Method comparison—a different approach

M A Pollock, S G Jefferson, J W Kane, K Lomax, G MacKinnon and C B Winnard
From the Department of Clinical Biochemistry, Hope Hospital, Eccles Old Road, Salford M6 8HD, UK

SUMMARY. The commonly accepted method of analysing data from method comparison studies is regression analysis, a method which has limitations. This study illustrates the use of a graphical presentation of data, the difference plot, which can be used as an alternative to least squares regression analysis. The data from comparison studies performed on five methods were analysed both by Deming's regression analysis, with calculation of the correlation coefficient, and by the difference plot. The results show that in most cases much more relevant information was obtained from the difference plot.

Additional key phrases: difference plot; regression analysis; correlation coefficient

In clinical biochemistry it is often necessary to compare a new method with an established one to see whether they agree sufficiently for the new to replace the old. Method comparison studies must include assessments of precision, sensitivity and specificity. Analysis of data from such studies should also address the questions of how much the new method differs from the old, whether there is a concentration related bias, and whether either method measures the true concentration of the analyte.

Currently, method comparison data are usually analysed by regression analysis and correlation coefficients. This approach is inappropriate for the following reasons:

1. For simple least squares regression analysis the points should be normally distributed which can be difficult to achieve where a range of values is tested. The least squares method also assumes that the x variable has no error although allowance can be made for this by use of Deming's regression method.[1]
2. The slope of the regression line can give some information about the agreement between two methods but is substantially affected by the range of values chosen.
3. The degree of correlation also depends on the range of results in the sample. A wider range will give a better correlation but not necessarily better agreement. In addition a high r value (e.g. 0·95) may hide wide differences since data which show poor agreement visually can produce high correlation coefficients.
4. The test of significance of r will show that the methods are related. Two methods developed to measure the same analyte would, however, be expected to be related.

An alternative approach to the use of linear regression and correlation has been described for assessing agreement between two methods of clinical measurement.[2] We have previously described the use of this approach—the difference plot—in investigating bias in luteinizing hormone (LH) immunoassays.[3] This study compared a radio-immunoassay (RIA) for LH to an immunoradiometric assay (IRMA) and showed that the RIA was positively biased with respect to the IRMA particularly at low concentrations of LH. The bias was not apparent in the regression analysis but was clearly seen in the difference plot. The present study extends the use of the difference plot to other method comparisons in clinical biochemistry.

METHODS

Five method comparisons were performed: glycated haemoglobin measured by an ion exchange method versus an immunoenzymatic method (enzyme-linked immunoadsorbent assay; ELISA); ^{125}I-labelled RIA of androstenedione versus ^{3}H-labelled RIA; turbidimetric method for lipoprotein a (LPa) versus two-site IRMA; high

Correspondence: Dr M A Pollock.

TABLE 1. *Deming regression equations and correlation coefficients for five method comparisons*

Comparison	Deming regression line	Correlation coefficient (r)
Glycated Hb ion exchange (x) versus glycated Hb ELISA (y)	$y=1{\cdot}12x-3{\cdot}96$	0·83
^{125}I-androstenedione RIA (x) versus ^{3}H-androstenedione RIA (y)	$y=0{\cdot}93x-0{\cdot}18$	0·97
IRMA LPa (x) versus turbidimetric LPa (y)	$y=0{\cdot}334x+78$	0·68
HDL-cholesterol Fara versus (x) HDL-cholesterol Parallel (y)	$y=0{\cdot}87x+0{\cdot}03$	0·99
BCG albumin (x) versus turbidimetric albumin (y)	$y=0{\cdot}97x5{\cdot}65$	0·95

ELISA = Enzyme-linked immunoadsorbent assay; RIA = radio-immunoassay; IRMA = immunoradiometric assay; LPa = lipoprotein a; HDL = high density lipoprotein; BCG = bromo-crestol green.

density lipoprotein (HDL)-cholesterol measured on a Cobas Fara versus an American Monitor Parallel; a bromo-crestol green (BCG) method for albumin versus turbidimetric method.

