Norman Wen Xuan Koh, Corey Markus, Tze Ping Loh* and Chun Yee Lim

Comparison of six regression-based lot-to-lot verification approaches

https://doi.org/10.1515/cclm-2022-0274
Received March 22, 2022; accepted April 29, 2022;
published online May 16, 2022

Abstract

Objectives: Detection of between-lot reagent bias is clinically important and can be assessed by application of regression-based statistics on several paired measurements obtained from the existing and new candidate lot. Here, the bias detection capability of six regression-based lot-to-lot reagent verification assessments, including an extension of the Bland–Altman with regression approach are compared.

Methods: Least squares and Deming regression (in both weighted and unweighted forms), confidence ellipses and Bland–Altman with regression (BA-R) approaches were investigated. The numerical simulation included permutations of the following parameters: differing result range ratios (upper:lower measurement limits), levels of significance (alpha), constant and proportional biases, analytical coefficients of variation (CV), and numbers of replicates and sample sizes. The sample concentrations simulated were drawn from a uniformly distributed concentration range.

Results: At a low range ratio (1:10, CV 3%), the BA-R performed the best, albeit with a higher false rejection rate and closely followed by weighted regression approaches. At larger range ratios (1:1,000, CV 3%), the BA-R performed poorly and weighted regression approaches performed the best. At higher assay imprecision (CV 10%), all six approaches performed poorly with bias detection rates <50%. A lower alpha reduced the false rejection rate, while greater sample numbers and replicates improved bias detection.

Conclusions: When performing reagent lot verification, laboratories need to finely balance the false rejection rate (selecting an appropriate alpha) with the power of bias detection (appropriate statistical approach to match assay performance characteristics) and operational considerations (number of clinical samples and replicates, not having alternate reagent lot).

Keywords: between-reagent lot; bias; drift; reagent lot; shift.

***Corresponding author: Tze Ping Loh**, Department of Laboratory Medicine, National University Hospital, 5 Lower Kent Ridge Road, 119074 Singapore, Singapore, Phone: (+65) 67724345, Fax: (+65) 67771613, E-mail: tploh@hotmail.com

Norman Wen Xuan Koh and Chun Yee Lim, Engineering Cluster, Singapore Institute of Technology, Singapore, Singapore

Corey Markus, Department of Chemical Pathology, New South Wales Health Pathology, Prince of Wales Hospital, Sydney, Australia.
https://orcid.org/0000-0002-5594-9737

Introduction

To guide clinical decision making, quantitative laboratory results are often compared against fixed clinical decision limits, reference intervals or are monitored longitudinally and spurious shifts in patient results adversely compromising these clinical decisions [1, 2]. Due to differing manufacturing, storage and shipping conditions, a reagent lot change has the potential to alter the analytical performance of any measurement procedure. The between-reagent lot verification process aims to ensure that the performance of a new (candidate) reagent lot has not altered to such an extent that it will cause a significant shift in patient results in comparison to the previous reagent lot. There are implementation challenges with all lot-to-lot evaluation processes and there have been several recent reports of the failure of routine between-reagent lot verification processes to detect clinically significant analytical shifts, leading to renewed interest and consideration of how the performance of these procedures can be improved by laboratories [1, 3–5].

The Clinical and Laboratory Standards Institute (CLSI) has a well-defined reagent lot verification procedure (EP-26A) [6]. Under this procedure, the means of several concentrations of patient samples, measured in parallel using the existing and candidate reagent lots are statistically compared, with samples at clinically important concentrations ideally evaluated. If the mean difference between patient samples exceeds a predefined threshold, the verification assessment is considered as failed. While this approach is relatively simple to perform, it only detects significant changes at the selected concentrations evaluated. It does not provide information about the nature

of the shift across the entirety of the analytical measurement range. Hence, it may not inform the laboratory if any shift beyond the evaluated concentrations is present and, even if so, whether it is within acceptable limits.

An alternative approach to between-reagent lot verification involves the evaluation of the regression coefficients (slope and intercept) between the existing and candidate reagent lots. Under this approach, a set of patient samples is measured in parallel using the existing and candidate reagent lots [1, 7, 8]. Twenty to thirty patient samples are typically recommended for evaluation, but this is often not achievable within the resource constraints of a routine laboratory. Additionally, a greater number of samples may be required in the presence of small range ratios and the desired statistical power [8]. A range ratio is the ratio of the upper to lower limits of the analytical measurement range. With regression analysis, the characteristics of the shift, i.e. proportional and constant biases, over the entire analytical measurement range can be evaluated.

