When the linearity verification data show essentially the same repeatability SD throughout the concentration interval for which linearity is being verified, OLS regression analysis is also appropriate.

4.2 Samples and Preparation

Five (or more) samples are needed to perform the linearity verification study. Sample volumes should be sufficient for performing at least two measurements per sample. When only samples HIGH and LOW are supplied or procured by the laboratory, (at least) three intermediate samples should be prepared, as discussed in Subchapter 4.2.2. When the samples (either all samples or just samples HIGH and LOW) are supplied by the manufacturer or a third party, they should be accompanied by assigned concentration values. However, it suffices for the laboratory to know their **relative** concentration values (see Figure 22).

4.2.1 Samples HIGH and LOW

Ideally, when the manufacturer's linearity claim is verified (as in the first two scenarios listed in Subchapter 4.1.1), the concentrations of the samples LOW and HIGH should match the LLLI and ULLI claimed in the measurement procedure's label. In general, as noted in Subchapter 4.1.1, these limits coincide with the measurement procedure's stated analytical measuring interval. It is essential that the samples yield explicit numerical measurement results in the study, as opposed to results expressed as "greater than" or "less than" some concentration.

The choice of samples depends in part on the interval over which they return explicit numerical results. This interval can be broader, at one or both extremes, than the stated analytical measuring interval that has been approved for external reporting of (undiluted) patient sample results. That is, for some measurement procedures, the laboratory is responsible for re-expressing patient results beyond this reporting interval as greater or less than some concentration when reporting them to physicians, patients, or others outside the laboratory. For performance verification studies, it is often desirable for a measurement procedure to report numerical results well beyond its stated analytical measuring interval, thereby making it possible to fully characterize performance (eg, precision, linearity) at the upper and lower limits of the stated analytical measuring interval. Many measurement procedures are designed such that their numerical value reporting interval coincides with or is insufficiently broader than the stated analytical measuring interval. In such cases, for a linearity verification study, the chosen samples HIGH and LOW must have concentrations as close as possible to the claimed limits but adjusted inwards just enough to avoid results expressed merely as greater or less than some concentration (ie, measurement results that cannot be used in the subsequent data analysis).

Tables 16 and 17 provide concentrations appropriate for samples HIGH and LOW with known values, given robust estimates of the measurement procedure's repeatability at those two concentrations. The data in these tables are based on the assumption that the measurement procedure's numerical reporting interval is identical to the stated analytical measuring interval, ie, that numerical values are returned only for measurement results in the interval defined by the LLoQ and ULoQ. Tables 16 and 17 indicate how far samples HIGH and LOW, respectively, should be adjusted inwards, as a function of the measurement procedure's imprecision, to ensure that two individual measurements of those samples return numerical results 95% of the time. (See Appendix J for a justification of the tabulated adjustments.)

Table 16. Adjustments for Sample HIGH

% CV for Repeatability	Sample HIGH Concentration Adjustment: Percent Below the ULoQ
% CV ≤ 1%	–2%
1% < % CV ≤ 2%	–4%
2% < % CV ≤ 3%	–5%
3% < % CV ≤ 4%	–7%
4% < % CV ≤ 5%	–10%
5% < % CV ≤ 10%	–15%
10% < % CV ≤ 15%	–20%

Abbreviations: % CV, coefficient of variation expressed as a percentage; ULoQ, upper limit of quantitation.

Table 17. Adjustments for Sample LOW

% CV for Imprecision for LLoQ	Sample LOW Concentration Adjustment: Percent Above the LLoQ
5	10%
10	15% to 20%
15	25% to 30%
20	30% to 40%

Abbreviations: % CV, coefficient of variation expressed as a percentage; LLoQ, lower limit of quantitation.