The methods compared were chosen to represent the range of method comparisons likely to be encountered in the routine laboratory.

The data from each comparison were analysed by linear regression by Deming's method[1] with calculation of the correlation coefficient, and by the use of a difference plot as described by Bland and Altman.[2] In this analysis the difference between the results, expressed as a percentage of the mean of the two measurements for an individual sample, is plotted against the mean. A 10% difference between the results means that one result is approximately 1·1 times the other. Similarly a 66% difference means that one is twice the other and a 100% difference that one is three times the other.

RESULTS

Table 1 shows the Deming regression equations and correlation coefficients obtained for the five method comparisons. Figures 1 to 5 show the regression plots and difference plots.

Regression analysis for the glycated haemoglobin comparison shows that the results obtained from the ELISA method are lower than those from the ion exchange method (Fig. 1a). However, the difference plot shows clearly that the percentage difference between the two results decreases with increasing glycated haemoglobin concentration (Fig. 1b). This concentration dependent difference in results is not apparent from the regression analysis.

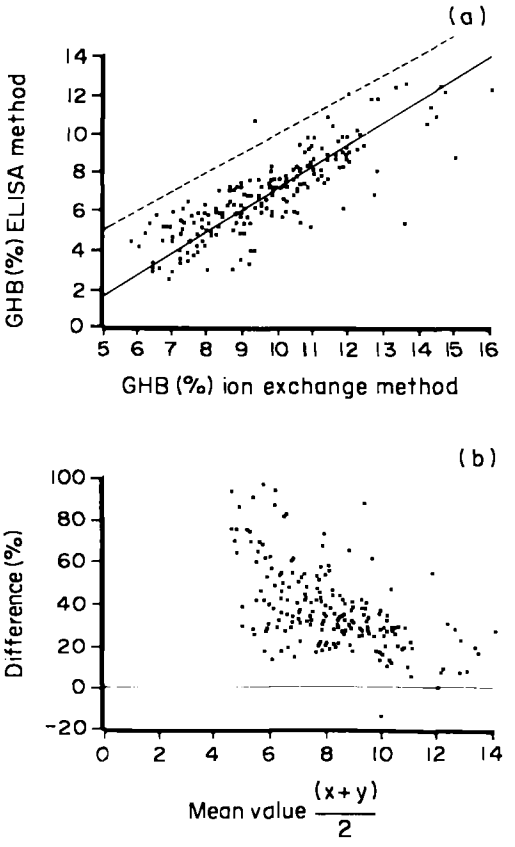

FIGURE 1. *Deming regression analysis (a) and difference plot (b) for glycated haemoglobin comparison GHB (in exchange) (x) versus GHB (ELISA) y. The dotted line represents the line of equivalence. Per cent of difference* = $[(x-y)/(x+y)/2] \times 100$.

In the comparison of two RIA methods for androstenedione using different radioactive labels the regression and correlation analysis suggests acceptable agreement between the methods (Fig. 2a). In contrast, the difference plot shows a large degree of scatter of results obtained although there is no constant or concentration related difference evident between the methods (Fig. 2b).

Conventional analysis of the data from the LPa method comparison shows a wide scatter between

