Florian Dufey*

Derivation of Passing–Bablok regression from Kendall's tau

<https://doi.org/10.1515/ijb-2019-0157>

Received December 13, 2019; accepted April 20, 2020; published online August 12, 2020

Abstract: It is shown how Passing's and Bablok's robust regression method may be derived from the condition that Kendall's correlation coefficient tau shall vanish upon a scaling and rotation of the data. If the ratio of the standard deviations of the regressands is known, a similar procedure leads to a robust alternative to Deming regression, which is known as the circular median of the doubled slope angle in the field of directional statistics. The derivation of the regression estimates from Kendall's correlation coefficient makes it possible to give analytical estimates of the variances of the slope, intercept, and of the bias at medical decision point, which have not been available to date. Furthermore, it is shown that using Knight's algorithm for the calculation of Kendall's tau makes it possible to calculate the Passing–Bablok estimator in quasi-linear time. This makes it possible to calculate this estimator rapidly even for very large data sets. Examples with data from clinical medicine are also provided.

Keywords: Kendall's tau; method comparison; Passing–Bablok regression; robust regression; Theil–Sen regression.

Acronyms

CI	Confidence interval
cPBR	classical Passing–Bablok regression
ePBR	equivariant Passing–Bablok regression
LPR	least product regression
LSR	least square regression
PB	Passing–Bablok
PBR	Passing–Babkok regression
TSR	Theil–Sen regression

1 Introduction

The comparison of measurements obtained with different instruments or test formats are a biometrical problem familiar to every scientist. While disciplines differ considerably with respect to the preferred methods, regression is of general importance [1]. Since both measurements to be compared have some level of imprecision, least squares regression (LSR) is not the best choice, as it assumes the abscissa values to be error free. This drawback is avoided by techniques like Deming regression (also known as major axis regression) or least product regression (LPR) [1–5]. However, the results of these methods may be seriously biased by a small number of outlying measurements. To overcome this problem, in the field of clinical chemistry, Passing and Bablok [6] proposed a regression method, to be called classical Passing–Bablok regression (cPBR) in the following, which may be considered a variant of Theil–Sen regression (TSR) [7, 8], adapted to take into account the variation not only of the ordinate but also of the abscissa.

*Corresponding author: Florian Dufey, Roche Diagnostics GmbH Standort Penzberg, Biostatistics, DXREBA, Nonnenwald 2, Penzberg, 82377, Germany, E-mail: florian.dufey@roche.com

While these authors gave only a heuristic justification for their method, they showed empirically the unbiasedness of the results obtained [9] and the robustness of the method. Although it has become one of the most often used statistical tools in clinical chemistry [5], lamentably, the method remains widely unknown outside this field.

Due to the limited analytical understanding of cPBR, formulas for the variance and confidence intervals of the slope and intercept are also not well developed. Furthermore, the cPBR is limited to slopes near 1, although, in a follow up paper [10], a method applicable for any value of the slope was proposed, which in the following shall be called equivariant PBR (ePBR). cPBR may be shown to be an approximation to ePBR. In the following, if both cPBR and ePBR are in scope, the term Passing–Bablok regression (PBR) will be used.

It is the aim of this paper to show that all the aforementioned regression techniques and specifically PBR result from the condition that either Pearson's or Kendall's correlation shall diminish to 0 upon a suitable slope-dependent scaling and rotation of the original variables. This extends the ideas of Sen [8] on the derivation of TSR to PBR. Along the same lines, a robust variant of the Deming regression procedure is developed, which is found to be equivalent to the determination of the circular median of the doubled slope angle – a well-known estimator in the theory of directional statistics.

The paper is structured as follows: in Section 2.1, the methods within scope of the paper are shortly recapitulated. The estimators obtained with the different methods are compared using an example from clinical chemistry (Figure 1). Section 2.2 contains the theoretical derivation of the Passing–Bablok estimator and is relevant mainly for the reader more interested in the theory. In Section 2.3, the main equations for the calculation of the PB estimators are laid out, while in the following Section 2.4, expressions for the circular median estimator are given. It also contains a numerical comparison of the different methods. In Section 2.5, it is shown how the PB estimators can be calculated efficiently. Section 3 contains a derivation of the confidence interval for the slope and an efficient algorithm for its numerical calculation. Finally, in Section 4, new equations for the confidence intervals of the intercept and bias are derived. Their coverage is then studied by simulations. Generally, examples or simulations more relevant for the practitioner are included at the end of the sections or in the figures and their captions. Recommendations on the use of the methods are put together in the discussion (Section 5).

Figure 1: To exemplify the performance of the different regression schemes for measurements with very different ranges, results from a comparison of a clinical chemistry tests, a PCR assay (concentration [IU/ml]) and an immunoassay (cutoff index, COI), respectively, are shown. Red lines are for TSR of concentration on cutoff index (dashed), cutoff index on concentration (dotted)¹, ePBR (solid). Blue lines are for LSR of concentration on cutoff index (dashed), cutoff index on concentration (dotted) and LPR solid. Green curve is cPBR. On the left, all curves are forced through the center of mass of concentration and cutoff index. On the right, the intercept for the TSR and PBR curves is obtained as proposed by Passing and Bablok.

2 Derivation of regression from correlation

2.1 Synopsis of the relevant regression methods

In the following, a caret will designate the estimated value, e.g., of the slope $\hat{m}$, and a bar the mean, e.g., $\bar{x} = \sum_i x_i/n$. Furthermore, $X = \{x_i\}$, and $Y = \{y_i\}$.

It will be shown how the regression schemes mentioned in the introduction, namely, LSR, LPR, Deming regression, and notably PBR, may be derived from the condition that either Pearson's or Kendall's correlation diminishes to 0 for transformed input data.

The following assumptions are made, although for some methods further generalizations are possible. For LSR and Deming regression, both weighted and unweighted variants exist. For the ease of presentation, only the unweighted variants will be discussed.