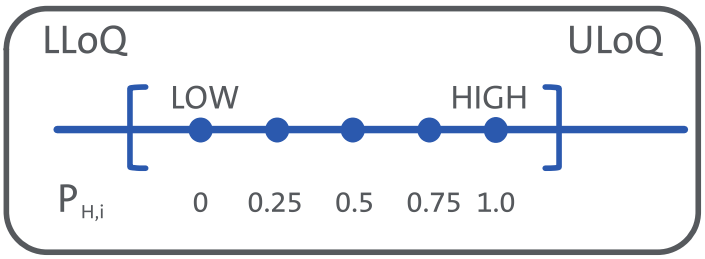
NOTE: In the first and third scenarios described in Subchapter 4.1.1, concentrations for samples HIGH and LOW can reasonably be treated as having known values, because the manufacturer or third party providing the samples likely has rigorously controlled their composition and evaluation. However, in the second scenario, the laboratory's uncertainty as to the sample concentrations should also be accounted for by adjusting the target values somewhat further inwards, depending on the uncertainties of the sample concentration estimates. Sources of variability relevant to these uncertainties but beyond those included in repeatability variation (eg, run-to-run sources) should also be assessed. In all three scenarios, adjustments might be needed to accommodate uncertainties in the measurement procedure's imprecision near the limits of the interval being tested.

4.2.2 Intermediate Samples

In addition to samples HIGH and LOW, three (or more) intermediate samples are needed to complete the linearity verification study's sample panel. These samples can be supplied by the manufacturer or a PT/EQA provider, with either assigned concentrations or proportions (P_H). When values are not supplied, the laboratory can easily generate them by mixing samples HIGH and LOW. For P_H = 0, 0.25, 0.5, 0.75, and 1, five equally spaced samples S_i, for i = 1 ... 5, can be produced by mixing samples HIGH and LOW in the following proportions:

$$S_i = P_{H,i}(\text{HIGH}) + (1 - P_{H,i})(\text{LOW}) \quad (55)$$

Although equal spacing is not required, it is convenient. Figure 22 depicts a sample panel produced in this manner, starting with samples LOW and HIGH spanning an interval just slightly narrower than the analytical measuring interval bracketed by the LLoQ and ULoQ.



Abbreviations: LLoQ, lower limit of quantitation; $P_{H,i}$, proportion of sample HIGH in sample i; ULoQ, upper limit of quantitation.

Figure 22. Sample Panel Spanning an Interval Just Slightly Narrower than the Analytical Measuring Interval Bracketed by the LLoQ and ULoQ

4.2.3 Goals

Determining the ADL is a critical step in linearity verification. Although the process for ADL determination is beyond this guideline's scope, several points should be considered. Because deviation from linearity is one component of systematic error (ie, bias), the ADL should be no greater than a fraction of the allowable bias. Although there is no consistent relationship between the ADL and allowable total error (ATE), it is unusual for the ADL to be greater than 50% of the ATE. Unusually high ADLs (eg, a level that exceeds 50% of ATE) should be supported by an explicit justification.

The ADL should be specified in a form applicable to each of the samples. Depending on the measurement procedure or (more generally) on the measurand and the test's intended clinical use, the ADL can be a constant (ie, fixed) value, in units appropriate to the measurand (eg, 0.3 ng/L); a relative value (eg, 6%); or some combination thereof (eg, the larger of 0.3 ng/L or 6%) (see Subchapter 2.9).

4.2.4 Precision

Before beginning the linearity verification study, the laboratory should verify its ability to perform the measurement procedure with acceptable repeatability (ie, within-run precision) consistent with the manufacturer's claims.

To be prepared for the study's data analysis, the laboratory should have at least a basic understanding of the repeatability expected for the measurement procedure throughout the interval being tested for linearity, ie, even at the extreme concentration levels represented by samples HIGH and LOW. For this data analysis, the measurement procedure's label, including its precision table and LLoQ claim, might or might not suffice, because the information about precision at concentrations close to the ULoQ might not be available.

4.3 Study Procedure

At least five samples are tested (at least) in duplicate in a single run.

When more than one run is needed to complete all measurements, every sample should be represented in each run. In this situation, run-to-run imprecision is relevant, not just repeatability.

4.4 Inspection for Data Integrity

Owing to the limited number of samples and replicates, formal testing for outliers or repeatability is not recommended. However, the laboratory should perform an informal data integrity check by preparing a simple tabular summary and scatterplot of the data. Examples are provided in Table 18 and Figure 23. The table should include the mean, SD, and CV, where the CV = 100(SD / mean). The recommended graph is a plot of the individual measurements on the y-axis and either RCs or assigned values on the x-axis. These two summaries can be examined for any obvious problems with imprecision (ie, repeatability), gross outliers, pronounced nonlinearity, or other questionable patterns. When any of these problems are detected, the data evaluation should not proceed, because the data are not reliable enough for a linearity assessment.