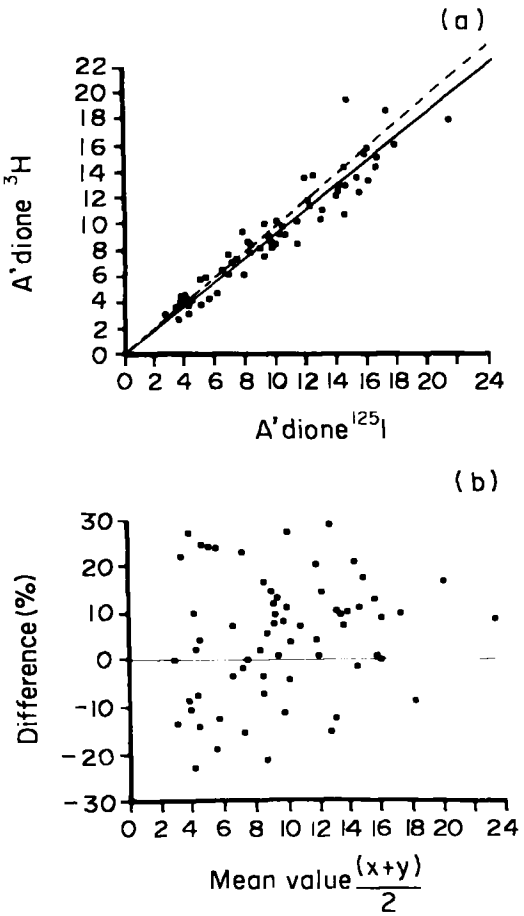

FIGURE 2. *Deming regression analysis (a) and difference plot (b) for androstenedione (A'dione) method comparison ^{125}I androstenedione (x) versus ^{3}H androstenedione (y). The dotted line represents the line of equivalence.*

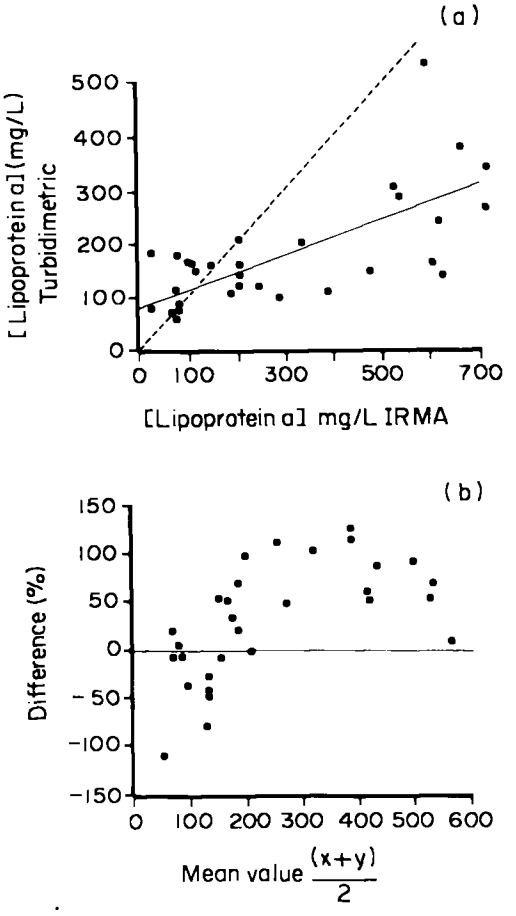

FIGURE 3. *Deming regression analysis (a) and difference plot (b) for lipoprotein a method comparison IRMA lipoprotein a (x) versus turbidimetric lipoprotein a (y). The dotted line represents the line of equivalence.*

methods and suggests poor agreement (Fig. 3a). The difference plot (Fig. 3b), however, shows marked concentration-related differences in results which are not at all evident from the regression plot.

The HDL cholesterol example compares the same chemistry on two instruments. The difference plot confirms that the Parallel method gives lower results than the Fara (Fig. 4b) although again the degree of scatter is greater than that suggested by the regression and correlation analysis (Fig. 4a).

In the comparison of albumin methods the difference plot (Fig. 5b) shows a wider scatter of results at low albumin concentrations than at higher levels. The concentration-related difference between the methods is not apparent from the regression plot (Fig. 5a).

For comparison Figures 6 and 7 show the difference plots of the previous two examples plotted according to the original paper of Bland and Altman.[2] The HDL cholesterol concentrations (Fig. 6) show larger differences with increasing HDL cholesterol values although the percentage differences between the methods are constant (Fig. 4b). The albumin comparison (Fig. 7) shows constant absolute differences between the concentrations whereas the percentage differences increase at lower albumin values (Fig. 5b).

