Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data?

Dietmar Stöckl, Katy Dewitte, and Linda M. Thienpont*

We compared the application of ordinary linear regression, Deming regression, standardized principal component analysis, and Passing–Bablok regression to real-life method comparison studies to investigate whether the statistical model of regression or the analytical input data have more influence on the validity of the regression estimates. We took measurements of serum potassium as an example for comparisons that cover a narrow data range and measurements of serum estradiol-17β as an example for comparisons that cover a wide data range. We demonstrate that, in practice, it is not the statistical model but the quality of the analytical input data that is crucial for interpretation of method comparison studies. We show the usefulness of ordinary linear regression, in particular, because it gives a better estimate of the standard deviation of the residuals than the other procedures. The latter is important for distinguishing whether the observed spread across the regression line is caused by the analytical imprecision alone or whether sample-related effects also contribute. We further demonstrate the usefulness of linear correlation analysis as a first screening test for the validity of linear regression data. When ordinary linear regression (in combination with correlation analysis) gives poor estimates, we recommend investigating the analytical reason for the poor performance instead of assuming that other linear regression procedures add substantial value to the interpretation of the study. This investigation should address whether (a) the x and y data are linearly related; (b) the total analytical imprecision \( s_{\text{a,tot}} \) is responsible for the poor correlation; (c) sample-related effects are present (standard deviation of the residuals \( \gg s_{\text{a,tot}} \)); (d) the samples are adequately distributed over the investigated range; and (e) the number of samples used for the comparison is adequate.

Linear regression, when applied to method comparison data, provides useful information about proportional, constant, and random error via, respectively, the slope, intercept, and standard deviation of the residuals \( S_{y|x} \) (1). Linear regression data may be used for calibrating a new method against an established one or validating the utility of a method in relation to analytical quality specifications (1, 2). The classical linear regression model is referred to here as ordinary linear regression (OLR).\(^1\) The regression line is calculated by minimizing the squared residuals in the y direction (“least squares”). OLR assumes an error-free x variable and a constant analytical imprecision \( s_a \) of the y variable (also called “homoscedastic” variance), both of which are seldom met in practice. To compensate for variable \( s_a \), weighted forms of OLR have been introduced; however, they have not received much attention in practice. For the case in which x is not error-free, alternatives to OLR have been proposed that consider variances in both variables. One of the most important of those alternatives is the so-called Deming regression (DR) (3). In its most simple form, DR assumes equal x and y variances and minimizes the distance of the data points orthogonal to the regression line (3). When the variances in x and y differ considerably (but their ratio is constant), the distance of the data point is minimized at an angle to the regression line that is dependent on the ratio of the two variances (4). Note that the regression procedures that consider variances in x and y are also known under the name principal component analysis (4). When a constant ratio of the variances is assumed, the technique is called standardized principal component analysis (SPCA). In

\(^1\)Nonstandard abbreviations: OLR, ordinary linear regression; \( s_a \), analytical imprecision; DR, Deming regression; SPCA, standardized principal component analysis; PBR, Passing–Bablok regression; \( s_{\text{a,tot}} \), total analytical imprecision; and IQC, internal quality control.
addition to the parametric alternatives, nonparametric versions of linear regression were established in method comparison studies, of which the Passing–Bablok regression (PBR) probably is the most widely used variant (5–7). PBR is based on the rank principle and assumes a constant ratio of the variances. The procedure is claimed less sensitive to outliers than its parametric counterparts.

Note that, regardless of the model used, the uncertainty in the regression estimates increases when (a) the magnitude of $S_{y|x}$ relative to the data range is high; (b) the data are not adequately distributed over the investigated range; (c) the number of data points is low; or (d) the relation between the data is not linear. In the last case, the model itself is wrong and should be replaced by a nonlinear one. (Nonlinear procedures are sometimes referred to as second-order linear regression.) For the above reasons, it is recommended that method comparison data always be accompanied by graphical presentations (8, 9).

