CLIN. CHEM. 39/3, 424–432 (1993)

Evaluation of Regression Procedures for Methods Comparison Studies

Kristian Linnet

Using simulation, I evaluated five regression procedures that are used for analyzing methods comparison data: ordinary least-squares regression analysis, weighted least-squares regression analysis, the Deming method, a weighted modification of the Deming method, and a rank procedure. I recorded the following performance measures: plus or minus bias of the slope estimate, the root mean squared error of the slope estimate, and correctness of hypothesis testing. I evaluated the unweighted regression procedures by using a simulated comparison of two electrolyte methods; only the Deming method gave unbiased slope estimates. Using the jackknife method, a computer program that estimates the standard error, I showed that hypothesis testing was correct for the Deming method. I used all regression procedures on data from a simulated comparison of two measurement methods with proportional analytical errors dispersed over one decade. Bias of the slope estimates was not a problem for these cases. The weighted least-squares regression analysis and the weighted Deming method were most efficient (lowest root mean squared error); the other procedures required 1.6 to 2.2 times as many observations to attain the same precision for the slope estimate. Hypothesis testing was correct by the weighted Deming method with the jackknife principle for standard error computation; the other methods rejected the null hypothesis 1.4 to 4.4 times too frequently. In conclusion, it is preferable to use an alternative to ordinary least-squares regression analysis for methods comparison studies.

Indexing Terms: *statistics · intermethod comparison*

It is often necessary in clinical chemistry to look for a systematic difference between the measurements of two methods. To do this, a set of samples is analyzed by both methods and, on the basis of a regression analysis, any systematic difference is evaluated. Ordinary least-squares regression analysis, which is readily available in pocket calculators and statistical packages (*1*), is commonly used. This procedure presumes that the measurements for one of the methods are without random error, i.e., the analytical standard deviation is zero, and that the analytical standard deviation for the other method is constant throughout the measurement range. Both assumptions are rarely fulfilled. Therefore, it is reasonable to consider alternatives to ordinary least-squares regression analysis for evaluation of methods comparison data. In weighted least-squares regression analysis (*2*), the analytical standard deviation for the method that is subject to random errors need not be constant. The Deming procedure (*3–10*) realistically assumes that the measurements by both methods are subject to random errors, and a weighted modification of the Deming method (*11*) takes into account nonconstant analytical standard deviations for both methods. For example, it can be assumed that both methods are subject to proportional measurement errors (constant coefficients of variation), which is a very common situation in clinical chemistry. Finally, a regression procedure based on the rank principle, which also takes random errors for both methods into account, may be considered (*12–14*). In this paper, I evaluate the regression procedures mentioned above, using simulations of typical situations in clinical chemistry. I focus on the shortcomings of ordinary least-squares regression analysis, and how these can be overcome by using an alternative procedure.

Department of Clinical Biochemistry KB 76.4.2, Rigshospitalet University Hospital, Tagensvej 20, DK-2200 Copenhagen N, Denmark.

Received March 4, 1992; accepted September 21, 1992.

Comparison of Two Clinical Chemistry Methods by Regression Analysis

Because every method in clinical chemistry is subject to some random measurement error, we must distinguish between the measured value (x_i) and the true or target value (X_i). The latter is the average of all the values that we would obtain if we repeated the measurement of a given sample an indefinite number of times. A particular measured value is likely to deviate from the target value by some small "random" amount (ϵ or δ). Given two clinical chemistry methods, we have for the ith sample

$$x_i = X_i + \epsilon_i$$

$$y_i = Y_i + \delta_i$$

The standard deviation of the dispersion of measured values about the target value is the analytical standard deviation (Figure 1). The analytical standard deviation may be constant, i.e., independent of the target value, but usually it increases with increasing target values. For many compounds, the analytical standard deviation is approximately proportional to the concentration level, corresponding to a constant coefficient of variation (Figure 2*a*) (*15*). In some cases, e.g., hormone assays, the analytical standard deviation becomes constant in the low range, resulting in an increased coefficient of variation in this area (Figure 2*b*) (*16*). One must consider the relation between analytical standard deviation and concentration value when one chooses the best regression analysis for a method comparison study (see below).

From a set of paired measured values $\{(x_i; y_i)\}$ the relation between the target values $\{(X_i; Y_i)\}$ is to be elucidated. The null hypothesis of no difference (H_o)

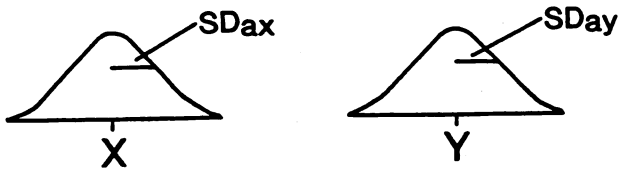

Fig. 1. Distributions of measured values about the target values
The analytical standard deviations SD_{ax} and SD_{ay} are schematically shown