If errors are suspected, the laboratory should conduct and document an investigation. If warranted, the study procedure in Subchapter 4.3 can be repeated. However, a justification for study repetition needs to be documented.

In this example, data are collected with a single lot of reagent, on a single instrument, in a single run. Six concentration levels covering the claimed linearity interval are tested with two replicates. The RC between samples HIGH (ie, HIGH pool) and LOW (ie, LOW pool) is not known, so sample LOW is assigned a proportion of zero.

Table 18. Example Linearity Data From Six Concentration Levels

Proportion of HIGH Pool	Rep 1, mg/dL	Rep 2, mg/dL	Mean, mg/dL	SD, mg/dL
1	3350	3293	3321.5	40.31
0.75	2500	2377	2438.5	86.97
0.5	1650	1653	1651.5	2.12
0.25	777	791	784.0	9.90
0.1	338	341	339.5	2.12
0	36	35	35.5	0.71

Abbreviations: Rep, replicate; SD, standard deviation.

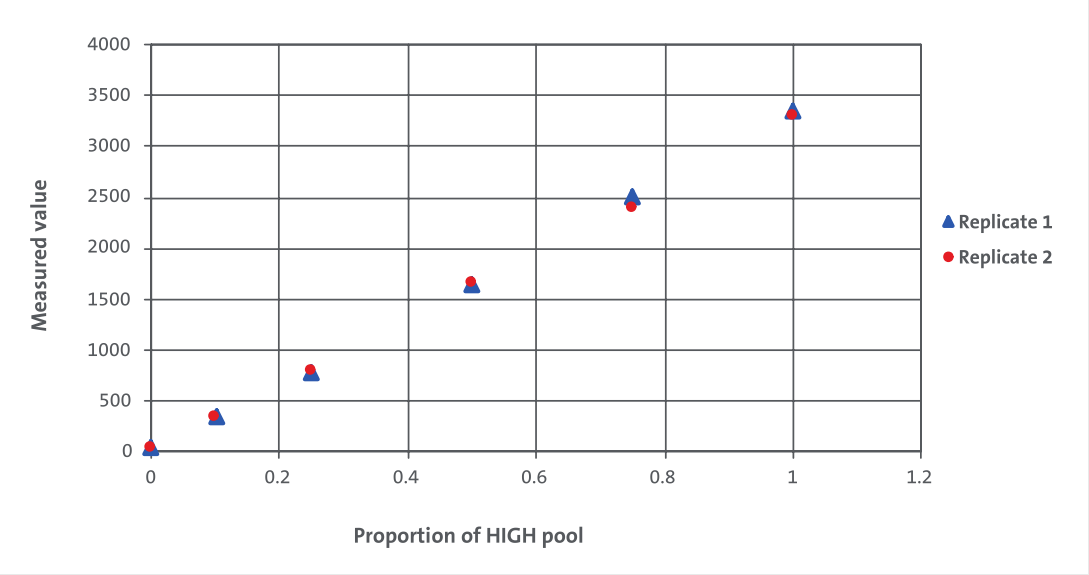


Figure 23. Graph of Example Linearity Data From Six Concentration Levels

4.5 Data Analysis: Weighted Least Squares Regression Analysis With Confidence Intervals

A straight line should be fit to the data using WLS regression. Owing to the small number of replicates used for linearity verification, estimation of the repeatability SDs (used in the weighting), σ_i, from the replicates results in estimates with very few degrees of freedom, df = R – 1, R = number of replicates (in this case, R = 2 and df = 1). A small number of degrees of freedom for any single pool creates a high amount of uncertainty in the localized SD. To create a more confident estimate of SD, a precision profile is needed. A precision profile is a linear model of the SD of replicates for each pool vs the mean measurand value for the pool.

Normally, a measurement procedure is assumed to have a relatively constant % CV across its analytical measuring interval. However, even with this assumption, there is a range, as values approach zero, in which the % CVs start to increase rapidly. The concentration of the LOW pool is often in this range. For this reason, the LOW pool is not included in the precision profile. However, the actual SD from the LOW pool is used in the weight calculation.

If the LOW pool does not fall within the range of rapidly increasing % CVs, this point should be included in the precision profile. Also, if the measurement procedure does not have a relatively consistent % CV, or if for some reason this method cannot be applied, other weighting schemes may be used.

