Research Article

Statistics in Medicine

Received 24 July 2008, Accepted 24 September 2009 Published online 8 December 2009 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3769

Comparing methods of measurement: Extending the LoA by regression

Bendix Carstensen[*†]

Method comparison studies are usually analyzed by computing limits of agreement (LoA). If only one measurement by each method is taken on each person, and the difference across the range is not constant, it has been suggested (*Stat. Methods Med. Res.* 1999; 8:136–160) to regress the differences on the averages and use the resulting equation to construct LoA.

LoA can be converted to a prediction foumula for one method given a measurement by the other. The meaning of the regression of differences on means is clarified in the framework of a proper model and prediction equations linking one method to another are devised. The performance of this model based method is evaluated against the simple approach proposed earlier and against the Deming regression. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: method comparison; Bland–Altman plot; agreement; differences vs averages; non-constant difference; regression; prediction; calibration

1. Introduction

Comparing two methods of measurement is normally done by computing limits of agreement (LoA), i.e. prediction limits for a future difference between measurements with the two methods. When the difference is not constant it is not clear what this means, since the difference between the methods depends on the average; hence, unlike the case where the difference is constant, LoA cannot directly be translated into a prediction interval for a measurement by one method given that of another.

The main point in the paper by Bland and Altman [1] is however different from the outlook in this paper; Bland and Altman mainly discuss whether two methods of measurement can be used interchangeably and how to assess this with the help of proper statistical methods to derive LoA, i.e. prediction limits for differences between two methods.

This paper takes as starting point that the classical LoA can be converted to a prediction interval for one method given a measurement by the other (details in the next section). This sort of relationship can be shown in a plot as a line with slope 1 and prediction limits as lines also with slope 1; applicable for the prediction both from method 1 to method 2 and vice versa. In the case of non-constant difference it would be desirable to be able to produce a similar plot, usable both ways. Thus, the aim of this paper is to produce a conversion from one method to another that also applies in the case where the difference between methods is not constant.

In this paper, I set up a proper model for data for method comparison studies which in the case of constant difference between methods leads to the classical LoA, and in the case of linear bias gives a simple formula for the prediction. The paper only addresses the situation where only one measurement by each method is available, although replicate measurements by each method are desirable whenever possible [2]. Moreover, the situation with non-constant variance over the range of measurements is not covered either.

2. Model for LoA

The classical computation of LoA is based on a model for measurements y_{mi} by method $m=1,2$ on item $i=1\ldots I$ (item is here used generically for individual or sample) which can be written as:

$$y_{mi}=\alpha_m+\mu_i+e_{mi},\quad e_{mi}\sim\mathcal{N}(0,\sigma_m^2) \tag{1}$$

Steno Diabetes Center, Niels Steensens Vej 2, DK-2820 Gentofte, Denmark

*Correspondence to: Bendix Carstensen, Steno Diabetes Center, Niels Steensens Vej 2, DK-2820 Gentofte, Denmark.

†E-mail: bxc@steno.dk

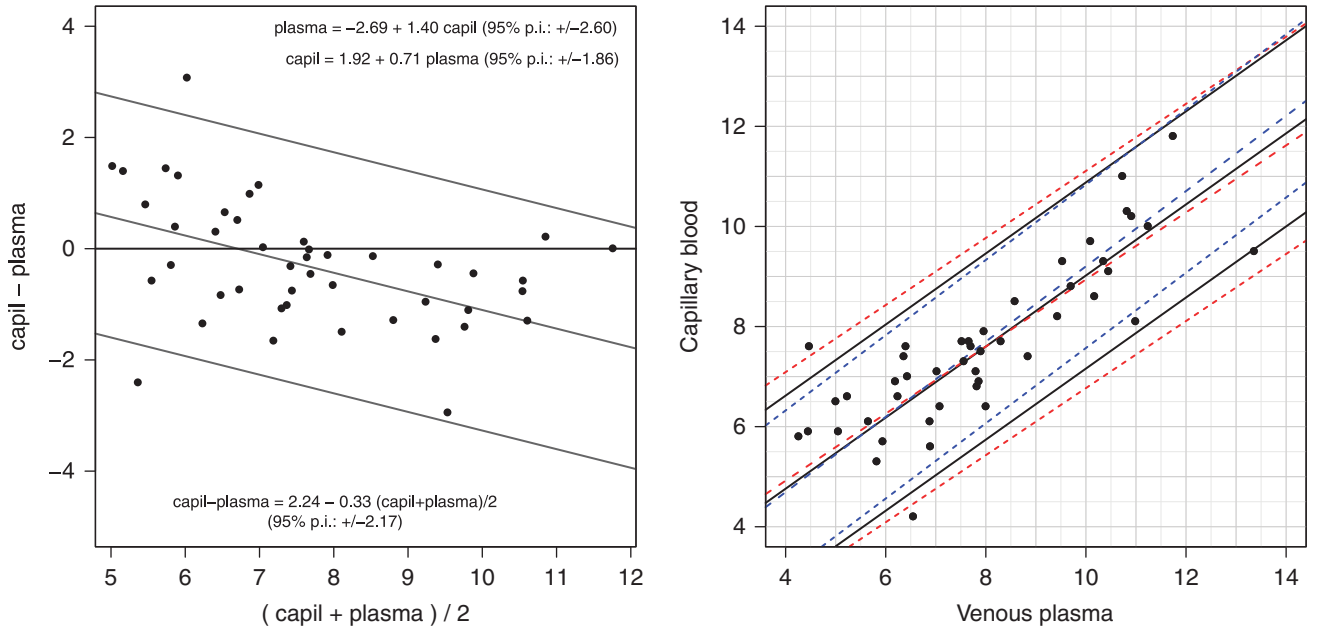

Figure 1. Left: Bland–Altman plot for the blood glucose data in Table I. Right: prediction limits between the two methods for the glucose data. The dashed lines are obtained using the regression of the Plasma–Capillary on the average and from regression of the opposite difference. The full black lines are the model-based lines proposed.

