Clinical and Experimental Pharmacology and Physiology (1997) **24**, 193–203

SPECIAL ARTICLE

COMPARING METHODS OF MEASUREMENT

John Ludbrook

Cardiovascular Research Laboratory, University of Melbourne Department of Surgery, Royal Melbourne Hospital, Parkville, Victoria, Australia

SUMMARY

1. The purpose of comparing two methods of measurement of a continuous biological variable is to uncover systematic differences not to point to similarities.

2. There are two potential sources of systematic disagreement between methods of measurement: fixed and proportional bias.

3. Fixed bias means that one method gives values that are higher (or lower) than those from the other by a constant amount. Proportional bias means that one method gives values that are higher (or lower) than those from the other by an amount that is proportional to the level of the measured variable.

4. It must be assumed that measurements made by either method are attended by random error: in making measurements and from biological variation.

5. Investigators often use the Pearson product–moment correlation coefficient (*r*) to compare methods of measurement. This cannot detect systematic biases, only random error.

6. Investigators sometimes use least squares (Model I) regression analysis to calibrate one method of measurement against another. In this technique, the sum of the squares of the vertical deviations of y values from the line is minimized. This approach is invalid, because both y and x values are attended by random error.

7. Model II regression analysis caters for cases in which random error is attached to both dependent and independent variables. Comparing methods of measurement is just such a case.

8. Least products regression is the reviewer's preferred technique for analysing the Model II case. In this, the sum of the products of the vertical and horizontal deviations of the x,y values from the line is minimized.

9. Least products regression analysis is suitable for calibrating one method against another. It is also a sensitive technique for detecting and distinguishing fixed and proportional bias between methods.

10. An alternative approach is to examine the differences between methods in order to detect bias. This has been recommended to clinical scientists and has been adopted by many.

11. It is the reviewer's opinion that the least products regression technique is to be preferred to that of examining differences, because the former distinguishes between fixed and proportional bias, whereas the latter does not.

Key words: correlation, Deming's method, differences, least products, least squares, model I regression, model II regression, random error, residuals, systematic bias.

Correspondence: Professor John Ludbrook, University of Melbourne Department of Surgery, Royal Melbourne Hospital, Parkville, Victoria 3050, Australia.

Received 1 July 1996; revision 4 October 1996; accepted 7 October 1996.

INTRODUCTION

A problem that sometimes confronts biomedical investigators is how to compare one method for measuring a biological variable with another. Most often they wish to compare the accuracy of a new method with that of an established one, the new method having the merit of being simpler, cheaper, quicker or less invasive. Sometimes they wish to calibrate one method against the other, and sometimes they wish to determine how reproducible a single method is over time or as used by different observers.

In their publications, investigators usually plot the observed x,y values as a scattergram. They may then adduce the value of *r*, the Pearson product–moment correlation coefficient, as evidence for or against there being good agreement between the methods. They may calculate a least-squares regression line that they assert describes the relationship between the two methods and they may use deviations of that fitted straight line from the 45° line of identity as evidence for or against adopting the new method or as evidence that a method performs consistently or otherwise when used by the same or different observers. It should be said now that if these analytical techniques are used, the conclusions reached may be seriously flawed.

Two techniques will be described for analysing experiments designed to compare methods of measurement. In a general way, both depend on an understanding of forms of linear regression analysis that rely on normal distribution theory. As background to the detailed descriptions of these techniques that will follow, it is useful to recall some of the ground rules for linear regression analysis.

SOME GROUND RULES FOR REGRESSION ANALYSIS

All biomedical scientists are familiar with fitting a straight line to experimental data by the ordinary least squares (OLS) technique, which involves minimizing the sum of the squares of

List of abbreviations:

α, β	Population parameters for least squares linear regression (Model I)
a, b	Least squares estimates (regression coefficients) of the Model I parameters
α', β'	Population parameters for least products linear regression (Model II)
a', b'	Least products estimates (regression coefficients) of the Model II parameters
Bias	Systematic bias between measurements: fixed, proportional or both
Error	Random error: biological variation, measurement error or both
E(Y)	Estimated (predicted) value of Y in a population of X,Y values
M1, M2	Values resulting from measurement of the same variable by two methods
Model I	Regression in which the X values are without error
Model II	Regression in which both Y and X values are attended by error
n	Number of x,y pairs in a sample
OLP	Ordinary least products (regression)
OLS	Ordinary least-squares (regression)
r	Pearson's product–moment correlation coefficient
Residuals	The deviations of y (or x) values from the estimated regression line
S_x, S_y	Sample standard deviations of x and y
WLP	Weighted least products (regression)
WLS	Weighted least-squares (regression)
X,Y	Values of X,Y pairs in the population
x,y	Values of x,y pairs in a random sample of a population of X,Y values

the vertical deviations of the observed y values from the line (y residuals; Fig. 1). The linear regression model for a population of X,Y values can be stated as $Y = \alpha + \beta X + \epsilon$, where the parameters α and β are, respectively, the 'true' Y intercept and the 'true' slope of the population regression of Y on X. The term ϵ refers to the random error associated with Y and will be discussed shortly. However, investigators have only their sample values to work with. The sample regression equation can be stated as $E(Y) = a + bx$, where E(Y) is the estimated value of Y in the population of X,Y values, x is an observed value and the coefficients a and b are estimates of the corresponding population parameters α and β.

Sources of random error

As stated earlier, the term ϵ in OLS regression refers to the random error or variation associated with Y in the population of X,Y values. This can originate from: (i) measurement error resulting from random imperfections of the measuring instrument or random error attributable to those who use it; and (ii) biological variation reflecting the effects of random biological variation in Y.

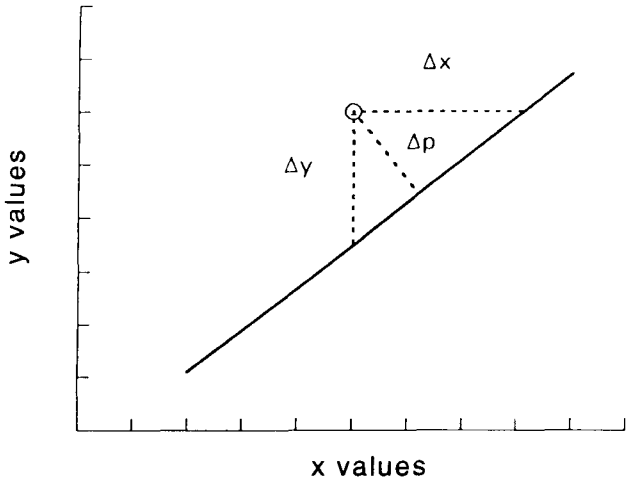

Fig. 1. Techniques of regression analysis. Δx, horizontal distance of x,y point from line; Δy, vertical distance of x,y point from line; Δp, perpendicular distance of x,y point from line. The functions that can be minimized include ordinary least squares regression, $\Sigma(\Delta y)^2$, ordinary least products regression, $\Sigma(\Delta y)(\Delta x)$, and major axis regression, $\Sigma(\Delta p)^2$.