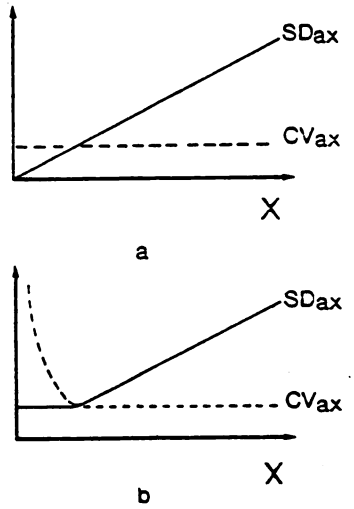

Fig. 2. Examples of relations between analyte concentration (X), analytical standard deviation (SD_{ax}), and coefficient of variation (CV_{ax})

$$H_0: Y_i = \alpha_0 + \beta X_i \quad (\alpha_0; \beta) = (0; 1)$$

is tested against the alternative hypotheses (H_{a1} and H_{a2}):

H_{a1}: A linear relationship exists between X and Y.

$$Y_i = \alpha_0 + \beta X_i \quad (\alpha_0; \beta) \neq (0; 1)$$

H_{a2}: Y is a nonlinear function of X or totally independent of X.

The hypotheses are evaluated by some type of regression analysis, supported by graphical plots. If a linear relationship seems plausible from graphical plots or statistical tests of linearity, we must decide between H_0 and H_{a1}. By determining whether the slope deviates significantly from 1 (proportional systematic error) and whether α_0 is different from 0 (constant systematic error), we decide between these two hypotheses.

Each regression method requires certain assumptions concerning data distribution and analytical errors. If these assumptions are violated, the statistical analysis may be incorrect or nonoptimal. The regression method that requires the simplest computations, ordinary least-squares regression analysis, relies on the most rigid assumptions, whereas the more complicated methods operate with assumptions that are more likely to be fulfilled in reality (*Appendix*).

Ordinary least-squares regression analysis requires that one of the methods (corresponding to x) is without random measurement error, that the measurement errors of the other method (y) have a gaussian distribution at a given level, and that the standard deviation is constant throughout the measurement range (Figure 3) (*1*). The computation of an ordinary least-squares regression analysis is based on minimization of the squared deviations from the line in the vertical direction. It can be shown that this procedure is optimal under the stated conditions.

Weighted least-squares regression analysis (*Appendix*) (*2*) allows for a nonconstant standard deviation for the y method, but it is still presumed that the x method is without random measurement error. Weights are introduced that are inversely proportional to the squared analytical standard deviation of y measurements at a given concentration. For example, the procedure can be applied to the case with a proportional analytical standard deviation for y (constant coefficient of variation). The sum of weighted squared deviations from the line in the vertical direction is minimized. In the situation with proportional analytical standard deviation, the weights reduce the influence of observations at high concentrations. This is reasonable because the analytical standard deviation, and thus the uncertainty, increases with the concentration. The result is a more precise estimation of the regression line.

The Deming method (*Appendix*) (*3–10*) allows measurement errors for both methods, requiring that the ratio between the analytical standard deviations is known. The Deming method is primarily used when analytical standard deviations are constant. The sum of squared deviations from the line at an angle determined by the ratio between the standard deviations is minimized (Figure 4). Various procedures for estimation of standard errors of slope and intercept have been suggested (*5, 10*). If a computerized method such as the jackknife method is used, gaussian error distributions need not be assumed; i.e., the procedure becomes nonparametric (*11*).

A weighted modification of the Deming method takes into account nonconstant measurement errors for both methods (*Appendix*) (*11*). It still has to be presumed that the ratio between the analytical standard deviations is constant. This is true for the common situation with proportional analytical standard deviations (constant coefficients of variation) for both methods [and strictly, a regression line passing through (0; 0)]. The sum of

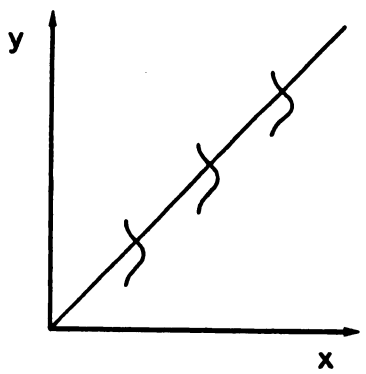

Fig. 3. The model assumed in ordinary least-squares regression analysis: no error in x and constant gaussian error distribution for y
In this example, the regression line is coinciding with the line of identity

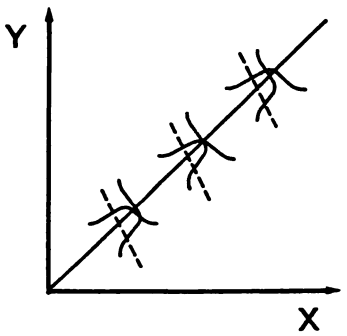

Fig. 4. The model assumed in the Deming regression procedure
The regression line is estimated by minimization of squared distances to the line at an angle determined by the ratio of the analytical standard deviations for the two methods

weighted squared deviations at an angle to the line is minimized.

Finally, a nonparametric rank method developed by Passing and Bablok is evaluated (*Appendix*) (*12–14*). The slope of the regression line is calculated as the median of all possible slopes. This estimation principle makes the method robust towards outliers, which is the main advantage of the method. Measurement errors for both chemistry methods are partially taken into account. Passing and Bablok claim that this method can be used when errors are proportional.

