MSACL Connect 2020

Method Comparison and Precision Experiments – Design, Analysis and Interpretation

Andrea Geistanger
Roche Diagnostics GmbH, Germany

Overview

- Method Comparison Experiments
  - Experimental Design
  - Analysis
  - Interpretation
- Precision – Variance - Components Experiments
  - Experimental Design
  - Analysis
  - Interpretation

Overview

- Method Comparison Experiments
  - Experimental Design
  - Analysis
  - Interpretation
- Precision – Variance - Components Experiments
  - Experimental Design
  - Analysis
  - Interpretation

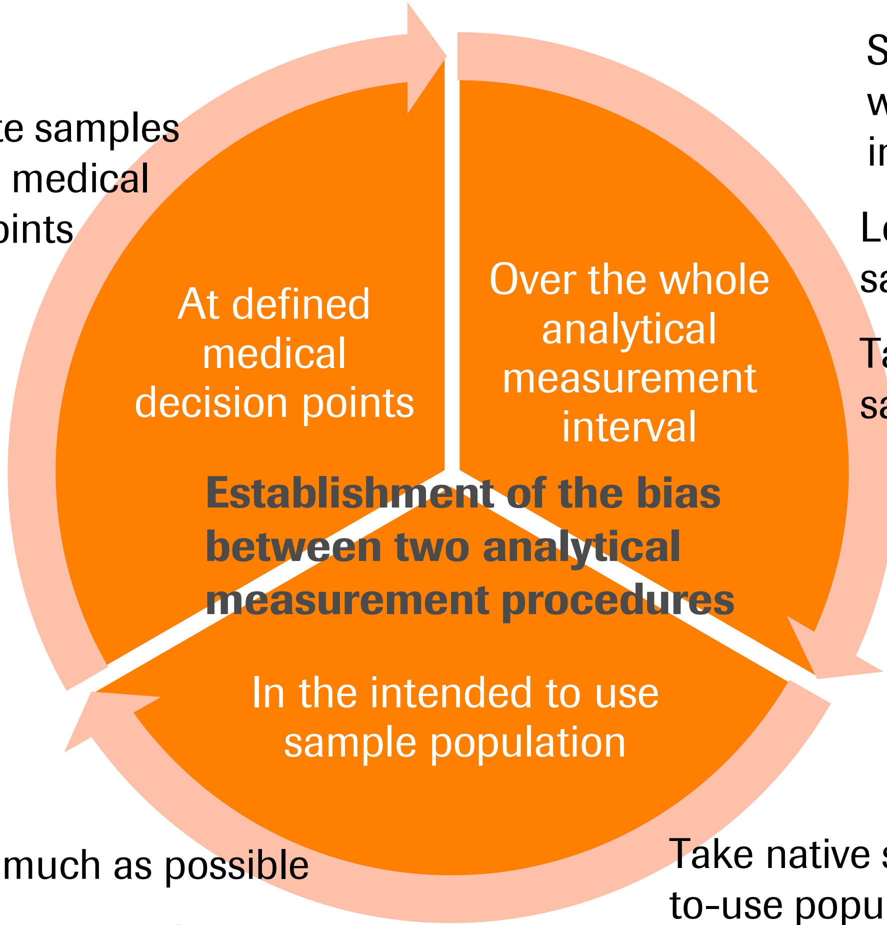
Method Comparison – Study Goals and Design

Concentrate samples around the medical decision points

Spread your samples over the whole analytical measurement interval

Look for uniform distribution of samples

Take high and low concentrated samples into account



Consider native samples as much as possible

Restrict modified samples to low numbers

Take native samples from the intended-to-use population, e.g. cancer patients, children…

MC – Experimental Design

- Number of samples:
  - CLSI EP9:
    - 100 samples for establishment of bias claims;
    - 40 samples for verification of bias claims
    - Restriction to max. 20% modified samples
- Number of replicates:
  - Measurement design should be the same as used in routine for both methods
- Consider
  - Improvement in the confidence of statistical estimates
  - Effects of unexpected interfering substances
  - Differences between patient populations

⇒ Might lead to additional number of samples

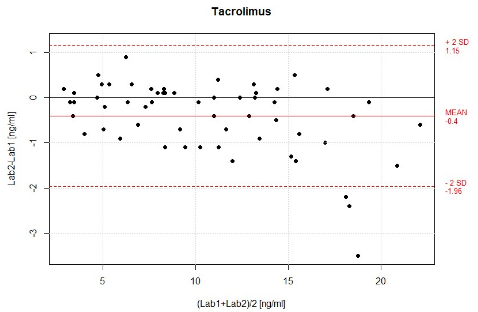
Example of Method Comparison Experiment

Measure the same samples in two measurement methods

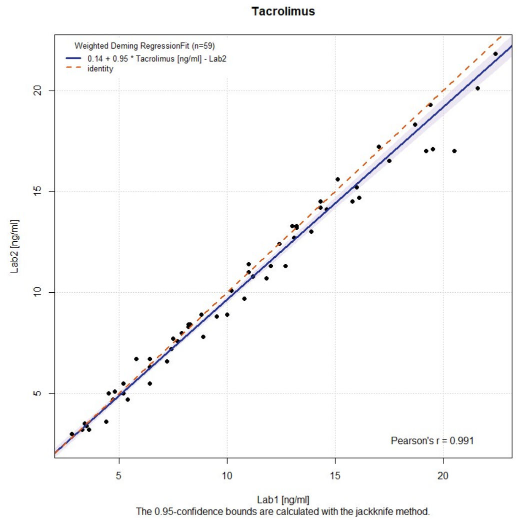
First observations from a method comparison experiment

Sample	Tacrolimus [ng/ml] - Lab1	Tacrolimus [ng/ml] - Lab2
T_01	14.2	14.3
T_02	17.0	19.2
T_03	11.3	12.7
T_04	10.1	10.2
T_06	7.8	8.9
T_09	7.7	7.5

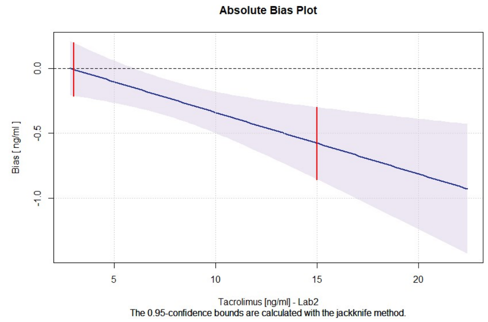
Graphical representations in terms of difference plots (Bland-Altmann Plots)

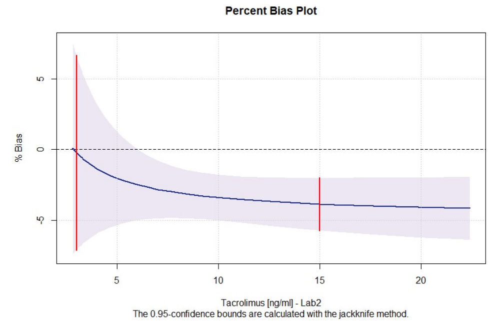


Scatterplot and Regression analysis – Deming, Passing-Bablok Regression



Bias Analysis

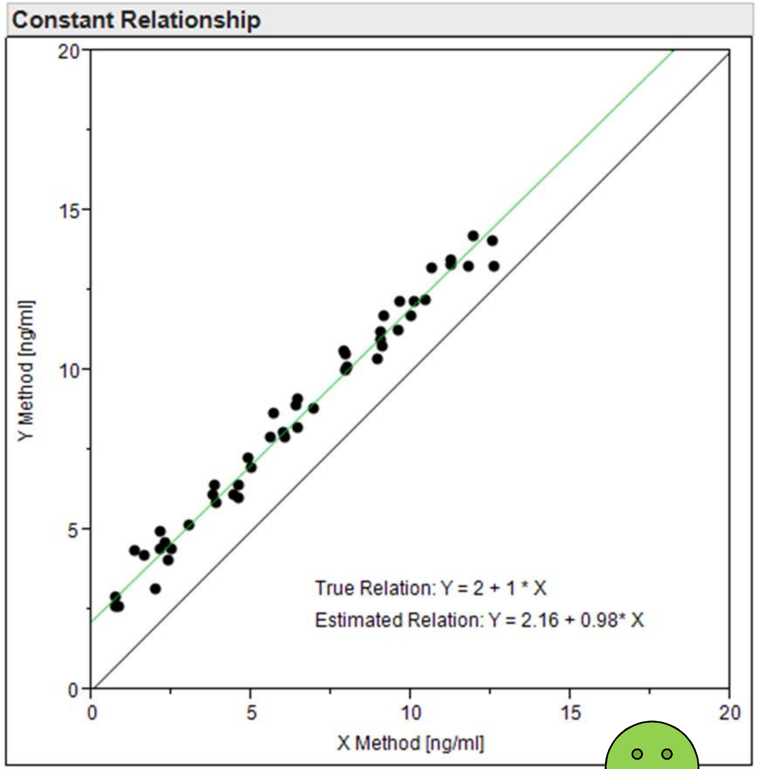




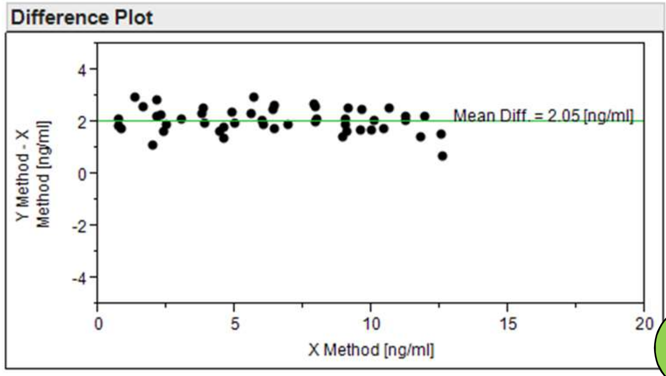
Regression Method	Parameter	Estimate	Std.Error	95LCL	95UCL
Deming	Intercept	0.34	0.20	-0.05	0.74
Deming	Slope	0.93	0.02	0.89	0.97
Weighted Deming	Intercept	0.14	0.14	-0.15	0.42
Weighted Deming	Slope	0.95	0.02	0.92	0.99
Passing Bablok	Intercept	0.24	NA	-0.06	0.69
Passing Bablok	Slope	0.95	NA	0.90	0.99

