Proper Ways of Comparison of Two Laboratory Methods*

[a]J. MOCÁK**, [a]B. BALLA, [b]A. BOBROWSKI, and [c]P. BLAŽÍČEK

[a]*Department of Analytical Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology, SK-812 37 Bratislava*
e-mail: mocak@cvt.stuba.sk

[b]*University of Mining and Metallurgy, PL-30 059 Krakow, Poland*

[c]*Hospital of the Ministry of Defence, Division of Clinical Laboratories, SK-833 31 Bratislava*

Received 16 October 2002

For statistically correct decision whether two laboratory methods provide concordant results when measuring the same objects, only those regression techniques are reliable, which respect the errors of both compared variables or are not influenced by them. Overview of such techniques is given in this paper. Ordinary least-squares regression may provide biased results leading to wrong decision making.

Statistical comparison of two analytical methods (or any laboratory methods) is a persistent task in an analytical laboratory and one of the most important steps in the method *validation* process. Such method comparison study is usually made by *regression*. A series of results for the same measurement objects is obtained by the *investigated* and the *reference* methods. The results found by the reference method are usually plotted on the horizontal (X) axis, and those related to the investigated method are plotted on another axis, usually vertical (Y). Then the *slope* is tested with respect to the theoretical value 1.0 by the *t-test*. If a significant difference is found, then the *proportional systematic error* is indicated.

The t-test is also used to prove whether the *intercept* is or is not significantly different from zero. A significant difference is the indication of the *constant systematic error*. Neither of the mentioned statistical tests can reveal by itself which one of the compared methods is correct or incorrect.

In this paper we will consider that the only violation of the obligatory assumptions on the least-squares method is that both regression variables are subject to errors, which enables us to concentrate solely on this problem. We have proved by means of used software that in the problem of comparison of two cholesterol methods demonstrated in further text, only this type of violation occurred.

Common ordinary linear least-squares regression (OLS) is often used for the mentioned tests but in general it provides biased estimates of the regression parameters and their standard deviations, necessary for the t-test, which leads finally to an improper decision. It is due to the violation of the presupposed condition in the OLS that one of the methods, representing independent variable, is error-free. Consequently, a better regression method has to be applied. It should either respect the errors present in both variables, or to be uninfluenced by them, which is typical for a nonparametric or robust approach. The objective of this work is to give an overview of mostly applied statistical techniques and relevant software, capable to fulfill properly the methods comparison task. Such techniques and software can be of course used for all other purposes at which all considered variables are subject to random errors.

EXPERIMENTAL

From the values c (mmol dm^{-3}) of data on the concentration of total cholesterol (Chol), HDL-cholesterol (HDL), and triacylglycerols (Tg) in blood of 288 patients, the corresponding LDL-cholesterol (LDL) values were calculated according to the Friedewald formula

$$\mathrm{LDL_{calc}} = \mathrm{Chol} - \mathrm{HDL} - \mathrm{Tg}/2.2 \qquad (1)$$

All data were measured by two automatic analyzers, namely Konelab 20 and Hitachi 911. Thus, two series of the calculated values resulted: $\mathrm{LDL_{calc}}$-K for Konelab and $\mathrm{LDL_{calc}}$-H for Hitachi. Determinations

*Presented at the XVIth Slovak Spectroscopic Conference, Košice, 23—27 June 2002.
**The author to whom the correspondence should be addressed.

in serum were based on enzymatic assays for Chol and Tg (Roche) and HDL (Genzyme) with spectrophotometric final measurement. Since the database was used retrospectively, duplicate measurements were not available.