Currently, there is no definitive guidance on how the rejection limits for the regression coefficients should be determined. Hence, these limits are often arbitrarily set in practice, instead of having well-defined analytical performance specifications [9]. For example, the Clinical Laboratory Improvement Amendments (CLIA) 1988 guidelines recommend a total allowable error of 10% for glucose [6]. Assuming an analytical variation (coefficient of variation, CV) of 3%, the allowable bias (allowable bias=total allowable error – 1.65 imprecision, all expressed in percentages) can be computed as 10% – 1.65 × 3%=5%. Hence, Linnet [8] proposed rejecting candidate lots when the slope deviates from 1.0 by more 5% (i.e. less than 0.95 or greater than 1.05). For example, the clinically important fasting plasma glucose concentration of 7.0 mmol/L is used to diagnose diabetes. If the absolute intercept coefficient deviates from zero by greater than 5% × 7 mmol/L=0.35 mmol/L, the proposed candidate reagent lot should be rejected. A limitation of the existing rejection limits is that these evaluate proportional and constant shifts in isolation.

However, proportional and constant shifts can be present simultaneously and their effects may be cumulative. To understand why this can be an issue, consider the glucose example described in the previous paragraph. Based on the rejection limits proposed by Linnet [8], there is a reasonable probability that a candidate lot with a proportional shift of 5% and a constant shift of 0.35 mmol/L be accepted. At the medical decision limit of 7.0 mmol/L, the combined bias of the candidate lot is $5\% + \frac{0.35}{7} \times 100\% = 10\%$, which is within the allowable bias of 10% based on the CLIA 1988 guidelines.

The joint confidence region is a generalization of confidence intervals when more than one regression parameter is evaluated simultaneously, and also considers the correlation between these parameters [10]. The joint confidence region assumes the shape of an ellipse when two parameters of the regression line (i.e. slope and intercept) are considered concurrently. Recently, the joint confidence ellipse approach has been applied to detect the bias between two clinical laboratory methods [11]. The rejection rule is based on whether the elliptical area encloses the point of slope=1.0 and intercept=0. Under this approach, proportional and constant shifts are considered simultaneously and improve the power of between-lot verification exercises.

In this simulation study, we propose an extension of the approach based on Bland–Altman analysis that jointly considers proportional and constant shifts. In particular, we propose an acceptance/rejection rule based on whether the confidence band over the analytical measuring range derived from linear regression modelling, encloses zero difference from Bland–Altman analysis. We compared the performance of this proposed approach against five other regression-based approaches used for lot-to-lot assessments.

Methods

Linear regression assumptions

Let X_i and Y_i denote the true value of sample i under reagent lots X (existing) and Y (candidate), respectively. Furthermore, let:

$$Y_i = (1 + \Delta_p)X_i + \Delta_c \quad (1)$$

where Δ_p and Δ_c are proportional and constant biases, respectively, and which describe the difference between the measurements under the two reagent lots. Let x_i and y_i denote the measured value of sample i under reagent lots X and Y, respectively:

$$x_i = X_i + \epsilon_i \quad (2)$$
$$y_i = Y_i + \delta_i \quad (3)$$

where $\varepsilon_i \sim N(0, (X_i \cdot \mathrm{CV})^2)$ and $\delta_i \sim N(0, (Y_i \cdot \mathrm{CV})^2)$ represent the random errors for X_i and Y_i respectively, which are normally distributed with mean of zero and standard deviation equal to the product of X_i and Y_i with CV, and CV is the analytical coefficient of variation for the measurement procedure.

Conventional linear regression approaches

Four conventional linear regression approaches, namely ordinary least square regression (OLS), weighted least squares regression (WLS), Deming regression (DR) and Weighted Deming regression

(WDR) are applied on the two sets of measurements to determine the estimated slope b and intercept a. The implementations for these linear regression approaches follow the standard formulations as described by Linnet [8]. The weight for WLS is set to $w_i = 1/x_i^2$ whereas the weight for WDR is defined as $w_i = 1/[(x_i + y_i)/2]^2$.

Bias between the two reagent lots is considered present if the estimated slope differs significantly from 1.0, or if the intercept deviates significantly from 0. The statistical significance of the deviation is determined based on two-tailed t-tests with a null hypothesis stating that the slope or intercept coefficients are equal to 1.0 or 0 respectively by:

$$t_b = (b - 1)/\mathrm{SE}(b) \quad (4)$$

$$t_a = (a - 0)/\mathrm{SE}(a) \quad (5)$$

where SE(b) and SE(a) are the standard errors for the estimated slope and intercept, respectively. The standard errors for the slope and intercept can be estimated by simple formulae [8] on the assumption that the slope is close to unity and the standard deviations of the measurements for the two reagent lots are approximately equal. The critical t-value for rejection of the null hypothesis, which lead to bias detection is dependent on the choice of significance level alpha (α). Three variations of between-lot rejection criteria were investigated, namely rejection based on slope or intercept in isolation and rejection based on either slope or intercept.