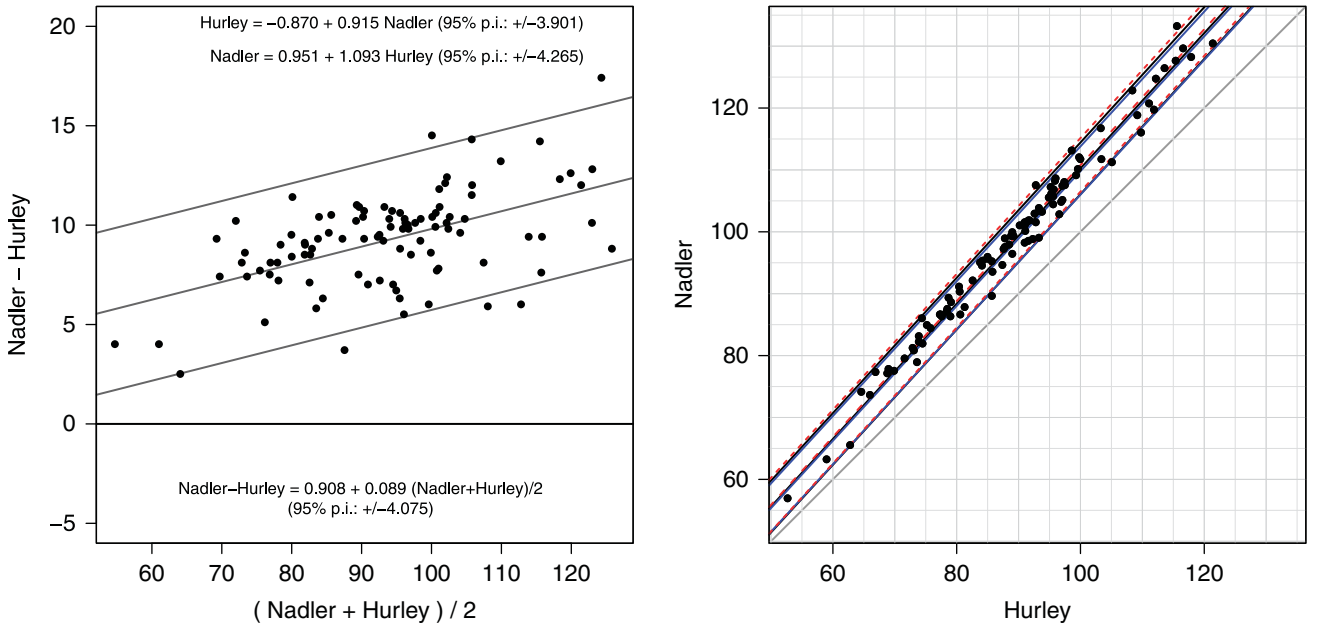

Figure 2. Left: Bland–Altman plot for the plasma volume data with a regression line of differences on averages. Right: prediction limits between the two methods for the plasma volume data. The dashed lines are obtained using the regression of the Hurley–Nadler on the average, and from regression of the opposite difference. The full lines are the model-based limits proposed.

Prediction limits for a difference between measurements by the two methods on a future item, 0, say, are based on the distribution of the difference under the model:

$$D_0 = y_{10} - y_{20} = \alpha_1 - \alpha_2 + e_{10} - e_{20} \sim \mathcal{N}(\alpha_1 - \alpha_2, \sigma_1^2 + \sigma_2^2)$$

Even though the separate variances cannot be identified, the sum of them can be estimated by the empirical variance of the differences. Likewise the separate αs cannot be estimated, only their difference can be estimated as $\bar{D}_\cdot$.

The standard error of $\bar{D}.$ is $\sqrt{(\sigma_1^2+\sigma_2^2)/I}$; hence, a 95 per cent prediction interval for the difference between two future measurements on the same item may be computed as:

$$\bar{D}. \pm 1.96 \times \text{s.d.}(D_i)\sqrt{1+\frac{1}{I}}$$

LoA are usually computed using 2 instead of 1.96, partly compensating for the omitted $1/I$ and partly for ignoring the estimation of the variance:

$$\bar{D}. \pm 2 \times \text{s.d.}(D_i)$$

Specifically, the correct factor to use is $\pm t_{0.975}(I-1)\sqrt{(\widehat{\sigma_1^2+\sigma_2^2})(I+1)/I}$, which is 2.08 for $I=30$ and less than 2 if $I>85$ (and converges to 1.96 for $I\to\infty$; for $I>668$ it rounds to 1.96).

This can be converted to a prediction interval for a future measurement by method 2, y_{20}, given a measurement by method 1, y_{10}:

$$y_{20}=y_{2|1}=y_{10}-\bar{D}. \pm 2 \times \text{s.d.}(D_i)$$

The assumptions underlying this is that the mean and the variance of the differences are constant over the range of measurements. To check these assumptions it is customary to make a so-called Bland–Altman plot of the differences versus the means, as exemplified in the left panels of Figures 1 and 2.