Because of the availability of different linear regression procedures, the question of which is the most appropriate for method comparison studies has been investigated by several groups, e.g., by Cornbleet and Gochman (10), Wakkers et al. (11), Linnet (12, 13), Payne (14), and Lawton et al. (15). From these investigations, which focused mainly on the error in slope estimation and the consequent statistical unreliability of hypothesis testing, many prefer DR over OLR. Others make use of OLR dependent on the value calculated for the product–moment correlation coefficient ($r$). For example, Wakkers et al. (11), Westgard and Hunt (1), and the EP9 protocol from the NCCLS (8) restrict the use of OLR to those cases where $r \geq 0.99$ (11) or $\geq 0.975$ (1, 8). With respect to PBR, its relevance still deserves better demonstration because some put its use as an alternative to OLR into perspective by revealing inadequacies in simulation tests (12), whereas others advocate it over DR (14). In the most extreme, it is argued that linear regression should not be applied at all in connection with method comparison studies (16), but that graphical techniques (bias plots) (9, 17, 18) be used instead.

To elucidate the relevance of these recommendations in practice, we here compare the usefulness of different linear regression variants (OLR, DR, SPCA, and PBR) on the basis of data obtained from real-life studies (19, 20). Our main emphasis was to clarify whether the quality of the analytical input data or the particular regression method used has the greater influence on the validity of the regression data. We discuss, in particular, the importance of correlation analysis in connection with regression analysis and the relevance of $S_{y|x}$.

The discussion of whether bias plots or linear regression procedures are more appropriate for interpretation of method comparison studies is beyond the scope of this study. The reader is referred to reports by Stöckl (9), Lawton et al. (16), Bland and Altman (17), and Hyltoft Peterson et al. (18) for this subject. Note also that the data of the method comparison studies here serve only the purpose of evaluating the different regression procedures. The analytical and clinical relevance of the method comparison studies have been described elsewhere (19, 20).

Materials and Methods

The data sets used for our investigation came from recently published method comparison studies for serum estradiol-17β (19) and serum potassium (20) (using 60 samples). The estradiol-17β study used a relatively low number of samples because it investigated the possibility of recalibration of routine methods in terms of a reference method.

For the different regression procedures, OLR was performed with Microsoft EXCEL® or the EVAL KIT (21) software. DR was performed with Microsoft EXCEL [using the respective formulae used by Cornbleet and Gochman (10) and Linnet (12)] or the EP Evaluato (22) software. SPCA and PBR were performed with the EVAL KIT software. Note that the latter calculates the SPCA slope as the geometric mean of the two slopes that result from OLR, performed with $y$ and with $x$ as the independent variable (4). Robust OLR was performed with SYSTAT 7.0® software (23). From the several variants available, the trimming procedure was chosen, with a factor of 0.1 for discriminating outlying residuals.

Results and Discussion

OVERVIEW

Tables 1 and 2 summarize the regression and correlation data for the two method comparison studies. Table 1 contains the data for the serum estradiol-17β study (19), where measurements by 12 different immunoassays ($y$ variables) were compared with those of a reference method based on isotope dilution gas chromatography–mass spectrometry ($x$ variables). Table 2 contains the data for serum potassium, obtained by comparison of results from 12 laboratories using four different routine methods ($y$) and a reference method based on ion chromatography ($x$) (20). The substantial difference between these two cases lies in the investigated concentration ranges. In the case of estradiol-17β, the concentrations ranged from 0.015 to 3.33 nmol/L (i.e., ~3 decades), whereas in the case of potassium they ranged from 3.56 to 5.42 mmol/L (i.e., 1 decade). The comparisons are ordered in both tables according to increasing values of $S_{y|x}$, as calculated from OLR.

THE $S_{y|x}$ PROBLEMATIC: CONNECTION BETWEEN $S_{y|x}$ AND $r$

Like the correlation coefficient $r$, $S_{y|x}$ indicates the magnitude of total random error of the method comparison, including non-linearity, drift or shift, total analytical imprecision ($s_{\text{tot}}$), and sample-related effects. However, in our experience when several methods are compared with each other at the same time, $S_{y|x}$ is a better indicator for total random error than $r$, particularly when the data range is wide, as illustrated in Table 1. There, the $S_{y|x}$
values increase >10-fold, i.e., from 0.017 to 0.192 nmol/L, whereas the corresponding r values decrease only from 1 to 0.98. On the other hand, r is more useful for a general expression of the magnitude of total random error because it is independent of the units of x and y. However, r is dependent on the data range. For wide data ranges (e.g., 3 decades), much greater values of r are needed to describe the same quality of correlation than for small data ranges (e.g., 1 decade). This might explain the difference in what past studies have considered a “good” r value in method comparisons, namely r ≥ 0.975 (8) or r ≥ 0.99 (11).

