How to use linear regression and correlation in quantitative method comparison studies

P. J. Twomey, M. H. Kroll

SUMMARY

Linear regression methods try to determine the best linear relationship between data points while correlation coefficients assess the association (as opposed to agreement) between the two methods. Linear regression and correlation play an important part in the interpretation of quantitative method comparison studies. Their major strength is that they are widely known and as a result both are employed in the vast majority of method comparison studies. While previously performed by hand, the availability of statistical packages means that regression analysis is usually performed by software packages including MS Excel, with or without the software program Analyze-it as well as by other software packages. Such techniques need to be employed in a way that compares the agreement between the two methods examined and more importantly, because we are dealing with individual patients, whether the degree of agreement is clinically acceptable. Despite their use for many years, there is a lot of ignorance about the validity as well as the pros and cons of linear regression and correlation techniques. This review article describes the types of linear regression and regression (parametric and non-parametric methods) and the necessary general and specific requirements. The selection of the type of regression depends on where one has been trained, the tradition of the laboratory and the availability of adequate software.

Introduction

The main purpose of method comparison studies is to assess not whether the two methods agree but, as healthcare professionals treat individual patients, how well the two methods agree from the point of individual specimens across the whole data range found in clinical practice (1). Linear regression methods try to determine the best linear relationship between data points while correlation coefficients assess the association (as opposed to agreement) between the two methods. They need to be employed in a way that compares the agreement between the two methods examined and more importantly, because we are dealing with individual patients, whether the degree of agreement is clinically acceptable. Despite their use for many years, there is a lot of ignorance about the validity as well as the pros and cons of linear regression and correlation techniques.

Before the introduction of difference plots by Bland and Altman to statisticians (2) and to medics (3) in the mid-1980s, the interpretation of quantitative method comparison studies relied heavily on linear regression and correlation. While difference plots have added to regression methods, the latter are important for validating the utility of a method in relation to analytical quality specifications (4,5) and are still employed in the vast majority of published papers on quantitative method comparison studies. We are limiting the scope of this review to the use of linear regression and associated plots.

When investigating the relationship between two methods, all individual samples from 1, 2, ... n in the population being studied are analysed by both analytical methods with one method being assigned x (horizontal axis) and the other method being assigned y (vertical axis). Usually, the x variable is called the independent, explanatory or predictor variable and the y variable the dependent, response or outcome variable. For example, taller individuals tend to be heavier and thus height would be the independent variable while weight would be the dependent variable. In quantitative method comparison studies, this is not strictly the case but where a method has a higher hierarchy, it is usual to assign it as x. Thus, a ‘gold standard’ or reference method would be assigned as x; however, many studies...
compare two methods of equal status and accordingly \( x \) and \( y \) can be assigned as decided by the investigators.

To see if there is a relationship between the two methods, the points are plotted on a scatter diagram, that is, the concentrations for one method are plotted on the x-axis against their corresponding concentrations from the second method on the y-axis of a two dimension diagram (Figure 1). As both methods purport to measure the same analyte, the relationship can be assigned as decided by the investigators.

Another linear regression output is the \( S_{\text{res}} \) (\( S_{\text{res}} \) is utilised by some authors). This is the standard deviation (SD) of the residuals and provides us with a quantification of the spread of the data around the regression line. While statistics packages may produce a value for the \( S_{\text{res}} \), its validity from a quantitative method comparison study viewpoint depends on whether the SD of the residuals is constant or not across the whole of the data range. If it is constant, then the \( S_{\text{res}} \) concept is valid for the whole data range; however, frequently, the SD of the residuals increases with increasing concentration and accordingly, the \( S_{\text{res}} \) is not representative of all the data and an alternative method of evaluating the data such as difference plots is advisable (5).

**Types of linear regression**

Two main types of linear regression techniques exist: parametric and non-parametric. Parametric methods include ordinary (or least squares) linear regression (OLR) and Deming’s linear regression (DR) (7) and the non-parametric methods include Passing Bablok linear regression (PBR) (8–10). For all linear regression methods, the uncertainty and/or error in the regression estimates increase when:

- The data are not adequately/evenly distributed over the investigation range.
- The number of data points is low.
- The samples employed are not independent, for example, analysing specimens more than once and using more than one data set per sample results in dependent samples.
- The relationship between the data is not linear.
- The magnitude of the SD of the residuals (\( S_{\text{res}} \) – the average deviation from the best-fit line) relative to the data range is high.