Performance of the Regression Methods

To compare the regression methods, I used each regression analysis in two simulated methods comparison studies; one compared two methods for determination of an electrolyte, and the second, two methods for determination of a metabolite.

Electrolyte Study

Serum concentrations of electrolytes are tightly regulated, which means that the range of concentrations is very small in healthy individuals, and deviations associated with disease are small to moderate. For example, the ratio between the upper and lower reference limits (range ratio) for serum sodium is 146 mmol/L:136 mmol/L = 1.07 (*17*). In a study that included samples from diseased subjects, Cornbleet and Gochman (*5*) found a range ratio of 1.15 for serum sodium. For my model I suppose a gaussian distribution of target values with a mean of 135.5 mmol/L and a standard deviation of 3.8 mmol/L. If the minimum and maximum values are 2.5 standard deviations below and above the mean, respectively, the range ratio is 1.15 (145 mmol/L:126 mmol/L). The analytical coefficients of variation are supposed to be 0.01 and 0.015 for methods 1 (x) and 2 (y), respectively. For this case, it makes little difference whether we suppose constant or proportional analytical errors; I selected constant analytical errors with the above-mentioned coefficients of variation at the mean value. Consequently, I did not need to use the weighted least-squares regression analysis or the weighted Deming method. To test the null hypothesis, that the target values for the two methods are identical, I selected a sample size of 50 duplicate measurements by each method. Using duplicate measurements, one can estimate the analytical standard deviations and compute their ratio. This ratio is then used for computing the slope by the Deming method. In the simulation process, pseudorandom gaussian numbers are generated according to the specified model and regression equations are calculated. By repeating the procedure many times (=nrun), e.g., 2000–5000 times, one obtains various performance measures for each regression method. These are listed in Table 1. The average slope ($\bar{b}$) indicates whether the slope estimate is biased or not. The root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\Sigma(b-1)^2/\text{nrun}} = \sqrt{\text{BIAS}^2 + \text{SE}^2}$$

is an estimate of the total error of the slope and includes both the random error, i.e., the standard deviation of the dispersion of b around $\bar{b}$, and the systematic error or bias. The *real* standard error of the slope [SE($\bar{b}$ − 1)] is the standard error observed in the simulation study, i.e., the standard deviation for the distribution of b in the simulation runs. The average *estimated* standard error of the slope, on the other hand, refers to the standard error derived by statistical analysis. In each simulation run, an estimated standard error is computed so as to carry out a test of the null hypothesis. A prerequisite for a correct test is that the estimated standard error, on average, agrees with the real one. In the rank method, hypothesis testing is carried out in another way, and no average estimated standard error is given. Lastly, I evaluated the performance of hypothesis testing by comparing the observed and expected number of rejections of the null hypothesis. Under the null hypothesis, we expect 50 rejections out of 1000 trials, given a nominal type I error of 0.05. Any deviation from this number may be expressed as a factor. For example, if 100 rejections are recorded, the factor is

$$f = 100/50 = 2$$

The ideal hypothesis test factor is, of course, 1. Figure 5 shows the average regression lines obtained by the three procedures for this example. The scatter points of one simulation run are also shown.

Ordinary least-squares regression analysis and the

Table 1. Performance Characteristics of Three Regression Procedures for a Simulated Comparison of Two Electrolyte Methods

	Least-Squares reg. anal.[a]	Deming reg. anal.	Rank reg. anal.
Average slope ($\bar{b}$)	0.940[b]	1.001	1.031[b]
Root mean squared error	0.088	0.069	0.072
Real SE(b)	0.064	0.069	0.065
Average estimated SE(b)	0.058	0.070	—
Hypothesis test factor (f)	4.1	1.0	3.8

The true slope β is equal to 1. Sample size: 50 duplicate measurements by each method. Based on 2000–5000 simulation runs.
[a] Ordinary least-squares regression analysis.
[b] Significantly different from 1 ($P < 0.001$).

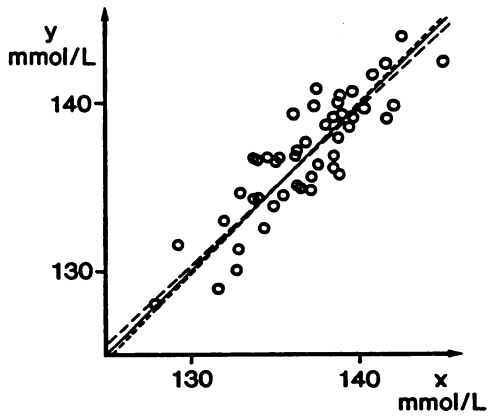

Fig. 5. The simulated electrolyte case

Fifty scatter points (means of duplicate measurements by each method) are shown for one simulation run. The regression lines are averages of 5000 simulation runs. (——) the Deming procedure; (— —) ordinary least-squares regression analysis; (- - -) the rank procedure