The mathematical relationship between $S_{y|x}$ and r can be delineated by the formula of $S_{y|x} = N - 1 \left( \frac{N-2}{N-2} (S_y^2 - b_{y|x} \cdot r \cdot S_x \cdot S_y) \right)$ (10). By substituting the slope $b_{y|x} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$, and after simplification ($N - 1 \approx N - 2$), $r^2 = 1 - \frac{S_{y|x}}{S_y}$ is obtained. Thus, the greater the ratio $S_{y|x}$/data range, the lower the r value.

Note that the higher the ratio $S_{y|x}$/data range or the lower the value of r, the greater the uncertainty in the estimates of slope and intercept (this holds true for all regression procedures).

THE $S_{y|x}$ PROBLEMATIC: CONNECTION BETWEEN $S_{y|x}$ ANALYTICAL IMPRECISION, AND SAMPLE-RELATED EFFECTS

The importance of calculating and interpreting the value for $S_{y|x}$ follows the fact that $S_{y|x}$ is influenced by two effects, namely, by $s_{a,tot}^2$ (which equals $\sqrt{s_{a,x}^2 + s_{a,y}^2}$) and by sample-related effects. Note, however, that $S_{y|x}$ may be inflated by nonlinearity and drift/shift during the comparison. Because of the latter, we recommend that method comparison studies should be done with particular care.

### Table 1. Regression data from method comparison for serum estradiol-17b measurement: isotope dilution gas chromatography–mass spectrometry reference method (x) and routine immunoassays (y).

| Test | $S_{y|x}$ from OLR | r   | OLR | DR | SPCA | PBR |
|------|-------------------|-----|-----|----|------|-----|
| 1    | 0.017             | 1.00| 0.648| 0.648| 0.648| 0.651|
| 2    | 0.047             | 0.999| 0.839| 0.840| 0.840| 0.849|
| 3    | 0.067             | 0.996| 0.955| 0.958| 0.958| 0.936|
| 4    | 0.072             | 0.996| 0.750| 0.752| 0.753| 0.686|
| 5    | 0.086             | 0.996| 0.860| 0.863| 0.864| 0.840|
| 6    | 0.088             | 0.993| 0.674| 0.677| 0.679| 0.677|
| 7    | 0.096             | 0.996| 0.997| 1.001| 1.001| 0.982|
| 8    | 0.100             | 0.991| 0.681| 0.685| 0.687| 0.726|
| 9    | 0.109             | 0.994| 0.899| 0.905| 0.905| 0.844|
| 10   | 0.110             | 0.990| 0.730| 0.735| 0.737| 0.779|
| 11   | 0.181             | 0.984| 0.919| 0.933| 0.935| 0.852|
| 12   | 0.192             | 0.980| 0.890| 0.906| 0.908| 0.839|

* $S_{y|x}$ and intercept are in nmol/L.

### Table 2. Regression data from method comparison for serum potassium measurement: ion chromatography reference method (x) and routine immunoassays (y).

| Test | $S_{y|x}$ from OLR | r   | OLR | DR | SPCA | PBR |
|------|-------------------|-----|-----|----|------|-----|
| 1    | 0.0373            | 0.996| 1.014| 1.018| 1.018| 1.014|
| 2    | 0.0404            | 0.995| 1.015| 1.020| 1.020| 1.014|
| 3    | 0.0410            | 0.995| 1.010| 1.016| 1.015| 1.018|
| 4    | 0.0441            | 0.994| 0.989| 0.995| 0.995| 0.990|
| 5    | 0.0475            | 0.994| 1.026| 1.033| 1.033| 1.020|
| 6    | 0.0478            | 0.994| 1.062| 1.069| 1.068| 1.058|
| 7    | 0.0494            | 0.993| 1.039| 1.046| 1.046| 1.039|
| 8    | 0.0569            | 0.990| 0.962| 0.972| 0.972| 0.962|
| 9    | 0.0766            | 0.983| 0.992| 1.010| 1.010| 1.000|
| 10   | 0.1199            | 0.954| 0.929| 0.973| 0.974| 0.970|
| 11   | 0.2686            | 0.871| 1.162| 1.390| 1.334| 1.244|
| 12   | 0.2942            | 0.652| 0.616| 0.918| 0.946| 0.955|

* $S_{y|x}$ and intercept are in mmol/L.
for internal quality control (IQC). Thus, in addition to the common interpretation of systematic differences between methods reflected by slope and intercept, we advise using the information content of the regression analysis about random error by comparing the observed value of $S_{y|x}$ with the one predicted from $S_{a,tot}$ (18). However, only the $S_{y|x}$ value obtained from OLR can be used for this purpose (8). Naturally, as already discussed, the higher the $S_{y|x}$ values, the more uncertain the regression estimates for systematic differences between methods will be.