Some assumptions in linear regression analysis

The assumptions that underlie OLS regression analysis are given in all general texts of statistics. Berry, in his specialized monograph, lists eight.[1] Most biomedical investigators will be aware of these, but two are of such importance that they are restated below.

Random error in Y and X

It is an important assumption of OLS regression that whereas the values of Y in the population that has been sampled are attended by error, those of X are not. Strictly, this can be so only if the X values are categorical: for instance, conditions, treatments or places. However, most statistical theorists follow the advice of Berkson, which is that if the X values have been fixed in advance by the experimenter (e.g. by specifying times, doses or settings of a pump), then they can be regarded for practical purposes as error-free.[2] When X is error-free, Model I regression analysis is the proper form to use. It includes the well-known OLS regression technique as well as modifications of it, such as weighted least squares (WLS) regression.[3]

When both X and Y are free to vary and are attended by error, some statisticians allow that Model I regression analysis may still be used if it is certain, on biological grounds, that Y must depend on X and never the reverse.[4–6] This is the case, for instance, in dose- or stimulus–response relationships. Even then, the Model I regression line should be used for empirical, rather than explanatory, purposes.[7,8]

When investigators plan experiments to compare methods of measurement, they must assume that both Y and X will be attended by random error. Moreover, it is impossible to decide which method should be regarded as dependent and which independent and because of this it is wrong to use Model I regression analysis. Instead, one or another form of Model II regression analysis must be used. These are described later.

Normality

In OLS (Model I) regression, it is assumed that at each level of X in the population, the errors in Y conform to the same

normal distribution. That is, at each level of X, the vertical deviations of the Y values from the population regression line: (i) are normally distributed; (b) have a mean of zero; and (c) the normal distributions have the same variance. Most statisticians urge that these assumptions should always be tested by graphical examination of the sample residuals.[3] In the case of OLS regression, these are the vertical deviations of the observed values of y from the estimated regression line (Fig. 1). These y residuals should be examined by means of a normal probability plot and by plotting them against the estimated values of Y. If both X and Y are attended by random error (Model II regression), then the horizontal as well as the vertical deviations from the line should be examined.

As in other forms of statistical analysis based on normal distribution theory, a moderate degree of non-normality can be tolerated. However, it is generally agreed that although progressive increase in scatter of the sample residuals (often described as proportional error) causes only moderate bias in the estimates of α and β in the parent population of X,Y values, it can seriously affect the confidence intervals that surround those estimates. In biological systems, proportional error seems to be the rule rather than the exception. Its origins have not been closely studied, but there are probably two. It may be a genuine biological phenomenon, in the sense that a biological response may become more erratic as the strength of the stimulus increases. But, more often, it occurs because measurement error increases in proportion to the level of the measured variable.

THE GOAL OF COMPARING METHODS OF MEASUREMENT

It stands to reason that using two different methods for measuring the same physiological variable or repeated use of the same method by the same or different observers is likely to produce similar results. Investigators should not be looking for similarities but should be looking for systematic differences or biases between the methods. Systematic measurement biases can take either or both of two forms (Fig. 2): (i) fixed bias, meaning that one method gives values that are consistently higher (or lower) than those from the other by a constant amount over the whole range of measurements; or (ii) proportional bias, meaning that one method gives values that are higher (or lower) than those from the other by an amount that is proportional to the level of the measured variable.

There are two recognized techniques for detecting and quantitating fixed and proportional biases between methods of measurement. One is based on Model II regression analysis and has been used by clinical chemists for more than 20 years. The other was proposed just over 10 years ago for use by clinical scientists and can be called the method of differences.[9,10] But, first, it is necessary to dismiss from further consideration the still-popular techniques of correlation and Model I regression analysis.

CORRELATION

In reports published in biomedical journals it is still common practice to plot the values resulting from one method against those resulting from the other and to calculate the Pearson product–moment correlation coefficient, r. The value of r and a P value for the hypothesis that $r = 0$ are cited to indicate that the two methods give either very similar or very dissimilar results.

There are several serious criticisms of this practice, but the most important is as follows. Given that the distribution of X,Y values in the population takes the form of a bivariate normal distribution and that there is a linear relationship between these two variables, Pearson's r merely indicates the scatter of values around the line of best fit, regardless of whether the slope of that line differs from unity (proportional bias) or whether its intercept differs from zero (fixed bias; Fig. 2). It does no more than indicate the strength of the linear association between the X and Y variables in the population. The information provided by r is, therefore, of no value in detecting systematic biases between methods. But, regrettably, biomedical investigators still sometimes use it for this purpose.[11,12]

MODEL I REGRESSION ANALYSIS

Sometimes investigators superimpose on a scattergram of the observed x,y values a putative line of best fit. This line has generally been constructed by Model I (OLS) regression analysis. It may be presented to show how close it is to the 45° line of identity. Or the investigators may plan to use the line to calibrate one method against the other. Neither of these goals can be attained by Model I regression. One easily understood defence of this statement is that not one but two lines can be drawn by OLS regression (Fig. 3). In one, the sum of the squared y residuals is minimized, while in the other the sum of

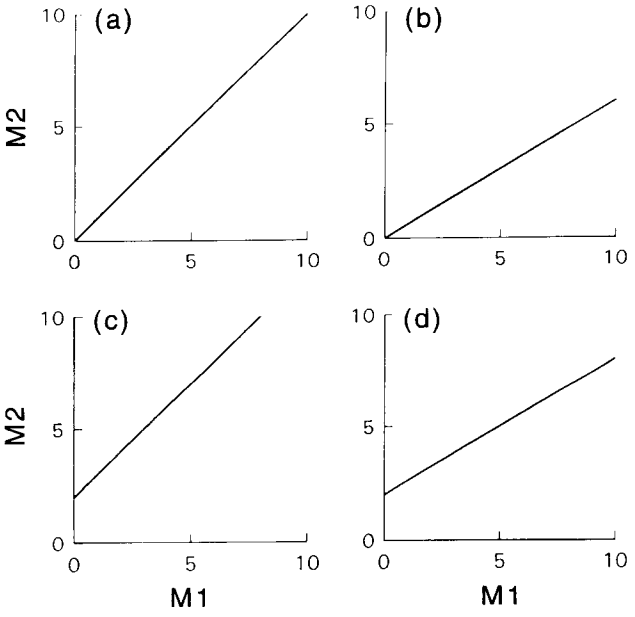

Fig. 2. Systematic bias between methods of measurement, illustrated by regression lines of the form $E(M2) = a' + b'(M1)$ from Model II regression analysis. (a) There is neither proportional nor fixed bias, as $b' = 1.0$ and $a' = 0$. (b) There is proportional but not fixed bias because $b' = 0.6$ but $a' = 0$. (c) There is fixed but not proportional bias as $a' = 2$ but $b' = 1.0$. (d) There is both proportional and fixed bias as $b' = 0.6$ and $a' = 2$.

the squared x residuals is minimized. Neither line is a good illustration of the interrelation of Y and X in the population because both methods are attended by error. Only a Model II regression line (Fig. 3; OLP) can properly be used for calibration or to detect bias between the methods.