1. There exists a structural relationship between the expected values of random variables x_i and y_i , denoted as x_i^* and y_i^* , respectively, of the form $y^* = b + mx^*$. Here, $x_i = x_i^* + \xi_i$ and $y_i = y_i^* + \eta_i$. For LSR and TSR, $\xi_i = 0$.
2. Error terms for points $i \neq j$ are independent. ξ_i and η_i are independent, too.
3. $E(\xi_i) = E(\eta_i) = 0$, for all i .
4. The ratio of the variances is independent of i , $\gamma^2 = \sigma_{\eta_i}^2/\sigma_{\xi_i}^2 = \text{constant}$.
5. The distribution functions of the error terms may depend on i , but for fixed i , the distributions of ξ_i and η_i are equal up to scaling with the standard deviation², i.e., $\forall i, x : P(\xi_i/\sigma_{\xi_i} < x) = P(\eta_i/\sigma_{\eta_i} < x)$.

For a detailed description of the estimation procedures to be considered, the reader is referred to the literature [1–4, 6]. Here, only a very condensed synopsis of the main assumptions and formulas for the estimated slope are given for the methods of interest, as follows:

- **LSR**: In least squares regression of Y on X , the estimate of the slope $\hat{m}$ is obtained by minimizing the sum of the squared deviations of the y_i from the estimated values, $\sum_i (y_i - \bar{y} - m(x_i - \bar{x}))^2$ as

$$\hat{m} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2}.$$

- **LPR**: In least product regression, the estimate $\hat{m}$ is obtained minimizing the sum of the products of the deviations of the y_i and x_i from their estimated values, $\sum_i (y_i - \bar{y} - \hat{m}(x_i - \bar{x}))(x_i - \bar{x} - (y_i - \bar{y})/\hat{m})$ as

$$\hat{m} = \left(\frac{\sum_i (y_i - \bar{y})^2}{\sum_i (x_i - \bar{x})^2} \right)^{1/2},$$

where the sign of the root is chosen equal to the sign of Pearson's correlation coefficient of x and y . Sometimes, LPR is called geometric mean regression, as the estimate $\hat{m}$ may also be expressed as the geometric mean of the slope estimate from a LSR of y on x and the reciprocal slope estimate of regression of x on y . Also the terminology “reduced major axis regression” is common.

- **Deming Reg.**: In major axis, or Deming regression, the slope is determined by minimization of both the sum of the squared deviances of x and y from the estimated true values $\hat{x}_i^*$ and $\hat{y}_i^* = \bar{y} + \hat{m}(\hat{x}_i^* - \bar{x})$,

² Especially in the case of Deming and linear regression, it is often assumed that the errors are Gaussian and homoscedastic. These assumptions will not be made here. While this means that the estimators are not necessarily of minimal variance, they are still consistent.

$$\sum_i \left((y_i - \hat{y}_i^*)^2 + \gamma^2 (x_i - \hat{x}_i^*)^2 \right),$$

where the sum is minimized with respect to both $\hat{m}$ and the $\hat{x}_i^*$. The slope estimate is

$$\hat{m} = \frac{1}{2s_{xy}} \left((s_{yy} - s_{xx}\gamma^2) + \sqrt{(s_{yy} - s_{xx}\gamma^2)^2 + 4s_{xy}^2\gamma^2} \right),$$

where $s_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$, $s_{xx} = \sum_i (x_i - \bar{x})^2$, and $s_{yy} = \sum_i (y_i - \bar{y})^2$. It can be seen that the slope estimators of LSR and LPR arise as special cases of the Deming regression estimate. Namely, $\gamma = \infty$ corresponds to regression of y on x and $\gamma = 0$ to regression of x on y . Assuming that $\gamma = m$, replacement of γ by its estimator $\hat{m}$ in the last expression, yields the LPR expression for $\hat{m}$.

- *Roos Reg.*: As first discussed by Roos [11] and Xu [12], LSR may be considered as a special case from a family of regressions where the squared distance of the points from a regression line is minimized along lines with slope κ , e.g., LSR is a special case with $\kappa = \infty$. The slope estimate is

$$\hat{m}_\kappa = \frac{s_{yy} - \kappa s_{xy}}{s_{xy} - \kappa s_{xx}}.$$

- *TSR*: According to Sen [8], in Theil–Sen regression the slope can be estimated from the condition that Kendall's correlation coefficient τ between the X and the intercepts $Y - mX$ shall be 0,

$$\tau(X, Y - \hat{m}X) = N \sum_{i < j} \text{sgn}(x_j - x_i) \text{sgn}(y_j - y_i - \hat{m}(x_j - x_i)) = 0.$$

N is a normalization factor. When there are no ties, $N = 2/n(n-1)$. The function $\text{sgn}(x)$ is the sign of x , if $x \neq 0$, and 0, if $x = 0$. If the x_i are ordered so that $x_j > x_i$ if $j > i$, then

$$\hat{m} = \text{median} \left(\underbrace{\frac{y_j - y_i}{x_j - x_i}}_{=: S_{ij}} \right),$$

where S_{ij} is the slope of the line connecting points i and j .

- *cPBR*: In classical Passing–Bablok regression [6] proposed to modify the TSR estimator, to take account of the X not being free of error. If K is the number of the slopes $S_{ij} < -1$ and $N = n(n-1)/2$ the total number of pairs, then

$$\hat{m} = F_S^{-1} \left(\frac{1}{2} + \frac{K}{N} \right),$$

where F_S is the empirical cumulative distribution function of the $\{S_{ij}\}$. Furthermore, Passing and Bablok made the assumption that $m \approx 1$.

- *ePBR*: In a follow-up article Bablok et al. [10] proposed an alternative method to cPBR which is applicable also for slopes $\neq 1$. They gave a recursive definition which can be shown [13] to be equivalent to

$$\hat{m} = \text{med}(|S_{ij}|).$$

In Figure 1, some of these methods are applied to a data set from clinical chemistry containing paired measurements obtained with both PCR assay and an immunoassay. Due to different units, the numerical ranges of the results differ by several orders of magnitude. It can be seen that the LSR, TSR and also the cPBR are sensitive to the choice of the dependent and independent variables, while the LPR and ePBR yield results which are more centered than those of the other methods.