More specific requirements include:

- Parametric methods – the distribution of the residual values (and not the data points themselves) in the \( y \) axis has a Gaussian distribution.
- Unweighted parametric methods – the variability of the distribution of the \( y \) values in the population is the same for all values of \( x \), that is, the variance \( \sigma^2 \) is constant.

![Figure 1](A) Scatter plot of method A (mmol/l) vs. method B (mmol/l) with the ordinary linear regression line \( y = 0.87x + 0.227, r = 0.998 \). (B) Scatter plot of method A (mmol/l) vs. method B (mmol/l) with the line of equality \( x = y \) for the same data as (A). The line of equality graphically shows the slope < 1.0 for the data.
• For OLR, the x variable can be measured without error (model I regression technique) – this is not a prerequisite for DR or PBR (model II regression techniques).

The availability of statistical packages means that regression analysis is usually not performed by hand. Once employed appropriately, such packages usually produce fewer errors than manual calculation. A good statistics package will calculate the 95% confidence interval for the slope and intercept as well as tests of linearity and outlier detection.

The regression line is calculated in OLR by minimising the squared residuals in the y direction leading to its other name of ‘least squares’ regression (Figure 2). It assumes that the x variable is error free (no constant, proportional or random error), something that is not met even by reference methods because of random measurement errors. However, any differences because of this effect decrease as the range of data points increase but the type I error (also called α error or false-positive) may still be high (11). The usual form of OLR (unweighted OLR) assumes a constant analytical imprecision (SD$_Y$) of the y variable – the ‘homoscedastic’ variance. If the imprecision of the y variable is not constant, then weighted forms can be employed to take account of the variable imprecision. Views differ regarding the unweighted and weighted forms of OLR, with Cornbleet and Gochman (12) believing that unweighted is satisfactory when SD$_Y$ is proportional to x, while Linnet (11) believes that the standard error of the slope may be underestimated by 40% on average. If the constant variance assumptions are in doubt, we recommend that unweighted OLR should not be used.

Another issue with OLR is the effect of outliers which generate large squared residuals which can shift the calculated regression line towards the errant point(s). Some authors recommend looking for outliers where r < 0.975 (13) or < 0.99 (14). Draper and Smith (5) have recommended that points which result in residuals > 4 SD$_Y$ should be omitted from OLR analysis and Cornbleet and Gochman (12) have shown how residual analysis about the regression line can produce a criterion for rejecting spurious values. Irrespective of the exact value of the cut-off employed, the use of such a cut-off is to be recommended to improve the accuracy of the data derived by OLR. Another concept is Cook’s distance which is based on the residual of the data point – the larger the residual, the larger Cook’s distance and thus the more likely the data point may have an influence on the linear regression equation. Thus, the higher the Cook distance, the more like a point is an outlier. The use of Cook’s distance or equivalent is also recommended when regression analysis is used to identify points that influence the regression line (15).

Deming’s linear regression minimises the distances of the data points orthogonal (at right angles) to the regression line (7) as opposed to OLR which does so in the y (vertical) axis (Figure 2). Like OLR, DR also has unweighted and weighted forms with the former assuming equal x and y variances throughout the measurement range and the latter being more efficient when there is proportional measurement uncertainty (16). The distance of the data points are minimised at an angle to the regression line when the variances of x and y differ with a constant ratio and this is independent of this ratio (17). Again, if the constant variance assumptions are in doubt, we recommend that unweighted DR should not be employed. As DR is also a parametric regression method, it is sensitive to outliers in a similar fashion to OLR. Thus, we recommend that the use of a residual cut-off is to be recommended to improve the accuracy of the data derived by DR.

The non-parametric PBR (8–10) is believed by Jones and Payne (18) to have no bias in its estimates of the slope and intercept when variances change at different rates because of increases in analyte concentrations. As PBR is based on the rank principle with each pair of results being given equal weight in the calculation of the regression line, it is believed to be less sensitive to outliers. One disadvantage of PBR is that it has no SD$_Y$.