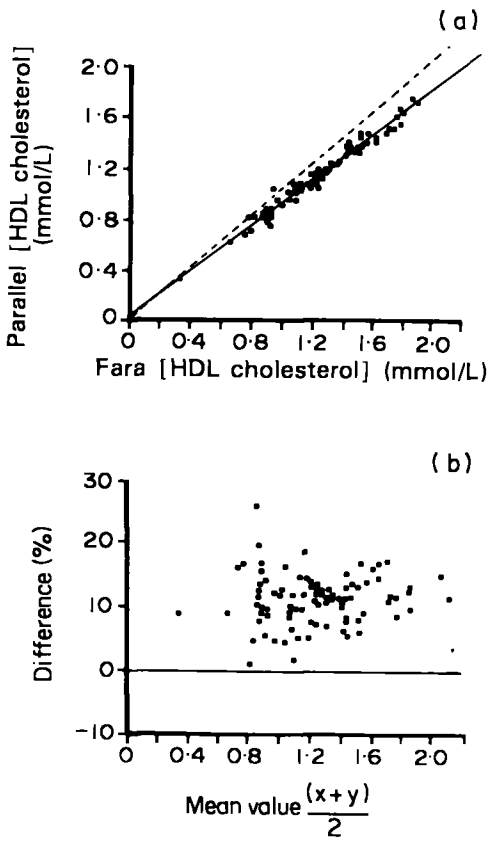

FIGURE 4. *Deming regression analysis (a) and difference plot (b) for high density lipoprotein (HDL) cholesterol method comparison HDL Fara (x) versus HDL parallel (y). The dotted line represents the line of equivalence.*

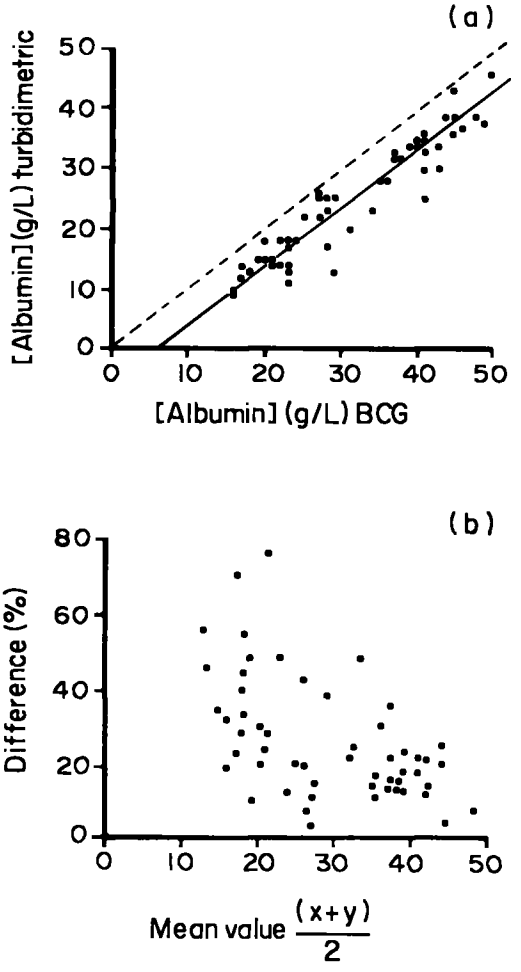

FIGURE 5. *Deming regression analysis (a) and difference plot (b) for albumin method comparison BCG (bromo-crestol green) albumin (x) versus turbidimetric albumin (y). The dotted line represents the line of equivalence.*