Joint parameter confidence ellipse approach

The joint parameter confidence ellipse (CE) approach proposed by Sadler [11] was also implemented for the between-lot bias detection, by jointly evaluating the slope and intercept of the regression line. Briefly, the slope and intercept are first obtained from OLS and WLS regression methods. The respective CE for OLS and WLS are then defined based on the formulation in Sadler [11] for the selected significance level α. Bias is detected when the CE does not enclose the point where slope=1.0, intercept=0.

Bland–Altman analysis with regression

Here, we propose an alternative approach to achieve joint parameter detection of between-lot bias by considering Bland–Altman (BA) analysis. The horizontal axis u_i and the vertical axis v_i of the analysis are defined as:

$$u_i = \frac{x_i + y_i}{2} \quad (6)$$

$$v_i = \frac{y_i - x_i}{0.5(y_i + x_i)} \quad (7)$$

where u_i is the average of the individual measurements from the two reagent lots while v_i represents the proportional difference between the individual measurements from the two lots, scaled by their average.

An OLS regression model is then determined on scaled differences v_i predicted by the average of individual differences from BA analysis (i.e. v_i=a + bu_i). In the absence of any between-lot bias, the regression line is expected to be a horizontal line which passes through zero difference (i.e. slope and intercept of the line are both zero). A confidence band is defined around the regression line-of-best fit, with the upper band, $v_\alpha^{\mathrm{upper}}$ and lower band $v_\alpha^{\mathrm{lower}}$ given by the equations below [12]:

$$v_\alpha^{\mathrm{upper}}(u_i) = a + bu_i + t_{\alpha/2} \cdot \sqrt{\frac{\sum_{i=1}^{n} (v_i - \hat{v}_i)^2}{n-2}} \cdot \sqrt{\left(\frac{1}{n} + \frac{(u_i - \overline{u})^2}{\sum_{i=1}^{n} (u_i - \overline{u})^2}\right)} \quad (8)$$

$$v_\alpha^{\mathrm{lower}}(u_i) = a + bu_i - t_{\alpha/2} \cdot \sqrt{\frac{\sum_{i=1}^{n} (v_i - \hat{v}_i)^2}{n-2}} \cdot \sqrt{\left(\frac{1}{n} + \frac{(u_i - \overline{u})^2}{\sum_{i=1}^{n} (u_i - \overline{u})^2}\right)} \quad (9)$$

where n is the number of pairs of measurements, $t_{\alpha/2}$ is the critical t-value for (n – 2) degrees of freedom, v_i is the value of v predicted by the regression line for the corresponding u_i and $\overline{u}$ is the mean of u_i values.

If the confidence band includes zero over the entire range of interest, the difference in the measurements between the two lots is not statistically significant at the level of 1 – α, and the candidate reagent lot can be accepted (see Figure 1A). On the contrary, if the confidence band does not include zero at any point in the measurement range, then a between-lot bias is considered to be present. To assess the enclosure of zero line by the confidence band, the following conditions are inspected across the entire concentration range (i.e. u_i values):

- if the smallest $v_\alpha^{\mathrm{upper}}(u)$ is less than zero, then the upper confidence band will intersect the zero line (see Figure 1B).
- if the largest $v_\alpha^{\mathrm{lower}}(u)$ is greater than zero, then the lower confidence band will intersect the zero line (see Figure 1C).

Either of the two scenarios will indicate that the confidence band does not enclose zero difference and the presence of between-lot bias can be inferred. In addition, the presence of a very large positive or negative constant bias will also shift the entire band away from zero difference without enclosing it, violating one of the two conditions above (without any intersection with the zero line).

Numerical simulation

The numerical simulation included combinations of differing parameters: results range ratios of 1:10 and 1:1,000, analytical coefficients of variation of 3 and 10%, differing number of replicates and sample sizes, different alpha (level of significance) ranging from 2.5 to 10%, levels of constant bias (0.0, 0.05, 0.1, 0.2) and proportional biases of 0, 2, 5 and 8%. The sample concentrations simulated were uniformly distributed across the concentration range. A baseline scenario in the absence of any constant or proportional bias was also simulated to determine the false rejection rate. For each scenario, 10,000 rounds of simulation were performed, and the average proportion of simulations where bias was detected (i.e. probability of bias detection) were summarised and reported.