There is no agreement when it comes to which linear regression method should be employed in...
quantitative method comparison studies. The research into which regression method to use in quantitative method comparison studies focused mainly on the slope estimation error. Many evaluators prefer DR over OLR; however, some employ OLR depending on the Pearson coefficient despite the lack of supporting evidence. For example, they restrict the use of OLR to those cases where $r \geq 0.975$ (13) or $\geq 0.99$ (14). PBR also has its advocates over DR (19) whereas others say that it is an alternative to OLR (11). In addition to not having $S_{yr}$, Linnet (11) has shown that PBR has inadequacies in simulation tests. When OLR and correlation analysis provide poor estimates, Stockl et al. (20) recommend the investigation of the analytical reasons for the poor estimations rather than the use of other linear regression methods. This also applies when there are significant differences between the different regression methods. Where possible, the finding of an outlier should be handled by repeating the assay to rule out a gross analytical error, for example, because of mis-sampling. The greater the difference between the two values relative to the coefficient of variation, the greater the likelihood that there is a gross analytical error. Sometimes, a third analyses may be necessary to confirm which value is the outlier. When a gross analytical error is discovered, one of the correct estimations and not the mean should be used. When there is a focus on the accuracy of an analytical method, the choice of a regression method can be quite important. The unknown true slope has a 95% chance of being within two SDs of the slope estimate. Accordingly, the achievement of good precision for the slope estimate is as important as eliminating bias because if an unbiased slope estimate has a poor precision, possibly caused by the use of a non-optimal regression method, the result may still differ significantly from the true value.

The selection of the type of regression depends on where one has been trained, the tradition of the laboratory and the availability of adequate software. Outside of the clinical laboratory, almost all evaluators use OLR. It has the advantage that because it is parametric and derived directly from mathematical principles its limitations are clearly known and understood. If one is careful about using appropriately collected data with a sufficient spread of distribution the results are reproducible and clear to interpretation. However, even with the best of data, regression analysis has its limits. The $y$ intercept possesses a great amount of uncertainty if there are no data points near it, for example, in comparing methods for serum sodium. If all of the points hover near the lowest 25% of reference interval values, the slope, especially at higher values is suspect, whether or not one uses parametric or non-parametric methods. If the range of values approaches two orders of magnitude, predictions and assessments at the lower end are more suspect.

**Types of correlation**

Two main types of correlation techniques also exist: parametric and non-parametric. The Pearson $r$ product moment correlation coefficient is the parametric method and the Spearman’s $\rho$ ($\rho$) rank is the non-parametric method. For the Pearson correlation method, the uncertainty and/or error in the correlation estimate increases when:

- The data are not adequately/evenly distributed over the investigation range or contains subgroups.
- The number of data points is low.
- The samples employed are not independent.
- The relationship between the data is not linear.
- The data contains outliers.
- The distribution of the values in both the $x$ and $y$ axes has a non-Gaussian distribution.

The use of ‘uncertainty’ above includes bias, as in the case of inadequate data distribution, increased error irrespective of any estimates as in the case of a low number of data points, and inappropriate methods as with the use of dependent samples, non-linear data or non-Gaussian distributed data. Should there be any concern repoints 1–6, the Spearman correlation method should be employed. Additionally, it should always be employed when at least one of the two variables is measured on an ordinal scale.

**Preanalytical planning**

Preanalytical planning is essential for valid comparison as it determines the quality of all subsequent stages. The first step is to have an appropriate statistical package available – such a package must be used appropriately to produce valid results and interpretation.

The data range employed in the comparison study should ideally reflect the whole analyte range found in clinical practice with an even distribution throughout so that conclusions can be made for all values encountered in clinical practice. The study must be adequately powered to detect a significant difference between the two methods. Such a difference needs to be decided before the study is carried out and should ideally have a foundation in evidence-based medicine before performing power analysis to ensure that sufficient samples are analysed in the study. You need the SD of both methods across the whole analyte range to be able to calculate the required number of patients (21). With increasing
analyte concentration, you will see if the SDs of both methods are the same, change but have the same ratio or change and have a changing ratio. Such information may be important in deciding which type of regression to use and whether $S_{y|x}$ is representative throughout the data range employed in the study. If the constant variance assumptions are in doubt, unweighted parametric methods should not be employed.

Because the quality of the data may have a significant influence on the regression procedure, it is important that internal quality control (IQC) are acceptable. However, some authors (20) believe that more IQC should be run more often than usual to ensure that the quality of the data is satisfactory by reducing the likelihood of assay drift.