When $S_{y|x} \gg S_{a,tot}$ sample-related effects are present. When the values for $x$ come from a hierarchically higher reference method, then it is obvious that the routine method ($y$) caused the problems. When two routine methods are involved, either or both methods might be affected by sample-related effects.

When $S_{a,tot}$ and $S_{y|x}$ are of similar magnitude, one should still check whether the correlation coefficient $r$ has a “reasonable” value (e.g., $0.975$ or $0.99$; see the discussion below). When $r$ is considerably lower and $S_{a,tot} \approx S_{y|x}$ method precision is too poor for sufficiently reliable estimation of slope or intercept. In such a case, there is a strong need to minimize the influence of imprecision by performing replicate analyses, checking IQC data (e.g., for drifts or shifts), or measuring all samples within the same analytical run. We want to stress that point that method comparison studies should be done with particular care for IQC, which means that many more IQC measurements should be performed when carrying out a method comparison study than when using an assay in routine operation.

**The $S_{y|x}$ Problematic: Different Regression Procedures Calculate Different Values**

It is important to notice that the different regression procedures give different values for $S_{y|x}$. Note that PBR does not calculate values for $S_{y|x}$ at all, which we consider a major disadvantage of this regression method. The formula given by Cornbleet and Gochman (10) calculates “true” $S_{y|x}$ values, in the sense of orthogonal distances of the data pairs from the regression line (those authors used the term $S_{y|x}$ instead of $S_{y|x}$). Generally, they are smaller than those from OLR. However, they become identical to those of OLR in the case of a zero slope, whereas they are smaller by $\sqrt{2}$ when the slope is 1.

Furthermore, the reader should be aware that commercial software packages also calculate different values for $S_{y|x}$: DR, as performed with the program of Rhoads (22), gave values for $S_{y|x}$ that were nearly identical to those of OLR. We assume that the Rhoads program for DR calculates “usual” $S_{y|x}$ values; however, they may differ slightly from those of OLR when slope and intercept do not totally agree in both procedures. SPCA, as performed by the software we used (21), calculated values of $S_{y|x}$ that were smaller by $\sqrt{2}$ than those of OLR. In this program, true $S_{y|x}$ values are calculated by assuming an equal imprecision of both methods and, hence, dividing the usual $S_{y|x}$ value by $\sqrt{2}$.

As noticed above, only the $S_{y|x}$ values, as calculated by OLR, can be used to compare the predicted variance across the regression line with the observed one (8).

**Interpretation of the Regression Data for Case 1**

(METHOD COMPARISONS FOR SERUM ESTRADIOL-17β MEASUREMENT): COMPARISON OF OLR AND DR/SPCA

The values for slope and intercept according to OLR and DR/SPCA are very similar for methods 1–10 (see Table 1). The biggest difference in the slope (intercept) value is 0.007 (0.006 nmol/L) for method 10. Additionally, the standard errors for slope and intercept differ <1% between OLR and DR. (The EVAL KIT program does not provide standard errors for SPCA.) Interestingly, all respective values for $r$ are $\geq 0.99$. These findings correspond with the restrictions implied before by Wakkers et al. (11) to apply OLR to method comparison data only when $r \geq 0.99$. On the contrary, for methods 11 and 12, the values for slope and intercept by OLR differ substantially from those by DR/SPCA. For these methods, $r$ values $< 0.99$ are observed. In consequence, in these cases DR/SPCA would be most appropriate (10–13). However, from the purely statistical point of view, one would prefer the application of DR/SPCA in all cases.