Results

At a lower range ratio of 1:10 and CV of 3%, the BA approach with regression had the highest bias detection rates for proportional, constant, and mixed biases. However, this higher detection rate is achieved at the expense of

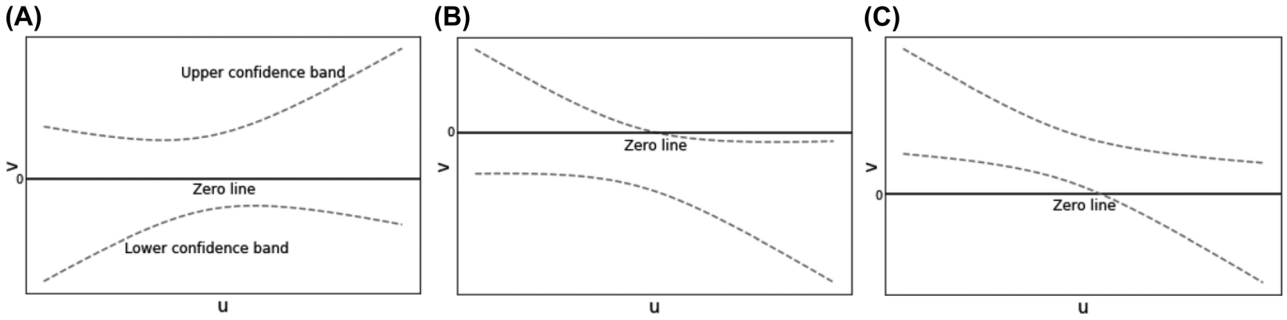

Figure 1: Confidence band in Bland–Altman plot showing (A) enclosure of zero difference, (B) intersection of upper confidence band with zero difference and (C) intersection of lower confidence band with zero difference.

a greater false positive rate that is 2.5 times more than the specified alpha (Table 1). The false positive rate for the BA approach with regression can be mitigated by using a smaller alpha e.g. α=0.025, while maintaining a high overall bias detection capability (Supplementary Tables 1 and 2). The CE with WLS approach performed second best overall, followed by WDR that were able to detect the largest bias with >80% probability. Of the two conventional regression approaches with and without weighting, the weighted regression models had better bias detection capability if slope and intercept were considered simultaneously for bias detection, when compared to the non-weighted versions.

At an increased range ratio of 1:1,000 and CV of 3%, the constant bias detection capability of the BA approach with regression deteriorated considerably, although it was still able to detect proportional biases (Table 2, Supplementary Table 3). In this scenario, WLS and WDR had overall best performance closely followed by the CE. The non-weighted regression approaches (OLS and DR) all performed relatively poorly.

From Tables 3 and 4, when the analytical imprecision was 10%, the bias detection capability of all six regression-based approaches deteriorated noticeably, particularly for detection of constant bias. All the regression-based methods generally had a power of <50% for detection of any bias. In this scenario, the BA approach with regression performed the best although still accompanied by a higher false positive rate.

Next, the BA approach with regression was further examined for the effects of differing analytical imprecision (CV=1, 3, 5, 10%) and range ratio (1.25, 2, 10, 100). In general, the bias detection capability deteriorated with increasing imprecision and range ratio (Table 5, Supplementary Table 4). The poorest detection capability was found at the combination of the highest range ratio and analytical imprecision. On the other hand, increasing the number of samples and the number of replicates improved the detection capability of the BA approach with regression (Table 6, Supplementary Table 5). Increasing the number of replicates had a larger impact on increasing the bias detection capability compared to increasing the number of samples, but also (mildly) increase the false rejection rates (Table 6).

Discussion

Recently between-lot reagent verification has received increasing attention [13] owing to some well publicised cases where clinically unfit reagent lots were made available for routine patient testing resulting in suboptimal patient management [1]. While multiple approaches have been used to verify new reagent lots, including a recent proposal by a European Federation of Clinical Chemistry and Laboratory Medicine working group [14], regression-based approach are still practiced in some laboratories.

Bland–Altman analysis is a familiar tool used in laboratory practice to assess the magnitude of a difference (analytical bias) between two different measurement procedures [15]. In this instance, Bland–Altman analysis is used to assess the bias between a current or existing reagent lot and a new candidate reagent lot. The application of a regression line across the differences between paired measurements produced by the two reagent lots indicates the direction and magnitude of bias. When a confidence band is constructed around the regression line of best fit, this can be used as a statistical assessment for the presence of bias. When the confidence band does not enclose zero difference throughout the analytical measuring range, this indicates the presence of statistically significant bias and considers both proportional and constant biases simultaneously.

Table 1: Comparison of bias detection rate between the six different regression approaches for range ratio=1:10. (Number of samples=10, range ratio=1:10, CV=3%, α=0.05) (OLS=ordinary least squares, WLS=weighted least squares).