**Visual data analysis**

As in examining a patient, the first step is visual inspection. This allows you to see if visually determine whether there is a relationship between the data. Visual inspection can quickly help you identify problems with your data and thus can potentially save time by identifying problems, for example, the relationship may actually be curved but this does not prevent software producing values for the slope and intercept. From the outset, you must decide what method to place on the x axis. If the study includes a ‘gold standard’ or reference method, this should be assigned as the x variable. Many studies compare two methods of equal status and accordingly it is not important which method is chosen as x and y variables; however, if one of the methods is your current laboratory method, you may want to choose it as the x variable. Ideally, the scales for both the x and y axes should be he same and the graph should include the point (0, 0). However, software packages tend to produce graphs with a longer x axis and a shorter y axis and this now seems to be the norm; resolution can also be an issue and thus the (0, 0) may be dropped depending on the data.

The next step is to examine the relationship between the data. Clearly, a linear regression model is only valid if the data has a linear relationship. Accordingly, quantitative method comparison studies should be accompanied by graphical presentations (13,22) to prove that the data is linear. A simple visual examination of the data is the sign sequence – all one has to do is to count the number of consecutive points on the same side of the regression line. The greater the number, the less likely the data is linear. As a result of performing the sign sequence, it may be possible to see non-linear relationships (Figure 3A, B).

![](image)

**Figure 3** (A) Visual inspection of the data points around the linear regression line shows that at concentration on the x axis of 6 mmol/l and below, that all the points are under the regression line; at greater concentrations they cross the regression line until a concentration of > 20 mmol/l when they all fall below the regression line. In the middle section of the data, approximately 12 points are above the regression line. The residual plot will clearly show you that the data is not linear. In addition, statistical packages and the Runs test can be also used to confirm the visual impression that this is not what you would expect from linear data. (B) Shows that the data has a better fit with a polynomial curve (visually, more points closer to the line and mathematically, r is closer to 1.0).

The residual plot (Figure 4A, B) is a more complex graphical interpretation of the analytical data. A residual is defined as the difference between the observed and fitted values of the dependent variable in a regression analysis ($y = y - \hat{y}$); basically, it is the distance between the data point and the regression line. In OLR, the distance is measured vertically to the regression line where as in DR it is measured perpendicular to the regression line. The y axis is freely scalable and the x axis consists of the comparative method and bisects the y axis at zero. The residual plot is useful for the judgement of linearity as per the sign sequence – if there is an even spread of points around the $y = 0$ throughout the x axis this implies that there is linearity between the two methods (Figure 4A, B).
The next step is visual examination of the scatter and residual plots for outliers and influential points. Outliers are points ‘whose discordancy from the majority of the sample is excessive in relation to the assumed distributional model for the sample, thereby leading to the suspicion that it is not generated by this model’ (23). Unfortunately, the visual use of the scatter plot (Figure 1A) is not the most sensitive means of detecting outliers although the residual plot (Figure 4A) may help. Therefore, it is recommended that graphical techniques such as absolute and relative difference plots should also be employed as these often have better resolution (24).

Finally, the residual plot (Figure 4A, B) can be employed to verify whether the SD is constant or not. Where the absolute residuals are not uniform throughout the concentration range, for example, if the data spread is similar to the letter ‘v’ on its side, the SD is not representative throughout the data range. Appropriate forms of linear regression should be employed when the SD is not constant and it should be remembered that $S_{xy}$ is not representative through out the data range.

**Statistical data analysis**

The polynomial test for linearity represents an objective and statistically robust method of assessing the linearity of a data set (25–28). Basically, the use of polynomial regression (5) (such as $y = x^2 + 5x + 2$) produces a better fit for the data as judged by a correlation coefficient nearer to 1.0 than can a straight line ($y = mx + c$). The Runs test (an objective statistical form of the sign sequence) for sample randomness can also be employed to check for linearity (29). If the residuals (as opposed to the original data points) do not have a Gaussian distribution about the x-axis, then non-parametric methods such as PBR should be employed.