MODEL II REGRESSION ANALYSIS

This technique is designed for cases in which both X and Y values are attended by random error, especially when it is impossible to decide which should be regarded as dependent. The deviations of both y and x values from the fitted line must be minimized. Whereas in Model I regression there are two lines that could describe the X,Y relationship, in Model II there is only one (Fig. 3; OLP), which is sometimes called the line of symmetry.[5] Advanced statistical texts consider this matter,[4] but among the elementary general statistical texts only Sokal and Rohlf devote much space to it.[6] Brace gives an account of it for physiologists, which is strongly recommended reading.[5] Richter gives a comprehensive and lucid review, but it is not easily accessible.[13] There are several techniques for taking into account the errors attached to both X and Y in linear regression analysis,[4–6,13,14] but only two are in common use.

Major axis regression

This is also known as principal component regression or the perpendicular distance method. In it the sum of the squared perpendicular distances of the x,y values from the line is minimized (Fig. 1). However, it is proper to use this only if: (i) the slope of the line is 1.0; (ii) the standard deviations of the x and y values are identical ($s_x = s_y$); and (iii) the scales of measurement of X and Y are the same. Otherwise, the scales of

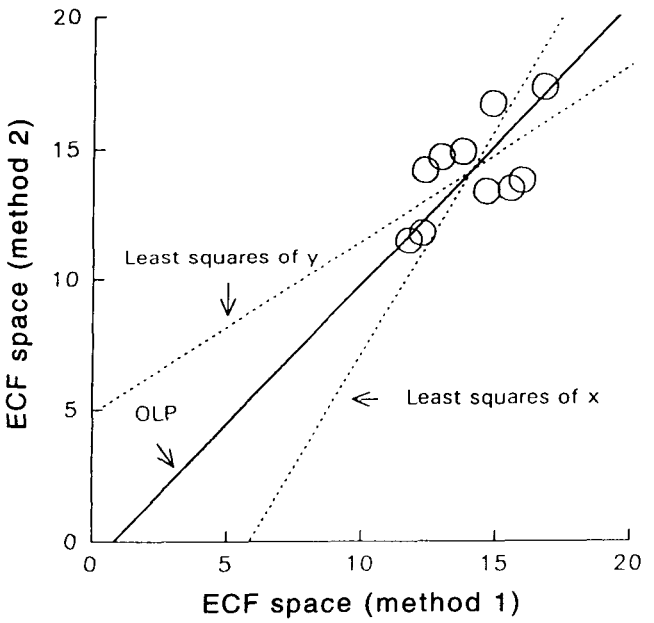

Fig. 3. Model I and model II regression lines. Hypothetical example of two methods for measuring extracellular fluid space (ECF). Dotted lines: model I regression by ordinary least squares of y and ordinary least squares of x techniques. Solid line: model II regression by ordinary least products technique.

X and Y must be standardized, usually by multiplying the y values by (s_x/s_y).[5,13–15] These scaled values of y and the original values of x are used in the major axis regression analysis, but then a and b must be back-transformed by dividing them by (s_x/s_y). This technique and variants of it are popular among clinical chemists, who describe it as Deming's method or standard major axis regression.[14–17] If this rather tortuous process is followed, it turns out that standard major axis regression is identical to the more easily understood least products regression, described below.

Ordinary least products regression

This is also known as reduced major axis regression, standardized principal component regression or geometric mean regression. The term ordinary least products (OLP) regression is preferred here because it describes explicitly how it is executed. In OLP regression, the sum of the products of the vertical and horizontal distances of the x,y values from the line is minimized (Fig. 1). It has been popular for over 30 years among marine scientists and zoologists, especially with respect to allometry.[13,18,19] However, despite the encouragement given by Brace to physiologists and pharmacologists,[5] they rarely use OLP regression analysis.

In order to compare two methods of measurement, two pieces of information are required: (i) estimates of the slope and intercept of the OLP regression line for the population of X,Y values (the regression coefficients b' and a'); and (ii) measures of confidence in those estimates in the form of standard errors of the regression coefficients a' and b' or confidence intervals for the corresponding population parameters α' and β'.

The coefficient for the slope of the OLP regression $E(Y) = a' + b'x$ can be calculated by hand in any of three ways. First, it can be derived from the values of b in the OLS equations $E(Y) = a + bx$ and $E(X) = a + by$ when they are estimated by the least squares of y and x, respectively. Then $b' = \sqrt{(b_{y.x})(1/b_{x.y})}$; hence the term geometric mean regression. Second, it is given by $(b_{y.x})/r$, where r is the Pearson product–moment correlation coefficient. Third, it is given by s_y/s_x, the ratio of the standard deviations of y and x values.

Because the OLP regression line passes through the point $\bar{x},\bar{y}$ (mean x, mean y), the intercept a' can be obtained from the formula $a' = \bar{y} - b'\bar{x}$.

Calculating the 95% confidence intervals (CI) for the population parameters β' and α' is more difficult. The author has tried three techniques: (i) iterative estimation using a specified LOSS function in a microcomputer statistics package (this provides asymptotic standard errors (ASE) for the regression coefficients and, in some cases, 95% CI; it is the preferred technique); (ii) Jolicoeur and Mosimann,[20] reproduced more accessibly by Richter,[21] give a formula to calculate approximate 95% CI for the population parameters α' and β'; and (iii) bootstrapping, by means of a microcomputer spreadsheet.

A detailed description of how to perform these calculations is in Appendix 1.

