.88   0.381    -.0900519    .0346019
------------------------------------------------------------------------------
              -> regression-based SD:           -.027725 +  .1992128 × Mean(A,B)
              -> regression-based 95% LLOA:  -.0015729 + -.4349947 × Mean(A,B)
              -> regression-based 95% ULOA:  -.1102529 +  .3459053 × Mean(A,B)

PERCENTAGE DIFFERENCES (using Mean as denominator)...
Calculation              N        Mean          SD       Interval(s)
100*(A-B)/[(A+B)/2]    206   -10.07663    17.10901

. regress percentmean ln_mean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      0.10
       Model |  29.4792833         1  29.4792833   Prob > F        =    0.7518
    Residual |  59977.7646       204  294.00865    R-squared       =    0.0005
-------------+----------------------------------   Adj R-squared   =   -0.0044
       Total |  60007.2439       205  292.718263   Root MSE        =    17.147

------------------------------------------------------------------------------
    __000005 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __00000A |   1.027934   3.246288     0.32   0.752    -5.372644    7.428513
       _cons |  -10.04945   1.197744    -8.39   0.000      -12.411   -7.687909
------------------------------------------------------------------------------
              -> regression-based bias:  -10.04945 +  1.027934 × ln(Mean(A,B))

. regress adj_abs_resid ln_mean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      1.33
       Model |   239.33826         1   239.33826   Prob > F        =    0.2503
    Residual |  36737.0171       204  180.083417   R-squared       =    0.0065
-------------+----------------------------------   Adj R-squared   =    0.0016
       Total |  36976.3554       205  180.372465   Root MSE        =     13.42

------------------------------------------------------------------------------
    __00000E | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
```

```
-------------+----------------------------------------------------------------
    __00000A |   2.928957   2.540644     1.15   0.250    -2.080331    7.938246
       _cons |   16.74617   .9373909    17.86   0.000     14.89795    18.59439
------------------------------------------------------------------------------
          -> regression-based SD:         16.74617 +  2.928957 × ln(Mean(A,B))
          -> regression-based 95% LLOA:  -42.87134 + -4.712716 × ln(Mean(A,B))
          -> regression-based 95% ULOA:   22.77244 +  6.768585 × ln(Mean(A,B))

PERCENTAGE DIFFERENCES (using Logarithmic Mean as denominator)...
Calculation               N        Mean          SD       Interval(s)
100*(A-B)/LMean(A,B)    206   -10.16368    17.29902

. regress percentlmean ln_gmean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      0.24
       Model |  72.0980071         1  72.0980071   Prob > F        =    0.6247
    Residual |  61275.3688       204  300.369455   R-squared       =    0.0012
-------------+----------------------------------   Adj R-squared   =   -0.0037
       Total |  61347.4668       205  299.255936   Root MSE        =    17.331

------------------------------------------------------------------------------
    __000006 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __000009 |   1.609633   3.285436     0.49   0.625    -4.868133    8.087398
       _cons |   -10.1131   1.211925    -8.34   0.000     -12.5026   -7.723594
------------------------------------------------------------------------------
          -> regression-based bias:      -10.1131 +  1.609633 × ln(GMean(A,B))

. regress adj_abs_resid ln_gmean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      0.78
       Model |  145.256099         1  145.256099   Prob > F        =    0.3771
    Residual |  37812.2272       204  185.354055   R-squared       =    0.0038
-------------+----------------------------------   Adj R-squared   =   -0.0011
       Total |  37957.4833       205  185.158455   Root MSE        =    13.614

------------------------------------------------------------------------------
    __00000G | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __000009 |   2.284716    2.58087     0.89   0.377    -2.803884    7.373317
       _cons |   16.89376   .9520259    17.75   0.000     15.01669    18.77084
------------------------------------------------------------------------------
          -> regression-based SD:        16.89376 +  2.284716 × ln(GMean(A,B))
          -> regression-based 95% LLOA: -43.22427 + -2.868329 × ln(GMean(A,B))
          -> regression-based 95% ULOA:  22.99807 +  6.087595 × ln(GMean(A,B))