According to Jones and Payne (18), it is reasonable to use parametric methods if the tests for normality give probability values of 0.1 or greater. For non-linear data, there are two options: first, fit the data to a polynomial regression or curvilinear regression; secondly, the data can be transformed to a linear relationship, such as using log or reciprocal transformations. Logarithmic transformation usually is sufficient to meet this criterion but this cannot be assumed; furthermore, it may be associated with other problems too (30). Unfortunately, linear regression techniques are often inappropriately applied and therefore the limits of agreement determined are not valid across the whole range of values. Accordingly, incorrect data interpretation and/or conclusions may result.
Various mathematical and statistical techniques are available for the detection of outliers (31). Influential points are points that substantially alter the estimates of the slope, intercept or correlation coefficient when included in the analysis. For example, the correlation coefficient is weighted towards data points with large numerical values, especially when analysing data that is numerically small. The inclusion of a single data point of a high concentration can result in a large improvement in both the r value and the regression data (Figure 5A, B). The easiest way of investigating the effect of such points is to perform linear regression and calculation of the correlation coefficient with and without such points and to note the differences between the calculated values and their respective confidence intervals. As a general rule, there is less concern about \( r \geq 0.99 \) and some authors would even be happy with \( r \geq 0.975 \) (13). As the correlation coefficient, \( r \), measures the amount of association between the two variables, \( x \) and \( y \), a good correlation coefficient does not necessarily mean that there is good agreement between the two methods. For example, if the results of \( y \) were double each value of \( x \), \( r \) would be 1.0 despite the clear differences between the two methods. Thus, correlation coefficients alone have little use in determining agreement between two methods.

**Interpreting the regression data**

In the first instance, the slope and intercept can be compared with 1.0 and 0.0 respectively. Graphically, the regression line and the line of equality \((x = y)\) can be added to the scatter plot to improve such an assessment (Figure 1A, B). Such methods are sometimes not enough as resolution of the scatter plot can be poor at low concentrations (Figure 1A, B), but it is relatively easy to plot another scatter plot with the relevant data such that the resolution problem is overcome (Figure 6A, B).

The next step is to examine the slope and intercept with their respective 95% limits of agreement. If the 95% limits of agreement for the slope include 1.0, then the regression line slope is not statistically different from the slope in the line of equality; similarly, if the 95% limits of agreement for the intercept include 0.0, then the regression line intercept is not statistically different from the intercept in the line of equality. However, interpretation of the regression data is more than simply recording the slope, intercept and their 95% limits of agreement and comparing them to 1.0 and 0.0 respectively. Such statistical methods simply quantify the spread of the data; they do not put the data into a clinical perspective which is essential in quantitative method comparison studies.

---

**Figure 5** (A) The method comparison between methods A and B is not good for concentrations < 12 mmol/l. (B) However, the addition of an extra data point where both methods give the same result at a much higher concentration (100 mmol/l) results in much improved regression and correlation data.

**Figure 6** (A) Scatter plot of method A (mmol/l) vs. method B (mmol/l) using values < 10.0 mmol/l with the ordinary linear regression line \((y = 0.87x + 0.227, r = 0.998)\). (B) Scatter plot of method A (mmol/l) vs. method B (mmol/l) with the line of equality \((x = y)\) for the same data as Figure 3A. The line of equality graphically shows the slope < 1.0 for the data
To put the data into a clinical perspective, you need to compare the data spread with the data error that would be allowable without compromising the interpretation and medical use of the assay. Thus, you should be looking at the data of the individual samples studied as opposed to the central tendency of the population studied. For example, two glucose methods might have a linear relationship of approximately $y = 1.0x$ with an $r$ of 0.91 but have a difference range between the two methods of approximately 54% (Figure 7A, B). A fasting plasma glucose of 7.0 mmol/l on one method could thus have a glucose concentration between 5.1 and 9.0 mmol/l on the other method. Such a 95% limit of agreement between the two methods for the fasting diagnostic cut-off of 7.0 mmol/l would be 5.12–8.96 mmol/l. This is unacceptable from a clinical perspective as the spread is far too large.

To put the data into a clinical perspective, you need to compare the data spread with the data error that would be allowable without compromising the interpretation and medical use of the assay. Thus, you should be looking at the data of the individual samples studied as opposed to the central tendency of the population studied. For example, two glucose methods might have a linear relationship of approximately $y = 1.0x$ with an $r$ of 0.91 but have a difference range between the two methods of approximately 54% (Figure 7A, B). A fasting plasma glucose of 7.0 mmol/l on one method could thus have a glucose concentration between 5.1 and 9.0 mmol/l on the other method. Such a 95% limit of agreement is clinically unacceptable. Accordingly, methods can have poor agreement despite a slope of 1.0 and an intercept of zero.