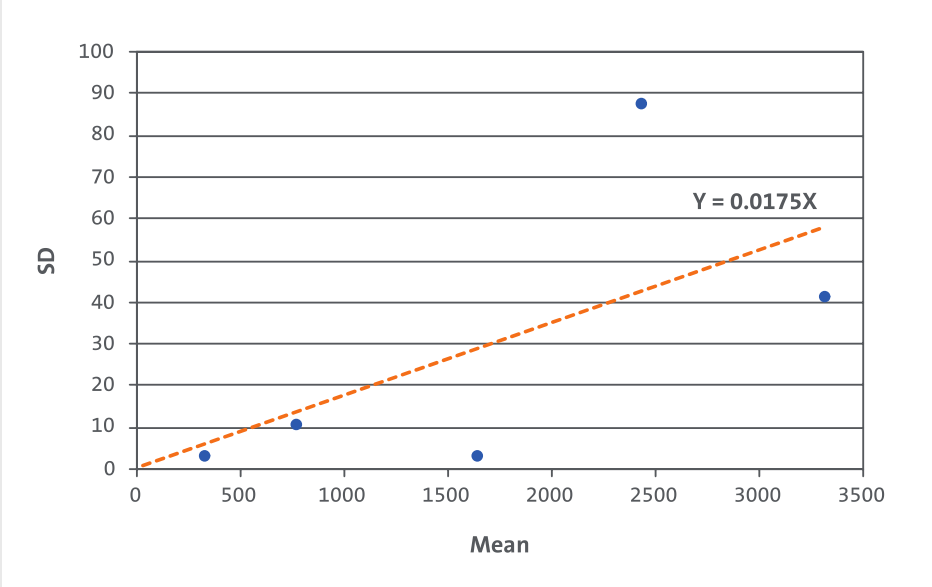
4.5.1 Example

The precision profile shown in Table 19 and Figure 24 is produced from the test data in Subchapter 4.4, pools 1 to 5, using unweighted regression and a forced intercept of zero. For this example, the % CV is assumed to be relatively constant across the analytical measuring interval of the instrument.

Table 19. Example Precision Profile From an Unweighted Regression and Zero Intercept

Pool	Proportion of HIGH Pool	Replicate 1	Replicate 2	Mean	SD	Sigma Value (based on precision profile)
1	1	3350	3293	3321.50	40.305	58.178
2	0.75	2500	2377	2438.50	86.974	42.712
3	0.5	1650	1653	1651.50	2.121	28.927
4	0.25	777	791	784.00	9.900	13.732
5	0.1	338	341	339.50	2.121	5.947
6	0	36	35	35.50	0.707	

Abbreviation: SD, standard deviation.



Abbreviation: SD, standard deviation.

Figure 24. Graph of Example Precision Profile From an Unweighted Regression and Zero Intercept

The model from the precision profile (SD = 0.0175 • Mean) is used to estimate sigma values that are in turn used to calculate weights for each mean. Weights are calculated as:

$$\text{Weight} = \frac{1}{\sigma^2} \tag{56}$$

The proportion of the HIGH pool is calculated from the samples HIGH and LOW dilution scheme. The proportion of the HIGH pool should not be confused with the concentration or the RC. A straight-line WLS model fits to the proportion of the HIGH pool and the mean values of two replicates. When the LOW pool is not

included in the precision profile, its replicate SD is still used to calculate a weight. For this example, because the % CV is assumed to be relatively constant across the analytical measuring interval of the instrument, the sigma value from the precision profile is used to calculate the weight for the LOW pool. Results are shown in Table 20.

Table 20. Example Calculation of Weights From Estimated SDs

Pool	Rep 1	Rep 2	Mean	SD	Sigma Value (based on precision profile)	Proportion of HIGH Pool	Weight (based on precision profile and SD of the LOW pool)
1	3350	3293	3321.5	40.305	58.178	1	$1 / 58.178^2$
2	2500	2377	2438.5	86.974	42.712	0.75	$1 / 42.712^2$
3	1650	1653	1651.5	2.121	28.927	0.5	$1 / 28.927^2$
4	777	791	784.0	9.900	13.732	0.25	$1 / 13.732^2$
5	338	341	339.5	2.121	5.947	0.1	$1 / 5.947^2$
6	36	35	35.5	0.707		0	$1 / 0.707^2$

Abbreviations: Rep, replicate; SD, standard deviation.