Assumptions for OLP regression analysis

Ordinary least products regression analysis depends on a similar set of assumptions to OLS, except that both X and Y can be

attended by random error. Therefore, not only the vertical deviations of Y from the regression line, but also the horizontal deviations of X, must conform to normal distributions (a bivariate normal distribution) and the scatter of Y axis deviations from the line must be the same at all levels of X (and vice versa). Unfortunately, in almost every branch of biology, random error tends to increase in proportion to the levels of X and Y. Just as in least squares regression, this proportional error can be dealt with either by data transformation or by weighting the residuals. This will be discussed further later.

Example

This is taken from Daniel's set of data in which two methods of systolic blood pressure measurement are compared.[22] The data are given in Table 1 and are plotted in Fig. 4. Altman and Bland used the same set of data to illustrate the method of differences,[9] which will be discussed later.

Table 1. Data set for comparison of one method of measuring systolic blood pressure with another (from Daniel[22])

Patient	Method 1 (M1)	Method 2 (M2)
1	132	130
2	138	134
3	144	132
4	146	140
5	148	150
6	152	144
7	158	150
8	130	122
9	162	160
10	168	150
11	172	160
12	174	178
13	180	168
14	180	174
15	188	186
16	194	172
17	194	182
18	200	178
19	200	196
20	204	188
21	210	180
22	210	196
23	216	210
24	220	190
25	220	202

Patients are described as having essential hypertension. It should be assumed that the measurements (in mmHg) on each patient were made in random order and in the same setting.

The coefficients of the least products regression M2 = $a' + b'$ (M1) and the 95% CI for α' and β' were calculated by each of the three techniques mentioned earlier and described in the Appendix. The results are given in Table 2. There is good agreement among the techniques, except that the confidence interval for α' is narrower by bootstrapping. From Table 2, the following statistical inferences can be made from all three techniques: (i) the hypothesis $\alpha' = 0$ (no fixed bias between the methods) is not rejected because the confidence interval for α' includes zero; and (ii) the hypothesis $\beta' = 1$ (no proportional bias between the methods) is rejected at the 5% level because the 95% CI for β' does not include 1. That is, method 1 gives readings that are progressively higher than those of method 2 as the level of systolic blood pressure (SBP) increases (Fig. 4).

In summary, by using OLP linear regression analysis one can calibrate one method against another by means of the regression

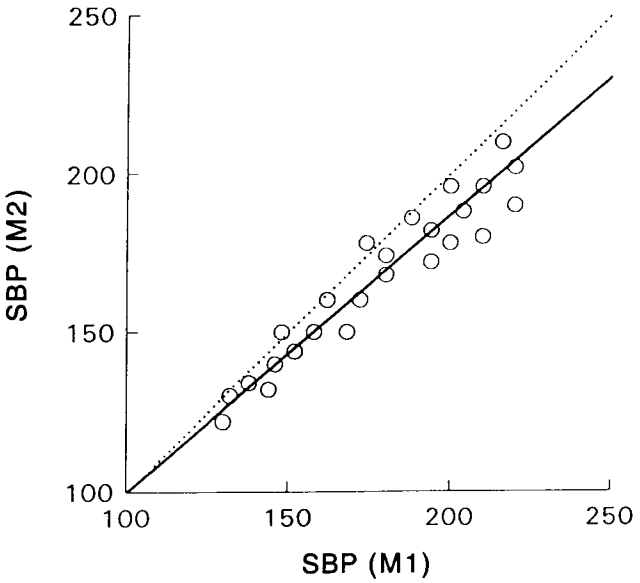

Fig. 4. Scattergram of values for systolic blood pressure (SBP) measured by two different methods (M1, M2; see Table 1). Solid line: weighted least products (model II) regression line E(M2) = 12.868 + 0.867(M1; see Table 2). Dotted line: 45° line of identity. Note the increasing scatter of M1 and M2 around the regression line as blood pressure increases, which has been allowed for by weighting the M1 and M2 values.

Table 2. Outcomes of Model II (least products) regression analysis for the example of Table 1

Technique	Method of execution	a'	95% CI	b'	95% CI
OLP	Calculator	13.951	−6.893–32.197	0.861	0.758–0.978
OLP	Computer	13.951	−6.166–34.067	0.861	0.749–0.973
OLP/BTS	Spreadsheet	14.010	−0.826–31.559	0.861	0.755–0.956
WLP	Calculator	12.830	−6.369–29.897	0.867	0.771–0.976
WLP	Computer	12.868	−4.783–30.518	0.867	0.764–0.971

The regression equation is E(M2) = $a' + b'$ (M1). Regression coefficients: a', M2 intercept; b', slope of M2 on M1. 95% confidence intervals (CI) are for the population parameters α' and β'. OLP, ordinary least products regression; WLP, weighted least products regression; OLP/BTS, based on 1000 bootstrap resamples of s_y/s_x. See Appendix 1 for details of computational techniques.

equation so that results from one method can be compared with or converted to results from the other, test whether there is a proportional bias between the methods (i.e. whether the slope of the regression of Y on X differs significantly from unity) and test whether there is a fixed bias between the two methods (i.e. whether the intercept of the regression of Y on X differs significantly from zero).

Proportional error

There is, however, a problem in accepting the inferences arrived above. In Fig. 4 it is clear that the scatter of x,y points around the regression line increases in approximate proportion to the level of SBP. This is even more obvious if an analysis of residuals is performed (not shown). This proportional error breaches one of the important assumptions of linear regression analysis and it commonly occurs in method comparison studies.[9,10,14,15,23,24] There are two popular ways of coping with it.

Log–log transformation

The simplest approach is to logarithmically transform both x and y values before undertaking the regression analysis. This is easy to execute, it usually evens out scatter around the line and it provides a more accurate tool for calibrating one method against another. However, it leads to difficulties in making inferences about the presence of proportional or fixed bias.

Weighted regression

An alternative is to weight the values of x and y in an analogous fashion to weighted least squares (WLS) regression in the context of Model I regression analysis.[3] Monte Carlo simulation studies of standard major axis regression show that type I error in testing for proportional bias is better controlled if weighting is done.[14,15] It is likely that this also holds true for the closely related OLP regression. A description of how to execute weighted least products (WLP) regression is in Appendix 1. The outcome of working the example by this technique is given in Table 2. In this case, the WLP regression coefficients are not very different from those by OLP, but the population CI are consistently narrower. The inferences from hypothesis testing are unchanged.