But what is the analytical relevance of these findings? Does the application of, for example, DR in place of OLR add a value to the method comparison for cases 11 and 12? From the analytical point of view, we would doubt this. Interestingly, when recalibration of the routine methods on the basis of their correlation with the reference method was proposed (19), these cases were intuitively excluded because of the poor correlation and the high values of $S_{y|x}$. In other words, even when recalibrated for individual samples, those methods would reveal differences from the reference method that were too large. Note also that the uncertainty of the slope (95% confidence level) was $\sim 0.08$ for cases 11 and 12, which would introduce a considerable calibration uncertainty when those methods would be recalibrated by use of the method comparison. Therefore, despite its statistical justification, application of DR instead of OLR makes little analytical sense for these cases.

From these first observations, we confirm that OLR is a valid regression procedure when $r \geq 0.99$ for method comparison studies that cover a wide data range. This holds true for the estimation of slope and intercept and their respective confidence intervals. Consequently, under the restriction that $r$ be $\geq 0.99$ for data that cover a wide range, OLR can be applied for calibration purposes as well as for hypothesis testing. When $r < 0.99$, one should investigate whether a different regression procedure really solves the analytical problem.
INTERPRETATION OF THE REGRESSION DATA FOR CASE 1
(METHOD COMPARISONS FOR SERUM ESTRADIOL-17β
MEASUREMENT): INVESTIGATION OF PBR

PBR corresponds very well to the other regression procedures for low to medium values of $S_{yx}$ (see methods 1–7), with the exception of method 4. However, PBR slope and intercept estimates differ from the other regression variants when $S_{yx}$ becomes high (in particular, in methods 11 and 12). It is obvious from this observation that PBR cannot be regarded as a substitute for DR or SPCA.

As addressed before, test 4 shows the peculiarity that the slopes for OLR and DR/SPCA differ from that of PBR, despite a relatively low value of $S_{yx}$ (Fig. 1A). When plotting the OLR residuals (Fig. 1B), it can be seen that this discrepancy originates from the fact that the method comparison data are not linearly related. This is also evidenced from the sign sequence of the $y$ residuals. PBR gives a sign sequence of $2 \times$ plus, minus, $3 \times$ plus, $2 \times$ minus, plus, $8 \times$ minus, and $5 \times$ plus, whereas OLR gives $7 \times$ plus, minus, plus, $8 \times$ minus, plus, minus, plus, minus, and plus. These sequences reveal that the middle block of results has a negative bias compared with the low and high block of results. On the basis of this observation, a nonlinear regression procedure (e.g., a quadratic one) may be more appropriate in this case.

From the fact that OLR and DR/SPCA were more closely related with each other than with PBR in all comparisons, we conclude that PBR should be applied with care to method comparisons that use medium sample sizes. PBR may treat too many data points as outliers. On the other hand, when discrepancies between different linear regression procedures are observed, this should be taken as a hint for an in-depth investigation of the underlying problem.

INTERPRETATION OF THE REGRESSION DATA FOR CASE 2
(METHOD COMPARISON DATA FOR SERUM POTASSIUM)

Table 2 shows that in the case of clustered method comparison data (by clustered we mean a small concentration range of the $x$ variables, here within 1 decade), the values for slope and intercept according to the four regression procedures are quite similar for low to medium values of $S_{yx}$ (see methods 1–9). Note that for these nine methods, $r$ values between 0.996 and 0.983 were found. This observation stands in relation to the previously mentioned restriction for using OLR dependent on $r \geq 0.975$ (1, 8). Interestingly, after logarithmic transformation of the data for estradiol-17β, the range comes to fall within 1 decade, and the “critical” $r$ value of 0.99 decreases to $\sim 0.975$. This indicates that the “$r \geq 0.975$ rule” (8) might be generally useful as a screening rule for valid application of OLR when data ranges are $< 1$ decade. However, as mentioned above, from the purely statistical point of view, DR would be preferable also in those cases.

For methods 10–12, the slope according to OLR differs distinctly from the slopes according to DR, SPCA, and PBR. Notice that for the latter methods $r$ values of 0.954 (method 10), 0.871 (method 11), and even 0.652 (method 12) were found. In those cases, are PBR, SPCA, or DR really the solution to the problem? Again, we would doubt such a statement. This can be substantiated by the graphical comparison of a “good” (method 7, $r = 0.993$), a “borderline” (method 10, $r = 0.954$), and a “poor” method comparison (method 11, $r = 0.871$; Fig. 2). Compared with method 7, the worse correlation of method 10 mostly seems to be associated with several outlying results. Indeed, robust OLR applying a trimming factor of 0.1 for outlying residuals (23) gives results nearly identical to those of PBR and DR (robust OLR, slope and intercept:...
0.973 and 0.115 mmol/L; PBR, slope and intercept: 0.970 and 0.129 mmol/L; DR, slope and intercept: 0.973 and 0.135 mmol/L. Alternatively, outliers could have been eliminated on the basis of the $4 \cdot S_{\text{u}x}$ rule (10). In consequence, one would look for reasons for the poor outcome. We know from our study (20), that it was not the method that caused the problems, but the performance of the laboratory. Clearly, in this case it is not a different regression procedure that is helpful, but an investigation of the reasons for poor analytical quality.