2.2 Coordinate transformations

In this section it will be shown that all these estimators may be derived from a common expression. Since LSR and LPR can be derived from Deming regression, the latter will be the focus. The idea is to rotate the coordinate system so that either Pearson's correlation coefficient r , or Kendall's correlation coefficient τ , vanishes [13]. However, this cannot be done directly because, upon rotation, the error terms will in general be no longer independent:

Let Σ be the covariance matrix of the $\{(x_i, y_i)^T\}$, Σ^* of the $\{(x_i^*, y_i^*)^T\}$ and Σ_i^{Ξ} that of $(\xi_i, \eta_i)^T$, then

$$\Sigma = \Sigma^* + \sum_i \Sigma_i^{\Xi},$$

where $\Sigma_{i,11} = E((x_i - x_i^*)^2)$, $\Sigma_{i,12} = \Sigma_{i,21} = E((x_i - x_i^*)(y_i - y_i^*))$, and $\Sigma_{i,22} = E((y_i - y_i^*)^2)$. Corresponding expressions hold for Σ^* , e.g., $\Sigma_{11}^* = \sum_i (x_i^* - \bar{x}^*)^2$, while $\Sigma_{i,11}^{\Xi} = \sigma_{\xi_i}^2$, $\Sigma_{i,12}^{\Xi} = \Sigma_{i,21}^{\Xi} = 0$, and $\Sigma_{i,22}^{\Xi} = \sigma_{\eta_i}^2$. The rotation angle, which diagonalizes Σ is

$$\phi = \frac{1}{2} \arctan \left(\frac{\Sigma_{12}}{\Sigma_{22} - \Sigma_{11}} \right) = \frac{1}{2} \arctan \left(\frac{\Sigma_{12}^*}{\Sigma_{22}^* - \Sigma_{11}^* + \sum_i \sigma_i^2 (y^2 - 1)} \right),$$

which is only consistent if $y^2 = 1$.

However, if new x -values are introduced scaling by y so that $x'_i = yx_i$, the covariance matrix of the new error terms $\xi' = y\xi$ and η will become a multiple of the unit matrix. Upon scaling, the original slope m will be transformed into $m' = m/y$. The coordinate system can now be rotated by the angle $\phi = -\arctan(m')$, whence

$$x'' = x' \cos(\phi) - y \sin(\phi) = \left(y + x \frac{y^2}{m} \right) \frac{m}{\sqrt{y^2 + m^2}},$$

and

$$y'' = y \cos(\phi) + x' \sin(\phi) = (y - mx) \frac{y}{\sqrt{y^2 + m^2}}.$$

Dropping irrelevant constant factors, $X'' = \{y_i + x_i y^2/m\}$, and $Y'' = \{y_i - mx_i\}$ are defined.

The condition that the Pearson correlation $r(X''(m), Y''(m))$ of the transformed points diminishes to 0 for $m = \hat{m}$ becomes

$$r(X''(\hat{m}), Y''(\hat{m})) = \frac{\sum_i (x''_i - \bar{x}'')(y''_i - \bar{y}'')}{\sqrt{\sum_i (x''_i - \bar{x}'')^2} \sqrt{\sum_i (y''_i - \bar{y}'')^2}} = 0, \quad (1)$$

which can be simplified to

$$\sum_i \left(y_i - \bar{y} + \frac{y^2}{\hat{m}} (x_i - \bar{x}_i) \right) (y_i - \bar{y} - \hat{m} (x_i - \bar{x}_i)) = 0. \quad (2)$$

It can be shown that the solution of this equation yields exactly the Deming regression estimate $\hat{m}$.

When Kendall's τ is substituted for Pearson's r in the preceding derivation, differences $\Delta x_{ij} = x_j - x_i$ and Δy_{ij} has to be considered. The errors of these differences, $\Delta \eta_{ij}$, and $\Delta \xi_{ij}$ will also be independent with the same ratio y^2 of the variances. Hence, the equivalent of Eq. (1) becomes

$$\tau(\hat{m}) = \tau(X''(\hat{m}), Y''(\hat{m})) = \left(\frac{n(n-1)}{2} \right)^{-1} \sum_{i < j} \text{sgn}(\Delta x''_{ij} \Delta y''_{ij}) = 0. \quad (3)$$

This can also be simplified to

$$\sum_{i < j} \text{sgn} \left[\left(\Delta y_{ij} + \frac{y^2}{\hat{m}} \Delta x_{ij} \right) (\Delta y_{ij} - \hat{m} \Delta x_{ij}) \right] = 0. \quad (4)$$

The solution of this equation yields an estimator³ which is the equivalent of Deming regression. For this slope estimate to be consistent, it is necessary that the expectation value of Eq. (4) with $\hat{m}$ replaced by m must be 0,

$$E\left(\sum_{i < j} \operatorname{sgn}\left[\left(\Delta y_{ij}^* + \frac{\gamma^2}{m} \Delta x_{ij}^* + \Delta \eta_{ij} + \frac{\gamma^2}{m} \Delta \xi_{ij}\right)(m \Delta \xi_{ij} - \Delta \eta_{ij})\right]\right) = 0. \quad (5)$$

If the x_i^* are not random variables, generically, the expectation values of each term in the sum has to be 0 on its own. This condition is fulfilled if either:

- $\gamma = |m|$ and the distributions of the $m \Delta \xi_{ij}$ and $\Delta \eta_{ij}$ are identical. The term whose expectation is taken in Eq. (5) is anti-symmetric under an interchange of the $m \Delta \xi_{ij}$ and $\Delta \eta_{ij}$ whence the expectation must be 0. These are exactly the assumptions made by Passing and Bablok.
- Or the $m \Delta \xi_{ij} - \Delta \eta_{ij}$ are statistically independent of the corresponding $\Delta y_{ij}^* + \frac{\gamma^2}{m} \Delta x_{ij}^* + \Delta \eta_{ij} + \frac{\gamma^2}{m} \Delta \xi_{ij}$ and their expectations be 0. For general m and γ , this implies that the distribution of the $\gamma \xi_i$ and η_i has to be rotationally invariant, which is only true if they are independent samples from a common Gaussian distribution with mean $E(\eta_i) = E(\xi_i) = 0$. Although requiring a Gaussian distribution may seem quite restrictive, one should note that the two terms become also approximately statistically independent when the errors are small compared to the mean distance between the points, since it follows that $\Delta y_{ij}^* + \frac{\gamma^2}{m} \Delta x_{ij}^* + \Delta \eta_{ij} + \frac{\gamma^2}{m} \Delta \xi_{ij} \approx \Delta y_{ij}^* + \frac{\gamma^2}{m} \Delta x_{ij}^*$ in most of the terms, which may therefore be pulled out of the expectation. The expectation $E(\operatorname{sgn}(m \Delta \xi_{ij} - \Delta \eta_{ij})) = 0$, if the distribution of the ξ_i and η_i is symmetric.