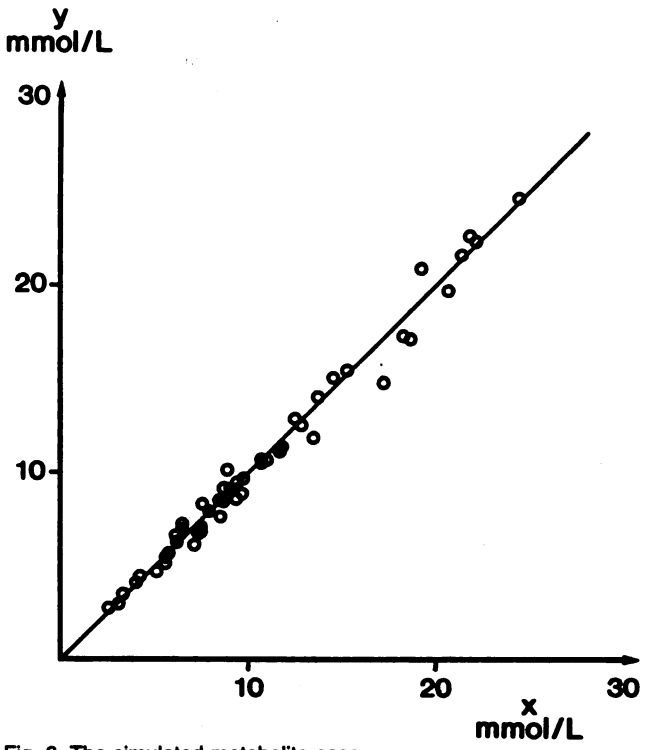

Fig. 6. The simulated metabolite case

Fifty scatter points are shown for one simulation run. The *solid line* is the diagonal, which corresponds to the nearly coinciding average regression lines of each of the five methods

rank method yield biased slope estimates, $\bar{b} = 0.94$ and 1.03, respectively, whereas the Deming method is without bias. The negative bias of the ordinary least-squares method is simply the squared ratio of the coefficients of variation of x errors and target value distributions ($0.01^2 \cdot 0.5/0.028^2 = 0.06$; the factor 0.5 enters because we have duplicate measurements). This example and other similar simulations show that the rank method does not provide unbiased slope estimates if the analytical standard deviations for methods 1 and 2 are different. This conclusion is also apparent from the study by Bablock and Passing (*13*). For both ordinary least-squares regression analysis and the rank procedure, I have observed that a biased slope estimate is accompanied by a biased intercept.

The root mean squared error is smallest for the unbiased Deming method. The jackknife principle of estimating the standard error results in correct hypothesis testing [hypothesis test factor (f) = 1.0]. In contrast, the null hypothesis is rejected too frequently by the other two procedures, $f = 3.8$ and 4.1, corresponding to rejection frequencies of ~20%. This is caused partly by bias and partly by incorrect estimation of the standard error.

Metabolite Study

Metabolites include commonly measured substances such as glucose, urea, and urate. The serum concentrations of these substances vary more than those of electrolytes, resulting in larger range ratios. I compared two methods for glucose determination and supposed that the glucose concentrations extended from 2.5 to 25 mmol/L, corresponding to a range ratio of 10. I assumed that a majority of the samples came from healthy subjects or patients with moderately elevated glucose concentrations; thus three-quarters of the observations were located on the lower half and one-quarter on the upper half of the interval (Figure 6). The measurement errors were proportional to the concentrations, with coefficients of variation of $CV_1 = 0.05$ (x) and $CV_2 = 0.075$ (y). The sample comprised 50 duplicate observations. The simulations show that ordinary and weighted least-squares regression analyses yield slope estimates with a negligible bias ($\bar{b} = 0.996$ and 0.993) (Table 2). The analytical coefficient of variation of method 1 is small when compared with the coefficient of variation of the target value distribution. However, ordinary least-squares regression analysis is not appropriate because the hypothesis testing is inaccurate; an actual type I error exceeds the nominal one by a factor of $f = 4.4$. This is caused by an underestimation of the standard error of the slope by ~40%, as evident from the simulation study. The weighted procedure performs better, with an error factor of $f = 1.7$. In addition, the slope, as calculated by the weighted procedure, has a smaller root mean squared error than it would if it were calculated by the ordinary least-squares regression analysis.

The Deming and the weighted Deming methods give unbiased slope estimates, as expected, and the rank method has only a negligible bias ($\bar{b} = 1.002$). The weighted Deming method performs hypothesis testing correctly, whereas the unweighted procedure and the rank method have moderately increased type I errors (f = 1.4 and 1.6, respectively). The average estimated standard error of the unweighted Deming procedure is correct, but the distribution of the test statistic does not conform exactly to the t-distribution, affecting the hypothesis testing. The weighted Deming method is more efficient than the unweighted modification, as indicated by the root mean squared error, and the rank method has an intermediate efficiency.