Difference Plots, given a constant bias: Y = 2 + X

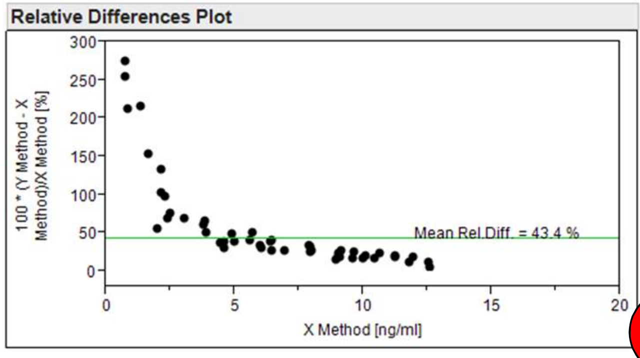
Simulated Data



The estimated regression line describes the relationship between the two methods well.



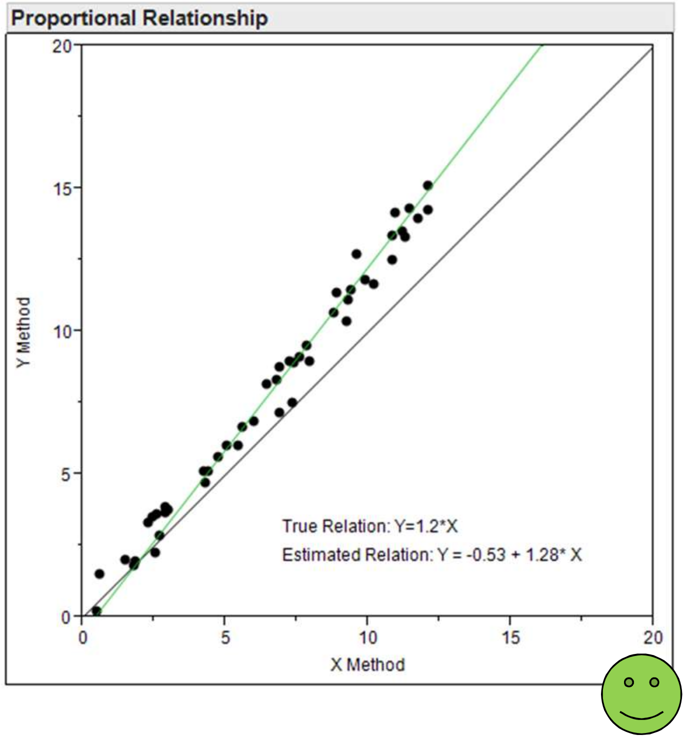
The mean of the differences describes the relationship between the differences and the concentration well.



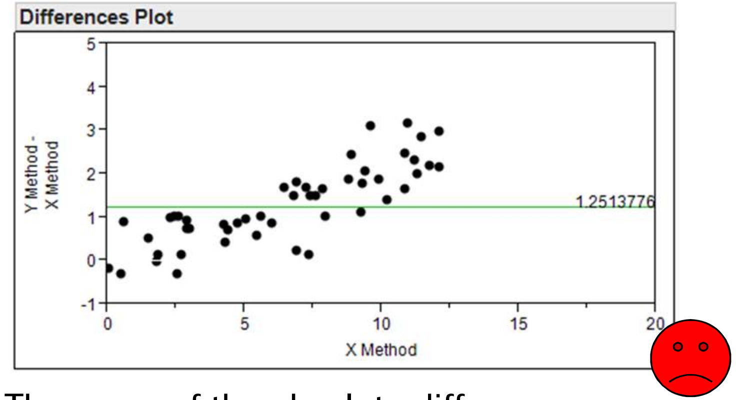
The mean of the relative differences, does not fit to the relative differences over the concentration range.

Difference Plots, given a proportional bias: Y = 1.2 * X

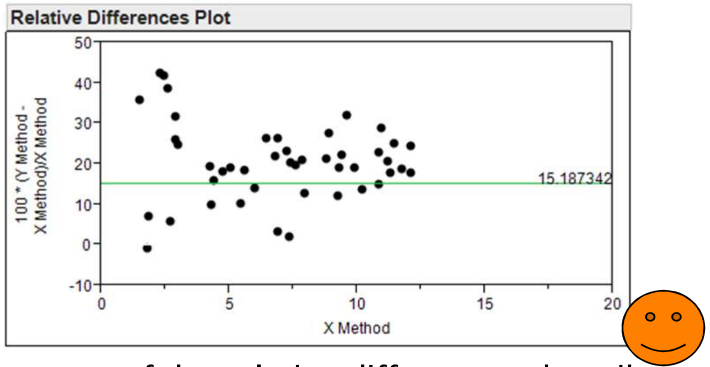
Simulated Data



The estimated regression line describes the relationship between the two methods well.



The mean of the absolute differences, does not fit to the absolute differences over the concentration range.



The mean of the relative differences describes the relationship between the relative differences and the concentration good enough.

The linear regression approach, accounts for absolute and proportional bias between two measurement methods.

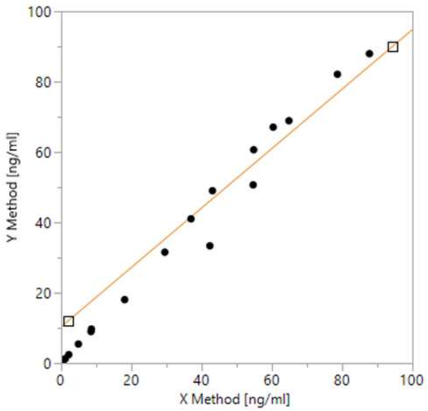
Summary measures of difference plots can only cope with one of the bias terms.

Regression Methods for Method Comparison (mcr R Package options)

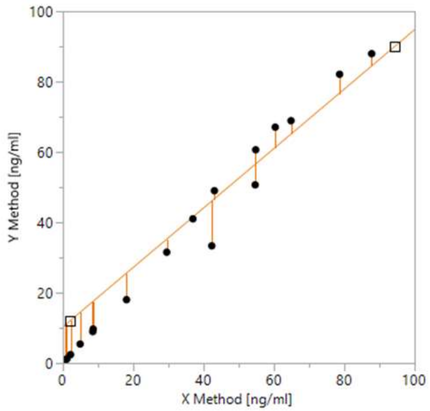
method.reg	Regression-Type	When to use	Recommendation for CIs
LinReg	Ordinary linear least-squares regression, residuals on the y-axis direction	Measurement method on the x-axis error-free; constant measurement error SD.	Analytical
WLinReg	Weighted linear least-squares regression, residuals on the y-axis direction	Measurement method on the x-axis error-free; increasing measurement error SD, constant CV	Analytical
Deming	Deming Regression; Residual direction depends on the error.ratio settings	Measurement method on the x-axis with error; constant measurement error SD.	Jackknife
WDeming	Weighted Deming Regression; Residual direction depends on the error.ratio settings	Measurement method on the x-axis with error; increasing measurement error SD, constant CV	Jackknife
PaBa	Passing-Bablok regression, robust regression method for method comparison data	Measurement method on the x-axis with error; Slope around 0.9-1.1	Quantile Bootstrap Don’t use jackknife!
PaBaLarge	Passing-Bablok regression for large datasets, computation time efficient for n > 150	Measurement method on the x-axis with error; Slope around 0.9-1.1; Large datasets	Quantile Bootstrap Don’t use jackknife!

Some words on least-squares estimation

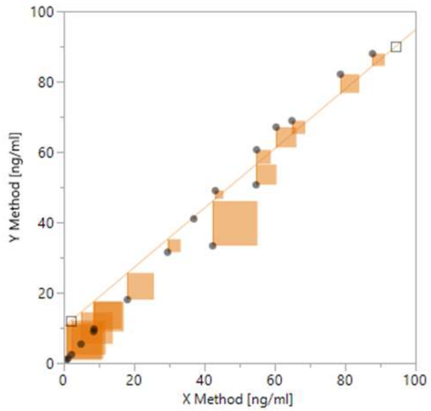
Let’s try to guess a line:



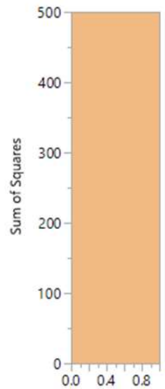
Data points, through which a line is drawn, but is this the best line for these data?



Y-Direction residuals: Distance from the point to the line, in y-direction.