The two cases shall be discussed separately in the following subsections. It is remarked that in both cases there are two possible solutions which only differ in the sign of the resulting slope. As in LPR, the sign may be fixed as that of Kendall's correlation between the original X and Y . In the following descriptions, it will be assumed that this sign is positive, which can always be achieved by eventually changing the sign of the X .

2.3 Passing–Bablok regression

The first case leads to ePBR:

Setting $\gamma^2 = m^2$, finding the $\hat{m}$ for which the Pearson correlation becomes 0 yields the LPR estimate. Analogously, the condition that Kendall's correlation shall diminish to 0 leads to an equation for $\hat{m}$

$$\sum_{i < j} \operatorname{sgn}\left[(\Delta y_{ij} + \hat{m} \Delta x_{ij})(\Delta y_{ij} - \hat{m} \Delta x_{ij})\right] = 0. \quad (6)$$

with

$$S_{ij} = \frac{\Delta y_{ij}}{\Delta x_{ij}},$$

this can also be written as

$$\tau(\hat{m}) \sim \sum_{i < j} \operatorname{sgn}(\hat{m}^2 - S_{ij}^2) = \sum_{i < j} \operatorname{sgn}(\hat{m} - |S_{ij}|) = 0,$$

or,

$$\hat{m} = \operatorname{median}(|S_{ij}|).$$

Unlike the cPBR estimator, a scaling of the S_{ij} with a factor will also scale the estimator $\hat{m}$ by the same factor. Therefore, this estimator will be called ePBR estimator in the following.

Alternatively, this estimator arises on introduction of the angles $\hat{\phi} = \arctan(\hat{m})$, and $\phi_{ij} = \arctan(S_{ij})$, as

³ The estimator is usually not unique. It may be made so by taking the mean of the largest and smallest possible solution. As a popular alternative, the tangent of the mean of the two slope angles is used.

$$\hat{\phi} = \text{median}(|\phi_{ij}|).$$

Therneau [13] could show that this expression corresponds to an estimator proposed by Bablok et al. [10], who gave a recursive approximation for this estimator. This recursion uses the cPBR as a starting point. It can be seen that if in the first bracket of Eq. (6), $\hat{m}$ is replaced by 1, this will yield exactly the cPBR estimator. If Pearson correlation is used instead of Kendall's, the equivalent of the cPBR corresponds to the Roos regression estimate with parameter $\kappa = -1$.

While Eq. (6) yields a consistent estimator, the bias of the ePBR estimator for small sample sizes remains to be determined. It is assumed that the true $m \neq 0$. Multiplying

$$\sum_{i < j} \text{sgn} \left(\Delta y_{ij}^2 - \left(\frac{\hat{m}}{m} \right)^2 (m \Delta x_{ij})^2 \right) = 0 \quad (7)$$

with $(m/\hat{m})^2$ yields

$$\sum_{i < j} \text{sgn} \left((m \Delta x_{ij})^2 - \left(\frac{m}{\hat{m}} \right)^2 \Delta y_{ij}^2 \right) = 0. \quad (8)$$

As $m \Delta x_{ij}$ and Δy_{ij} are distributed alike, it can be concluded that $\hat{m}/m$ and $m/\hat{m}$ also follow the same distribution. Therefore, $\ln(\hat{m})$ is symmetrically distributed around $\ln(m)$. As the variance of $\hat{m}$ is $O(1/n)$ (see below), $\ln(\hat{m})$ is an unbiased estimator of $\ln(m)$, and $\hat{m}$ is an asymptotically unbiased estimate of m . An equivalent result holds for the LPR estimator [14].

Considering the robustness properties of the ePBR estimator, it is easy to show that it has the same minimal breakdown point as the TSR estimator. For the TSR estimator the breakdown point is reached earliest, when the slopes between any two points of which at least one is an outlier make up 50% of all slopes and are either all larger or all smaller than all the slopes between any two points which are not outliers ([15], p. 67). Asymptotically, this is the case when the fraction of the outliers reaches $1 - 1/\sqrt{2} \approx 29\%$. The calculation remains true, if, in case of the ePBR, the slopes are replaced by their absolute value, whence the breakdown point of the ePBR is also 29 %.

Figure 2: A: Histogram of all unique pairwise slope angles for the example from Figure 1 after scaling with the PB slope estimate. Light grey: slopes which contribute +1 to Kendall's correlation function; Dark grey: slopes which contribute -1 to the correlation function. B: Histogram for the doubled slope angles (modulo 2π). An equal amount of points lie left (light grey) and right (dark grey) of the line passing through the circular median at $\pi/2$ and the origin of the circle.

2.4 Circular median of the doubled angle

Considering the general solution of Eq. (4) for fixed and known γ , setting $\phi = \arctan(m/\gamma)$, $\phi_{ij} = \arctan(\Delta y_{ij}/(\gamma \Delta x_{ij}))$, and making use of addition theorems for the trigonometric functions, one obtains

$$\sum_{i < j} \operatorname{sgn}(\sin(2(\hat{\phi} - \phi_{ij}))) = \sum_{i < j} -\operatorname{sgn}\left(2(\hat{\phi} - \phi_{ij}) - 2\pi \frac{2(\hat{\phi} - \phi_{ij})}{2\pi} + \pi\right) = 0.$$

Here, $\lfloor x \rfloor$ is the floor function, i.e., the largest integer less than or equal to x . $\hat{\phi}$ is a standard estimator of circular or, more generally, of directional statistics [16, 17] known as the “circular median” of the doubled angle. Doubling of the angle is standard in circular statistics for data showing diametrically bimodal distributions. The circular median has been shown to be a robust and unbiased location estimator of the slope angle [18]. To visualize the circular median, the (doubled) angles are represented as points on a unit circle and a diameter PQ is drawn which separates the circle into two halves containing the same number of points, cf. Figure 2. The point P , for which the majority of points have less distance than to Q , is then the circular median. For an odd number of points it coincides with a measured angle, while for an even number of points the mean of the nearest two points is used.

Estimation of the slope via the circular median seems to be an attractive and robust alternative to the usual Deming regression, especially if the ratio of the standard errors is known and different from the expected slope. Similar to the way the Deming slope estimator interpolates between LSR estimator for the slope of Y versus X and the inverse of X versus Y depending on the ratio γ , the circular median can be seen to interpolate between the corresponding TSR estimates. The calculation of the PBR estimate is also possible using the circular median algorithm.