For comparison, Table 2 shows the number of obser-

Table 2. Simulated Comparison of Two Glucose Methods

	Ordinary least-squares reg. anal.	Weighted least-squares reg. anal.	Deming reg. anal.	Weighted Deming reg. anal.	Rank reg. anal.
Average slope ($\bar{b}$)	0.996[a]	0.993[a]	1.001	1.000	1.002[a]
Root mean squared error	0.028	0.019	0.028	0.018	0.023
Real SE(b)	0.028	0.018	0.028	0.018	0.022
Average estimated SE(b)	0.018	0.016	0.028	0.018	—
Hypothesis test factor (f)	4.4	1.7	1.4	1.0	1.6
Sample sizes yielding the same precision of slope estimate, n	210	100	220	100	160

$\beta = 1$, $CV_1 = 0.05$, $CV_2 = 0.075$, 50 duplicate measurements, and 2000–5000 simulation runs.
[a] Significantly different from 1 ($P < 0.001$).

vations required by each of the methods to obtain slope estimates that have the same precision. Weighted least-squares regression analysis and the weighted Deming method each require a sample size of 100, with the other methods demanding up to 2.2 times as many observations.

Performance of Regression Methods in the Presence of Outliers and Nongaussian Error Distributions

In the previous section, I considered ideal situations, i.e., cases with gaussian error distributions and no outliers. Because skew error distributions and outliers may occur in real cases, it is relevant to evaluate the regression methods under these conditions. I compared the weighted Deming procedure, which is based on the least-squares computation principle, with the rank method. The least-squares computation procedure is sensitive towards outliers, in contrast to the rank principle, where a median value serves as a slope estimate.

I investigated the impact of outliers by using the metabolite model and introducing an outlier randomly in 1 of 50 observations. Here, an outlier is considered as an observation for which the target value deviates from the value expected under the given hypothesis. Under the null hypothesis, we expect $x_i = y_i$, but for the outlier the relation is $x_i = y_i + d$, where d is some constant. The target value of the outlier is shifted by one to eight analytical standard deviations at the given value from the identity line. The slope is estimated in each simulation run, and the root mean squared error is recorded for 1000–5000 repetitions. For the weighted Deming method, I calculated the root mean squared error both with and without application of a rejection rule for outliers. For each pair of observations (x_i, y_i), I calculated the standardized deviation from the estimated regression line, and if this deviation exceeded three standard deviations of the distribution of deviations, I rejected the observation (*11*). If a rejection rule is not used, the root mean squared error for the slope estimate of the weighted Deming method increases steadily as outlier deviation increases, whereas the root mean squared error of the rank method is almost unaffected (Figure 7). When the rejection rule is applied, however, the root mean squared error of the weighted Deming method increases only moderately up to a maximum and then declines towards the starting value. For all outlier positions, the root mean squared error of the weighted Deming procedure is smaller than that of the rank method. No rejection rule for outliers is specified for the rank method (*12*). Actually, identification of outliers is important, because they may indicate the presence of special matrix effects or interferents.

Skew measurement error distributions may be modeled by log-gaussian distributions (Figure 8). The coef-

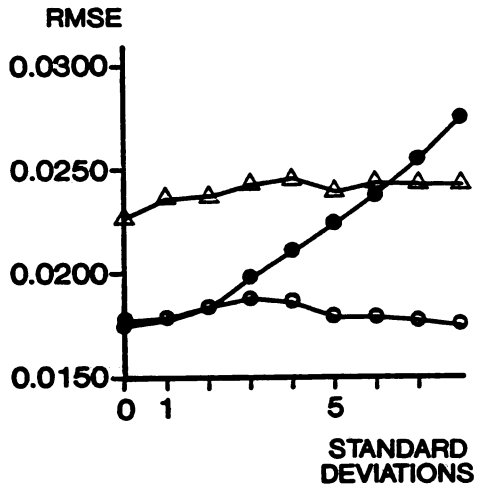

Fig. 7. The root mean squared error (*RMSE*) of the slope estimate as a function of outlier deviation expressed in standard deviations
●, the weighted Deming procedure; ○, the weighted Deming procedure with outlier rejection; △, the rank procedure

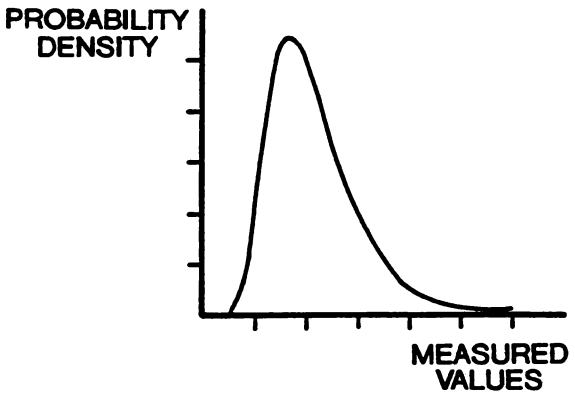

Fig. 8. The shape of the skew measurement error distributions
The coefficient of skewness is 1.3

ficient of skewness is 1.3 for this example. Table 3 gives average slopes, root mean squared errors, standard errors, and hypothesis test factors for the metabolite case with skew error distributions. Bias was not important for either method. The root mean squared error of the weighted Deming method is smaller than that of the rank method, and $f = 1$ only for the weighted Deming method.