The derivation of total allowable error has a hierarchy (32). Irrespective of how it is derived, it should be prespecified in the preanalytical phase. The data can be compared with the total allowable error graphically or mathematically – 95% of the specimens should be within the total error limits for a method to be clinically acceptable. Graphically, constant or proportional total error limits can be easily applied to scatter plots (Figure 8A). Resolution of the scatter plot may be limiting but rescaling can improve this as mentioned previously (Figure 8B). Mathematical comparison involves either comparing the residuals (calculated by using the regression equation) or the differences (absolute or relative) for each specimen in the study to the total allowable error.

$S_{y|x}$

$S_{y|x}$ is the SD of the residuals and is calculated by parametric regression methods, each with its own calculation. When performing quantitative method comparison studies, one is attempting to answer the question whether there is a proportional bias, a constant bias or random error. In linear regression, proportional bias is represented by the slope, constant bias by the $y$ intercept and the degree of total random error [including sample-related effects, total analytical precision ($S_{a, tot}$), drift or shift, and non-linearity] between the two methods by the $S_{y|x}$. While $S_{y|x}$ indicates the magnitude of total random error like the correlation coefficient $r$, it is independent of the data range and is also more likely to have a greater degree of change with increasing random error (20). Sample-related effects are possible and
should be suspected when $S_{yx} \gg S_{a,\text{tot}}$; accordingly, it is recommended that $S_{yx}$ is compared with $S_{a,\text{tot}}$ to investigate the existence of such an issue.

Usually, both methods contribute to $S_{yx}$ in method comparison studies; however, without formally comparing the SDs of both methods, it is not possible to say which method in the study might be contributing more to $S_{yx}$ or the 95% limits of agreement. The $F$-test (or variance ratio test) tests the null hypothesis that the variances of the two methods do not differ (18). Accordingly, irrespective of whether the $S_{yx}$ is representative or not across the concentration range, it is important that the $F$ test is applied to the data. Only then can one determine which method has the best coefficient of variation and accordingly is contributing least to the $S_{yx}$ and/or 95% limits of agreement. Where the $S_{yx}$ is representative across the whole data range, this value needs to be put into a clinical perspective; 95% of values will be within $\pm 2 S_{yx}$ and it is this range that must be compared with the clinical requirements of the test in a manner similar to the $\pm 2$ SDs as advocated by Altman and Bland (2,3). Unless objective total criteria are used by means of graphical and/or statistical techniques, regression methods in qualitative method comparison studies will not be employed as they ideally should be. This point is similar to the production of difference plots and not putting the two SD limits into a clinical context, for example, it is possible to have a linear regression equation of $y = x$, that is, a line with a slope of 1.0 and an intercept of 0.0, but yet have a large $S_{yx}$ and/or 95% limits of agreement such that the data exceeds the total allowable error (Figure 7A, B).

**Summary**

Linear regression plays an important part in the interpretation of quantitative method comparison studies. Its major strength is that it is widely known. However, as with all methods used to interpret such data, it is important that it is employed properly as failure to do so means that incorrect conclusions may be arrived at. Various techniques can be employed to overcome weaknesses. These include the addition of the line of equality, the regression line and total error criteria to scatter plots as well as the potential to regraph data using lower values to improve the resolution. However, if the spread of the differences is not equal throughout the data range, the $S_{yx}$ is not valid and in such cases, difference plots should be employed (24). One must evaluate the results of regression analysis based on the quality and of the data entered. However, regression analysis has its limitations even with the best of data because of the biological limitations of the data as outlined for sodium. Also, in cases with very wide ranges, points of discontinuity, or in the presence of small amounts of non-linearity, dividing the data into subsets for linear regression analysis remains a valid and useful approach.