While finding the circular median is as complex as finding the usual median as required in the practical computation of the PBR, the concept offers the advantage, that it can be extended to find a regression line when comparing more than two sets of data [19].

It is of interest to compare the ePBR and the circular median method numerically. To that end, 100 points were sampled 1000 times with $x^* = y^*$ from a uniform distribution, i.e., with $m = 1$, and a) equal coefficient of variance $cv_y = cv_x = 0.3$ and b) unequal coefficient of variance $cv_x = 0.3$ and $cv_y = 0.9$, with Gaussian error. In case a), both methods yield practically the same results with mean $\hat{m} = 1.005$ (sd = 0.067) for the ePBR and mean $\hat{m} = 1.007$ (sd = 0.084) for the median circular slope. However, in case b), the ePBR is biased with mean $\hat{m} = 1.606$ (sd = 0.160), while the mean median circular slope estimate is $\hat{m} = 1.007$ (sd = 0.084).

2.5 Numerical complexity of the calculation of the PBR and circular median estimators

The question remains how efficiently the TSR, PBR, and circular median estimate can be calculated. In case of the TSR, this problem has long been solved and represents a classical problem of computational geometry [20]. While the brute force approach to find the median of all the pairwise slopes between n points has a numerical complexity of $O(n^2)$, better algorithms have only a complexity of $O(n \ln n)$. The problem can be shown to be equivalent to that of computing Kendall’s tau, for which an efficient solution was given by Knight in 1966 [21]. It is based on the observation that the evaluation of Kendall’s $\tau(X'', Y'')$ is equivalent to that of counting the inversions needed to bring the set Y'' into the order of X'' , an operation which can be accomplished in $O(n \ln n)$ steps using a sorting algorithm like mergesort [22]. x_i'' and y_i'' can be interpreted as intercepts $b_i(m)$ of lines passing through the points (x_i, y_i) whose slope m is taken as an independent variable. This lends to an interpretation in terms of a duality transformation which maps points (x_i, y_i) to lines (parametrized by $(m, b_i(m))$) and lines to points. In the field of image analysis, this transformation is known as Hough transformation [23].

Table 1: Comparison of corresponding regression methods obtained by zeroing of either Pearson or Kendall correlation.

Assumed variance structure	X''	Y''	Correlation function	
			Pearson, $r(X'', Y'') = 0$	Kendall, $\tau(X'', Y'') = 0$
$\sigma_x = 0$	X	$Y - mX$	LSR	TSR
$\sigma_y/\sigma_x = m$	$Y + mX$	$Y - mX$	LPR	ePBR
$\sigma_y/\sigma_x = m \approx -\kappa$	$Y - \kappa X$	$Y - mX$	Roos	cPBR ($\kappa = -1$)
$\sigma_y/\sigma_x = \gamma (\neq m)$	$Y + \frac{\gamma^2}{m}X$	$Y - mX$	Deming Reg.	circular median of doubled slope angle

Since Knight's algorithm has been implemented in many statistical packages and these also take care of the handling of ties, it is an easy task to use these algorithms to calculate the PBR and circular mean estimators. Even in combination with a simple bisection algorithm for root finding it thus becomes possible to find the estimators for data sets with $n = 10^7$ on an ordinary laptop in a few minutes.

3 Confidence interval of the slope estimate

Passing and Bablok [6] gave arguments to justify a formula for the confidence intervals of the estimated slope. They noted that, in general, the variance of their slope estimator should depend on the distribution of the measured values. Numerical simulations showed that, in practice, their formula led to acceptable coverages. Nevertheless, they concluded that “the empirical derivation of this statement might seem unsatisfactory”. Their estimator can be justified by deriving the confidence intervals for the slope from that of Kendall's τ .

Writing $\tau(m) = \tau(X''(m), Y''(m))$, where X'' and Y'' may be any of the expressions listed in Table 1, the estimator for the slope $\hat{m}$ is the solution of the equation $\tau(\hat{m}) = 0$. Using the delta method,

$$\hat{\sigma}_{\hat{m}} \approx \hat{\sigma}_{\tau} \frac{dm}{d\tau} \approx \frac{\tau^{-1}(z_{1-\alpha/2}\sigma_{\tau}) - \tau^{-1}(-z_{1-\alpha/2}\sigma_{\tau})}{2z_{1-\alpha/2}}, \quad (9)$$

where in the last step, the differential quotient was regularized with a McKean–Schrader type finite difference approximation [24]. $z_{1-\alpha/2}$ is the quantile of level $\alpha/2$ of the normal distribution.

Hence, an asymptotic confidence bound for $\hat{m}$ is

$$\hat{m}_{u/l} = \tau^{-1}(\pm z_{1-\alpha/2}\hat{\sigma}_{\tau}).$$

If m is the true value of the slope, the variance of $\tau(m)$ is

$$\text{Var}(\tau(m)) = \sigma_{\tau}^2 = \frac{2(2n+5)}{9n(n-1)}, \quad (10)$$

if x_i'' and y_i'' are statistically independent, and all the y_i'' are from the same distribution. With these assumptions, the estimates for the slope's confidence intervals coincide with the result of Passing and Bablok.

However, these assumptions are quite restrictive, since they are only fulfilled if the original η_i and ξ_i are homoscedastic and from a Gaussian distribution. If the errors are small compared to the Δx_{ij}^* or when $\gamma = m$, a symmetric distribution of the errors will be sufficient, but heteroscedasticity is still an issue. In the Theil–Sen setting, bootstrap, and jackknife techniques have been recommended [25] as an alternative for the analytical confidence interval estimators given by Theil [7].

Alternatively, it is possible to derive analytical expressions for the confidence intervals of the PBR estimators without making restrictive assumptions about the distribution of the X and Y , using an expression given by Daniels and Kendall and Cliff [26] and Charlin [27] for the variance,

$$\widehat{\text{Var}}(\tau(m)) = \frac{n(n-1)\sum_j (\sum_i \tau_{ij})^2 - 1}{(n-2)(n-3)}, \quad (11)$$

where $\tau_{ij} = (n(n-1)/2) \text{sgn}(x_i'' - x_j'') \text{sgn}(y_i'' - y_j'')$. The estimator depends explicitly on the measured values. To evaluate this estimator efficiently, the order of the i, j is chosen so that, if $x_i'' < x_j''$, then $i < j$. Then, $\sum_j \tau_{ij}$ can be seen to be proportional to the difference of concordant and discordant pairs y_i'', y_j'' for all j . This number may be evaluated for all values of i in one sorting step of numerical complexity $O(n \ln n)$:

1. Initialize a vector q of length n to 0.
2. The vector Y'' , with elements in order of the i , is then ordered by decreasing size of the y_i'' using mergesort [22] where, during the merging of any two sub-lists, a counter is kept of the number of elements already taken from the left list.
3. When during merging an element y_i'' from the right list is selected, q_i is increased by the actual value of the counter.
4. When the complete list is sorted, then $\sum_j \tau_{ij} = 2(n - 2q_i)/n(n+1)$.