Discussion

Ordinary least-squares regression analysis is the simplest regression method and by far the most widely used one for methods comparison studies. However, modern computer technology has made more complicated methods generally accessible because computations can be automated. Therefore, it appears reasonable to evaluate whether more complicated regression methods offer advantages over ordinary least-squares regression analysis. Earlier, we noted that apart from the electrolyte case, ordinary least-squares regression analysis gives a slope estimate with only a modest bias, even though both analytical methods are subject to random measurement errors. A more serious problem is that hypothesis testing is unreliable when errors are proportional. For the metabolite case studied here, the standard error of the slope was on average underestimated by 40%, and thus the null hypothesis was rejected too frequently and confidence intervals were underestimated. Also, the efficiency of ordinary least-squares regression analysis may be low for large range ratios when there are proportional errors.

Weighted least-squares regression analysis is efficient for the metabolite case, and the bias is small. Hypothesis testing is better than for ordinary least-squares regression analysis, but the type I error is still somewhat high because the random measurement errors of x are neglected. Simulations show that if x is without error, hypothesis testing becomes correct.

The unweighted Deming method yields an unbiased slope estimate in both the metabolite and electrolyte method comparisons and, with use of the jackknife principle, hypothesis testing becomes correct for the electrolyte case with constant analytical standard deviations. Cornbleet and Gochman (*5*) suggest estimating the standard error of the slope by using the same formula as for ordinary least-squares regression analysis. Using this simple principle, I found that hypothesis testing was not as good as that achieved by using the jackknife method to derive standard errors ($f = 1.5$ for the electrolyte case and 3.8 for the metabolite case, compared with $f = 1.0$ and 1.4, respectively, for the jackknife method).

Table 3. Performance of the Weighted Deming and the Rank Regression Procedures on Skew Error Distributions

	Weighted Deming reg. anal.	Rank reg. anal.
Average slope ($\bar{b}$)	1.000	0.997[a]
Root mean squared error	0.017	0.022
Real SE(b)	0.017	0.022
Average estimated SE(b)	0.017	—
Hypothesis test factor (f)	1.0	2.2

[a] Significantly different from ($P < 0.001$).

Given proportional measurement errors for both methods, the weighted Deming method is more efficient than the usual unweighted approach. The larger the range ratio, the greater the advantage. For a range ratio of 10, the weighted approach requires less than half the sample size of the unweighted one to provide a slope estimate with the same precision. Achieving high precision for the slope estimate is as important as eliminating bias. In a particular methods comparison study, the slope estimate is likely to be within plus or minus two standard errors from the true, unknown slope value. Thus, if an unbiased slope estimate has a low precision, perhaps caused by the use of nonoptimal regression technique, the result may still deviate considerably from the true value.

For both modifications of the Deming method, it is presumed that the ratio (λ) between the analytical standard deviations is constant. This condition is satisfied if both standard deviations are constant, i.e., independent of the concentration, or if both methods have a constant coefficient of variation and a proportional relationship exists between the methods ($\alpha_0 = 0$). Although in reality λ is often not constant, it is frequently approximated. If the ratio is not quite constant, a small bias may arise. Studies have shown that the Deming model is robust towards misspecification of λ (*18*). The bias is, at most, of the same order of magnitude as that of ordinary or weighted least-squares regression analysis.

The rank method of Passing and Bablok (*12–14*) gives an unbiased slope estimate under the null hypothesis only if the analytical standard deviations for the two methods to be compared are of equal size. Hypothesis testing is not correct for the rank method, and because the rank method is, in principle, an unweighted method, it is not optimal for the case with proportional errors. The advantage of the method is its resistance towards outliers, but the method needs further development to eliminate the inadequacies revealed by the simulations performed here.

In a time when there is an increasing focus on accuracy in analytical methods, the choice of an optimal regression technique for methods comparison studies is important. For example, an inaccuracy of <3% has been recommended for cholesterol measurement methods in clinical laboratories (*19*). If an inappropriate or nonoptimal regression procedure is used to analyze the data for a cholesterol methods comparison study, a bias of perhaps 1% or an unnecessarily large standard error for the slope estimate may soon become critical.

Appendix

Computation Principles for the Regression Methods

Ordinary least-squares regression analysis (*1*). On the basis of k paired observations (x_i; y_i), the regression line is estimated from sums of squared deviations and cross-products:

$$u = \sum (x_i - \bar{x})^2; \quad q = \sum (y_i - \bar{y})^2;$$

$$p = \sum (x_i - \bar{x})(y_i - \bar{y})$$

$$b = p/u; \quad a = \bar{y}$$

$$\hat{Y}_i = a + b(x_i - \bar{x}) = a_0 + bx_i; \quad (a_0 = a - b\bar{x})$$

The caret of $\hat{Y}_i$ denotes that we obtain an estimate of the "true" value for a given x_i.