Because the RC between the HIGH and LOW pools is not known, the regression is completed with an intercept. Figure 25 shows the WLS linear regression of the mean pool values vs the proportion of the HIGH pool values. The WLS linear regression results in a line with the equation:

$$\text{Mean} = 35.30 + 3149.739(\text{Proportion of HIGH Pool}) \quad (57)$$

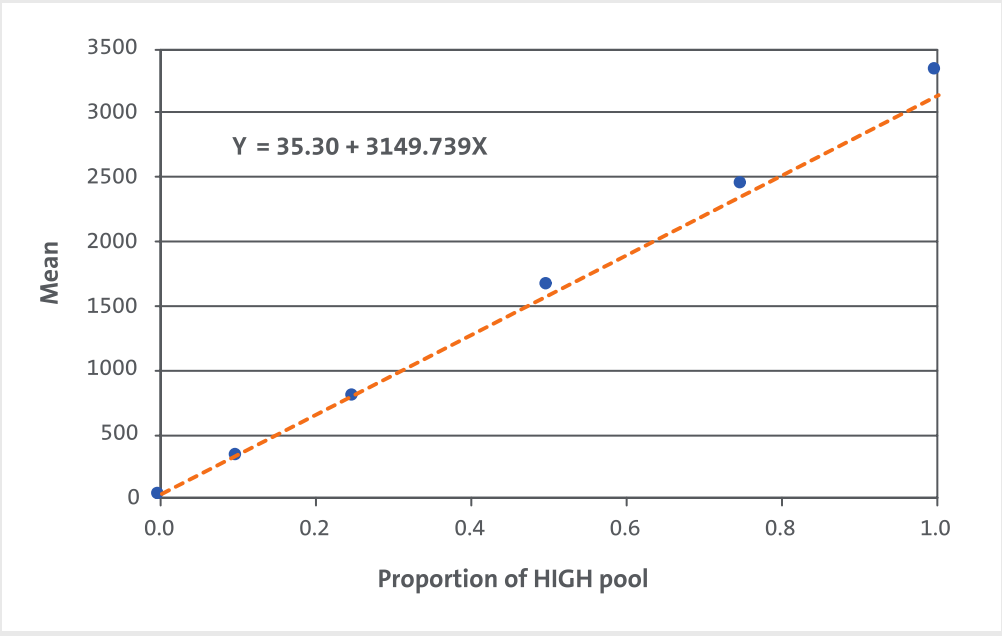


Figure 25. WLS Linear Regression of Mean Pool Values vs Proportion of HIGH Pool Values

Finally, the deviation from linearity, ie, the difference between the measured value (ie, the mean of the replicate measurements) and the predicted value from the linear model, is calculated along with the confidence interval (CI) for each difference:

$$\text{Deviation from Linearity}_{\text{pool}} = \text{Measured Value}_{\text{pool}} - \text{Predicted Value}_{\text{pool}} \quad (58)$$

The CI is based on the same sigma values used in the weighting. The $(1 - \alpha)100\%$ lower and upper confidence limits, L_i, U_i, around the deviation of the mean of R replicates (ie, measured value) from the fitted straight line (ie, predicted value) at the *i*th concentration level is calculated as:

$$[L_i, U_i] = (\bar{y}_i - \hat{y}_i) \pm Z_{1-\frac{\alpha}{2}} \left(\frac{\sigma_i}{\sqrt{R}} \right), \quad (59)$$

where:

- L_i = lower limit of the CI
- U_i = upper limit of the CI
- $\bar{y}_i$ = mean of R replicates (ie, measured value) at the *i*th concentration level
- $\hat{y}_i$ = concentration predicted from the fitted straight line (ie, predicted value) for the *i*th concentration level
- σ_i = sigma value or SD used to calculate the weight at the *i*th concentration level
- $Z_{1-\frac{\alpha}{2}}$ = $(1 - \alpha / 2)$-quantile of the normal distribution
- R = number of replicates

Alpha is adjusted to produce confidence limits around individual means that compensate for multiple individual evaluations, so that $(1 - \alpha_n)100\%$ confidence limits apply to the overall assessment of linearity verification:

$$\alpha = 1 - (1 - \alpha_n)^{\frac{1}{n}} \quad (60)$$

Adjusted quantile $Z_{1-\frac{\alpha}{2}}$ values for 90% confidence limits are listed in Table 21. These values are used in equation (58) to calculate confidence limits for deviations of mean test results from the fitted straight line.