α=0.05		Probability of rejection based on statistical approach														
Shift scenarios		OLS			WLS			Deming regression			Weighted deming regression			Confidence ellipse [11]		Bland–Altman plot
Proportional bias, Δ_p	Constant bias, Δ_c	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	OLS	WLS	OLS with confidence band
0	0	0.07	0	0.07	0.05	0.05	0.08	0.06	0	0.06	0.02	0.02	0.04	0.08	0.05	0.12
0	+0.05	0.07	0.01	0.07	0.05	0.14	0.17	0.06	0.01	0.06	0.02	0.09	0.10	0.10	0.16	0.30
0	+0.10	0.07	0.03	0.08	0.05	0.39	0.42	0.06	0.01	0.06	0.02	0.33	0.34	0.15	0.51	0.64
0	+0.20	0.07	0.15	0.18	0.05	0.91	0.91	0.06	0.09	0.12	0.02	0.92	0.92	0.44	0.98	0.98
2%	0	0.10	0	0.10	0.11	0.05	0.14	0.12	0	0.12	0.09	0.02	0.10	0.20	0.15	0.35
5%	0	0.31	0	0.32	0.50	0.05	0.51	0.38	0	0.39	0.54	0.03	0.57	0.70	0.70	0.91
8%	0	0.60	0	0.60	0.87	0.05	0.88	0.73	0	0.74	0.94	0.03	0.95	0.96	0.98	1
2%	+0.2	0.10	0.14	0.23	0.11	0.89	0.92	0.11	0.09	0.20	0.08	0.91	0.95	0.77	1	1
5%	+0.1	0.31	0.03	0.34	0.49	0.37	0.72	0.39	0.02	0.41	0.54	0.32	0.81	0.89	0.99	1
8%	+0.05	0.60	0.01	0.62	0.87	0.13	0.90	0.74	0.01	0.74	0.94	0.09	0.97	0.98	1	1

Table 2: Comparison of bias detection rate between six different regression approaches for range ratio=1:1,000. (Number of samples=10, number of replicates=1, range ratio=1:1,000, analytical imprecision (CV)=3%, α=0.05) (OLS=ordinary least squares, WLS=weighted least squares).

α=0.05		Probability of rejection based on statistical approach													
Shift scenarios		OLS			WLS			Deming regression			Weighted deming regression			Confidence ellipse [11]		Bland–Altman plot
Proportional bias, Δ_p	Constant bias, Δ_c	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	OLS	WLS	OLS with confidence band
0	0	0.08	0	0.07	0.05	0.05	0.10	0.08	0	0.08	0.02	0.02	0.04	0.09	0.05	0.12
0	+0.05	0.08	0	0.08	0.05	0.15	0.18	0.07	0	0.08	0.02	0.11	0.13	0.09	0.11	0.15
0	+0.10	0.08	0	0.08	0.05	0.44	0.47	0.08	0	0.08	0.02	0.43	0.44	0.09	0.32	0.19
0	+0.20	0.07	0	0.08	0.05	0.95	0.95	0.08	0	0.08	0.02	0.97	0.98	0.09	0.86	0.24
2%	0	0.14	0	0.13	0.19	0.05	0.23	0.17	0	0.16	0.17	0.02	0.20	0.21	0.15	0.35
5%	0	0.42	0	0.42	0.80	0.05	0.81	0.53	0	0.52	0.88	0.02	0.88	0.68	0.70	0.91
8%	0	0.72	0	0.72	0.99	0.05	0.99	0.86	0	0.86	1	0.03	1	0.95	0.98	1
2%	+0.2	0.14	0	0.13	0.19	0.94	0.95	0.17	0	0.16	0.19	0.97	0.98	0.21	0.93	0.55
5%	+0.1	0.41	0	0.42	0.80	0.42	0.88	0.52	0	0.53	0.87	0.41	0.95	0.70	0.92	0.95
8%	+0.05	0.73	0	0.73	0.99	0.13	0.99	0.87	0	0.87	1	0.11	1	0.95	0.99	1

Table 3: Comparison of bias detection rate between six different regression approaches for range ratio=1:10 and CV=10%. (Number of samples=10, number of replicates=1, range ratio=1:10, analytical imprecision (CV)=10%, α=0.05) (OLS=ordinary least squares, WLS=weighted least squares).

α=0.05		Probability of rejection based on statistical approach														
Shift scenarios		OLS			WLS			Deming regression			Weighted deming regression			Confidence ellipse [11]		Bland–Altman plot
Proportional bias, Δ_p	Constant bias, Δ_c	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	OLS	WLS	OLS with confidence band
0	0	0.08	0.01	0.08	0.08	0.05	0.11	0.07	0.02	0.13	0.02	0.02	0.04	0.08	0.07	0.13
0	+0.05	0.08	0.01	0.08	0.08	0.07	0.11	0.07	0.02	0.15	0.02	0.03	0.05	0.09	0.07	0.14
0	+0.10	0.09	0.01	0.08	0.08	0.1	0.13	0.06	0.02	0.14	0.02	0.04	0.06	0.08	0.08	0.18
0	+0.20	0.08	0.03	0.09	0.07	0.19	0.21	0.07	0.03	0.14	0.02	0.1	0.12	0.11	0.17	0.35
2%	0	0.07	0.01	0.07	0.06	0.05	0.09	0.08	0.03	0.16	0.03	0.03	0.04	0.08	0.05	0.15
5%	0	0.07	0.01	0.07	0.06	0.05	0.09	0.12	0.03	0.2	0.06	0.03	0.08	0.12	0.07	0.25
8%	0	0.08	0.01	0.08	0.09	0.05	0.12	0.17	0.03	0.27	0.13	0.03	0.15	0.21	0.14	0.41
2%	+0.2	0.07	0.03	0.08	0.05	0.19	0.21	0.08	0.03	0.16	0.03	0.11	0.13	0.13	0.21	0.46
5%	+0.1	0.06	0.01	0.07	0.05	0.09	0.13	0.12	0.03	0.2	0.06	0.04	0.1	0.17	0.16	0.43
8%	+0.05	0.08	0.01	0.09	0.09	0.06	0.14	0.18	0.03	0.28	0.13	0.03	0.15	0.25	0.19	0.51