3. Non-constant difference

3.1. Model

If it is observed that the assumption of constant difference between methods is violated, i.e. if there is clear slope in the Bland–Altman plot, the differences can be regressed on the averages (or sums) and the results used to construct prediction intervals of the difference between two future measurements.

The obvious extension of model (1) is a model where measurements by each of the methods are related linearly to the 'true' mean, μ_i:

$$\begin{aligned} y_{1i} &= \alpha_1+\beta_1\mu_i+e_{1i}, \quad e_{1i}\sim\mathcal{N}(0,\sigma_1^2) \\ y_{2i} &= \alpha_2+\beta_2\mu_i+e_{2i}, \quad e_{2i}\sim\mathcal{N}(0,\sigma_2^2) \end{aligned} \tag{2}$$

and where e_{1i} and e_{2i} are assumed independent. In this model, prediction of y_2 from y_1 is derived by isolating μ_i from the first equation, because the observation of y_{10} on a new item is the only data available for the estimation of the corresponding unknown μ_0: $\mu_0=(y_{10}-\alpha_1-e_{10})/\beta_1$. Inserting this in the second equation gives:

$$y_{20}=y_{2|1}=\alpha_2+\beta_2\mu_0+e_{20}=\left(\alpha_2-\alpha_1\frac{\beta_2}{\beta_1}\right)+\frac{\beta_2}{\beta_1}y_{10}-\frac{\beta_2}{\beta_1}e_{10}+e_{20} \tag{3}$$

Hence, the parameters of interest are:

$$\begin{aligned} \text{Intercept:} \quad & \alpha_{2|1} = \alpha_2-\alpha_1\frac{\beta_2}{\beta_1} \\ \text{Slope:} \quad & \beta_{2|1} = \frac{\beta_2}{\beta_1} \\ \text{Prediction variance:} \quad & \sigma_{2|1}^2 = \left(\frac{\beta_2}{\beta_1}\right)^2\sigma_1^2+\sigma_2^2 \end{aligned} \tag{4}$$

Note that equation (3) cannot be used as a basis for regressing y_2 on y_1, because the covariate (y_1) in the regression is strongly correlated with the error term; the correlation is $\rho=-1/\sqrt{1+[(\beta_1/\beta_2)/(\sigma_1/\sigma_2)]^2}$. In most practical circumstances, the βs and the σs are (pairwise) of the same order of magnitude, in which case ρ is about $-1/\sqrt{2}=-0.7$.

The 95 per cent prediction limits for the value by method 2, y_{20}, given a measurement by method 1, y_{10}, on a new item are then

$$\alpha_{2|1}+\beta_{2|1}y_{10} \pm 2\times\sigma_{2|1}$$

The next section deals with practical estimation of these parameters.

3.2. Regression of differences on averages

Now, consider the differences $D_i = y_{1i} - y_{2i}$ and the averages $A_i = (y_{1i} + y_{2i})/2$. The assumptions in model (2) lead to:

$$D_i = (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\mu_i + e_{1i} - e_{2i}$$

$$A_i = (\alpha_1 + \alpha_2)/2 + (\beta_1 + \beta_2)\mu_i/2 + (e_{1i} + e_{2i})/2$$

If we isolate μ_i from the expression for A_i and insert in the expression for D_i we obtain a linear relationship:

$$D_i = a + bA_i + e_i, \quad e_i \sim \mathcal{N}(0, \tau^2) \tag{5}$$

with the following relationship to the parameters in (4):

$$\left.\begin{aligned} a &= (\alpha_1 - \alpha_2) - (\alpha_1 + \alpha_2)\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \\ b &= 2\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \\ \tau^2 &= \left(\frac{2\beta_1}{\beta_1 + \beta_2}\right)^2 \left(\frac{\beta_2^2}{\beta_1^2}\sigma_1^2 + \sigma_2^2\right) \end{aligned}\right\} \quad \Leftrightarrow \quad \left\{\begin{aligned} \alpha_{2|1} &= \frac{-a}{1 + b/2} \\ \beta_{2|1} &= \frac{1 - b/2}{1 + b/2} \\ \sigma_{2|1} &= \frac{\tau}{(1 + b/2)} \end{aligned}\right. \tag{6}$$

The formulae for predicting method 1 from method 2 follows from symmetry; if the opposite differences $y_2 - y_1$ were regressed on the means, opposite values of a and b would result; τ would remain unchanged.

So from the estimates from a linear regression of the differences on the averages, it is possible to compute the parameters needed to provide a prediction equation linking the two methods.