For homoscedastic Gaussian errors, this variance estimator reduces asymptotically to Eq. (10).

A numerical comparison (cf. the Section 4) shows that the variance calculated with Eq. (11) for heteroscedastic, Gaussian errors with constant coefficient of variation is only 9.4 % greater than the variance for homoscedastic error and the corresponding difference in the standard deviation is 4.6 %. For moderate sample sizes, this difference will be below the statistical uncertainty of Eq. (11) and probably justifies the use of the simpler expression used by Passing and Bablok in most practical situations. Nevertheless Daniels and Kendall [26] derived a tight upper bound $2/n$ for the variance from Eq. (11) which, asymptotically for large n , is 4.5 times larger than the variance from Eq. (10).

Finally, in the context of estimating the variance of $\hat{m}$, the efficiency of the PBR estimation shall be addressed. The additional condition to the general assumptions made in Section 2.1 for the estimate Eq. (10) to hold, is that the x_i^*, y_i^*, η_i and ξ_i follow a normal distribution. Under this condition, the “twin” method LPR of the ePBR, which is derived here from the condition of Pearson correlation diminishing to 0, can also be derived from maximization of likelihood [28] and therefore has maximal asymptotic efficiency. The relative asymptotic efficiency of ePBR is then expressible as the ratio $\text{Var}(\hat{m}_{\text{LPR}})/\text{Var}(\hat{m}_{\text{ePBR}})$ of the two estimators [8]. The standard deviation of the LPR estimate can be calculated as the corresponding standard deviation of the ePBR,

$$\sigma_{\hat{m}_{\text{LPR}}} = \sigma_r \frac{dm}{dr},$$

where $r(X''(m), Y''(m))$ is Pearson’s coefficient of correlation and σ_r its standard deviation.

Asymptotically [26], $\sigma_r = 1/\sqrt{n}$. Also, for normally distributed X and Y ,

$$r = \sin(\pi\tau/2),$$

which allows relating the derivatives of $\hat{m}$ with respect to r and τ . This yields the relative asymptotic efficiency

$$\frac{\text{Var}(\hat{m}_{\text{LPR}})}{\text{Var}(\hat{m}_{\text{ePBR}})} \sim \frac{9}{\pi^2} \approx 0.91.$$

So, at least for normally distributed measurements and errors, the asymptotic relative efficiency will be nearly optimal. The same relative asymptotic efficiency is also found for Theil–Sen regression [8].

4 Confidence intervals for the intercept and bias

The estimators for the CI of the intercept proposed by Passing–Bablok

$$\hat{b}^{\text{low/up}} = \text{median}\{y_i - \hat{m}^{\text{up/low}} x_i | i = 1, \dots, n\}$$

do not seem acceptable, since with a shift of the origin of the variable x , the two limits can be made to coincide, so that the CI has length 0. One would merely expect the width of the CI to run through a minimum near the center of gravity of the x_i , as in simple linear regression.

Passing and Bablok propose the following estimator for the intercept:

$$\hat{b} = \text{median}\{y_i - \hat{m}x_i | i = 1, \dots, n\},$$

or with,

$$\zeta(b, \hat{m}) = \sum_i \text{sgn}(y_i - \hat{m}x_i - b) = n(2F_{Y''}(b) - 1)$$

as a solution of

$$\zeta(\hat{b}, \hat{m}) = 0 \text{ or } F_{Y''}(\hat{b}) = \frac{1}{2}. \quad (12)$$

The intercept b is the bias of y at $x = 0$. Here, $F_{Y''}$ is the empirical cumulative distribution function of the $Y'' = Y - \hat{m}X$. Neglecting a small possible dependence of the y''_i for different i due to the estimation of $\hat{m}$, the variance of $\zeta(b, \hat{m})$ for fixed b is $\widehat{\text{Var}}\zeta = \hat{\sigma}_\zeta^2 = n$ and that of $F_{Y''}(b, \hat{m})$ is $\widehat{\text{Var}}(F_{Y''}) = 1/2n$.

In clinical chemistry, the bias at the medical decision point x_c ,

$$\hat{B}_c = y_c - x_c = \hat{b} + (\hat{m} - 1)x_c,$$

is of special interest.

For the calculation of its standard error, the delta method is proposed

$$\widehat{\text{Var}}(\hat{B}_c) = (1x_c)Q^{-1}V(Q^{-1})^T(1x_c)^T,$$

with V being the covariance matrix of $\zeta(b, \hat{m})$ and $\tau(m)$. It has entries $V_{11} = \sigma_\zeta^2$, $V_{22} = \hat{\sigma}_\tau^2$ (cf. Eq. (10)) and

$$V_{12} = V_{21} = \widehat{\text{cov}}(\zeta, \tau) = \sum_{i < j} (\text{sgn}(y''_i) + \text{sgn}(y''_j))\tau_{ij} = : \widehat{\text{cor}}(\zeta, \tau)\hat{\sigma}_\zeta\hat{\sigma}_\tau.$$

This expression for the covariance can be derived using the fact that both ζ and τ are U-statistics [29]. It can be seen to quantify heteroscedasticity of the data and can also be expressed in terms of Kendall correlation coefficients,

$$\widehat{\text{cov}}(\zeta, \tau) = 2 \frac{\binom{n_+}{2}\tau_+ - \binom{n_-}{2}\tau_-}{\binom{n}{2}},$$

where $n_\pm$ are the numbers of positive and negative y''_i , respectively, and $\tau_\pm = \sum_{i < j} \tau_{ij}$ where either the sum is extended over those i , for which either $y''_i > 0$ or $y''_i < 0$, respectively.