MODEL II VERSUS MODEL I REGRESSION ANALYSIS

A matter on which all serious exponents of method comparison techniques agree is that Model I regression analysis gives biased regression coefficients and may lead to false inferences. The discrepancy between the two models is greatest when measurements are made over a narrow range of values located some distance from zero (Figs 3,4), and when random error is large (Fig. 3).[14,15] The poor performance of both OLS and WLS regression analysis of the example of two methods for measuring SBP is apparent in Table 3. The slope of M2 on M1 from Model I (OLS or WLS) analysis is less than that from their Model II counterparts and from OLP and WLP analysis (Tables 2,3). This is always the case. Thus, the Model I regression equation would provide inaccurate calibration of one method against the other. Of greater importance, the Model I approach invites the false inference that α, the Y intercept, is greater than zero and conveys the implication that there is fixed, as well as proportional, bias between the methods.

Table 3. Outcomes of Model I (least squares) regression analysis for the example of Table 1

Technique	Method of execution	a	95% CI	b	95% CI
E(M2) = $a+b$(M1)					
OLS	Computer	20.888	1.000–40.775	0.822	0.711–0.933
WLS	Computer	17.847	0.342–35.352	0.839	0.737–0.942
E(M1) = $a+b$(M2)					
OLS	Computer	−7.410	−32.564–17.743	1.109	0.960–1.258
WLS	Computer	−8.557	−30.616–13.502	1.116	0.979–1.252

Least squares regressions of M2 on M1 (above) and M1 on M2 (below). Regression coefficients: a, intercept; b, slope; 95% confidence intervals (CI) are for the population parameters α and β. OLS, ordinary least squares regression analysis; WLS, weighted least squares regression analysis. Regression analysis was performed with LOSS function on a computer (see Appendix 1).

Table 4. Outcomes of method of differences analysis for the example of Table 1

	Test statistic	Value	n or $d.f.$	P
Original data				
(i) Slope of (M1 − M2) on (M1 + M2)/2	r	0.450	$n = 25$	0.024
(ii) Mean (M1 − M2) = 10.72 versus zero	t	5.97	$d.f. = 24$	<0.0001
(iii) 95% population CI for (M1 − M2)		−7.8–29.3		
Log-log transformed data				
(i) Slope of (logM1 − logM2) on (logM1 + logM2)/2	r	0.298	$n = 25$	0.147
(ii) Mean (logM1 − logM2) = 1.06 versus zero	t	6.48	$d.f. = 24$	<0.0001
(iii) 95% population CI for M1/M2		0.96–1.17		

r, Pearson product–moment correlation coefficient; CI, confidence interval; t, one-sample t statistic. For (i), P value tests the hypothesis $r = 0$. For (ii), P value tests the hypothesis mean (M1 − M2) = 0.

ANALYSIS OF DIFFERENCES

A strategy that is very different to that of regression analysis or correlation has been proposed by Altman and Bland,[9,10] referred to here as the method of differences. Using the notation M1 and M2 to describe values obtained by the two methods, the starting point is to calculate the difference (M1 − M2) for each pair of values. These differences are then plotted against the corresponding means, (M1 + M2)/2. As Altman and Bland point out, this procedure resembles the analysis of residuals used to test whether the assumptions governing OLS regression analysis are satisfied. Thus, the values of (M1 − M2) should be normally distributed and the scatter of the values of (M1 − M2) should be uniform for all levels of (M1 + M2)/2. These are the same requirements as for residuals in OLS regression.[3]

Altman and Bland propose the following analyses of the plot of (M1 − M2) against (M1 + M2)/2.[9,10] If there is no proportional bias between the two methods, then the regression of differences on means should have a slope of zero. They recommend calculating the Pearson product–moment correlation coefficient and testing the hypothesis that $r = 0$. They do not say so explicitly, but the use of r implies that the data points in the plot come from a bivariate normal distribution (as they should do, because both variables are attended by random error). If there is no fixed bias between the two methods (or, as Altman and Bland put it, no relative bias), then the mean of (M1 − M2) should be zero. They recommend this hypothesis be tested by a one-sample t-test. They suggest that the 95% CI for the population of (M1 − M2) differences can be used to decide whether a new method is an acceptable substitute for, or alternative to, the established one.

Example

In their original description of the method of differences, Altman and Bland used the same data set for two methods of SBP measurement as was used earlier (Table 1).[9] The outcome of using their technique to analyse these data is given in Fig. 5 and Table 4. It can be seen that: (i) there appears to be an upward trend of the differences as the mean increases, confirmed by $r \neq 0$ ($P = 0.024$); (ii) the differences are generally located above the zero line and a one-sample t-test strongly rejects the hypothesis that the mean of (M1 − M2) = 0 ($P<0.0001$); and (iii) there is a wide 95% CI for the population of (M1 − M2) values (−7.8 to 29.3 mmHg).

Thus, Altman and Bland would invite the inferences that the upward trend of (M1 − M2) on (M1 + M2)/2 indicates that there is a proportional bias between the two methods, (i.e. method I gives progressively higher readings than method 2 as the level of SBP increases), that mean (M1 − M2) > 0 indicates that there is a fixed bias between the two methods (i.e. over the range of SBP encountered method 1 gives readings that are higher by a constant amount than those from method 2) and the wide 95% CI for (M1 − M2) provides the basis for an opinion that methods 1 and 2 could not be safely interchanged in clinical practice.

These inferences differ from those resulting from least products regression, which were that there was a proportional bias but not a fixed bias (Table 2). As it happens, they coincide with those that could be made from Model I regression analysis (Table 3).

Difficulties with the method of differences

There are two difficulties with the Altman–Bland method that are not fully explored by its originators although they are alluded to by Chinn.[23] First, the tests for proportional and fixed bias are not independent. Using Pearson's r to test for proportional bias is a reasonable approach, given that both M1 and M2 values are subject to random error, but it is wrong to attribute a difference between mean (M1 − M2) and zero to the presence of a fixed bias. It can equally well be accounted for by a mixture of proportional and fixed bias, both acting in the same direction. Conversely, if no difference were found between