RESULTS AND DISCUSSION

Deming method is based on *structural model* in which the *observed* (measured) variables X_i and Y_i are composed of the corresponding latent *expected* ("true") values x_i and y_i and random errors u_i and e_i, respectively, where u_i and e_i are supposed to be independent and normally distributed with a zero mean value and a constant standard deviation [1]. Then

$$X_i = x_i + u_i \tag{2}$$

$$Y_i = y_i + e_i \tag{3}$$

and the linear regression model can be expressed by two equivalent relationships

$$y = \beta_0 + \beta_1 x = \bar{y} + \beta_1(x - \bar{x}) \tag{4}$$

where $\bar{y}$ and $\bar{x}$ are the means of all coordinates y_i and x_i, respectively, and regression parameter $\beta_0 = \bar{y} - \bar{x}\beta_1$. The estimates b_0 and b_1 of the regression parameters β_0, β_1 are calculated by means of the parameter δ or its reciprocal value λ, representing the ratio of the error variances

$$\delta = V(e)/V(u) = s_e^2/s_u^2 \tag{5}$$

$$\lambda = V(u)/V(e) = s_u^2/s_e^2 \tag{6}$$

The case when $\delta = 1$ and $V(e) = V(u)$ is of a special importance and is called *orthogonal regression* (or Deming regression in a narrow sense), at which the least-squares minimization is made in the direction perpendicular to the regression line. When $\lambda = 0$ and $V(u) = 0$, then the OLS method is valid as the limiting case of Deming regression; when $\delta = 0$ and $V(e) = 0$, then another limiting case arises and the OLS is valid for the $X = f(Y)$ dependence.

If auxiliary variable L is defined as $L = (S_{YY} - \delta S_{XX})/(2S_{XY})$, where S_{YY} and S_{XX} are the corresponding variances for Y and X, respectively, and S_{XY} is the corresponding covariance, then

$$b_1 = L + (L^2 + \delta)^{1/2} \tag{7}$$

$$b_0 = \bar{Y} - b_1\bar{X} \tag{8}$$

Relatively complicated and not easily accessible is the *calculation of standard errors* of the regression parameters, s_{b1} and s_{b0}. However, for normally distributed measurement errors of both variables and assuming homoscedasticity (constant error variances), *Fuller* [1] gives the unambiguous estimates of s_{b1} and s_{b0}, which, as we have proved, is consistent with the results obtained by the *Analyse-it* software [2]. The mentioned two conditions need not be met when a nonparametric alternative of the standard errors estimates is made by the *jackknife* method, which is described in detail in Ref. [3]. *Method Validator* [4] as well as *CBstat* [5], both available on the web, use just this calculation approach.

Bivariate least-squares regression (BLS) is the generic name for a set of techniques used for regressing *bivariate* data, *i.e.* whenever a regression method is applied to data containing errors in both axes [6]. Specifically, the BLS derived in Refs. [6, 7] is based on the approach described in [8] *minimizing the sum, S, of the weighted residuals*, defined as

$$S = \sum_{i=1}^{n} w_{Ri} R_i^2, \quad i = 1, \ldots, n \tag{9}$$

$$R_i = [Y_i - f(X_i, b_j)], \quad j = 1, \ldots, m \tag{10}$$

$$w_{Ri} = 1/s_{Ri}^2 = 1/[s_{Yi}^2 + s_{Xi}^2 - 2b_1 \operatorname{cov}(X_i, Y_i)] \tag{11}$$

The residuals were expressed originally for m regression parameters [8], but weights w_{Ri} are expressed here for $m = 2$, needed for the method comparison. Minimization, leading to nonlinear equations, is made in the iterative way [6—9]. The BLS is effectively used in software package Calibro 2000 [9, 10] and the MATLAB code is listed in [7]. Despite the BLS definition [7], by which *e.g.* the Deming method can also be considered a BLS technique, only the variants described in [6—9] are designated in this way. The BLS can treat even a heteroscedastic case (with nonconstant variance) and the user of Calibro 2000 can put different weights not only for each regression variable, but also for every individual point i, if desired. On the other hand, in many practical situations the determination of the weights is difficult and even obtaining the variances ratio for compared variables is cumbersome or impossible. In such situations, equal weights should be adopted, consistent with the recommended choice $\delta = 1$ in Deming regression when the ratio of variances is not accessible.