RATIOS...
Calculation               N       GMean         GSD       Interval(s)
A/B                     206    .9033576    1.188854

. regress ln_ratio ln_gmean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      0.24
       Model |  .007209808         1  .007209808   Prob > F        =    0.6247
    Residual |  6.12753704       204  .030036946   R-squared       =    0.0012
-------------+----------------------------------   Adj R-squared   =   -0.0037
       Total |  6.13474685       205  .029925594   Root MSE        =    .17331

------------------------------------------------------------------------------
    __000008 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __000009 |   .0160963   .0328544     0.49   0.625    -.0486813     .080874
       _cons |   -.101131   .0121192    -8.34   0.000     -.125026   -.0772359
------------------------------------------------------------------------------
          -> regression-based GMean Ratio:      .9038146 × GMean(A,B)^ .0160963
```

```
. regress adj_abs_resid ln_gmean

      Source |       SS           df       MS      Number of obs   =       206
-------------+----------------------------------   F(1, 204)       =      0.78
       Model |  .014525613         1  .014525613   Prob > F        =    0.3771
    Residual |   3.7812229       204  .018535406   R-squared       =    0.0038
-------------+----------------------------------   Adj R-squared   =   -0.0011
       Total |  3.79574851       205  .018515846   Root MSE        =    .13614