Bland and Altman [1] wrote: 'Of course, in clinical practice, when only one method is being used, the observed value by that method provides the value of a (the average of the two methods—*my addition*)'. This can be read as a proposal to regress the differences on the averages and use the regression line and the residual standard deviation from this (τ) directly in producing LoA. This is a correct procedure for assessing how differences depend on the measurements; however, for the prediction of one method from the other it is problematic since it refers to the prediction of the difference for a given average, and the average is rarely known unless both measurements are known. If this is used naïvely one would just plug in y_{10} in the place of A and get:

$$y_{2|1} = -a + (1 - b)y_1 \pm 2\tau$$

This approach will however give lines relating the two methods that are different depending on how the differences are formed. If the opposite differences were used the residual standard deviation would be the same, but the regression coefficients would be the opposite; hence, in terms of the coefficients from regressing $y_1 - y_2$ on $(y_1 + y_2)/2$ the result would be:

$$y_{1|2} = a + (1 + b)y_2 \pm 2\tau$$

which will not give the same line. Moreover, the procedure gives a smaller slope in the prediction of method 2 from method 1, than the model-based approach:

$$\beta_{2|1} = \frac{1 - b/2}{1 + b/2} = \frac{(1 - b/2)^2}{(1 + b/2)(1 - b/2)} = \frac{1 - b + (b/2)^2}{1 - (b/2)^2} \geqslant 1 - b$$

The last inequality is because the subtraction of a positive constant from the numerator and the addition of a positive constant to the denominator (in this case $(b/2)^2$) make the fraction smaller.

The prediction standard deviation by this method is τ, which is too small if the prediction slope is larger than 1 ($b < 0$), and too large if the prediction slope is smaller than 1.

Hence, the formulae given in (6) are preferable to just plugging the measurement by method 1 for the average in the regression of the differences on the averages.

4. Worked examples

4.1. Blood glucose

Table I and Figure 1 show data from the measurements of glucose based on venous plasma and capillary whole blood on 46 non-diabetic, obese persons at 120 min after a 75 g oral glucose challenge. The data are a subset of a data set thoroughly reported and analysed in [3], which compare different ways of measuring blood glucose. The entire data set is available as `glucose` in the R-package `MethComp`.

Regression of the difference between methods Plasma and Capillary on the average gives $D = -2.24 - 0.33A$ with a residual standard deviation of 1.08 and hence a prediction interval for the differences of size ± 2.17.

Table I. Measurements of glucose (mmol/l) in 46 persons, 120 min after a glucose challenge, based on venous plasma or capillary blood drawn at the same time on each person.

	Plasma	Capillary		Plasma	Capillary		Plasma	Capillary
1	10.15	9.7	17	9.59	9.3	33	5.11	5.9
2	6.30	6.6	18	11.30	10.0	34	7.76	7.6
3	8.90	7.4	19	7.08	7.1	35	6.61	4.2
4	7.72	7.7	20	11.80	11.8	36	6.94	6.1
5	10.79	11.0	21	5.29	6.6	37	9.49	8.2
6	8.36	7.7	22	5.06	6.5	38	6.95	5.6
7	10.41	9.3	23	7.86	7.1	39	6.25	6.9
8	5.88	5.3	24	6.00	5.7	40	7.62	7.3
9	7.92	6.9	25	5.71	6.1	41	7.88	6.8
10	7.58	7.7	26	4.51	5.9	42	7.14	6.4
11	7.96	7.5	27	6.49	7.0	43	10.51	9.1
12	8.06	6.4	28	13.42	9.5	44	6.42	7.4
13	10.23	8.6	29	8.64	8.5	45	4.53	7.6
14	9.76	8.8	30	8.02	7.9	46	4.32	5.8
15	10.97	10.2	31	6.46	7.6			
16	10.88	10.3	32	11.05	8.1			

Data are a subset of the data analysed in Carstensen *et al.* [3].

Table II. Measurements of plasma volume expressed as a percentage of normal in 99 subjects, using two alternative sets of normal values due to Nadler and Hurley. Adapted from Table 2 in [1].

	Nadler	Hurley		Nadler	Hurley		Nadler	Hurley
1	56.9	52.9	34	93.5	86.0	67	104.8	97.1
2	63.2	59.2	35	94.5	84.3	68	105.1	97.3
3	65.5	63.0	36	94.6	87.6	69	105.5	95.1
4	73.6	66.2	37	95.0	84.0	70	105.7	95.8
5	74.1	64.8	38	95.2	85.9	71	106.1	95.5
6	77.1	69.0	39	95.3	84.4	72	106.8	95.9
7	77.3	67.1	40	95.6	85.2	73	107.2	95.4
8	77.5	70.1	41	95.9	85.2	74	107.4	97.3
9	77.8	69.2	42	96.4	89.2	75	107.5	97.7
10	78.9	73.8	43	97.2	87.8	76	107.5	93.0
11	79.5	71.8	44	97.5	88.0	77	108.0	97.6
12	80.8	73.3	45	97.9	88.7	78	108.2	96.1
13	81.2	73.1	46	98.2	91.2	79	108.6	96.2
14	81.9	74.7	47	98.5	91.8	80	109.1	99.5
15	82.2	74.1	48	98.8	92.5	81	110.1	99.8
16	83.1	74.1	49	98.9	88.0	82	111.2	105.3
17	84.4	76.0	50	99.0	93.5	83	111.7	103.6
18	84.9	75.4	51	99.3	89.0	84	111.7	100.2
19	86.0	74.6	52	99.3	89.4	85	112.0	100.0
20	86.3	79.2	53	99.9	89.2	86	113.1	98.8
21	86.3	77.8	54	100.1	91.3	87	116.0	110.0
22	86.6	80.8	55	101.0	90.4	88	116.7	103.5
23	86.6	77.6	56	101.0	91.2	89	118.8	109.4
24	86.6	77.5	57	101.5	91.4	90	119.7	112.1
25	87.1	78.6	58	101.5	93.0	91	120.7	111.3
26	87.5	78.7	59	101.5	91.2	92	122.8	108.6
27	87.8	81.5	60	101.8	92.0	93	124.7	112.4
28	88.6	79.3	61	101.8	91.8	94	126.4	113.8
29	89.3	78.9	62	102.8	96.8	95	127.6	115.6
30	89.6	85.9	63	102.9	92.8	96	128.2	118.1
31	90.3	80.7	64	103.2	94.0	97	129.6	116.8
32	91.1	80.6	65	103.8	93.5	98	130.4	121.6
33	92.1	82.8	66	104.4	95.8	99	133.2	115.8