The matrix Q is the Jacobian

$$Q = \begin{pmatrix} \frac{\partial \zeta}{\partial \hat{b}} \Big|_{\hat{m}} & \frac{\partial \zeta}{\partial \hat{m}} \Big|_{\hat{b}} \\ 0 & \frac{d\tau}{d\hat{m}} \end{pmatrix}, \text{ with inverse } Q^{-1} = \begin{pmatrix} \frac{\partial \hat{b}}{\partial \zeta} \Big|_{\hat{m}} & -\frac{d\hat{m}}{d\tau} x_0 \\ 0 & \frac{d\hat{m}}{d\tau} \end{pmatrix}, \quad (13)$$

where $x_0 := \frac{\partial \zeta}{\partial \hat{m}} \Big|_{\hat{b}} \frac{\partial \hat{b}}{\partial \zeta} \Big|_{\hat{m}} = -\frac{\partial \hat{b}}{\partial \hat{m}} \Big|_{\zeta=0}$.

For all appearing derivatives, McKean–Schradler type regularized approximations are used:

– For $\hat{\sigma}_\tau d\hat{m}/d\tau = \hat{\sigma}_{\hat{m}}$, the expression derived in the previous section.

Figure 3: A: A set of 200 points with $x^* = y^*$ from a uniform distribution in $(0, 1000)$ with a constant coefficient of variation $cv_x = cv_y = 0.1$. B: grey: 1000 simulated values of the bias $B_c^{\min}$ as a function of the cutoff at medical decision point x_c . Dashed: standard deviation of the simulated bias values. Dotted: standard deviation of the bias calculated from the means of $\widehat{\text{Var}}(\hat{m})$, $\widehat{\text{Var}}(\hat{B}_c^{\min})$, and of $x_c^{\min}$ from the simulations. The analytical expression not only approximates the width of the confidence intervals from simulation very well, but also reproduces the position where the variance attains its minimum.

Table 2: Comparison of standard deviations calculated with the formulas derived in the text to simulated values, and of the coverage of the bias B_c for values of the medical decision point $x_c \in (0, 250, 500, 750, 1000)$. The coverage of the slope and bias are generally good, only near $x_c^{\min}$ the coverage is rather conservative.

parameter	Theoretical	Simulation
$sd(\hat{m})$	0.0179	0.0178
$sd(\hat{b})$	3.44	3.21
$x_c^{\min}$	123	127
coverage m	95%	94.4%
coverage $b = B_0$	95%	97.4%
coverage B_{250}	95%	94.9%
coverage B_{500}	95%	94.0%
coverage B_{750}	95%	94.0%
coverage B_{1000}	95%	94.4%

$$\hat{\sigma}_{\hat{b}} = \left. \frac{\partial \hat{b}}{\partial \zeta} \right|_{\hat{m}} \approx \frac{F_{Y''}^{-1}\left(\frac{1}{2} + z_{1-\alpha/2} \hat{\sigma}_{F_{Y''}}\right) - F_{Y''}^{-1}\left(\frac{1}{2} - z_{1-\alpha/2} \hat{\sigma}_{F_{Y''}}\right)}{2z_{1-\alpha/2}}.$$

$$x_0 = -\left(\frac{\partial \hat{b}}{\partial \hat{m}}\right)_{\zeta=0} \approx -\frac{\hat{b}(\hat{m} + z_{1-\alpha/2} \hat{\sigma}_{\hat{m}}) - \hat{b}(\hat{m} - z_{1-\alpha/2} \hat{\sigma}_{\hat{m}})}{2z_{1-\alpha/2} \hat{\sigma}_{\hat{m}}}.$$

Thus the final expression for the variance of the bias at the medical decision points becomes

$$\widehat{\text{Var}}(\hat{B}_c) = \widehat{\text{Var}}(\hat{B}_c^{\min}) + \hat{\sigma}_{\hat{m}}^2 (x_c - x_c^{\min})^2.$$

Here, the variance is minimal for $x_c^{\min} = x_0 - \hat{\sigma}_{\hat{b}} / \hat{\sigma}_{\hat{m}} \widehat{\text{cor}}(\zeta, \tau)$ where it attains the value

$$\widehat{\text{Var}}(\hat{B}_c^{\min}) = \hat{\sigma}_{\hat{b}}^2 (1 - \widehat{\text{cor}}(\zeta, \tau)^2).$$

Hence, confidence intervals for the bias at medical decision point can be obtained as

$$\hat{B}_c^{\text{up/low}} = \hat{B}_c \pm z_{1-\alpha/2} \sqrt{\widehat{\text{Var}}(\hat{B}_c)}.$$

The method has been tested in numerical simulations. Results are shown (Figure 3) for a uniform distribution of 200 “true” $x^* = y^*$ values (i.e., $m = 1$ and $b = 0$) in the range (0, 1000) with a Gaussian error in x and y with constant coefficient of variance $cv_x = cv_y = 0.1$. The simulation was repeated 1000 times to estimate the standard deviations of the estimated slope, intercept and bias at medical decision point, as well as the coverage of the true parameter values, cf. Table 2. Furthermore, in Figure 3, the simulated estimates of the bias are shown together with their standard deviation and the theoretical curve for the standard derivation derived from the mean values of $\widehat{\text{Var}}(\hat{m})$, $\widehat{\text{Var}}(\hat{B}_c^{\text{min}})$, and χ_c^{min} .

5 Discussion

Most classical regression methods – including the ones which allow for imprecision in both variables, such as Deming or geometric mean regression – can be derived from the condition that Pearson’s correlation diminishes to 0 for a specific scaling and rotation of the data. The slope of interest is thereby a parameter of the scaling or rotation. It has been shown that a replacement of Pearson’s correlation by Kendall’s tau yields interesting robust regression estimates. While this is well known for TSR, the argument can be extended to PBR; the ePBR [10] can be derived this way, and it has been shown to be the analog of geometric mean regression. It is easier to calculate than the cPBR estimate, since it is merely the median of the absolute values of the slope $\hat{m} = \text{median}(|S_{ij}|)$, and it has the advantage that the estimator is equivariant under a scaling of the axes, while the cPBR estimator could only be determined for slopes approximately equal to 1.

The cPBR estimator appears as an approximation to the ePBR and has been shown to be the equivalent of Roos regression. Therefore, use of the ePBR seems to be preferable over that of the cPBR. Furthermore, as the PBR methods are derivable from Kendall correlation, it would be natural to report also Kendall’s τ as a measure of correlation rather than Pearson’s r .