Squared residuals for each data point, sum all of them together

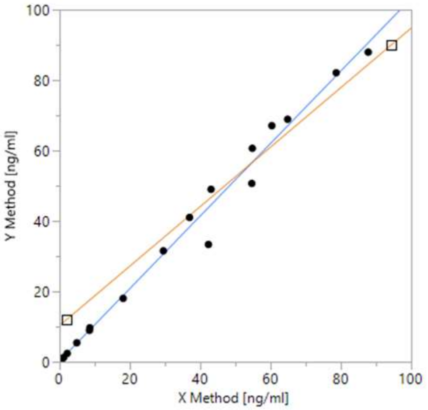


Obtain the sum of the squares. → Quite high for my guessed line. → **Can math do it better?**

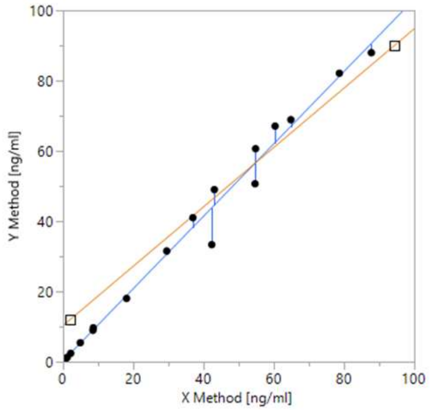
Linear least-squares regression

Best regression line is the one, which minimizes the sum of squared residuals in y-direction

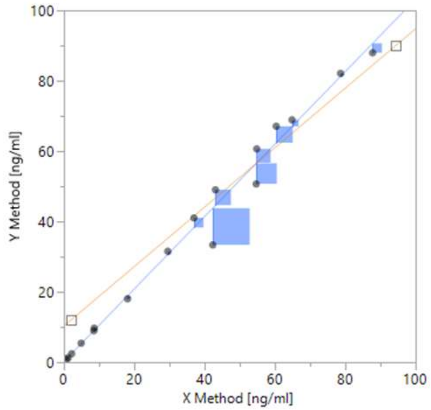
Assumption: Points are not on the line, because there is measurement error only in the y-data!



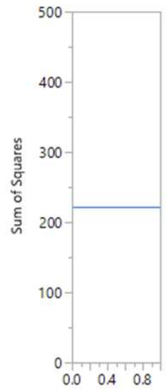
Blue: Best – least-squares regression line



Residuals of the LS Line: Distance from the point to the line, in y-direction.



Squared residuals for each data point, sum all of them together

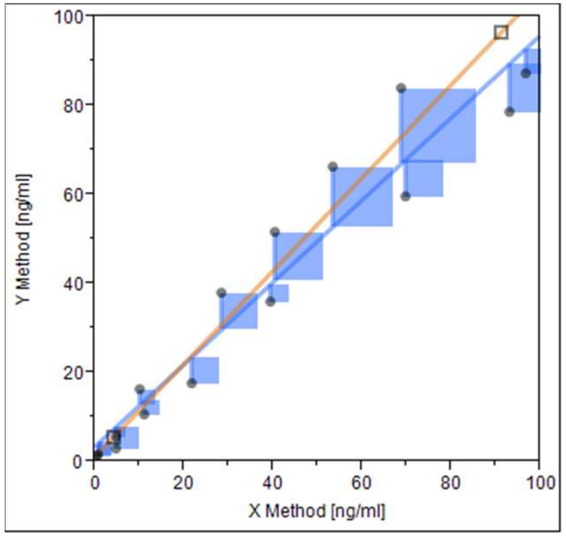


The minimal sum of squares is obtained by this line.

→ **Math can do it better!**

Weighted linear regression

- For measurement methods with increasing absolute measurement error (Standarddeviation), the residuals will be larger in the higher concentration range



Linear regression without weighting:

Datapoints in the higher measurement range dominate the estimation of the regression line, as all residuals enter equally the sum-of-squares.

Linear regression with weighting:

Residuals in the higher concentration range are down-weighted for the calculation of the sum-of-squares. Residuals in the lower concentration range are up-weighted for the estimation.

The weight of each datapoint is given by:

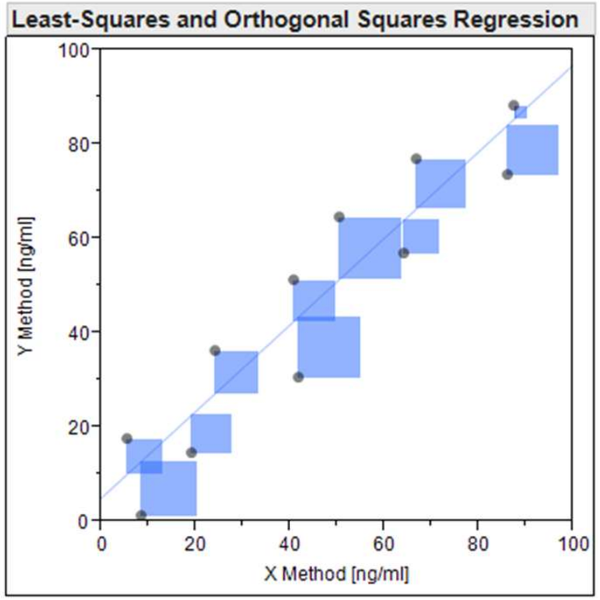
$$w_i = \frac{1}{X_i^2}$$

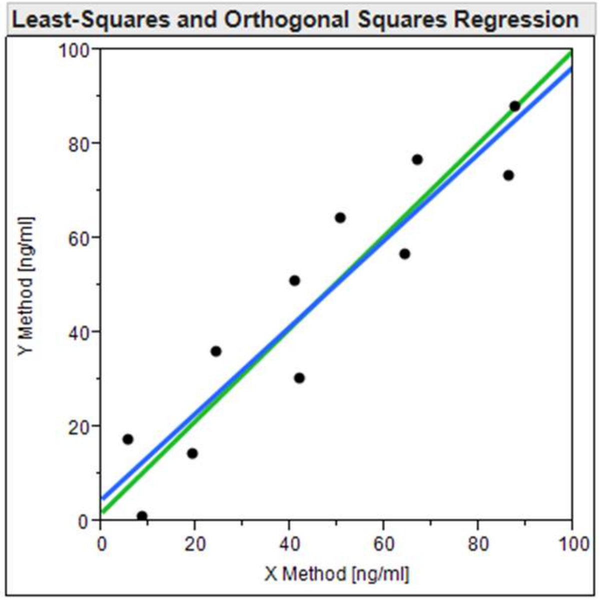
Least-squares regression for MC

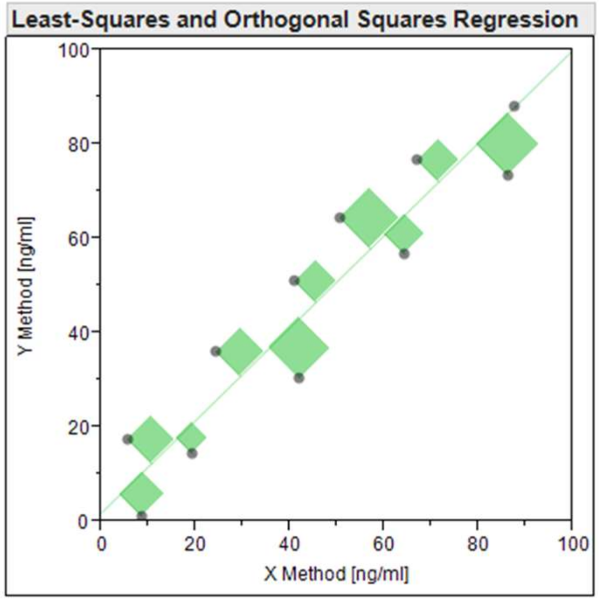
- The main assumption for linear least-squares regression is not true:
- *Assumption: Points are not on the line, because there is measurement error only in the y-data!*
- But:
- *Points are not on the line, because there is measurement error in the y-data and the x-data!*
- Both measurement methods are subject to measurement error.
- New definition of residuals is needed, to take the measurement error in the x-data also into account:

Deming regression

Linear and Deming regression







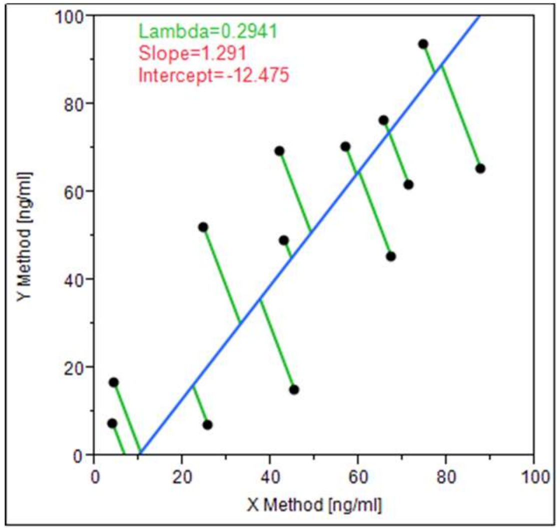
Linear regression

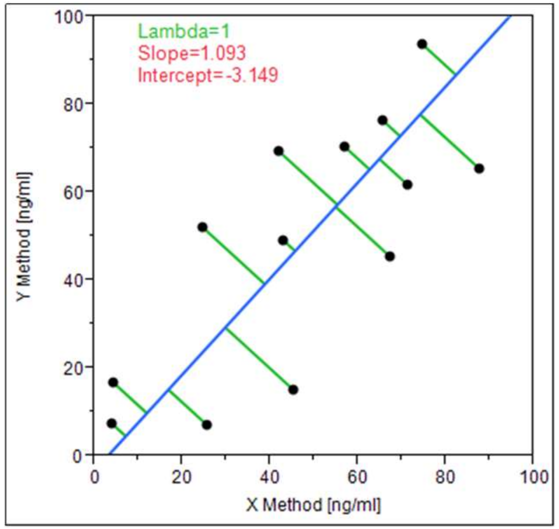
Minimize the squares in **y-direction**

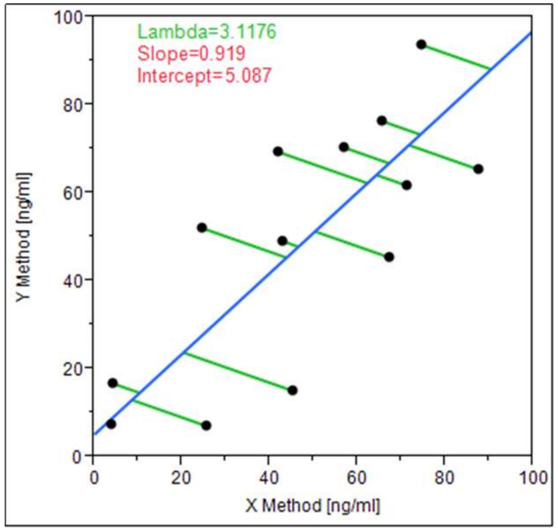
Deming regression

Minimize the squares **perpendicular to regression line**

Deming Regression – additional information needed







Lambda λ defines angle in Deming regression

$$\lambda = \frac{\sigma_x^2}{\sigma_y^2}$$

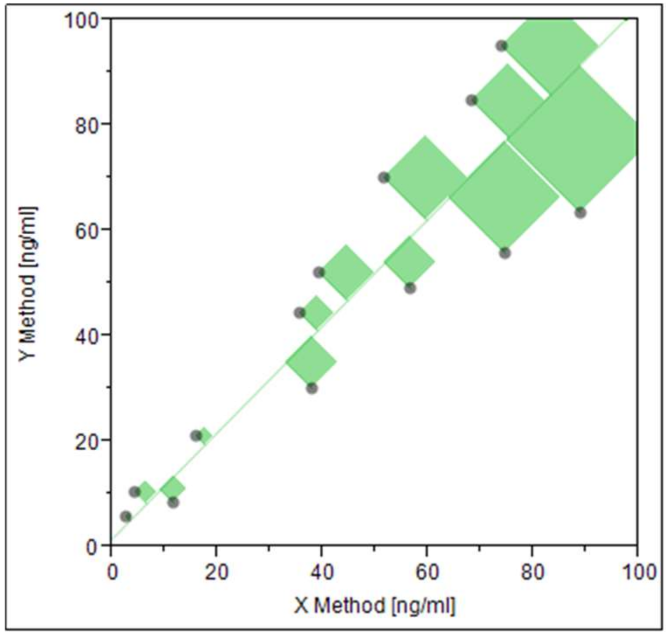
σ_x = standard deviation error of results x-method

σ_y = standard deviation error of results y-method

Regressionsmethoden

Weighted Deming Regression

- For measurement methods with increasing absolute measurement error (Standarddeviation), the residuals will be larger in the higher concentration range → Weighted version of Deming regression needed, similar as for linear regression.



Deming regression without weighting:

Datapoints in the higher measurement range dominate the estimation of the regression line, as all residuals enter equally the sum-of-squares.

→ **Deming regression with weighting**:

Residuals in the higher concentration range are down-weighted for the calculation of the sum-of-squares. Residuals in the lower concentration range are up-weighted for the estimation. Same logic as for linear regression.

The weight of each datapoint is given by:

$$w_i = \frac{1}{((X_i + Y_i)/2)^2}$$

An iterative procedure for the estimation and weight determination is used.

Passing Bablok Regression (I)

Robust non-parametric regression procedure

(not influenced by outlying observations, no assumptions on the distribution of the residuals)

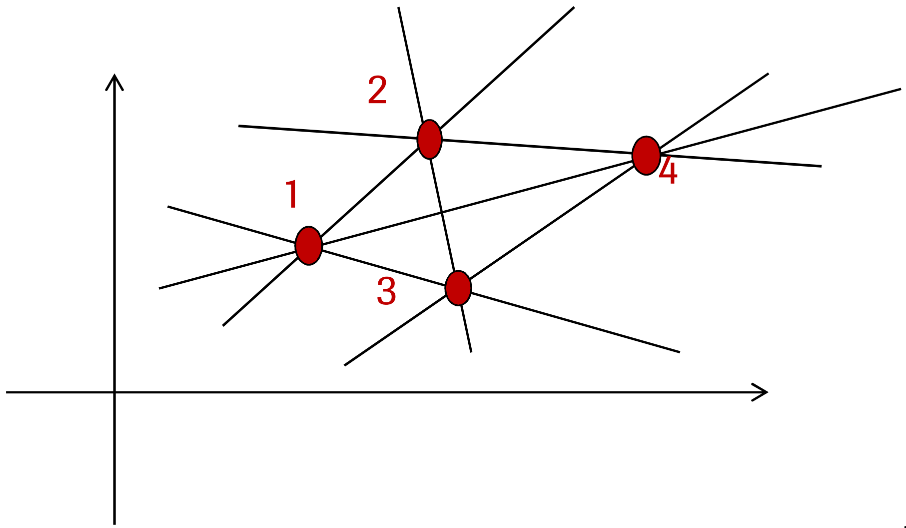
For 2 datapoints *i* and *j* the slope is calculated as:

$$b_{ij} = \frac{Y_i - Y_j}{X_i - X_j}; \qquad 1 \leq i < j \leq n$$

Result: b_{12}, b_{13}, b_{14}, b_{23}, b_{24}, b_{34}

The estimator of the slope is the median of all ordered slopes,

Shifted by K ranges, where K = (Number of slopes bij < -1)



@ H. Passing, W. Bablok,1983, Hoechst und Böhringer

Passing Bablok Regression (II)

Estimator of the Intercept

a1, a2, a3, a4,

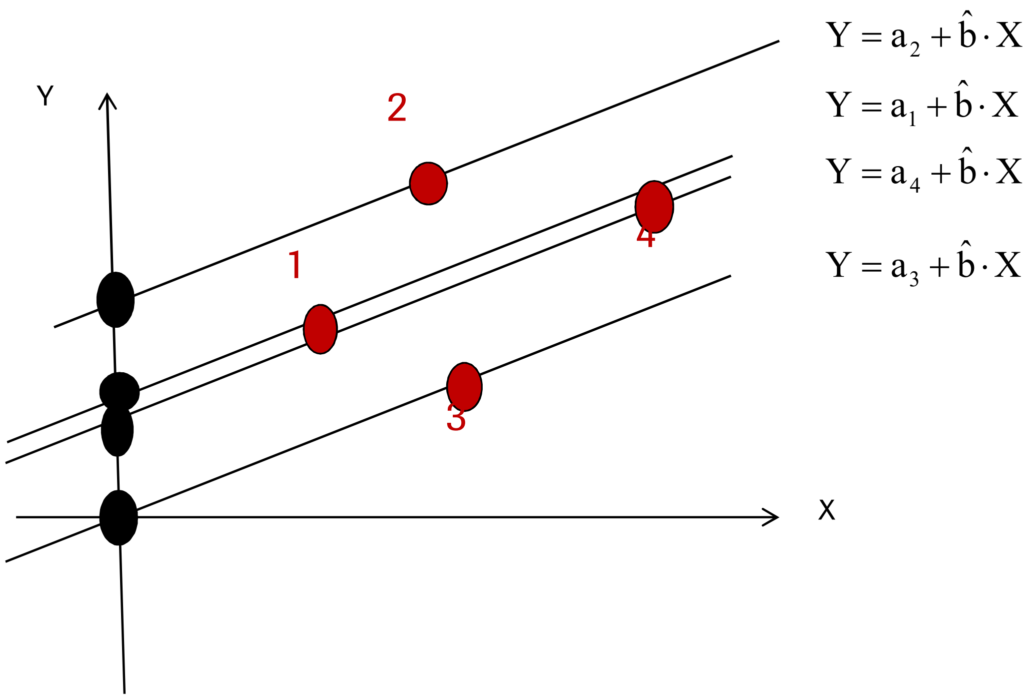
Based on the slope estimator b, obtain for

each datapoint an intercept of the parallel lines

Order according to the height:

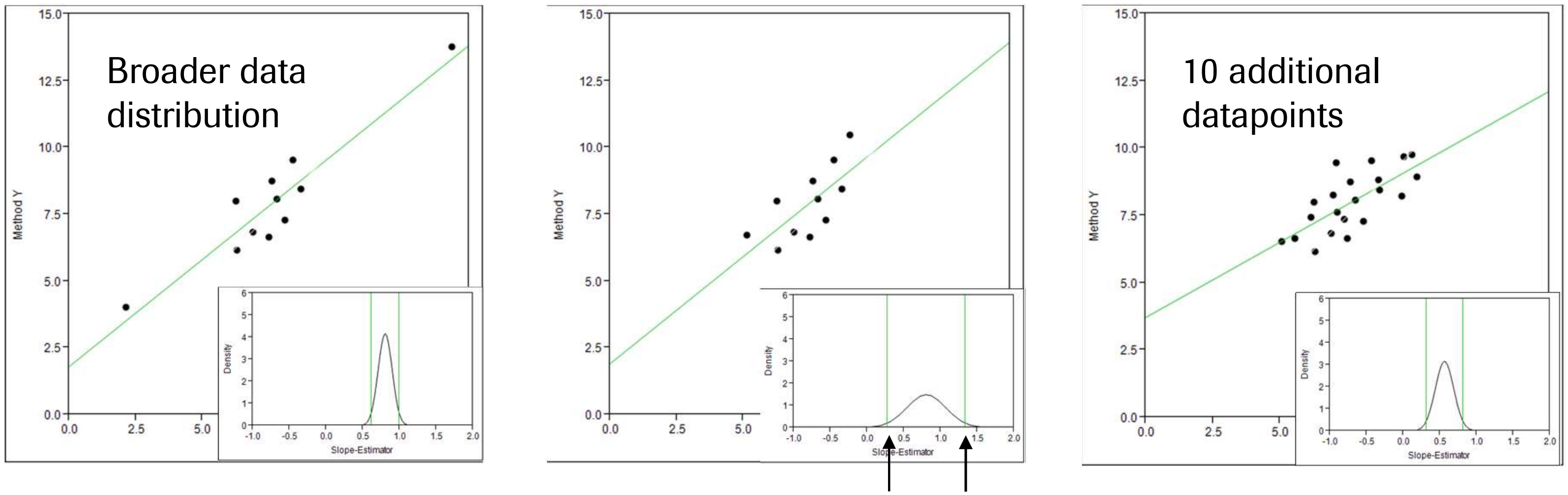
$a_{(1)} \leq a_{(2)} \leq a_{(3)} \leq a_{(4)}$

$$\hat{a} = \text{median}\{a_i,\ i = 1, \ldots n\}$$



Confidence Intervals (CIs)
Influencing Factors

- The width of the confidence intervals for intercept and slope depends on multiple factors:
  - Width of the distribution of the datapoints (often underestimated) → Broader data distribution → narrower CIs
  - Measurement errors of both measurement methods → Smaller measurement error → narrower CIs
  - Number of datapoints (often overestimated) → More data points→ narrower CIs
  - For intercept and bias in addition: Distance of the center of the data to the intercept or medical decision



Broadness of CI for slope

Result interpretation
Intercept and Slope

Regression Method	Parameter	Estimate	Std.Error	95LCL	95UCL
Deming	Intercept	0.34	0.20	-0.05	0.74
Deming	Slope	0.93	0.02	0.89	0.97
Weighted Deming	**Intercept**	**0.14**	**0.14**	**-0.15**	**0.42**
Weighted Deming	**Slope**	**0.95**	**0.02**	**0.92**	**0.99**
Passing Bablok	Intercept	0.24	NA	-0.06	0.69
Passing Bablok	Slope	0.95	NA	0.90	0.99

Mean measurement results of Lab 1 can be recalculated into mean measurement results of Lab 2, by the given regression equation

Lab2 = 0.14 + 0.95 * Lab1,

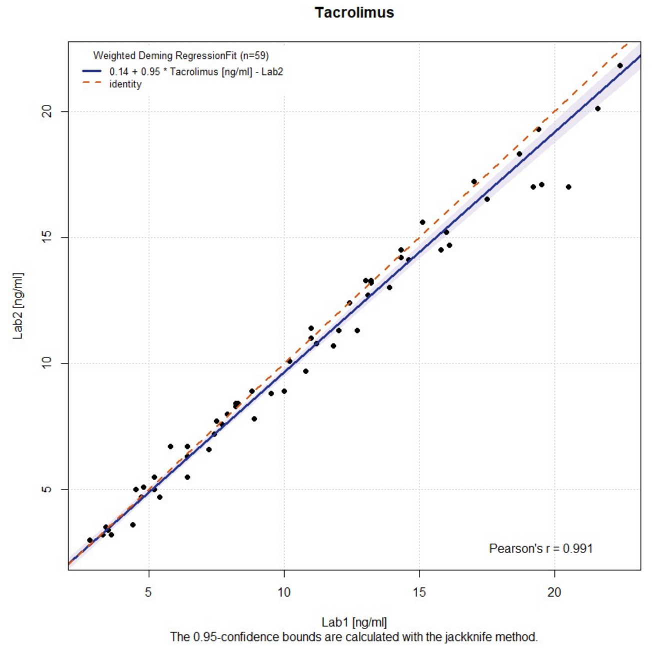
A mean result of 3 ng/ml in Lab1 results in a mean value of

0.14 + 0.95*3 = 2.99 in Lab2 → almost the same

and 15 ng/ml in **0.14 + 0.95*15 = 14.39** in Lab2. → higher difference

A slope < 1 is compensated in some concentration ranges by a positive intercept.

A slope > 1 is compensated in some concentration ranges by a negative intercept.



Result interpretation
Bias at medical decision points

- When talking about slope and intercept two values are needed for interpretation.
- The bias at the medical decision point is the difference between a mean value in Lab 1 and the predicted value in Lab 2

Lab2 = 0.14 + 0.95 * Lab1, eg.

a mean result of 3 ng/ml in Lab1 results in a mean values of **0.14 + 0.95*3 = 2.99** in Lab2

→ Bias = 2.99 ng/ml - 3 ng/ml = -0.01 ng/ml

and of 15 ng/ml in **0.14 + 0.95*15 = 14.39** in Lab2.

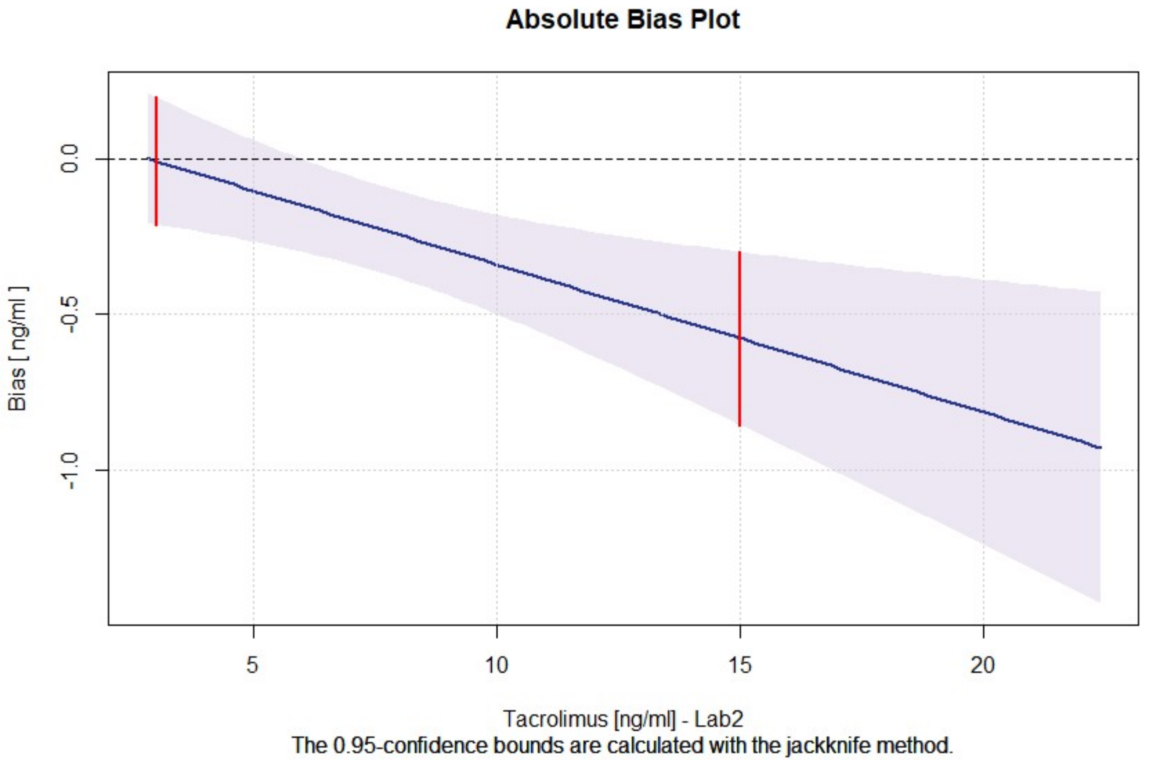
→ Bias = 14.39 ng/ml – 15 ng/ml = -0.58 ng/ml

Bias can be expressed as absolute or relative/percent bias.