Based on the results from the regression ($a=-2.24, b=0.33, \tau=1.08$) the prediction limits are:

$$\begin{aligned} y_{C|P} &= \frac{a}{1-b}+\frac{1+b}{1-b}y_N \pm 2\frac{\tau}{1-b} \\ &= 1.92+0.71\,y_N \pm 1.86 \\ y_{P|C} &= \frac{-a}{1+b}+\frac{1-b}{1+b}y_H \pm 2\frac{\tau}{1+b} \\ &= -2.69+1.40 y_H \pm 2.60 \end{aligned}$$

The prediction line and the prediction limits are shown in the rightmost panel in Figure 1 in black. The prediction equations and limits give identical lines in the graph, so the plot may be used for the prediction in both ways.

The major effect is on the size of the prediction interval; depending on whether the prediction is one way or the other, the interval will be too narrow or too wide. The advantage of using the proposed method is that the resulting prediction interval is applicable in both ways, and only one line is needed to convert between the methods.

4.2. Plasma volume

Table II shows data from measurements of plasma volume expressed as a percentage of normal in 99 subjects, using two alternative sets of normal values due to Nadler and Hurley. Data are from the paper by Bland and Altman [1]

Regression of the difference between method Hurley and Nadler on the average gives $D=-0.908-0.089A$ with a residual standard deviation of 2.037 and hence a prediction interval of ± 3.993 (using 1.96 instead of 2).

Based on the results from the regression ($a=-0.908, b=-0.089, \tau=2.037$) the resulting prediction limits are:

$$\begin{aligned} y_{H|N} &= \frac{a}{1-b}+\frac{1+b}{1-b}y_N \pm 2\frac{\tau}{1-b} \\ &= -0.870+0.915 y_N \pm 3.823 \\ y_{N|H} &= \frac{-a}{1+b}+\frac{1-b}{1+b}y_H \pm 2\frac{\tau}{1+b} \\ &= 0.951+1.093 y_H \pm 4.179 \end{aligned}$$

The prediction line and the prediction limits are shown in the right panel in Figure 2 in black. The prediction equations and limits give identical lines in the graph, so the plot may be used for the prediction in both ways.

In this case the methods are in close agreement and the slope relating the two methods is closer to 1; hence, the difference between the methods is much smaller than in the previous example. Bland and Altman [1] used a log-transform of data which remedied the non-constant differences.

5. Why it is wrong to use the regression of the differences on the averages

5.1. Substantially wrong

Model (2) is qualitatively different from (1) in the sense that it is invariant under arbitrary *scaling* of one of the methods. If, for example, all measurements by method 2 were multiplied by 2, it would just mean that α_2, β_2 and σ_2 would be twice as large. But the differences and averages would be irrelevant; any result can be obtained by rescaling measurements from one of the methods. Therefore, the proposed procedure is *only* relevant in situations where the two methods compared can be assumed to be on the same scale.

In the special situation where the relationship between methods is multiplicative, log-transformation of all measurements would result in a model of the form (2) for the transformed data. In this case, both methods would be on the same scale, namely the relative scale.

5.2. Statistically wrong

In the regression of the differences on the averages, it is implicitly assumed that the averages are independent of the error terms. But this is *not* the case under model (2):

$$\begin{aligned} \operatorname{cov}\left(A_i, e_{1i}\left(1-\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)-e_{2i}\left(1+\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)\right) &= \operatorname{cov}\left(\frac{\mathrm{e}_{1\mathrm{i}}+\mathrm{e}_{2\mathrm{i}}}{2}, \mathrm{e}_{1\mathrm{i}}\left(1-\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)-\mathrm{e}_{2\mathrm{i}}\left(1+\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)\right) \\ &= \frac{1}{2}\left\{\sigma_1^2-\sigma_2^2-\frac{\beta_1-\beta_2}{\beta_1+\beta_2}(\sigma_1^2+\sigma_2^2)\right\} \end{aligned}$$

This covariance is 0 iff:

$$\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2+\sigma_2^2}=\frac{\beta_1-\beta_2}{\beta_1+\beta_2} \quad \Leftrightarrow \quad \frac{\beta_2}{\beta_1}=\frac{\sigma_2^2}{\sigma_1^2}$$

i.e. if the slope relating one method to the other equals the ratio of the variances. In most practical situations both of these fractions are close to 1; hence, unlike in the case of equation (3) representing the usual regression of y_2 on y_1, we have approximate independence of the covariate and the residuals in this case.

Multiplying all the measurements from method 2 by a factor K changes the slope by a factor K, but the ratio of the variances by a factor K^2, so using the correct scaling of the methods relative to each other will make it possible to obtain uncorrelated sums and residuals. However, there is no data to estimate the scaling, which in any case would be an odd thing to do; estimation of the variance ratio should be done based on replicate measurements on each item by each method.

Hence, regressing differences on averages essentially assume that the estimated slope between methods is also the ratio of residual variances of the methods.