The standard deviation for the distribution about the line is as follows:

$$SDy \cdot x = \sqrt{\sum (y_i - \hat{Y}_i)^2/(k-2)}$$

The standard errors of a and b are as follows:

$$SE(a) = SDy \cdot x/\sqrt{k}$$

$$SE(b) = SDy \cdot x/\sqrt{u}$$

The null hypothesis of identity is tested by two independent *t*-tests:

$$t = (a - \bar{x})/SE(a); \quad t = (b - 1)/SE(b)$$

Weighted least-squares regression analysis (2). In principle, weights (w_i) are introduced in computations of sums of squared deviations and cross-products. The weights are inversely proportional to the squared analytical standard deviation of *y* at a given value. It is usual to assume that the analytical standard deviation is a function [$h(\cdot)$] of the "true" concentration (= x), i.e.,

$$SD_{ay} = ch(x_i)$$

where *c* is a proportionality factor. Then

$$w_i = 1/[h(x_i)]^2$$

If it is plausible from experimental data to assume that the coefficient of variation is constant, we have $SD_{ay} = c\,x_i$ (supposing that $\alpha_0 = 0$) and $w_i = 1/x_i^2$. For some analytes (as shown in Figure 2*b*), the analytical standard deviation becomes constant in the low region, below a limit L. We then have that $SD_{ay} = c\,L$ for $x_i < L$ and $SD_{ay} = c\,x_i$ for $x_i > L$. Correspondingly, $w_i = 1/L^2$ for $x_i \leq L$ and $w_i = 1/x_i^2$ for $x_i > L$.

The regression line is estimated as follows:
Weighted averages are computed:

$$\bar{x}_w = \sum w_i x_i / \sum w_i$$

$$\bar{y}_w = \sum w_i y_i / \sum w_i$$

Weighted squared deviations and cross-products are summed:

$$u_w = \sum w_i (x_i - \bar{x}_w)^2; \; q_w = \sum w_i (y_i - \bar{y}_w)^2$$

$$p_w = \sum w_i (x_i - \bar{x}_w)(y_i - \bar{y}_w)$$

Slope and location parameters are determined:

$$b = p_w/u_w; \quad a = \bar{y}_w$$

The line is estimated:

$$\hat{Y}_i = a + b(x_i - \bar{x}_w) = a_0 + bx_i \; (a_0 = a - b\bar{x}_w)$$

The proportionality factor (*c*) for the standard deviation about the line is computed:

$$c = \sqrt{\sum (y_i - \hat{Y}_i)^2 w_i/(k-2)}$$

The standard errors of a and b are determined:

$$SE(a) = c/\sqrt{k}; \quad SE(b) = c/\sqrt{u_w}$$

The null hypothesis of identity is tested by two independent *t*-tests:

$$t = (a - \bar{x}_w)/SE(a); \quad t = (b - 1)/SE(b)$$

The Deming regression analysis (also called: errors-in-variables regression analysis, a structural or functional relationship model) (*3–11*). In addition to the sums of squared deviations and cross-products, the ratio between the squared analytical standard deviations for the methods is needed:

$$\hat{\lambda} = SD_{ax}^2/SD_{ay}^2$$

Assuming constant standard deviations and given duplicate measurements (x_{1i} and x_{2i}; y_{1i} and y_{2i}), the analytical standard deviations are estimated as

$$SD_{ax}^2 = (1/2k) \sum (x_{1i} - x_{2i})^2$$

$$SD_{ay}^2 = (1/2k) \sum (y_{1i} - y_{2i})^2$$

The regression line is estimated as follows:

$$b = \left[(\hat{\lambda}q - u) + \sqrt{(u - \hat{\lambda}q)^2 + 4\hat{\lambda}p^2} \right] / 2\hat{\lambda}p$$

$$a = \bar{y}$$

$$\hat{Y}_i = a + b(\hat{X}_i - \bar{x}) = a_0 + b\hat{X}_i; \quad (a_0 = a - b\bar{x})$$

Rather complicated methods for computation of standard errors are given in the theoretical statistical literature (*10*). I used the computerized jackknife method (*11*), which has the advantage of being distribution-free.

Weighted Deming regression analysis. In the weighted Deming regression analysis, weights (w_i) are introduced

in computations of sums of squared deviations and cross-products, and the slope is calculated as described in the previous section (*11*). The weights are inversely proportional to the squared analytical standard deviations at a given value. Supposing that the analytical standard deviations are functions of the true concentration, we have

$$SD_{ax} = f_x h_x(X); \quad SD_{ay} = f_y h_y(Y)$$

To use a weighted modification of the Deming method, we must assume that the ratio SD_{ax}^2/SD_{ay}^2 is a constant ($= \lambda$). Given a proportional relationship between analytical standard deviations and concentrations, perhaps truncated at some lower limit L, and $\alpha_o = 0$, this condition holds true. Keeping in mind the symmetry of this model, we may then express the weights as

$$w_i = 1/[(X_i + Y_i)/2]^2$$

Because X_i and Y_i are unknown true values, we have to use estimates, which can be obtained by an iterative procedure, described in detail in reference *11*:

$$\hat{w}_i = 1/[(\hat{X}_i + \hat{Y}_i)/2]^2$$

Actually, it is possible that the most precise estimate may be the inverse of a weighted average of the estimated true values, i.e., $1/[(\hat{X}_i + \lambda\hat{Y}_i)/(1 + \hat{\lambda})]^2$. In the simulations performed here, only the simple average has been used.