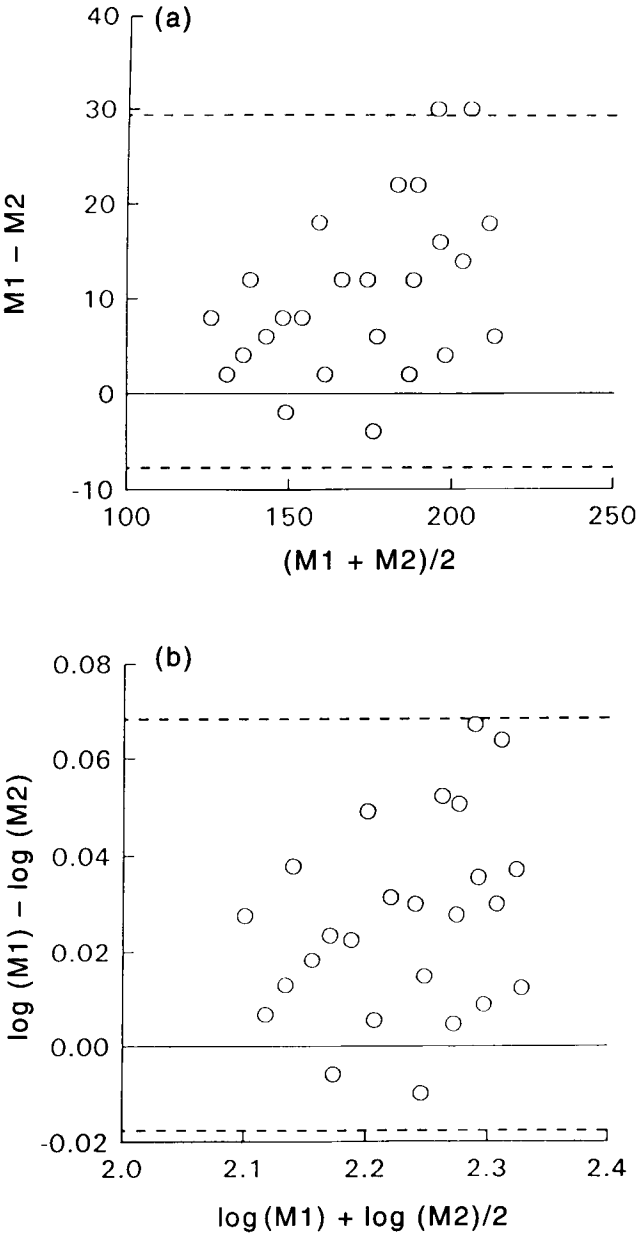

Fig. 5. The method of differences used on the same data as shown in Fig. 4. Dashed lines: 95% CI for the population. (a) Original values. Note the increasing scatter of (M1 − M2) as (M1 + M2)/2 increases. (b) Log_{10} transformed values of M1 and M2. Note the increase in scatter has been reduced, although not eliminated.

mean (M1 − M2) and zero, this could result from there being a proportional bias in one direction and a fixed bias in the other. Second, the scatter of the (M1 − M2) values increases *pari passu* with the value of (M1 + M2)/2 (Fig. 5a). This is to be expected from the proportional increase in error in the scatterplot of M2 on M1 (Fig. 4). This proportional error is commonly seen when the method of differences is used.[9,10,23–26] It introduces bias into estimates of the slope of (M1 − M2) on (M1 + M2)/2. Log-transformation of the original data has been suggested both by the originators of the method of differences[10] and by Chinn.[23] However, although this somewhat reduces the fan-shaped scatter of the differences (Fig. 5b), it creates new problems of interpretation. The term (logM1 − logM2) corresponds to log (M1/M2). This means that if there is proportional bias but no fixed bias between the methods, both the slope of the regression of log(M1/M2) on (logM1 + logM2)/2 and the correlation coefficient will be close to zero. However, if there is fixed but not proportional bias, the slope and r will not be zero. If the mean of log (M1/M2) differs from zero (Table 4), this suggests the presence of proportional bias (given no fixed bias). Furthermore, the antilogged 95% CI for log (M1/M2) is that for the ratio M1/M2 in the population. Thus, although log transformation renders the scatter of (M1 − M2) more uniform, it creates new problems of interpretation.

Finally, it is worth reinforcing a point that Bland and Altman made recently.[27] It is that even if investigators regard method 1 as a 'gold standard', the X axis should remain as (M1 + M2)/2 and should not be changed to M1. Some investigators have fallen into this trap.[28–30]

LEAST PRODUCTS REGRESSION VERSUS THE METHOD OF DIFFERENCES

The chief advantage claimed for the method of differences is that it is comprehensible to clinical investigators and that it is simple to execute. Its popularity can be gauged by its adoption by the British Hypertension Society as the preferred technique for comparing methods for measuring BP.[26] Yet the foregoing discussion indicates that although the method of differences is easy to execute, it is not easy to interpret the results in terms of identifying fixed or proportional bias.

Although least products regression analysis is more difficult to execute than the method of differences, the outcome may be easier for biomedical investigators to understand. They are accustomed to viewing scattergrams on which regression lines of best fit are superimposed and they are familiar with the use of CI. It is just that, in the past, Model I regression analysis has been used to fit the lines instead of Model II. It must be admitted that the Model II regression approach shares with the method of differences the problem of handling proportional error, in which the scatter of x,y points increases with the mean level, which is so common in method comparison studies. However, this can be catered for by using weighted least products regression analysis.

Both techniques rely on the assumption that the residuals around regressions are normally distributed. This is testable if the number of experimental x,y points is sufficiently large. If the normality assumption is not fulfilled or is in doubt, non-parametric bootstrapping techniques can be used to execute Model II regression analysis, especially for estimating CI,[14,31] although the intervals may be too narrow (Table 2).[32]

OTHER CONSIDERATIONS

Homogeneity of the observations

In the example that was analysed, one pair of observations was made in each of many subjects. Least products regression or the method of differences can be used equally well in the case that many measurements have been made in a single subject, provided they have been made independently in random order. What of the case that many measurements have been made in many subjects? This creates a very special problem of analysis, which has been reviewed for biomedical investigators by Feldman.[33] In short, it is not allowable to pool all the within- and between-subject x,y data points and perform Model II regression analysis. Nor, by extension, is it allowable to use the method of differences in this way, although this has sometimes been done.[26,29] However, the weighted least squares solution to this problem proposed by Feldman applies to Model I regressions and cannot be transferred directly to Model II regressions. In future, elegant, computer-intensive techniques may be devised to solve this problem. In the meantime, the best solution (although it is imperfect) may be merely to take the between-subject means of the coefficents a' and b', provided a sufficient number of observations (for instance 20) have been made in each subject to give some confidence in the coefficients.[34]

Non-linearity

So far it has been assumed that the relationship between one method of measurement and another is linear. This is not an unreasonable assumption because it is what usually happens in practice. But what if the relation is non-linear? It clearly indicates that there is a systematic bias between the two methods and that this should be picked up by examining a scattergram. If calibration, rather than detection of bias, is the goal, then it would be logical to use non-linear OLP or WLP regression analysis. This can be done as an extension of the LOSS function technique referred to earlier.

ACKNOWLEDGEMENTS

The author is supported by the National Health and Medical Research Council of Australia.