5.3. Why are the limits straight lines?

The prediction limits are straight lines because the estimation variance of $\alpha_{2.1}$ and $\beta_{2.1}$ is ignored. In well-designed studies these are so small that it would not make much difference whether they were included or not. If the study is so small than it matters it is advisable to do a better designed study.

5.4. What is the relation to standard regression?

The model (2) is *not* a classical regression model, because the individual levels (the μ_is) are explicit parameters in the model. Classical regression models are based on the conditional distribution of one method given that of another and the least-squares formulae may be derived from the bivariate normal as the conditional distribution of y, given x. Under the model (2) the covariate (method 1, say) is strongly correlated with the residuals and therefore the classical least-squares regression is wrong—classical regression introduces an asymmetry which is alien to the problem.

This is why the prediction in method comparison studies should be based on a model (like (2)) where the prediction distribution only contains what is likely to be similar between the calibration data set and an arbitrary future observation, whereas the parameters relating to the selection of the sample (the μ_is) are estimated. The basis of this argument is the same as that underlying the paired t-test and the construction of LoA, namely basing the inference on quantities (in this case the differences), which do not depend on the item-specific parameters. In the situation where the bias is constant ($\beta_1 = \beta_2 = 1$) it is tantamount to the principle of inference in the conditional distribution given the sufficient statistics for the nuisance parameters.

This is the reason that classical regression analysis is not appropriate for making predictions based on the method comparison studies.

5.5. What is the relation to Deming regression?

The regression model where both variables are subject to error can only be identified for data with one measurement per item and method if one is willing to assume a fixed value for ratio of the variances. This is called Deming regression; Cornbleet and Gochman [4] and Jensen [5] provide the maximum-likelihood estimator in the model (2) (with the usual bias correction for the variance), under the assumption that the ratio of the variances is known.

This highlights the desirability to have replicate measurements on each item by both methods, so that the method-specific variances, and hence their ratio can be estimated. Strictly speaking, one may get away with replicate measurements by one method, since Deming regression is also possible if one of the variances is known.

The prediction variance for predicting method 2 from method 1 is computed as $\sigma_{2|1} = \sqrt{(\beta_2/\beta_1)\sigma_1^2 + \sigma_2^2}$.

Hence, the Deming regression does not solve the prediction problem unless we are willing to assume a known value for the ratio of the variances. In studies without replicates there is no information about the ratio of the residual variances for the two methods.

6. How wrong is it to do it anyway?

To assess the possible discrepancy between the simple method based on regression of differences on sums and the method deriving results from model (2), and to see how well these reproduce the true values, a small simulation study was set up.

Clearly, the actual value of the intercept α is irrelevant for the outcome, only the slopes (i.e. their ratio) and the ratio of residual variances are of interest.

In three different scenarios 100 data sets were simulated with 50 items in each, the 'true' values (μ_is) uniformly distributed on $[-50, 50]$. For each data set estimates of $\alpha_{2|1}$, $\beta_{2|1}$ and $\sigma_{2|1}$ were computed, as well as the coverage. The coverage was estimated as the fraction of points from 10 new simulated data sets that fell within the 95 per cent prediction limits based on the parameter estimates, i.e. for each simulated data set used for estimation, 10 new data sets were simulated for computing the coverage.

The calculations were done using three different methods:

- The simple approach derived from the paper by Bland and Altman [1].
- Deming regression [4, 5], assuming equal variances.
- The model-based method proposed.

Figures 3 and 4 shows the results for $\beta_2 = 1.2$, $\beta_1 = 1$, i.e. $\beta_{2|1} = 1.2$. Simulations were done for five different values of $(\sigma_1, \sigma_2) = (3, 1), (3, 2), (3, 3), (3, 5), (3, 10)$. On the x-axis are the estimates by the method proposed, and on the y-axis the corresponding values by the two other methods.