The regression line is estimated as follows:
Weighted averages are calculated:

$$\overline{x_w} = \sum \hat{w}_i x_i / \sum \hat{w}_i; \quad \overline{y_w} = \sum \hat{w}_i y_i / \sum \hat{w}_i$$

Weighted sums of squares and cross-products are computed:

$$u_w = \sum \hat{w}_i (x_i - \overline{x_w})^2; \quad q_w = \sum \hat{w}_i (y_i - \overline{y_w})^2;$$

$$p_w = \sum \hat{w}_i (x_i - \overline{x_w})(y_i - \overline{y_w})$$

The proportionality factors f_x and f_y are estimated from duplicate measurements:

$$\hat{f}_x^2 = (1/2k) \sum (x_{1i} - x_{2i})^2/[(\overline{x_i} + \overline{y_i})/2]^2$$

$$\hat{f}_y^2 = (1/2k) \sum (y_{1i} - y_{2i})^2/[(\overline{x_i} + \overline{y_i})/2]^2$$

$\bar{x}_i$ and $\bar{y}_i$ as the means of the duplicate measurements. We obtain the following expression:

$$\hat{\lambda} = \hat{f}_x^2/\hat{f}_y^2$$

The slope estimate is calculated:

$$b = [(\lambda q_w - u_w) + \sqrt{[(u_w - \lambda q_w)^2 + 4\lambda p_w^2}]/2\lambda p_w$$

The intercept is determined:

$$a = \overline{y_w}$$

The usual Deming method is the special case where all weights are equal to 1. Standard errors are obtained by the jackknife method.

The weighted Deming regression analysis is evaluated in reference *11* for the null hypothesis situation. Supplementary evaluations with $\beta = 0.8$ and 1.2 (slopes are usually within this range in methods comparison studies) have confirmed that the procedure also is valid under alternative hypotheses.

A plot of standardized distances from the observed points to the line against the average value $[(\hat{X}_i + \hat{Y}_i)/2]$ is useful for evaluation of linearity and outliers. A practical rejection rule for outliers is three standard deviations of the distribution of standardized distances. A BASIC program that performs the weighted Deming regression analysis is available from the author.

Rank regression analysis (12–14). In the rank regression analysis, in principle, slopes between all possible sets of two points are calculated and ranked. The median expresses a slope estimate. Measurement errors for both methods are taken into account by a symmetry argument, which results in calculation of a shifted median that is the final slope estimate. Confidence intervals and hypothesis testing are established by using traditional rank techniques.

References

1. Snedecor GW, Cochran WG, eds. Statistical methods, 6th ed. Ames, IA: Iowa State University Press, 1967.
2. Hald A. Statistical theory with engineering applications. New York: Wiley, 1952.
3. Wakkers PJM, Hellendoorn HBA, Op De Weegh GJ, Heerspink W. Applications of statistics in clinical chemistry. A critical evaluation of regression lines. Clin Chim Acta 1975;64:173–84.
4. Lloyd PH. A scheme for the evaluation of diagnostic kits. Ann Clin Biochem 1978;15:136–45.
5. Cornbleet PJ, Gochman N. Incorrect least-squares regression coefficients in method-comparison analysis. Clin Chem 1979;25: 432–8.
6. 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.
7. Parvin CA. A direct comparison of two slope-estimation techniques used in method-comparison studies. Clin Chem 1984;30: 751–4.
8. Strike PW, Michaeloudis A, Green AJ. Standardizing clinical laboratory data for the development of transferable computer-based diagnostic programs. Clin Chem 1986;32:22–9.
9. Mandel J. The statistical analysis of experimental data. New York: Wiley, 1964.
10. Kendall MG, Stuart A. The advanced theory of statistics, Vol. 2. London: Charles Griffin, 1973:393–7.
11. Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. Stat Med 1990;9:1463–73.
12. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. J Clin Chem Clin Biochem 1983;21:709–20.
13. Passing H, Bablok W. Comparison of several regression procedures for method comparison studies and determination of sample sizes. J Clin Chem Clin Biochem 1984;22:431–45.
14. Bablok W, Passing H, Bender R, Schneider B. A general

regression procedure for method transformation. J Clin Chem Clin Biochem 1988;26:783–90.
15. Ross JW, Fraser MD. The effect of analyte and analyte concentration upon precision estimates in clinical chemistry. Am J Clin Pathol 1976;66:193–205.
16. Sadler WA, Smith MH, Legge HM. A method for direct estimation of imprecision profiles, with reference to immunoassay data. Clin Chem 1988;34:1058–61.
17. Tietz NW, ed. Textbook of clinical chemistry. Philadelphia: Saunders, 1986.
18. Beech DG. Some notes on the precision of the gradient of an estimated straight line. Appl Stat 1961;10:14–31.
19. Current status of blood cholesterol measurements in clinical laboratories in the United States: a report from the Laboratory Standardization Panel of the National Cholesterol Education Program. Clin Chem 1988;34:193–201.