Table 4: Comparison of bias detection rate between six different regression approaches for range ratio=1:1,000 and CV=10%. (Number of samples=10, number of replicates=1, range ratio=1:1,000, analytical imprecision (CV)=10%, α=0.05) (OLS=ordinary least squares, WLS=weighted least squares).

α=0.05		Probability of rejection based on statistical approach														
Shift scenarios		OLS			WLS			Deming regression			Weighted deming regression			Confidence ellipse [11]		Bland–Altman plot
Proportional bias, Δ_p	Constant bias, Δ_c	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	Rejection on slope	Rejection on intercept	Rejection on slope and intercept	OLS	WLS	OLS with confidence band
0	0	0.09	0	0.09	0.08	0.05	0.12	0.08	0.01	0.17	0.02	0.02	0.04	0.09	0.07	0.13
0	+0.05	0.09	0	0.09	0.08	0.04	0.12	0.09	0.01	0.17	0.02	0.03	0.05	0.1	0.06	0.13
0	+0.10	0.09	0	0.08	0.08	0.06	0.13	0.08	0.01	0.17	0.02	0.04	0.06	0.09	0.07	0.14
0	+0.20	0.09	0	0.09	0.08	0.14	0.2	0.09	0.01	0.16	0.02	0.12	0.15	0.1	0.11	0.15
2%	0	0.07	0	0.07	0.06	0.05	0.1	0.1	0.01	0.18	0.04	0.03	0.06	0.09	0.06	0.14
5%	0	0.08	0	0.08	0.08	0.05	0.11	0.15	0.02	0.25	0.11	0.02	0.13	0.13	0.07	0.24
8%	0	0.12	0	0.11	0.17	0.05	0.2	0.22	0.02	0.33	0.25	0.03	0.28	0.21	0.13	0.41
2%	+0.2	0.08	0	0.08	0.06	0.14	0.18	0.09	0.01	0.18	0.04	0.12	0.15	0.08	0.11	0.2
5%	+0.1	0.08	0	0.08	0.08	0.06	0.13	0.15	0.02	0.24	0.11	0.05	0.15	0.13	0.09	0.29
8%	+0.05	0.11	0	0.11	0.16	0.05	0.19	0.22	0.02	0.34	0.25	0.03	0.28	0.22	0.15	0.45

Table 5: Effect of range ratio and analytical imprecision (CV) on the bias detection capability of the Bland-Altman analysis with regression confidence band approach. (Number of samples=10, number of replicates=1, α=0.05).

α=0.05		Probability of rejection based on statistical approach															
Range ratio		1.25				2				10				100			
CV%		1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%	1%	3%	5%	10%
Proportional bias, Δ_p	Constant bias, Δ_c																
0	0	0.12	0.13	0.13	0.14	0.12	0.13	0.13	0.13	0.12	0.12	0.12	0.12	0.12	0.13	0.12	0.12
0	+0.05	1	0.87	0.51	0.23	1	0.71	0.38	0.19	0.90	0.30	0.18	0.14	0.30	0.16	0.15	0.13
0	+0.10	1	1	0.96	0.51	1	1	0.84	0.38	1	0.64	0.36	0.19	0.45	0.23	0.17	0.14
0	+0.20	1	1	1	0.95	1	1	1	0.83	1	0.98	0.76	0.34	0.63	0.36	0.25	0.16
2%	0	0.98	0.34	0.21	0.16	0.98	0.35	0.21	0.14	0.98	0.35	0.20	0.15	0.98	0.35	0.21	0.14
5%	0	1	0.91	0.57	0.25	1	0.92	0.56	0.24	1	0.91	0.57	0.24	1	0.90	0.57	0.24
8%	0	1	1	0.88	0.41	1	1	0.88	0.42	1	1	0.88	0.41	1	1	0.88	0.42
2%	+0.20	1	1	1	0.98	1	1	1	0.90	1	1	0.91	0.46	1	0.69	0.45	0.23
5%	+0.10	1	1	1	0.80	1	1	1	0.70	1	1	0.89	0.42	1	0.97	0.70	0.31
8%	+0.05	1	1	1	0.71	1	1	0.99	0.65	1	1	0.96	0.50	1	1	0.92	0.45