Absolute Bias Plot Example

Regression Method	Level [ng/ml]	Abs. Bias [ng/ml]	SE [ng/ml]	95LCL [ng/ml]	95UCL [ng/ml]
Deming	3	0.14	0.14	-0.14	0.41
Deming	15	-0.70	0.16	-1.02	-0.37
Weighted Deming	**3**	**-0.01**	**0.10**	**-0.21**	**0.20**
Weighted Deming	**15**	**-0.58**	**0.14**	**-0.86**	**-0.30**
Passing Bablok	3	0.10	NA	-0.10	0.43
Passing Bablok	15	-0.49	NA	-0.85	-0.10

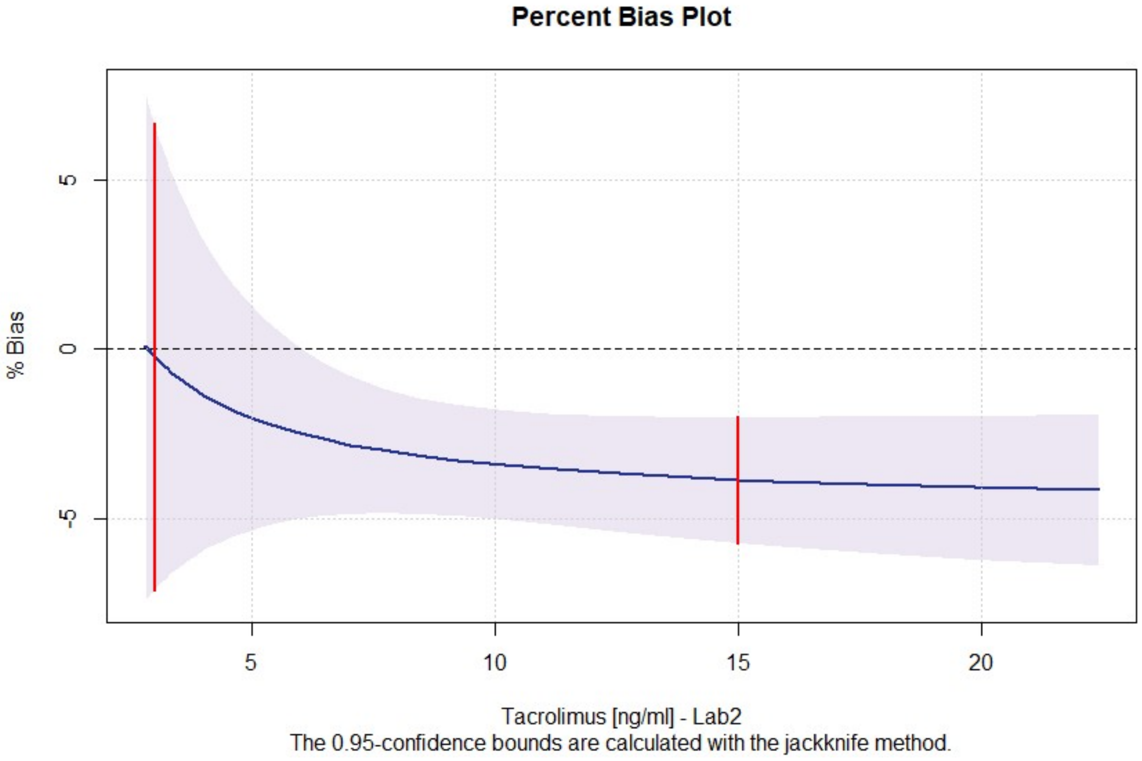
Absolute Bias Plot, based on Weighted Deming Regression



Relative Bias Plot - Example

Regression Method	Level [ng/ml]	Prop.bias(%)	SE(%)	95LCL(%)	95UCL(%)
Deming	3	4.53	4.56	-4.60	13.66
Deming	15	-4.64	1.08	-6.81	-2.48
Weighted Deming	**3**	**-0.22**	**3.43**	**-7.09**	**6.65**
Weighted Deming	**15**	**-3.84**	**0.93**	**-5.71**	**-1.97**
Passing Bablok	3	3.22	NA	-3.33	14.44
Passing Bablok	15	-3.29	NA	-5.65	-0.67

Percent Bias Plot, based on Weighted Deming Regression



Overview

- Method Comparison Experiments
  - Experimental Design
  - Analysis
  - Interpretation
- Precision – Variance - Components Experiments
  - Experimental Design
  - Analysis
  - Interpretation

Precision Experiment for LC/MS methods

Identification of sources of variability

- Calibrator preparation
- Calibration
- Sample preparation
- Injection and measurement on the LC/MS instrument

→ Precision experiment with dedicated variance-components [2]:

- Calibrator preparation per day, for 5 days
- Two calibrations of the instrument for each day
- Three sample preparations per calibration
- Two injections per sample preparation

5 x 2 x 3 x 2 = 60 measurements per sample

Precision experiments – Important Terms

- **Repeatability** – Closeness of the agreement between results of successive measurements of the same measure and **carried out under the same conditions of measurement**[1]
- **Reproducibility** - Closeness of the agreement between results of measurements of the same measurand **carried out under changed conditions of measurement**[1]
- **Intermediate precision** – Precision **under intermediate precision conditions**[2]
- **Intermediate precision conditions** – Where test results are obtained with the same method, on identical test items in the same test facility, under some different operating conditions; e.g. between-run, between-day, within-device[2]

[1]VIM – International Vocabulary of Basic and General Terms in Metrology

[2]EP5-A2

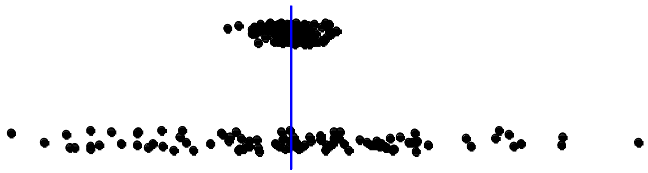
Measures of dispersion

- Results of precision experiments are given in standarddeviation (SD) or coefficient of variation (CV %)
- Both are known as measures of dispersion – describe the deviation of the datapoints from the center of the data distribution.

Same mean values

Low variability around the mean

High variability around the mean



Variance, SD, CV

- **Variance**: "mean of the quadratic distances from the mean"

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

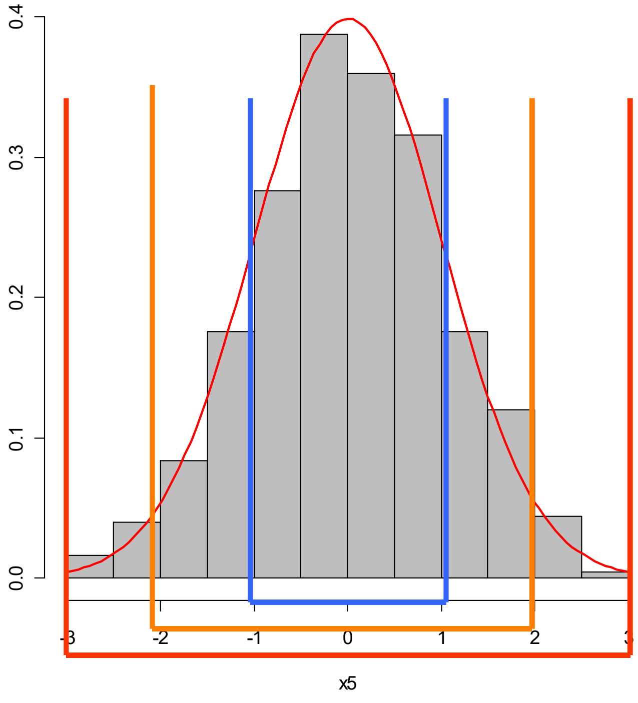
Why quadratic distance? $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

- **Standarddeviation (SD): s,** Square root of the variance
  - Same unit as your data: mg/L →Variance has unit $(mg/L)^2$, SD has unit mg/L

- **Coefficient of Variation (CV) SD(x)/MW(x) *100%**
  - Ratio of Standarddeviation to Mean → „Normalization"
  - Percent unit (%), good for comparison of variability among the measuring range.

Standard deviation, CV interpretation

- **68%** of the measurement values are around the mean value +- 1 SD
- **95.45%** of the measurement values are around the mean value +- 2 SD
- **99.73%** of the measurement values are around the mean value +- 3 SD



Interpretation for precision experiments:

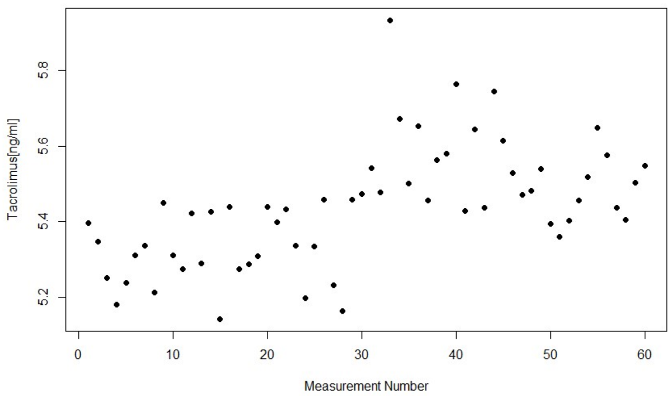
Intermediate precision CV: 6% → 95% of the measurement values are around the mean value ±2*6%

Measuring results in the lab can deviate up to 24% (Minimum to Maximum).

Precision Experiments

Quantification of relevant sources of variability

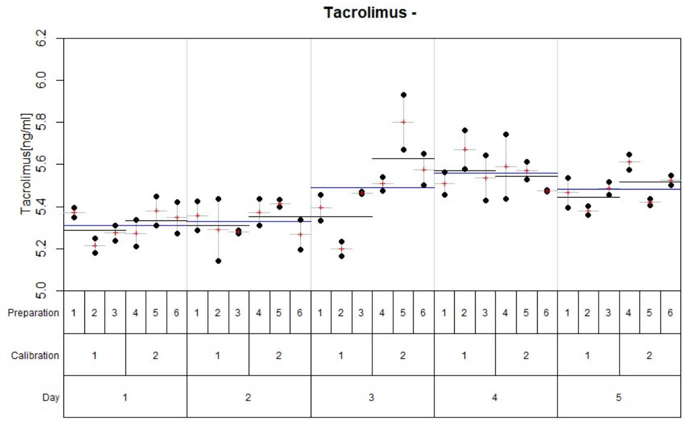
Unstructured look at the data



	Mean	Var	SD	CV [%]
	5.43	0.02	0.16	2.87

- Completely unstructured data
- Source of variability cannot be inferred without structure
- Best estimate of total variability is variance of all points, however this is a wrong estimate of total variability!