In addition, for what concerns the interpretation of linear regression data, investigation of the statistical reliability of the estimates is often overlooked. For example, for method 10 ($r = 0.954$), the 95% confidence limits for slope and intercept according to DR were $0.973 \pm 0.078$ mmol/L and $0.135 \pm 0.336$ mmol/L, respectively. It follows that from the statistical point of view, one would conclude that there is no difference between the routine and reference methods because the respective confidence intervals include a slope of 1 and an intercept of 0. Hence, recalibration of the routine method would not be necessary. (Generally, the higher the total random error of a study, the higher the chance that statistical hypothesis testing is passed.) On the contrary, from the analytical point of view, one would certainly consider recalibration of the routine method, especially because the lower limit of the slope is 0.895 ($0.973 - 0.078$). However, using the method comparison for recalibration, ~8% uncertainty would be added to the original calibration slope because of the uncertainty of the regression in method comparison alone. We consider such a value as too high for a potassium test. (The CLIA limit for total error is ~10% for potassium concentrations at the high end of the reference interval.) This demonstrates again that statistical considerations alone cannot give useful interpretation of method comparison studies.

**GENERAL RECOMMENDATIONS**

Present the data graphically and visually inspect them for adequacy of range and for outliers. Inspect the data for linearity:

- Use a residual plot and investigate the sign sequence of the residuals.
- If $x$ and $y$ are linearly related, perform correlation analysis.
- If $x$ and $y$ are not linearly related, perform nonlinear regression.

**Correlation analysis:**

- If $r < 0.99$ or $< 0.975$, perform outlier investigation.
- When $r$ increases satisfactorily, perform linear regression.
- When $r$ does not increase satisfactorily, perform OLR to obtain $S_{\text{u}x}$ and compare $S_{\text{u}x}$ with $s_{\text{a,tot}}$.
  - If $S_{\text{u}x} \approx s_{\text{a,tot}}$, reduce $s_{\text{a,tot}}$ (e.g., by performing replicates).
(b) If $S_{glx} \gg S_{a,tot}$ there is substantial analytical difference between the methods because of sample-related effects.

- Decide whether the difference is clinically relevant. If not, decide whether DR/SPCA or PBR add value to the interpretation of the study.
- If $r \geq 0.99$ (wide range), or $r \geq 0.975$ (small range), perform linear regression.

Linear regression:
- For estimation of slope and intercept, use OLR or DR/SPCA.
- For estimation of $S_{glx}$ use OLR.

Interpretation:
- For analytical interpretation, take into account the uncertainty of slope and intercept and compare $S_{glx}$ with $S_{a,tot}$ (see above). Apply proposed specifications wherever possible.
- Decide on clinical relevance when the slope differs considerably from 1, the intercept from 0, and $S_{glx} \gg S_{a,tot}$.

Special note:
- Perform the whole study with particular emphasis on IQC.

These recommendations are meant as help for interpreting method comparison studies. They will not work in every case; the skill, knowledge, and experience of the analyst are still the most important factor in adequate interpretation of, for example, regression estimates.

**Summary and Conclusion**

Our real-life data showed that the analytical input data have more influence on the reliability of the linear regression data than the particular regression procedure applied. Consequently, when OLR (in combination with correlation analysis) gives poor regression estimates, we recommend investigating the analytical reason for this performance instead of assuming that other linear regression procedures will improve the interpretation of the study. This investigation should address whether (a) the data for $x$ and $y$ are linearly related; (b) $S_{a,tot}$ is responsible for the poor correlation; (c) sample-related effects are present ($S_{glx} \gg S_{a,tot}$); (d) the samples are adequately distributed over the investigated range; and (e) the number of samples used for the comparison is adequate to the purpose of the application of linear regression.

**References**