REFERENCES

1. Berry WD. *Understanding Regression Assumptions.* Sage Publications, Newbury Park, 1993.
2. Berkson J. Are there two regressions? *J. Am. Stat. Assoc.* 1950; **45**: 164–80.
3. Neter J, Wasserman W, Kutner MH. *Applied Linear Statistical Models*, 3rd edn. Richard D Irwin Inc., Homewood, 1990.
4. Kendall MG, Stuart A. *The Advanced Theory of Statistics, Vol. 2. Inference and Relationship*, 3rd edn. Charles Griffin, London, 1961.
5. Brace RA. Fitting straight lines to experimental data. *Am. J. Physiol.* 1977; **233**: R94–9.
6. Sokal RR, Rohlf FJ. *Biometry*, 2nd edn. WH Freeman, New York, 1981.

7. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* John Wiley & Sons, New York, 1978.
8. Lehmann EL. Model specification: The views of Fisher and Neyman, and later developments. *Stat. Sci.* 1990; **5**: 160–8.
9. Altman DG, Bland JM. Measurement in medicine: The analysis of method comparison studies. *Statistician* 1983; **32**: 307–17.
10. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307–10.
11. Kirkeby OJ, Rise IR, Nordsletten L, Skjeldal S, Hall C, Risöe C. Cerebral blood flow measured with intracerebral laser–Doppler flow probes and radioactive microspheres. *J. Appl. Physiol.* 1995; **79**: 1479–86.
12. Tschakovsky ME, Shoemaker JK, Hughson RL. Beat-by-beat forearm blood flow with Doppler ultrasound and strain-gauge plethysmography. *J. Appl. Physiol.* 1995; **79**: 713–19.
13. Ricker WE. Linear regression in fishing research. *J. Fish. Res. Board Can.* 1973; **30**: 409–34.
14. Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin. Chem.* 1993; **39**: 424–32.
15. Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. *Stat. Med.* 1990; **9**: 1463–73.
16. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. *Clin. Chem.* 1979; **25**: 432–8.
17. Feldmann U, Schneider B, Klinkers H, Haeckel R. A multivariate approach for the biometric comparison of analytical methods in clinical chemistry. *J. Clin. Chem. Clin. Biochem.* 1981; **19**: 121–37.
18. Gould SJ. Allometry and size in ontogeny and phylogeny. *Biol. Rev.* 1966; **41**: 587–640.
19. Kuhry B, Marcus LF. Bivariate linear models in biometry. *Syst. Zool.* 1977; **26**: 201–9.
20. Jolicoeur P, Mosimann JE. Intervalles de confiance pour la pente de l'axe majeur d'une distribution normale dimensionelle. *Biom. Praxim.* 1968; **9**: 121–40.
21. Ricker WE. A note concerning Professor Jolicoeur's comments. *J. Fish. Board Can.* 1975; **32**: 1494–6.
22. Daniel WW. *Biostatistics: A Foundation for Analysis in the Health Sciences*, 2nd edn. John Wiley & Sons, New York, 1978.
23. Chinn S. The assessment of methods of measurement. *Stat. Med.* 1990; **9**: 351–62.
24. Close A, Hamilton G, Muriss S. Finger systolic pressure: Its use in screening for hypertension and monitoring. *BMJ* 1986; **293**: 775–8.
25. White SW, Quail AW, De Leeuw PW *et al.* Impedance cardiography for cardiac output measurement: An evaluation of accuracy and limitations. *Eur. Heart J.* 1990; **11** (Suppl. I): 79–92.
26. O'Brien E, Petrie J, Littler W *et al.* The British Hypertension Society protocol for the evaluation of automated and semi-automated blood pressure measuring devices with special reference to ambulatory systems. *J. Hypertens.* 1990; **8**: 607–19.
27. Bland M, Altman DG. Comparing methods of measurement: Why plotting difference against standard methods is misleading. *Lancet* 1995; **346**: 1085–7.
28. Miller R, Bartolo DCC, James D, Mortensen NJMcC. Air-filled microballoon manometry for use in anorectal physiology. *Br. J. Surg.* 1989; **76**: 72–5.
29. Vegfors M, Sjöberg F, Lindberg L-G, Gustafsson U, Lennmarken C. Basic studies of pulse oximetry in a rabbit model. *Acta Anaesthiol. Scand.* 1991; **35**: 596–9.
30. Mengden T, Bättig B, Vetter W. Self-measurement of blood pressure improves the accuracy and reduces the number of subjects in clinical trials. *J. Hypertens.* 1991; **9** (Suppl. 6): S336–7.
31. Jones G, Wortberg M, Kreissig SB, Hammock BD, Rocke DM. Application of the bootstrap to calibration experiments. *Analyt. Chem.* 1996; **68**: 763–70.
32. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* Chapman & Hall, New York, 1993.
33. Feldman HA. Families of lines: Random effects in linear regression analysis. *J. Appl. Physiol.* 1988; **64**: 1721–32.
34. Kelly GE. Use of the structural equations model in assessing the reliability of a new measurement technique. *Appl. Stat.* 1985; **34**: 258–63.

APPENDIX 1

There are two main goals in ordinary least products (OLP) and weighted least products (WLP) regression analysis. One is to make best estimates of the population parameters: that is, values for the coefficients in the linear regression equation $E(Y) = a' + b'x$ (where a' and b' are used to distinguish least products from least squares regression). The other is to obtain robust values for the standard errors attached to the regression coefficients or CI for the population parameters α' and β' (usually 95% intervals).

Three techniques for achieving these goals are described. The first uses the LOSS function in microcomputer statistics packages and should be regarded as the benchmark. The second is to use a hand-held calculator or spreadsheet program. The third uses non-parametric bootstrapping to obtain CI. The first two techniques give identical values for the regression coefficients. The problem lies in obtaining robust values for CI.

Microcomputer

This makes use of the LOSS function that is provided by microcomputer statistics packages that have non-linear regression modules. To use LOSS one has to specify the function that is to be minimized in order to fit a regression. The program then estimates the best-fitting linear regression according to this function by an iterative process and provides estimates of the regression coefficients and their standard errors. The 95% CI for the population parameters can be easily calculated from equations (10) and (11) below.

The example of Table 1 has been worked with SPSS Advanced Statistics 6.1.3/w (SPSS Inc., Chicago, IL, USA), STATISTICA 5.1/w (StatSoft Inc., Tulsa, OK, USA) and SYSTAT 5.0/DOS, 6.0/DOS and 6.0/w (SPSS Inc.). All these packages gave correct values for the regression coefficients (Tables 2,3). STATISTICA 5.1/w and SYSTAT 5.0/DOS gave correct values for standard errors and CI, whereas SYSTAT 6.0 (/DOS, /w) did not. SPSS 6.1.3/w used bootstrapping to arrive at standard errors and CI. The documentation of the bootstrapping technique was difficult to follow and the CI were narrower than those computed parametrically. Readers are advised to use STATISTICA 5.1/w or SYSTAT 5.0/DOS.