In the *Passing—Bablok* regression procedure for methods comparison (P-B), the slopes of the straight lines between any 2 points from the set of n regression points are calculated. The number of possible slopes is $N \leq n(n-1)/2$, since from all possible slopes those having $\pm\infty$ and 0 values have to be subtracted. All N slope values are calculated by the equation

$$S_{ij} = (y_i - y_j)/(x_i - x_j) \text{ for } 1 \leq i < j \leq N \tag{12}$$

and sorted in the increasing order. The *P-B slope* b_1 is then obtained as the *shifted median*

Table 1. Regression Parameters, their Standard Deviations and 95 % Confidence Intervals for the Regression Dependence LDL_{calc}-K = $b_0 + b_1$. LDL_{calc}-H (Model: $y = b_0 + b_1 x$)*

Way of calculation	b_1	b_0	s_{b1}	s_{b0}	b_{1L}	b_{1U}	b_{0L}	b_{0U}
OLS (Analyse-it)	0.9946	−0.0436	0.0095	0.0341	0.9759	1.0134	−0.1107	0.0236
Deming (Analyse-it)	1.0077	−0.0886	0.0097	0.0346	0.9987	1.0268	*−0.1566*	*−0.0206*
Deming (Meth. Validator)	1.008	−0.0886	–	–	0.986	1.029	*−0.1615*	*−0.0156*
BLS (Calibro 2000)	1.0077	−0.0886	0.0096	0.0342	0.9889	1.0266	*−0.1560*	*−0.0212*
P-B (Analyse-it)	1.0068	−0.0959	–	–	0.9863	1.0285	*−0.1741*	*−0.0305*
P-B (Meth. Validator)	1.007	−0.0958	–	–	0.986	1.028	*−0.1741*	*−0.0305*

*The symbols b_{1L} and b_{1U} denote the limits of the confidence interval for the slope, b_{0L} and b_{0U} are used to confine the intercept confidence limit. Variances ratio $\lambda = 1$ was used for the Deming and BLS methods. Confidence intervals not containing the value relevant to concordant methods are indicated in italics. Number of tabulated decimal digits is given by the respective software.

$$b_1 = S_{(N+1)/2+K} \text{ for } N \text{ odd} \qquad (13)$$

$$b_1 = \exp\{[\log(S_{N/2+K}) + \log(S_{N/2+1+K})]/2\} \text{ for } N \text{ even (geometric mean)} \qquad (14)$$

where K (representing the shift) is the number of S_{ij} values below −1. The *intercept* b_0 is obtained by calculating the median of all $(y_i - b_1 x_i)$ values. Confidence intervals for the slope and intercept are derived also by calculating the index of the sorted values, both for the upper and lower limits, as described in [11]. The standard errors are not obtained in this purely nonparametric procedure.

The Passing—Bablok procedure is applicable without the need to use any variance estimates. Its original version tests the agreement between two laboratory methods [11], therefore it is called *agreement P-B* variant. It works well even if the variables errors are not normally distributed, and the nonconstant variance over the sampling range does not have any harmful influence on results. However, *Linnet* [12] claims that the P-B results are without any bias only when a rather unusual condition $\delta = b_1^2$ is fulfilled (the corresponding ratio of the constant standard deviations equals the squared slope). Despite the mentioned drawback, the P-B procedure is often used, mainly in biochemical and clinical chemistry literature. An extremely useful feature of this procedure is its robustness with respect to the outliers.

We have made a full comparison of the results obtained by proper statistical techniques as well as the OLS regression using various analytes. Commercially available software and own QuickBasic programs, composed according to the above theoretical relationships, were used for this purpose. As an example, the results of indirect determination of LDL cholesterol are summarized in Table 1. Close inspection of this table reveals a constant systematic error since the intercept confidence interval does not contain zero (results are indicated in italics). However, the ordinary least-squares regression fails to give this decision. The mentioned results obtained by our programs and Analyse-it sofware were in a full agreement.