Table 21. $Z_{1-\frac{\alpha}{2}}$ Quantiles of Standardized Normal Distribution

Probability to Pass Verification	Number of Concentration Levels, n	
	5	6
90%	2.31	2.38

The pool-specific CIs are compared with the ADL. Linearity across the interval in question is verified only when all CIs overlap with the ADL interval(s).

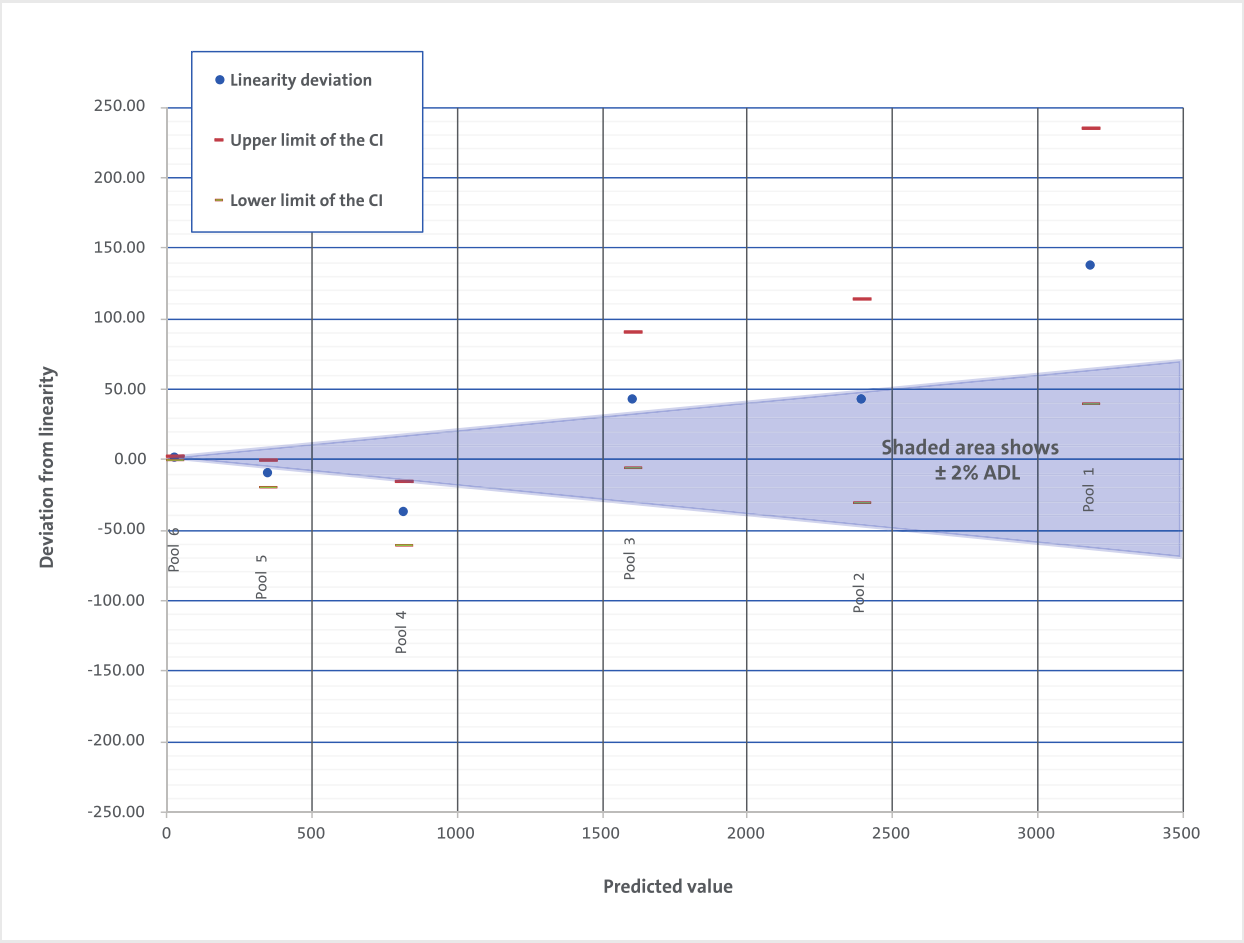
The CI is calculated using $\alpha_n = 0.10$ with $n = 6$, resulting in $Z_{1-\frac{\alpha}{2}}$ of 2.38. For this example, the ADL is set at ± 2% of the predicted value. Example calculations are shown in Table 22.