**References**

1. Altman DG, Bland JM. Commentary on quantifying agreement between two methods of measurement. *Clin Chem* 2002; 48: 801–2.
2. Altman DG, Bland JM. Measurement in medicine: the analysis of comparative studies. *Statistician* 1983; 32: 307–17.
3. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1: 307–10.
4. Westgard JO, Hunt MR. Use and interpretation of common statistical tests in method comparison studies. *Clin Chem* 1973; 19: 49–67.
5. Draper NR, Smith H. *Applied Regression Analysis*. New York: John Wiley and Sons, 1966, pp. 44–103.
6. Twomey PJ, Viljoen A, House IM, Reynolds TM, Wierzbicki AS. Problems with non-ceruloplasmin bound copper (‘free copper’ or copper index). *Clin Chem* 2005; 51: 1558–9.
7. Deming WE. *Statistical Adjustment of Data*. New York: Wiley, 1943; 184.
8. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *J Clin Chem Clin Biochem* 1983; 21: 709–20.
9. Passing H, Bablok W. Comparison of several regression procedures for method comparison studies and determination of sample sizes. Application of linear regression procedures for method comparison studies in clinical chemistry, Part II. *J Clin Chem Clin Biochem* 1984; 22: 431–45.
10. Passing H, Bablok W, Bender R, Schneider B. A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. *J Clin Chem Clin Biochem* 1988; 26: 783–90.
11. Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem* 1993; 39: 424–32.
12. Coraboeuf PJ, Gochman N. Incorrect least-square regression coefficients in method-comparison analysis. *Clin Chem* 1979; 25: 432–8.
13. National Committee for Clinical Laboratory Standards. *Method Comparison and Bias Estimation Using Patient Samples*. Approved guideline, NCCLS document EP9-A. Villanova, PA: NCCLS, 1995; 15: 1–36.
14. Wakkars PJ, Hellendoorn HB, Op De Weegh GI, Heerspink W. Applications of statistics in clinical chemistry. A critical evaluation of regression lines. *Clin Chim Acta* 1975; 64: 173–84.
15. Henderson AR. Information for authors: is the advice regarding the reporting of residuals in regression analysis complete? Should Cook’s distance be included? *Clin Chim Acta* 2006; 352: 1848–50.
16. Bablok W, Haackel R, Meyers W, Wosniok W. Biometrical methods. In: Haackel R, ed. *Evaluation Methods in Laboratory Medicine*. Weinheim: VCH Verlagsgessellschaft, 1993: 203–21.
17. Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. *Stat Med* 2000; 19: 1463–73.
18. Jones R, Payne B. *Clinical Investigation and Statistics in Laboratory Medicine*. London: Association of Clinical Biochemists, ACB Venture publications, 1997.
19. Payne RB. Method comparison: evaluation of least squares, Deming and Passing/Bablok regression procedures for method comparison studies. *Ann Clin Biochem* 1997; 34: 319–20.
20 Stockl D. Dewitte K. Thienpont LM. Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data? Clin Chem 1998; 44: 2340–6.

21 Linnet K. Necessary sample size for method comparison studies based on regression analysis. Clin Chem 1999; 45: 882–94.

22 Stockl D. Beyond the myths of difference plots. Ann Clin Biochem 1996; 33: 575–7.

23 Hawkins D. Outliers. In: Johnson N, Kotz S, eds. Encyclopedia of Statistical Sciences. New York: John Wiley & Sons, 1985: 539–43.

24 Twomey PJ. How to use difference plots in quantitative method comparison studies. Ann Clin Biochem 2006; 2: 124–9.

25 EP6-A, NCCLS. Evaluation of the Linearity of Quantitative Measurement Procedures: a Statistical Approach; Approved Guideline. Vol. 23, No. 16, NCCLS document EP6-A. Wayne, PA: NCCLS, 2003.

26 Kroll MH, Emancipator K. A theoretical evaluation of linearity. Clin Chem 1993; 39: 405–13.

27 Emancipator K, Kroll MH. A quantitative measure of linearity. Clin Chem 1993; 39: 766–72.

28 Kroll MH, Emancipator K, Floering D, Tholen D. An algorithm for finding the linear region in a nonlinear data set. Comput Biol Med 1999; 29: 289–301.

29 Sheskin DJ. Handbook of Parametric and Nonparametric Statistical Procedures, 2nd edn. Boca Raton: Chapman & Hall, CRC Press LLC, 2000.

30 Twomey PJ, Kroll MH. Errors in the assessment of estimated glomerular filtration rate. Clin Chem 2006; 52: 153.

31 Barnett V, Lewis T. Outliers in Statistical Data. New York: John Wiley & Sons, 1984.

32 Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Consensus Agreement. Scan J Clin Lab Invest 1999; 59: 585.

Paper received December 2007, accepted January 2008.