------------------------------------------------------------------------------
    __00000I | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    __000009 |    .0228472   .0258087     0.89   0.377    -.0280388    .0737332
       _cons |    .1689376   .0095203    17.75   0.000     .1501669    .1877084
------------------------------------------------------------------------------
        -> regression-based GSD Ratio:        1.184046 × GMean(A,B)^ .0228472
        -> regression-based 95% LLOA Ratio:  .6490518 × GMean(A,B)^-.0286833
        -> regression-based 95% ULOA Ratio:  1.258576 × GMean(A,B)^  .060876
```

The output for ratios makes use of the fact: $\exp(\alpha + \beta\ln(\text{GMean}(A,B))) = \exp(\alpha) \times \text{GMean}(A,B)^{\beta}$

A.2 Confidence intervals for LOA, prediction interval and tolerance intervals

Here we detail the calculation of confidence intervals for LOA, and the calculation and meaning of prediction and tolerance intervals produced by **blandaltman**.

It is assumed that differences (or percentage differences or log ratios) $\mathbf{y} = \{y_1, y_2, \dots y_n\}$ are randomly sampled from a normal distribution with mean μ, standard deviation σ, and cumulative distribution function F. We denote the sample size n, the sample mean $\bar{y}$, and the sample standard deviation s.

95% confidence intervals for LOA

Bland and Altman (1999) viewed $\bar{y} \pm 1.96$ s as estimates of:

lower limit of agreement (LLOA) = $\mu - 1.96\,\sigma$ (i.e. 2.5th percentile of population)

upper limit of agreement (ULOA) = $\mu + 1.96\,\sigma$ (i.e. 97.5th percentile of population)

They acknowledged that sampling error affects the estimates of μ, σ and LOA, and proposed calculating a 95% confidence interval for LOA. They described approximate methods assuming the sample size is large. However, there exists an exact method based on the noncentral t distribution (Carkeet 2015, Shieh 2018) which is implemented in **blandaltman**.[4] The formulae are:

for the ULOA: $\bar{y} + k_{inner} \times s$ to $\bar{y} + k_{outer} \times s$

for the LLOA: $\bar{y} - k_{outer} \times s$ to $\bar{y} - k_{inner} \times s$

where $k_{inner} = t_{n-1,1.96\sqrt{n},0.025}\sqrt{\frac{1}{n}}$ and $k_{outer} = t_{n-1,1.96\sqrt{n},0.975}\sqrt{\frac{1}{n}}$

and the quantities $t_{n-1,1.96\sqrt{n},0.025}$ and $t_{n-1,1.96\sqrt{n},0.975}$ are the 0.025 and 0.975 quantiles of the noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter $1.96\sqrt{n}$. In contrast to the approximate confidence intervals, these exact confidence intervals will not appear symmetric about the LOA.

Prediction interval

A two-sided 95% prediction interval for a single future observation y_{n+1} (Vardeman 1992, Meeker, Hahn, and Escobar 2017) is a random interval [L(**y**), U(**y**)] constructed such that

$$\text{Prob}\{L < y_{n+1} < U\} = 0.95$$

That is, if the process of {(i) gathering a sample of size n, (ii) constructing a 95% prediction interval, and (iii) gathering a single additional y_{n+1}} is repeated infinitely many times, then 95% of the prediction intervals will contain y_{n+1}.

The 95% prediction interval is calculated as: $\bar{y} \pm k_{PI} \times s$

where $k_{PI} = t_{n-1,0.975}\sqrt{1 + \frac{1}{n}}$

and the quantity $t_{n-1,0.975}$ is the 0.975 quantile of Student's t-distribution with $n - 1$ degrees of freedom.

Tolerance intervals

Tolerance intervals are statistical intervals that contain at least a specified percentage of a population, either (a) on average, or (b) with a stated confidence (Vangel 2005, Vardeman 1992).

<u>(a) 95%-expectation tolerance interval</u>

A two-sided 95%-expectation tolerance interval is a random interval [L(**y**), U(**y**)] constructed such that

$$E\{F(U) - F(L)\} = 0.95$$

That is, if the process of {(i) gathering a sample of size n, (ii) constructing a 95%-expectation tolerance interval, and (iii) calculating what percentage of the population is contained by the interval} is repeated infinitely many times, then the mean (i.e. expected) percentage will be 95%.

Mathematically, it is equivalent to the above-mentioned 95% prediction interval.

<u>(b) 95% tolerance interval with C% confidence</u>

A two-sided 95% tolerance interval with C% confidence is a random interval [L(**y**), U(**y**)] constructed such that

$$\Pr\{F(U) - F(L) \geq 0.95\} = C\%$$

That is, if the process of {(i) gathering a sample of size n, (ii) constructing a 95% tolerance interval with C% confidence, and (iii) calculating what percentage of the population is contained by the interval} is repeated infinitely many times, then C% of these intervals will contain at least 95% of the population.

There is no closed-form expression. **blandaltman** calculates an approximate two-sided 95% tolerance interval with C% confidence (Howe 1969): $\bar{y} \pm k_{TI} \times s$

where $k_{TI} = 1.96\sqrt{\left(1 + \frac{1}{n}\right)\left(\frac{n-1}{\chi^2_{n-1,1-C/100}}\right)\left(1 + \frac{n-3-\chi^2_{n-1,1-C/100}}{2(n+1)^2}\right)}$

and the quantity $\chi^2_{n-1,1-C/100}$ is the (1 – C/100) quantile of a χ^2 distribution with $n - 1$ degrees of freedom.

Table A1. Factors used to calculate (i) a 95% confidence interval for LOA, (ii) a 95% prediction interval and (iii) approximate 95% tolerance intervals with C% confidence. Intervals are calculated as described in A.2.

n	(i) k_{inner}, k_{outer}	(ii) k_{PI}	(iii) k_{TI}	
			C = 50	C = 95
10	1.16, 3.80	2.37	2.13	3.41
20	1.36, 3.01	2.14	2.04	2.76
50	1.55, 2.53	2.03	1.99	2.38
100	1.66, 2.34	1.99	1.98	2.23
∞	1.96, 1.96	1.96	1.96	1.96

[1] Confidence intervals are reported in the non-graphical output using **agree.**
[2] Factors 2 and 1.96 respectively are used in their 1986 and 1999 articles.
[3] Vock (2016) also considered an interval formed by the outer confidence limits for the LOA as an alternative.
[4] While this method is currently not implemented in **[R] centile**, it is now implemented in the recently revised community-contributed command **tolerance** (Chatfield 2021).