An important question is why and in which way a correction can be made in case when a systematic error is indicated in the performed method comparison study. For example, such a situation is common in clinical laboratories where often only one automatic analyzer is used in case of lower amount of samples, however, when the number of samples increases, then another analyzer (often of different brand or made by a different producer) comes into use. In order to receive harmonized results, it is then necessary to make a correction of the results obtained by one analyzer with respect to another. The most frequent approach is to make the correction directly by using the received regression equation. For the case treated in this paper, based on the regression coefficients for orthogonal (Deming) regression or the BLS (Table 1), the LDL results received by means of the Hitachi analyzer can be recalculated (*i.e.* corrected) to be valid for the Konelab analyzer by the following equation: LDL_{calc}-K = −0.0886 + 1.0077LDL_{calc}-H. If the opposite recalculation is needed, then the inverse linear equation has to be applied, namely LDL_{calc}-H = 0.0879 + 0.9924LDL_{calc}-K, which can be easily obtained by rearranging the previous equation (the intercept is now $-b_0/b_1$ and the slope is $1/b_1$).

Another approach for expressing the mentioned correction is based on the *bias plot* (*difference plot*) [13], well known in clinical chemistry, where the differences $(Y_i - X_i)$ are calculated for all pairs i, $i = 1, \ldots, n$. This way of calculation is equivalent to the regression model $y = b_0 + x$, with the fixed slope $b_1 = 1$, so that it can be used when the systematic error exists in the intercept (b_0 significantly different from zero), but it is incorrect for the cases when the slope is significantly different from 1. Bias plot is optional in several statistical software packages. In our case the significant bias equal to −0.062 was discovered by Analyse-it, which leads to the corrections LDL_{calc}-K = LDL_{calc}-H − 0.062 and LDL_{calc}-H = LDL_{calc}-K − 0.062; these corrections are constant in the whole range of the variable values.

CONCLUSION

The most frequent statistical techniques for com-

parison of two analytical methods are: *a*) Deming regression and orthogonal regression as its useful variant – when the estimates of the variances of compared methods are unknown, *b*) bivariate least-squares regression, and *c*) the Passing—Bablok rank method. Some software, namely Analyse-it, Calibro, Method Validator or CBstat make such calculations easy and provide also graphical outputs.

Acknowledgements. The authors (J. M. and B. B.) gratefully acknowledge the support of the Slovak Grant Agency VEGA, Project No. 1/9129/02, as well as the software licence provided by Analyse-it Software, Ltd.

REFERENCES

1. Fuller, W. A., *Measurement Error Models.* Wiley, New York, 1987.
2. *Analyse-it v. 1.61.* Analyse-it Software Ltd., Leeds, U.K., 2000, http://www.analyse-it.com.
3. Linnet, K., *Stat. Med. 9*, 1463 (1990).
4. Marquis, P., *Method Validator v. 1.1.9.* Metz, France, http://perso.easynet.fr/~philimar.
5. Linnet, K., *CBstat v. 4.1.0.* Risskov, Denmark, 2001, http://www.cbstat.com.
6. Riu, J. and Rius, F. X., *Anal. Chem. 68*, 1851 (1996).
7. Riu, J. and Rius, F. X., *Trends Anal. Chem. 16*, 211 (1997).
8. Lisý, J. M., Cholvadová, A., and Kutej, J., *Comput. Chem. 14*, 189 (1990).
9. *Calibro 2000 v. 1.0.* SynexChem LLC, Alexandria, VA, U.S.A., http://www.chemmea.sk.
10. Lisý, J. M., *Ropa Uhlie Plyn Petrochem. 42*, 32 (2000).
11. Passing, H. and Bablok, W., *J. Clin. Chem. Clin. Biochem. 21*, 709 (1983).
12. Linnet, K., *Clin. Chem. 445*, 1024 (1998).
13. Bland, J. M. and Altman, D. G., *Lancet 1986*, 307.