When compared to the other five regression-based lot-to-lot bias detection approaches, the BA with regression approach is a more sensitive approach when the range ratio is relatively small. This may offer laboratories an alternative regression-based approach for analytes with small range ratios where simple regression approaches may not be sensitive enough to detect small but critical biases [8]. The higher false positive rates compared to the other regression-based methods examined here, can be mitigated by selecting a smaller level of significance (alpha). However, in the setting of high range ratios and proportional biases the BA analysis with regression performs relatively poorly. A general guidance on the number of samples/replicates to include in the experiment design for the BA with regression approach can be found in Table 6, where practitioners can determine these parameters based on the desired statistical performance (probability of bias detection, false rejection rate) and the analytical characteristics (range ratio, imprecision profile) of the measurement procedure.

Both weighted version of OLS and DR performed better in the simulation conditions examined than the unweighted forms. The simultaneous consideration of both slope and intercept improved the detection capability compared to consideration of slope or intercept alone. The use of statistical null hypothesis with *a priori* defined alpha value (level of significance) helps the laboratory to balance the operational needs to avoid inappropriate reagent lot rejection, as well as the clinical risk of accepting an unfit reagent lot [16]. This also avoids the arbitrary setting of rejection criteria (e.g. slope of ±10%) that does not provide the laboratory with any indication of their likely performance in terms of rates for bias detection or false rejections.

Nonetheless, all regression-based approaches investigated here did not provide satisfactory power for bias detection in the presence of high analytical imprecision (10%). For such scenarios, a greater number of replicates or number of clinical samples may be used to improve the power. However, this may not necessarily be a practical solution for smaller laboratories. Instead, a networked-approach, where a network of laboratories using the same measurement procedure and reagent lots perform the reagent lot verification on a small number of clinical samples and pool their data for analysis to achieve higher statistical power of bias detection should be considered [16, 17].

Although it might appear logical to apply a weighted regression approach to determine the regression line and confidence band in the BA method to improve performance, the confidence band for WLS regression line is expected to be less sensitive for bias detection. This is because the width of the confidence band is proportional to the difference between the u value and its weighted mean. The higher weightages on the smaller u values will shift the weighted mean towards the lower end of the u-axis. This will cause the width of the confidence band to be enlarged drastically for larger u values (see Supplementary Figure 1), thus deteriorating bias detection capability.

Table 6: Effect of number of samples and number of replicates on the bias detection capability of the Bland–Altman analysis with regression confidence band approach. (CV=10%, α=0.05).

α=0.05		Probability of rejection based on statistical approach																							
Range ratio		1:10												1:1,000											
Sample size		5			10			20			40			5			10			20			40		
Replicates		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Proportional bias, Δ_p	Constant bias, Δ_c																								
0	0	0.10	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.14	0.13	0.11	0.12	0.13	0.12	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14
0	+0.05	0.11	0.14	0.17	0.14	0.17	0.19	0.16	0.20	0.24	0.19	0.27	0.33	0.10	0.13	0.15	0.12	0.14	0.15	0.12	0.14	0.14	0.14	0.14	0.14
0	+0.10	0.13	0.22	0.28	0.18	0.28	0.36	0.26	0.39	0.52	0.38	0.58	0.74	0.11	0.17	0.21	0.13	0.16	0.19	0.14	0.15	0.17	0.14	0.15	0.16
0	+0.20	0.20	0.44	0.61	0.35	0.62	0.79	0.55	0.84	0.95	0.81	0.98	1	0.14	0.26	0.39	0.15	0.24	0.31	0.15	0.20	0.23	0.14	0.17	0.21
2%	0	0.11	0.14	0.16	0.14	0.17	0.21	0.17	0.24	0.28	0.25	0.34	0.44	0.11	0.15	0.16	0.14	0.18	0.21	0.18	0.24	0.28	0.25	0.33	0.44
5%	0	0.14	0.24	0.33	0.24	0.40	0.54	0.41	0.65	0.80	0.64	0.89	0.97	0.14	0.24	0.33	0.25	0.40	0.54	0.40	0.66	0.80	0.65	0.90	0.97
8%	0	0.20	0.41	0.59	0.42	0.70	0.86	0.70	0.94	0.99	0.94	1	1	0.20	0.41	0.57	0.41	0.70	0.85	0.69	0.94	0.99	0.94	1	1
2%	+0.20	0.23	0.55	0.75	0.46	0.77	0.91	0.73	0.95	0.99	0.94	1	1	0.15	0.34	0.50	0.20	0.35	0.46	0.25	0.38	0.51	0.30	0.47	0.60
5%	+0.10	0.20	0.46	0.65	0.43	0.72	0.88	0.71	0.94	0.99	0.94	1	1	0.18	0.36	0.51	0.29	0.51	0.68	0.45	0.72	0.87	0.68	0.92	0.98
8%	+0.05	0.24	0.53	0.73	0.52	0.83	0.95	0.82	0.98	1	0.98	1	1	0.22	0.47	0.67	0.44	0.75	0.90	0.73	0.95	0.99	0.94	1	1