The corresponding regression method to Deming regression is shown to be equivalent to a standard problem in circular statistics, namely finding the circular median of the doubled slope angles. While this is a standard method in fields like geology or ecology, with proven robustness and un-biasedness, it has apparently not been applied in clinical chemistry. Consistency of this estimator has here only been proven for errors following a normal distribution. The effect of deviations from normality on the bias of the estimator needs to be evaluated in simulation studies; nevertheless it is expected to be lower than in Pearson regression. The method may be especially then a viable alternative to PBR, when the standard deviations are known and their ratio is very different from the expected slope.

Since the PBR estimator is derived from Kendall’s tau statistics, which is a member of the class of U-statistics, it has furthermore been possible to derive an analytical expression for the variance of the intercept. This is of special importance in clinical chemistry, where the intercept which is obtained after shifting the origin of the coordinate axes to the medical decision point is known as bias of the medical decision point.

Finally, the insight that the PBR estimator can be obtained as a root of Kendall’s τ yields a fast algorithm of numerical complexity $O(n \ln n)$, as compared to $O(n^2)$ for the original shifted median algorithm, so that PBR estimation also becomes possible for larger data sets. It is planned to include the algorithms described in a future version of the R [30] package “MCR” [31] to be deposited in the CRAN repository.

Acknowledgment: My sincere thanks are given to Dr. Terry Therneau for a thorough review and many comments which greatly helped to improve the paper. It was an honour to talk to Dr. Heinrich Passing and the family of the late Wolfgang Bablok about the collaboration which lead to the invention of the method which now bears their names. I am indebted to my dear colleagues, especially Ge Guo, Kendra Jones, Ekaterina Manuilova, Gabrielle Miles, and Crystal Williams, for valuable discussions and proofreading.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: None declared.

Conflict of interest statement: The author declares no conflicts of interest regarding this article.

References

1. Ludbrook J. Linear regression analysis for comparing two measurers or methods of measurement: but which regression? *Clin Exp Pharmacol Physiol* 2010;37:692–9.
2. Linnet K. Estimation of the linear relationship between the measurements of two methods with proportional errors. *Stat Med* 1990;9:1463–73.
3. Linnet K. Evaluation of regression procedures for methods comparison studies. *Clin Chem* 1993;39:424–32.
4. Ludbrook J. Special article comparing methods of measurement. *Clin Exp Pharmacol Physiol* 1997;24:193–203.
5. CLSI. Measurement procedure comparison and bias estimation using patient samples. CLSI guideline EP09c. Wayne, PA: Clinical and Laboratory Standards Institute; 2018.
6. Passing H, Bablok W. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, part I. *Clin Chem Lab Med* 1983;21:709–20.
7. Theil H. A rank-invariant method of linear and polynomial regression analysis. *Henri Theil's contributions to economics and econometrics*. Dordrecht: Springer; 1992:345–81 pp.
8. Sen PK. Estimates of the regression coefficient based on Kendall's tau. *J Am Stat Assoc* 1968;63:1379–89.
9. Passing H, Bablok W. Comparison of several regression procedures for method comparison studies and determination of sample sizes application of linear regression procedures for method comparison studies in clinical chemistry, part II. *Clin Chem Lab Med* 1984;22:431–45.
10. Bablok W, Passing H, Bender R, Schneider B. A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. *Clin Chem Lab Med* 1988;26: 783–90.
11. Roos CF. A general invariant criterion of fit for lines and planes where all variates are subject to error. *Metron* 1937;13:3–20.
12. Xu S. A property of geometric mean regression. *Am Stat* 2014;68:277–81.
13. Therneau T. Deming: deming, Theil–Sen and Passing–Bablok and total least squares regression. Rochester, Minnesota, USA: Mayo clinic; 2018, R package version 1.4. Available from: <https://CRAN.R-project.org/package=deming>.
14. Clarke M. The reduced major axis of a bivariate sample. *Biometrika* 1980;67:441–6.
15. Rousseeuw PJ, Leroy AM. Robust regression and outlier detection. Hoboken, New Jersey: John Wiley & Sons; 2003, Vol 516.
16. Jammalamadaka SR, Sengupta A Topics in circular statistics, Vol 5. Singapore: World Scientific; 2001.
17. Mardia KV, Jupp PE. Directional statistics. Hoboken, New Jersey: John Wiley & Sons; 2009, Vol 494.
18. He X, Simpson DG. Robust direction estimation. *Ann Stat* 1992;351–69. <https://doi.org/10.1214/aos/1176348526>.
19. Fisher N, Lunn AD, Davies SJ. Spherical median axes. *J Royal Stat Soc, Ser B* 1993;55:117–24.
20. Matoušek J. Randomized optimal algorithm for slope selection. *Inf Process Lett* 1991;39:183–7.
21. Knight WR. A computer method for calculating Kendall's tau with ungrouped data. *J Am Stat Assoc* 1966;61:436–9.
22. Goldstine HH, Von Neumann J. Planning and coding of problems for an electronic computing instrument; part II, Technical report. Princeton, NJ: Institute for Advanced Study; 1947, Vol II.
23. Ballester P Hough transform for robust regression and automated detection. *Astron Astrophys* 1994;286:1011–1018.
24. McKean JW, Schrader RM. A comparison of methods for studentizing the sample median. *Commun Stat - Simul Comput* 1984; 13:751–73.
25. Newson R. Confidence intervals for rank statistics: percentile slopes, differences, and ratios. *Stata J* 2006;6:497–520.
26. Daniels HE, Kendall MG. The significance of rank correlations where parental correlation exists. *Biometrika* 1947;34:197–208.
27. Cliff N, Charlín V. Variances and covariances of Kendall's tau and their estimation. *Multivar Behav Res* 1991;26:693–707, PMID: 26751027.
28. Fuller WA. Measurement error models. New York: John Wiley & Sons; 2009, Vol 305.
29. Lee AJ. U-statistics: theory and practice, statistics: a series of textbooks and monographs. Oxfordshire, United Kingdom: Taylor & Francis; 1990, Vol 110.
30. R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: <https://www.R-project.org/>.
31. Manuilova E, Schuetzenmeister A, Model F. MCR: method comparison regression; 2014, R package version 1.2.1. Available from: <https://CRAN.R-project.org/package=mcr>.