Table 22. Example Calculations for Linearity Evaluation

Pool	Rep 1	Rep 2	Mean	SD	Sigma Value (based on precision profile)	Proportion of HIGH Pool	$W = 1 / \Sigma^2$ (for the LOW pool, $W = 1 / SD^2$)	Predicted Value	Deviation From Linearity	Lower Limit of the CI	Upper Limit of the CI	± ADL
1	3350	3293	3321.5	40.305	58.178	1	$1 / 58.178^2$	3185.04	136.46	38.64	234.28	± 63.70
2	2500	2377	2438.5	86.974	42.712	0.75	$1 / 42.712^2$	2397.61	40.89	−30.92	112.71	± 47.95
3	1650	1653	1651.5	2.121	28.927	0.5	$1 / 28.927^2$	1610.17	41.33	−7.31	89.97	± 32.20
4	777	791	784.0	9.900	13.732	0.25	$1 / 13.732^2$	822.74	−38.74	−61.83	−15.65	± 16.45
5	338	341	339.5	2.121	5.947	0.1	$1 / 5.947^2$	350.27	−10.77	−20.77	−0.78	± 7.01
6	36	35	35.5	0.707		0	$1 / 0.707^2$	35.30	0.20	−0.99	1.39	± 0.71

Abbreviations: ADL, allowable deviation from linearity; CI, confidence interval; Rep, replicate; SD, standard deviation.
Symbols: Σ, sigma; *W*, weight.

The resulting statistics can be used to produce a graph of the analysis, as shown in Figure 26, although this step is optional.



Abbreviations: ADL, allowable deviation from linearity; CI, confidence interval.

Figure 26. Example of Linearity Analysis

In this example, linearity is verified because **all** pools have a CI that intersects with the ADL, as shown in Table 23 and Figure 26. If any of the individual pools does **not** show a CI that overlaps with the ADL, the measurement procedure linearity fails verification. The sample size used in this example is the minimum for linearity verification. Additional replicates give a more confident estimate of the mean and the uncertainty of the mean. They also most likely will decrease the size of the CI.

Table 23. Data Analysis Results

Group	Pools	Description
1	2 and 6	Pass: the deviation from linearity is within the ADL.
2	1, 3, 4, and 5	Pass: each CI overlaps the ADL.

Abbreviations: ADL, allowable deviation from linearity; CI, confidence interval.

NOTE: As shown in Figure 26, for some of the data points, although the calculated mean is **not** within the ADL, there is one confidence limit that is within the ADL. Therefore, there is **not** high certainty (ie, statistical proof) that the true mean is outside the ADL. In this case, the laboratory might choose to run additional data to obtain a more confident estimate of the mean and decrease the size of the CI.

4.5.2 Verification Example Conclusions

When observed results are outside the ADL, but CIs overlap with the ADL, the study should not be automatically accepted or automatically rejected. The results should be interpreted based on the overall performance characteristics of the measurement procedure, the allowable error, and the medical director's judgment. Possible study result scenarios are listed in Table 24.

When interpreting the results of the linearity verification, the laboratory should consider these factors:

- If the CI overlaps with the ADL, the study results do not disprove the linearity claim. When the overall performance characteristics are sufficient to satisfy clinical need, the laboratory is justified in accepting the results of the linearity study.
- When the overall performance characteristics are insufficient to satisfy clinical need (ie, the ATE is exceeded), the study should not be accepted. Additional investigation should be performed.
- When technical error or some other identifiable cause is suspected, repeating the study might be sufficient to resolve the uncertainty. Another option is to repeat the study with additional replicates, which is likely to provide a greater level of certainty.
- Repeating the study multiple times is not an appropriate or statistically valid approach. Ideally, the laboratory should take relevant and practical measures (eg, recalibration, general maintenance procedures) to remove any possible cause of the verification failure. Repeating a failed verification, without identifying and removing the cause of the failure, increases the risk of approving a truly nonlinear method. Failure of two consecutive verification attempts should be considered strong evidence that the method is truly nonlinear. No additional verification attempts should be conducted before the issue has been identified and resolved. Whenever a verification is reattempted, regardless of the reason, the rationale for the new attempt must be documented.