DISCUSSION

This study has demonstrated the usefulness of the difference plot in analysing data from a number of different types of method comparison. The results show that plotting the results obtained for one method against another gives some information but it is difficult to assess between-method differences. The difference plot gives clearer information about the degree of agreement between methods and is particularly useful for identifying concentration related bias. We have modified Bland and Altman's method[2] to calculate percent difference rather than absolute difference. This is more relevant to method comparison in clinical chemistry as large ranges of concentration are often measured for one analyte and the same difference in results may have far greater significance at one end of the range than at the other. An acceptable degree of agreement between the two methods must be decided upon before embarking on the comparison exercise. Bland and Altman[2] suggest that the inappropriate use of the correlation coefficient (r) may have arisen because of a misunderstanding of its use as a method of analysis of a scatter diagram, or the use of inappropriate data in statistical textbooks to illustrate the calculation of r. Once the correlation approach to method comparison was published other authors then used it to analyse similar data and so perpetuated the practice. The problem may be exacerbated as some referees for scientific

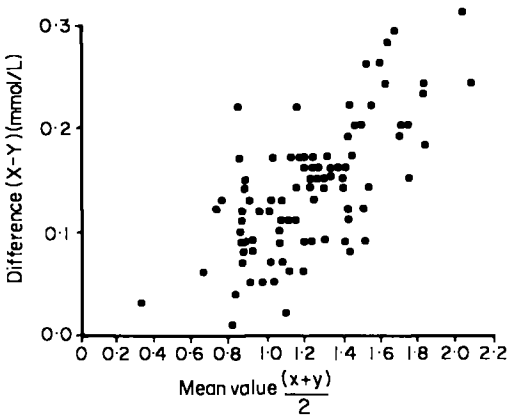

FIGURE 6. *Difference plot as described by Bland and Altman[2] for HDL (high density lipoprotein) cholesterol method comparison HDL Fara (x) versus HDL parallel (y).*

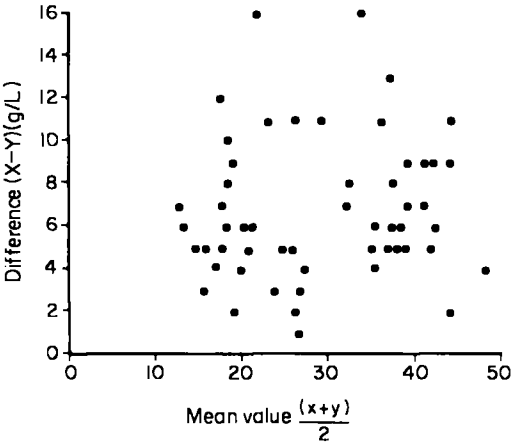

FIGURE 7. *Difference plot as described by Bland and Altman[2] for albumin method comparison bromocrestol green albumin (x) versus turbidimetric albumin (y).*

journals insist on the inclusion of correlation coefficients in method comparison studies.

The difference plot presented in this paper is a purely graphical method of data presentation which clearly demonstrates bias differences between methods. If required, calculation of the mean per cent difference and its standard deviation is very simple to perform and could be useful when deciding if a new method could replace the one currently in use in the laboratory.

Although Bland and Altman's paper was published in 1986 there has been little use of the difference plot in published method comparison studies. Recently, a comparison of assays for follicle stimulating hormone included difference plots.[4] We feel that it is important that method comparison studies should be analysed appropriately and that journals could promote such analysis by returning for revision, manuscripts which use inappropriate statistical techniques. This is likely to be a long process as many journals continue to publish method comparison studies which use least squares regression analysis without even Deming's correction being employed.

REFERENCES

1 Payne RB. Deming's regression analysis in method comparison studies (letter). *Ann Clin Biochem* 1985; **22**: 430

2 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **I**: 307–10

3 Pollock M, Bu'Lock D, Kane J. Use of the difference plot in determining bias in luteinising hormone (LH) immunoassays. *Ann Clin Biochem* 1991; **28**: 517–18

4 Jockenhovel F, Khan SA, Nieschlag E. Varying dose-response characteristics of different immunoassays and an *in vitro* bioassay for FSH are responsible for changing ratios of biologically active to immunologically active FSH. *J Endocrinol* 1990; **127**: 523–32

Accepted for publication 21 January 1992