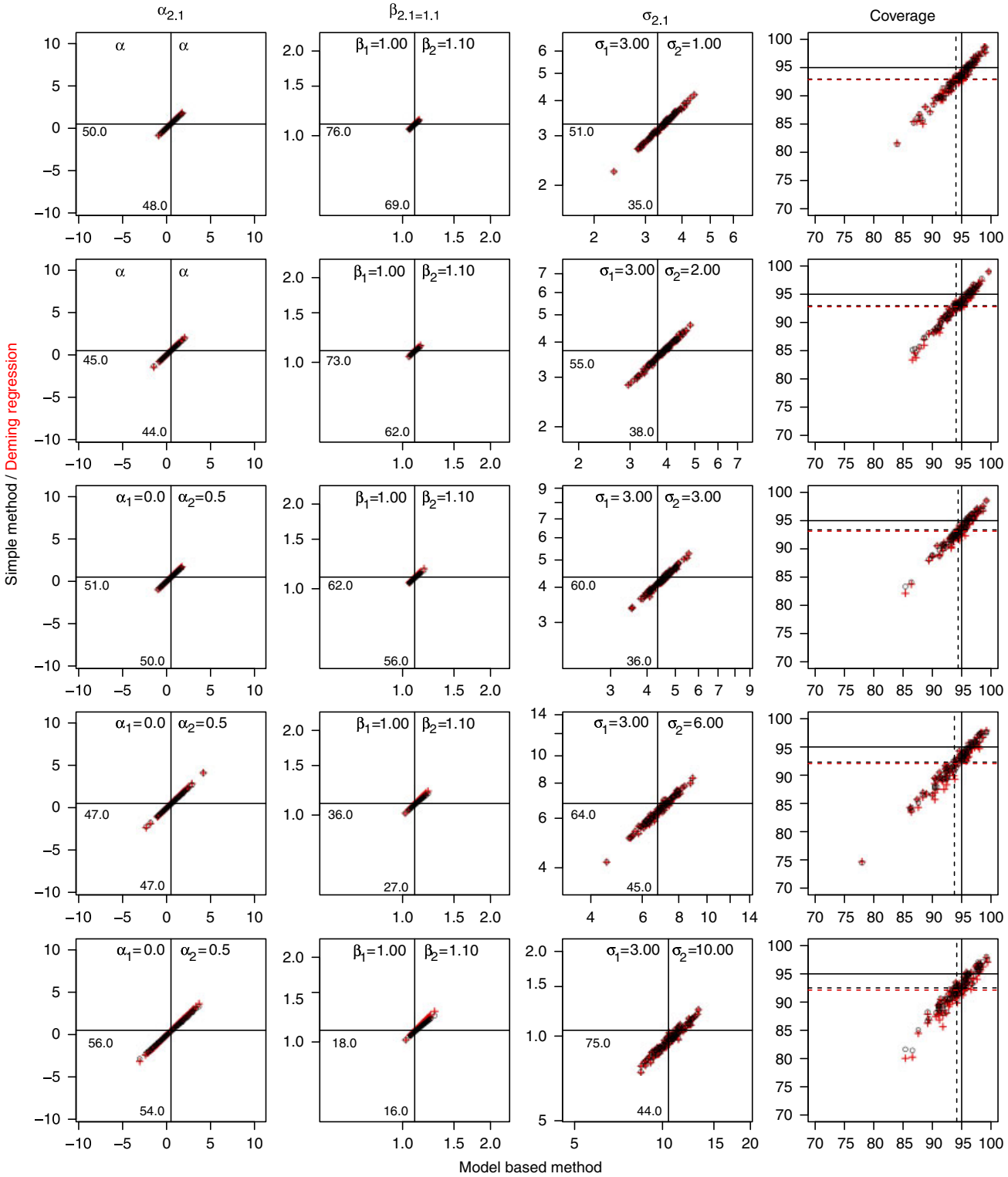

Figure 3. Comparison of the results from the model-based approach (*x*-axis) and the simple procedure and Deming regression (*y*-axis), for data simulated with $\beta_{2|1}=1.1$. The parameters compared are the intercept ($\alpha_{2|1}$), the prediction slope ($\beta_{2|1}$) and the prediction standard deviation ($\sigma_{2|1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels give the coverage computed as the average fraction of points from 10 independently simulated data sets that fall within the estimated prediction limits; the dashed lines indicate the mean coverage.