Practical application

This study presented the power of detection and false positive rates at different degrees of constant and proportional biases in order to provide an equitable performance comparison across all six regression models. In practice, analytical performance specification for bias is often applied in fixed percentages, for which an infinite combination of constant and proportional bias is possible. The constant, proportional and mixed bias selected in this study aim to provide a representative performance of the different approaches to help practitioners make informed decisions on the optimal statistical approach for their laboratory.

To do so, a practitioner should first determine the analytical characteristics of the measurement procedure (i.e. measurement range ratio, imprecision profile) and define the desired analytical performance specification for bias (e.g. using the total error model or using biological variation model). Subsequently, the laboratory can select the optimal statistical approach based on its risk tolerance, expressed as desired probability of bias detection and false rejection rate. Other practical factors such as resources and statistical capability should be considered when selecting a regression approach.

Conclusions

Inappropriate rejection of a valid reagent lot can deprive patients of timely clinical care as an alternate reagent lot may not be readily available [16]. Hence, equal consideration needs to be given to balance the rates false rejection (by selecting an appropriate level of alpha) with the power of bias detection (selection of the appropriate statistical approach for the performance characteristics of the measurement procedure) and operation considerations (the number of clinical samples and replicates and staff time). Finally, regardless of the regression approach used, it is important that the laboratory uses patient samples to ensure commutability, and that these patient samples span the entire measurement range.

Research funding: None declared.

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

Competing interests: Authors state no conflict of interest.

Informed consent: Not acceptable.

Ethical approval: Not acceptable.

References

1. Algeciras-Schimnich A, Bruns DE, Boyd JC, Bryant SC, La Fortune KA, Grebe SK. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. Clin Chem 2013;59:1187–94.
2. Loh TP, Lee LC, Sethi SK, Deepak DS. Clinical consequences of erroneous laboratory results that went unnoticed for 10 days. J Clin Pathol 2013;66:260–1.
3. Bais R, Chesher D. More on lot-to-lot changes. Clin Chem 2014;60: 413–4.
4. Liu J, Tan CH, Loh TP, Badrick T. Detecting long-term drift in reagent lots. Clin Chem 2015;61:1292–8.
5. Chen X, Wang J, Zhang W, Xie E, Zhang B, Xu HG. Failure of internal quality control in detecting significant reagent lot shift in serum creatinine measurement. J Clin Lab Anal 2019;33: e22991.
6. Clinical and Laboratory Standards Institute. User evaluation of between-reagent lot variation; approved guideline CLSI document EP26-A. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
7. Mazzo DJ, Connolly M. Analytical method comparison based upon statistical power calculations. Pharm Res (NY) 1992;9: 601–6.
8. Linnet K. Necessary sample size for method comparison studies based on regression analysis. Clin Chem 1999;45:882–94.
9. Thompson S, Chesher D. Lot-to-lot variation. Clin Biochem Rev 2018;39:51–60.
10. Draper NR, Smith H. Applied regression analysis. Hoboken, New Jersey: Wiley; 1998.
11. Sadler WA. Joint parameter confidence regions improve the power of parametric regression in method-comparison studies. Accred Qual Assur 2010;15:547–54.
12. Mendenhall WM, Sincich TL. Statistics for engineering and the science. London: CRC Press LLC; 2016.
13. Plebani M, Zaninotto M. Lot-to-lot variation: no longer a neglected issue. Clin Chem Lab Med 2022;60:645–6.
14. van Schrojenstein Lantman M, Çubukçu HC, Boursier G, Panteghini M, Bernabeu-Andreu FA, Milinkovic N, et al. An approach for determining allowable between reagent lot variation. Clin Chem Lab Med 2022;60:681–8.
15. Giavarina D. Understanding Bland Altman analysis. Biochem Med 2015;25:141–51.
16. Loh TP, Sandberg S, Horvath AR. Lot-to-lot reagent verification: challenges and possible solutions. Clin Chem Lab Med 2022;60: 675–80.
17. Tan RZ, Punyalack W, Graham P, Badrick T, Loh TP. Detecting reagent lot shifts using proficiency testing data. Pathology 2019; 51:711–7.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2022-0274).