Algebraic notation is used to formulate the LOSS functions given below, but this can easily be translated into the language required by an individual package.

Ordinary least squares regression analysis

As a simple example, OLS regression analysis can be executed by the LOSS function (although this is taking a sledgehammer to crack a nut!). The required LOSS function is:

$$(y - (a + bx))^2 \qquad (1)$$

This function represents the square of the vertical deviation of the observed y value from the regression line (the y residual). It is the sum of these squares that is minimized in OLS regression.

Weighted least squares regression analysis

When there is proportional error in Y (i.e. the scatter of observed y values increases in proportion to the level of x), WLS regression analysis is preferred. The required LOSS function is:

$$((y-(a+bx))/x)^2 \quad (2)$$

This function divides each y residual by the corresponding value of x, then squares it.

Ordinary least products regression analysis

In this case, it is the sum of the products of the y and x residuals that is miminized. The required LOSS function, arrived at by simple algebra, is:

$$(y-(a+bx))^2/|b| \quad (3)$$

where $|b|$ is the regression coefficent expressed as an absolute (positive) value.

Weighted least products regression analysis

In WLP regression analysis, the y residuals are divided by x, the x residuals are divided by y and then their products are minimized. The LOSS function can be worked out from equation (3) as:

$$((y-(a+bx))^2)/(|b|)(x)(y) \quad (4)$$

Calculator or spreadsheet

OLS regression analysis

How this can be done on a hand-held calculator is described in all general texts of statistics.

WLS regression analysis

Neter describes a complicated method for doing this by hand.[3] It is made much easier by using a computer spreadsheet:

$$b = \frac{\sum wxy - (\sum wx)(\sum wy)/\sum w}{\sum wx^2 - (\sum wx)^2/\sum w} \quad (5)$$

$$a = \frac{\sum wy - b\sum wx}{\sum w} \quad (6)$$

where x is the original x values, y is the original y values, w is the weighting factor (which is $1/x^2$), wx is the product of w and x (which corresponds to $1/x$), wy is the product of w and y, wxy is the product of w, x and y (which corresponds to y/x) and Σwx^2 corresponds to n, the number of x,y pairs, as $w = 1/x^2$.

The standard errors of the coefficients are calculated as:

$$SE_b = \sqrt{\frac{\text{Error MSq}}{\sum wx^2 - (\sum wx)^2/\sum w}} \quad (7)$$

$$SE_a = \sqrt{\frac{\text{Error MSq}}{\sum (wx - \overline{wx})^2}} \quad (8)$$

where $\overline{wx}$ is the mean wx and the notation is otherwise as given earlier. Error MSq (error mean square) is given by the formula:

$$\text{Error MSq} = \sum(\text{WRES})^2/d.f. \quad (9)$$

where E(Y) is the value of Y predicted from the corresponding value of x, using the coefficients for the WLS regression equation, RES is the vertical deviation (residual) of y from the regression line (y − E(Y)), WRES is the weighted residual (RES/x) and *d.f.* are the degrees of freedom ($n-2$).

The standard errors for a and b can be converted into 95% CI by the conventional formulae:

$$b \pm t\,(SE_b) \quad (10)$$

$$a \pm t\,(SE_a) \quad (11)$$

where the value of t is that corresponding to two-sided $P = 0.05$ at $(n-2)$ *d.f.* For the example, $t = 2.0687$. The values for a and b and for the 95% CI for α and β are almost identical to those obtained by using a LOSS function on a computer (see Table 3).

OLP regression

The coefficient for the slope of the OLP regression $E(Y) = a' + b'x$ can be calculated by hand in any of three ways. First, it can be derived from the values of b in the OLS equations $E(Y) = a + bx$ and $E(X) = a + by$ when they are estimated by ordinary least squares regression of Y on X, and X on Y, respectively. Then:

$$b' = \sqrt{(b_{y.x})(1/b_{x.y})}$$

Hence the phrase 'geometric mean regression' which is sometimes used. Second, the coefficient of the slope of the OLP regression is given by $(b_{y.x})/r$, where r is the Pearson product–moment correlation coefficient. Third, it is given by s_y/s_x, the ratio of the standard deviations of the y and x values.

Because the OLP regression line passes through the point $\overline{x},\overline{y}$ (mean x, mean y), the intercept a' can be obtained from the formula:

$$a' = \overline{y} - b'\overline{x} \quad (12)$$

Jolicoeur and Mosimann,[20] reproduced more accessibly by Richter,[21] give a formula to calculate approximate 95% CI for the population parameters α' and β'. The first step is to find the value of r for the least products regression. This is identical to the conventional Pearson product–moment correlation coefficient for x and y. The Jolicoeur formula depends on calculating B as:

$$B = F(1-r^2)/(n-2) \quad (13)$$

where the value of F corresponds to $P = 0.05$ at *d.f.* $(1, n-2)$.

Then, the upper and lower 95% confidence limits for β' are calculated, respectively, as:

$$b'(\sqrt{(B+1)} + \sqrt{B}) \quad (14)$$

and

$$b'(\sqrt{(B+1)} - \sqrt{B}) \quad (15)$$

These hand-worked limits for β' are similar, although not identical, to those given by the LOSS method (see Table 2).

The limits for α' can be obtained by the somewhat dubious device of substituting the upper and lower confidence limits for β' in equation (12). They are not very different from those obtained by the LOSS technique (Table 2).

WLP regression

Accurate values for b' can be obtained from the formula:

$$b' = \sqrt{(b_{y.x})(1/b_{x.y})} \quad (16)$$

where $b_{y.x}$ and $b_{x.y}$ are the WLS coefficients for the regression of y on x, and x on y, respectively (see Table 3).

The intercept a' can be obtained from equation (15). The accuracy of b' is guaranteed, while that of a' is a little less reliable (see Table 2).

Obtaining reliable 95% CI for α' and β' is more problematic. One approach is to average the two closely similar values of r resulting from WLS regression and enter them into equations (12) through to (15). This resulted in intervals that were of similar width to, but differently centred from, those obtained by using the LOSS function (Table 2).

Non-parametric bootstrapping

For OLP regression this was performed on a microcomputer spreadsheet by creating 1000 random resamples of the original x,y values of Table 1. For each resample, b' was calculated as s_y/s_x and a' from equation (15). The coefficients were ranked and the 25th and 975th values were taken as the 95% confidence limits. The CI for both α' and β' are narrower than those estimated parametrically (Table 2), as they were when SPSS was used (see earlier). This is an effect of the truncated tails of raw bootstrap distributions.[32]