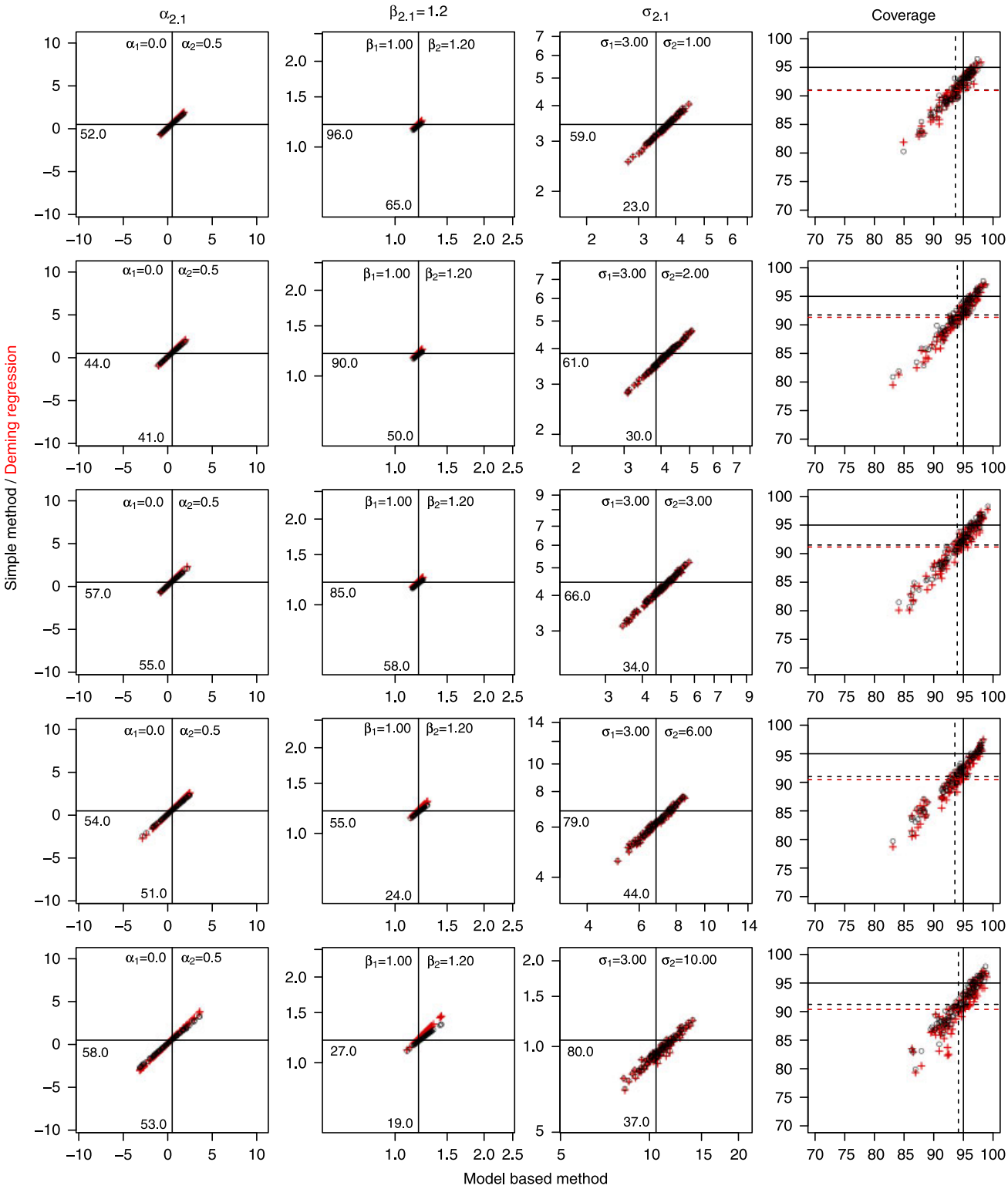

Figure 4. Comparison of the results from the model-based approach (*x*-axis) and the simple procedure and Deming regression (*y*-axis), for data simulated with $\beta_{2|1}=1.2$. The parameters compared are the intercept ($\alpha_{2|1}$), the prediction slope ($\beta_{2|1}$) and the prediction standard deviation ($\sigma_{2|1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels give the coverage computed as the average fraction of points from 10 independently simulated data sets that fall within the estimated prediction limits; the dashed lines indicate the mean coverage.

The simulations show that the different methods give largely the same results as far as the parameters $\alpha_{2|1}$, $\beta_{2|1}$ and $\sigma_{2|1}$ are concerned, but when it comes to the coverage of the prediction interval, the proposed method actually does a better job when judging by the average coverage as indicated by the broken lines in the plots. This is because the vertical broken lines (representing the median coverage of the model-based approach) are closer to the postulated 95 per cent level than the horizontal one.

7. Discussion

I have here proposed a simple twist to the results from regression of the differences on the sums in the case of a linear relationship between two methods of measurement. It is consistent with the obvious underlying model, and exploits the fact that although the parameters of the model cannot be estimated, those functions of the parameters that are needed for creating predictions can be estimated.

The prediction limits provided have the attractive property that if the prediction line with limits is drawn in a coordinate system, the chart will apply in both ways; hence, *both* the line *and* the limits are symmetric. Precisely as the prediction intervals derived from the classical LoA are in the case where the difference between methods is constant.

The drawback is that the regression of the differences on the means ignores that the averages are correlated with the residuals (i.e. the error terms), and therefore gives biased estimates if the slope linking the two methods is far from 1 or the residual variances are very different. However, both of these are rather uncommon in method comparison studies, so the method proposed here is widely applicable.

When considering LoA, the only feasible transformation is the log-transform, which gives LoA for the ratio of measurements, which is immediately understandable. If, for example, the measurements are fractions where some are close to either 0 or 1 a logit transform may be adequate. LoA would then be for (log) odds-ratios, not very easily understood. For other more arbitrarily chosen transformation the situation may be even worse. But if a plot with conversion lines and limits are constructed, then the plot is readily back-transformed to the original scale for practical use.

8. Recommendations

When comparing two methods of measurement with the aim of providing a conversion between them, and if only one measurement by each is available, the following is recommended:

1. Compute LoA and convert to a prediction interval for one method given the other.
2. Make a Bland–Altman plot to see if the basic assumptions of constant mean and variance of the differences are met. If they are, report the LoA or the conversion between methods based on a constant difference.
3. If there is non-constant difference but uniform variance, regress the differences $y_1 - y_2$ on the averages $(y_1 + y_2)/2$, obtain intercept a, slope b and residual standard deviation τ, and construct the prediction between the methods as:

$$y_{1|2} = \frac{a}{1-b/2} + \frac{1+b/2}{1-b/2} y_2 \pm 2\frac{\tau}{1-b/2}$$

$$y_{2|1} = \frac{-a}{1+b/2} + \frac{1-b/2}{1+b/2} y_1 \pm 2\frac{\tau}{1+b/2}$$

Plot the conversion line and limits with a background grid to facilitate the practical use of the derived relationship.

Acknowledgements

I am grateful to Lene Theil Skovgaard from Department of Biostatistics, University of Copenhagen for helpful discussions and valuable suggestions, and to Doug Altman for helpful review comments that substantially improved the paper.

References

1. Bland JM, Altman DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 1999; **8**:136–160.
2. Carstensen B, Simpson J, Gurrin LC. Statistical models for assessing agreement in method comparison studies with replicate measurements. *International Journal of Biostatistics* 2008; **4**(1):Article 16.
3. Carstensen B, Lindström J, Sundvall J, Borch-Johnsen K, Tuomilehto J, DPS Study Group. Measurement of blood glucose: comparison between different types of specimens. *Annals of Clinical Biochemistry* 2008; **45**(2):140–148.
4. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. *Clinical Chemistry* 1973; **25**(3):432–438.
5. Jensen AC. Deming regression. *Technical Report*, Vignette for the `MethComp` package for R, 2007. Available from: http://staff.pubhealth.ku.dk/~bxc/MethComp/Deming.pdf.