Account for the experimental structure



	DF	VC	%Total	SD	CV[%]
Total	17.04	0.03	100.00	0.16	2.98
Day	4.00	0.01	28.34	0.09	1.59
Calibration	5.00	0.01	22.20	0.08	1.40
Preparation	20.00	0.00	14.05	0.06	1.12
Error	30.00	0.01	35.42	0.10	1.77

- Experiment structure indicated; Source of variability visible
- Understand, where main variability comes from

Where does most of the variability come from?

The variability chart

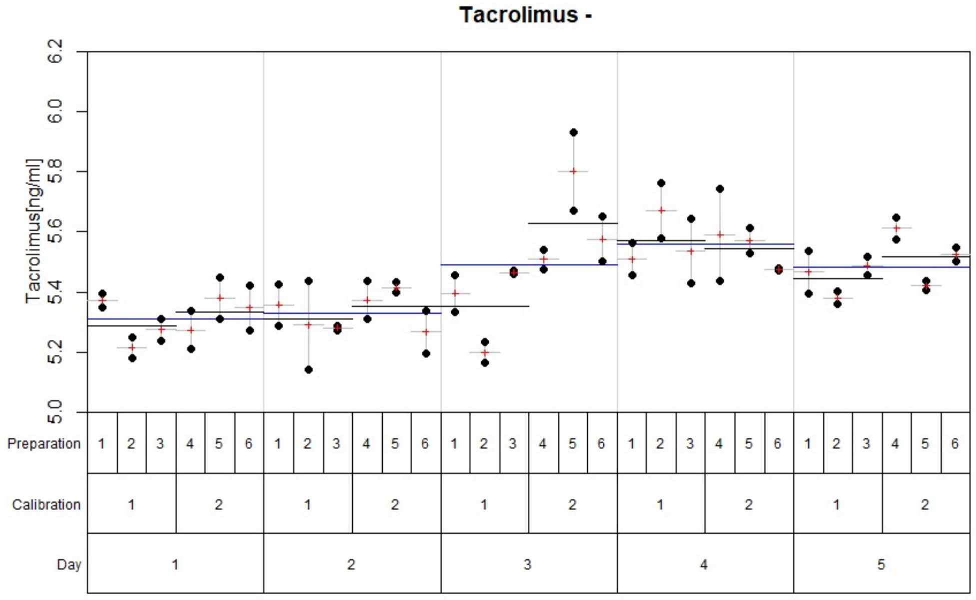
First observations of the dataset

Sample.Name	Analyte	Day	Calibration	Preparation	Injection	Concentration
Native A	Tacrolimus	1	1	1	1	5.39
Native A	Tacrolimus	1	1	1	2	5.34
Native A	Tacrolimus	1	1	2	1	5.25
Native A	Tacrolimus	1	1	2	2	5.18
Native A	Tacrolimus	1	1	3	1	5.23
Native A	Tacrolimus	1	1	3	2	5.31

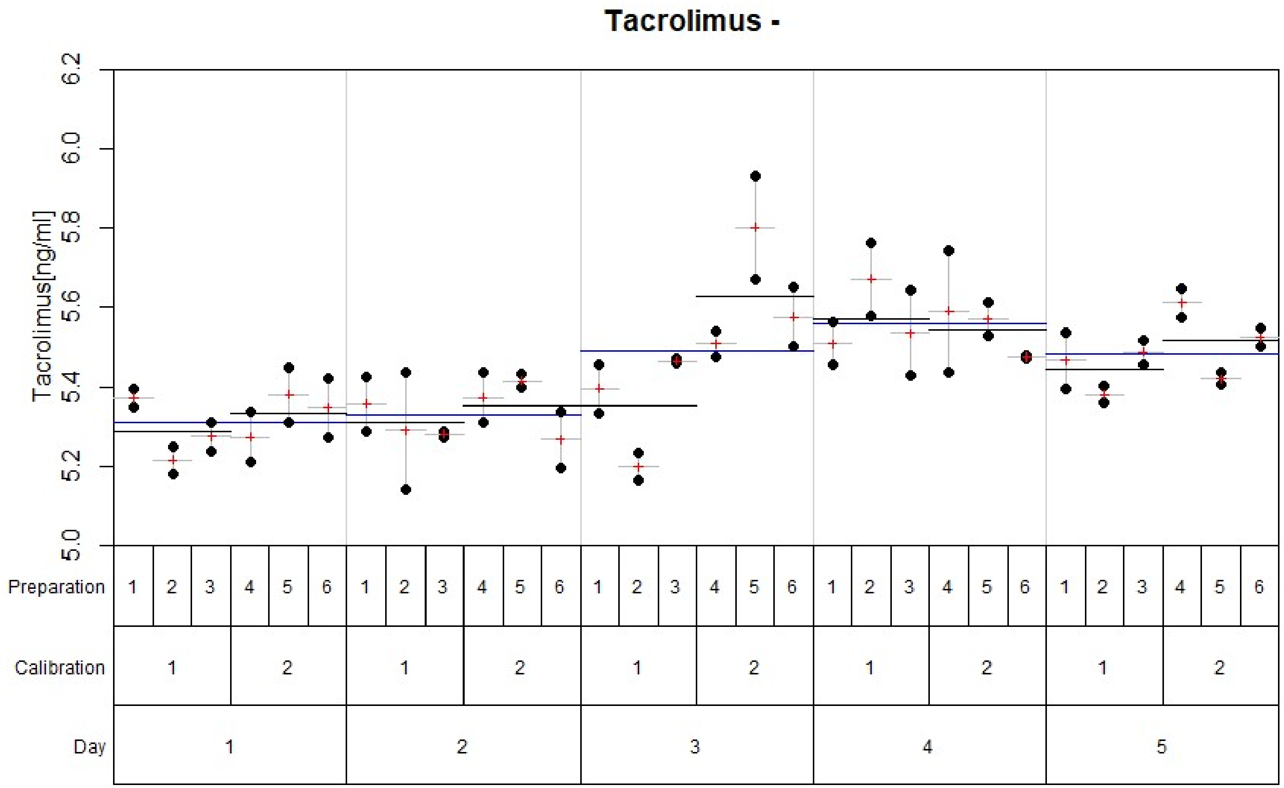
Concentration ~ Day/Calibration/Preparation

Look at this graph to

- Check your data,
- Identify the sources of variability,
- Identify missing or outlying observations



Variance-Components Estimation



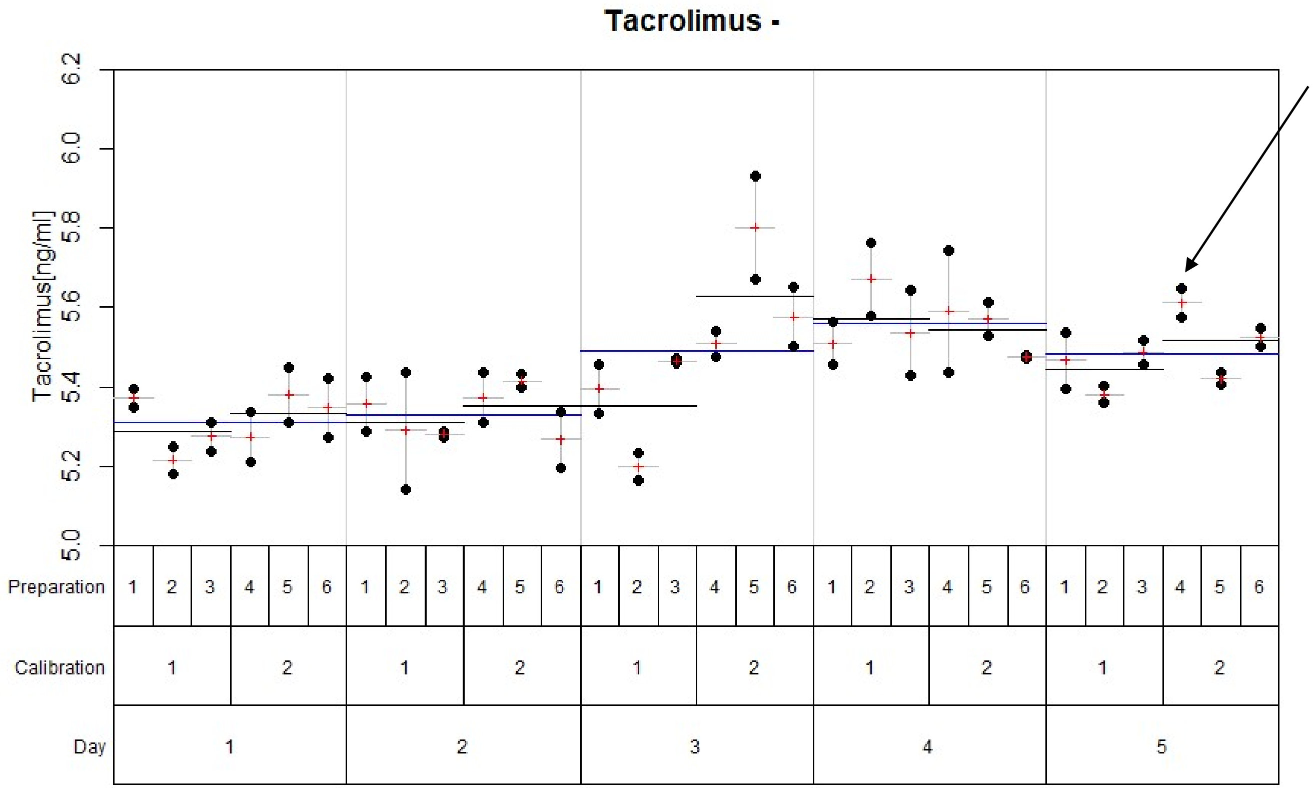
From this dataset the following variance-components can be estimated:

- Repeatability/Within-Sample Preparation variability
- Between-Sample Preparation Variability
- Between-Calibration Variability
- Between-Calibrator Preparation

== Between-Day Variability (in this experiment)

- Total Variability == Intermediate Precision

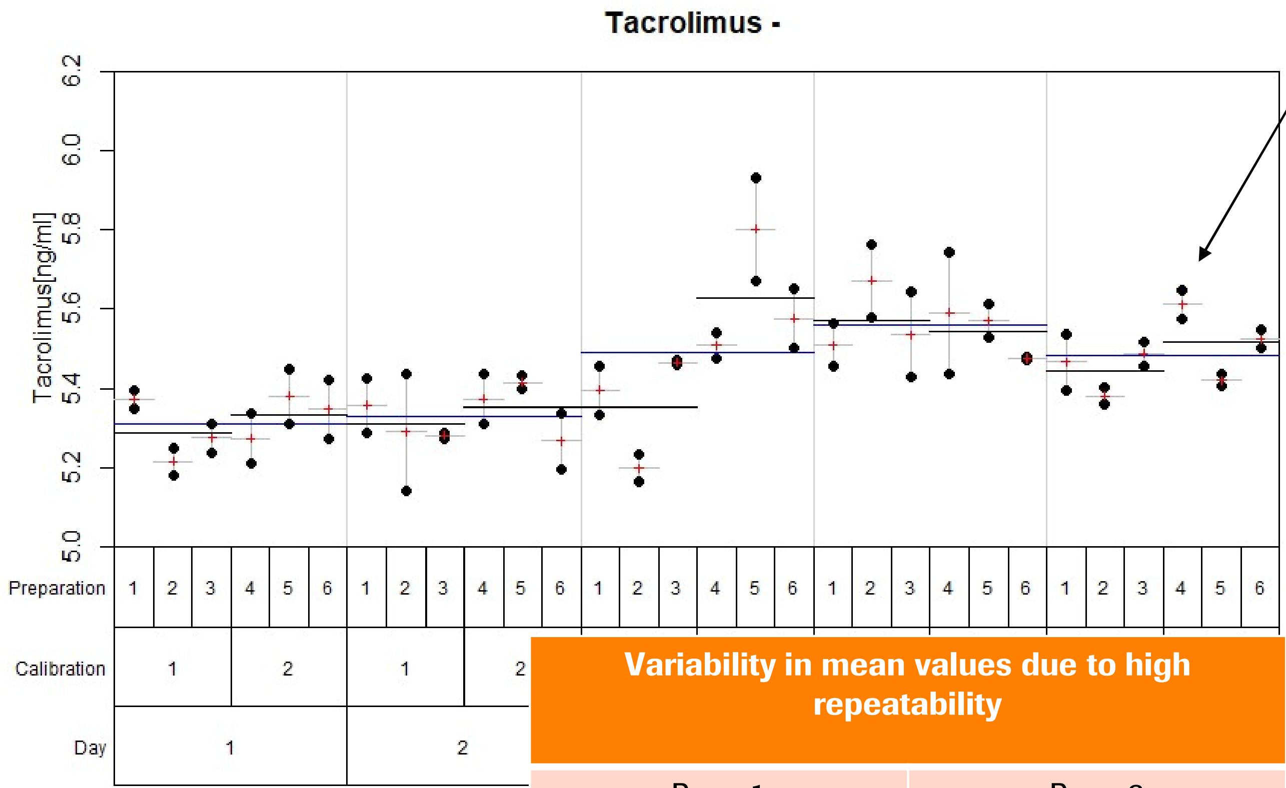
Repeatability Estimation



30 pairs of replicates on the lowest variability level

Repeatability is estimated as the mean variability of the 30 pairs of replicates.

Between-Sample Preparation Variability Estimation

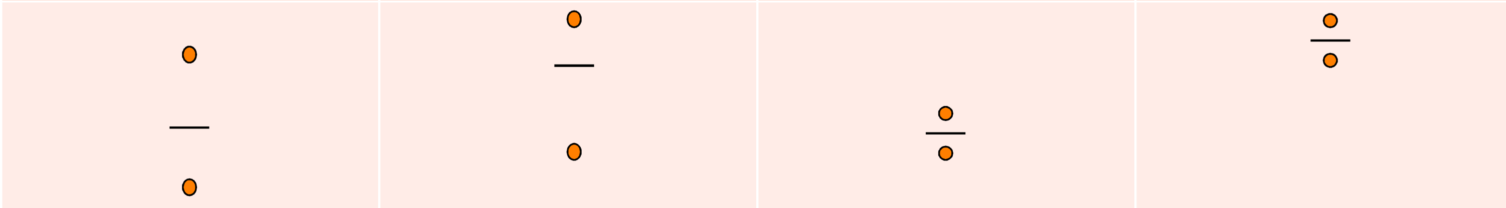


30 mean values from sample preparation events

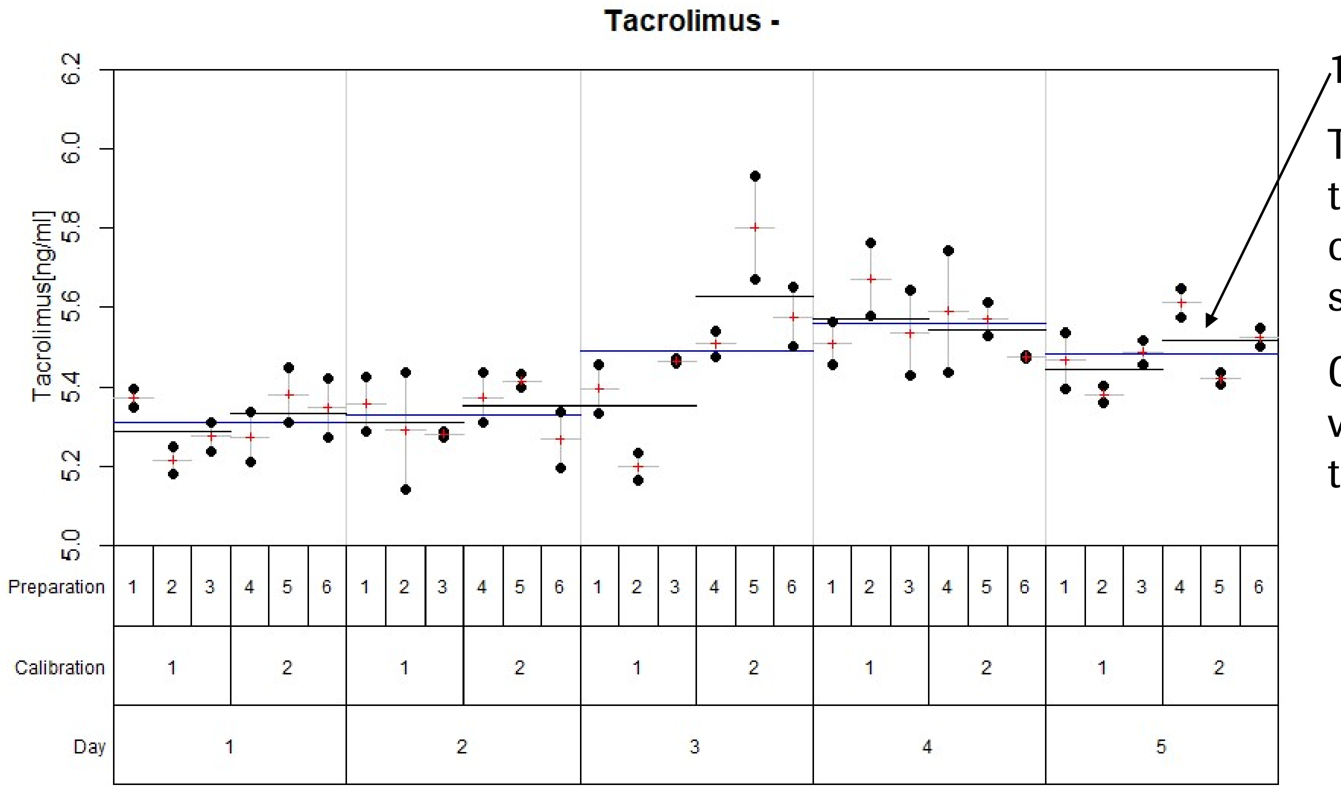
The between-sample preparation variability is the variability of the 30 mean values, corrected by the estimated repeatability.

Correction by repeatability ensures that only the variability part due to sample preparation is taken into account.

Variability in mean values due to high repeatability		Variability in mean values due to high between-preparation variability	
Prep. 1	Prep. 2	Prep. 1	Prep.2



Between-Calibration Variability Estimation

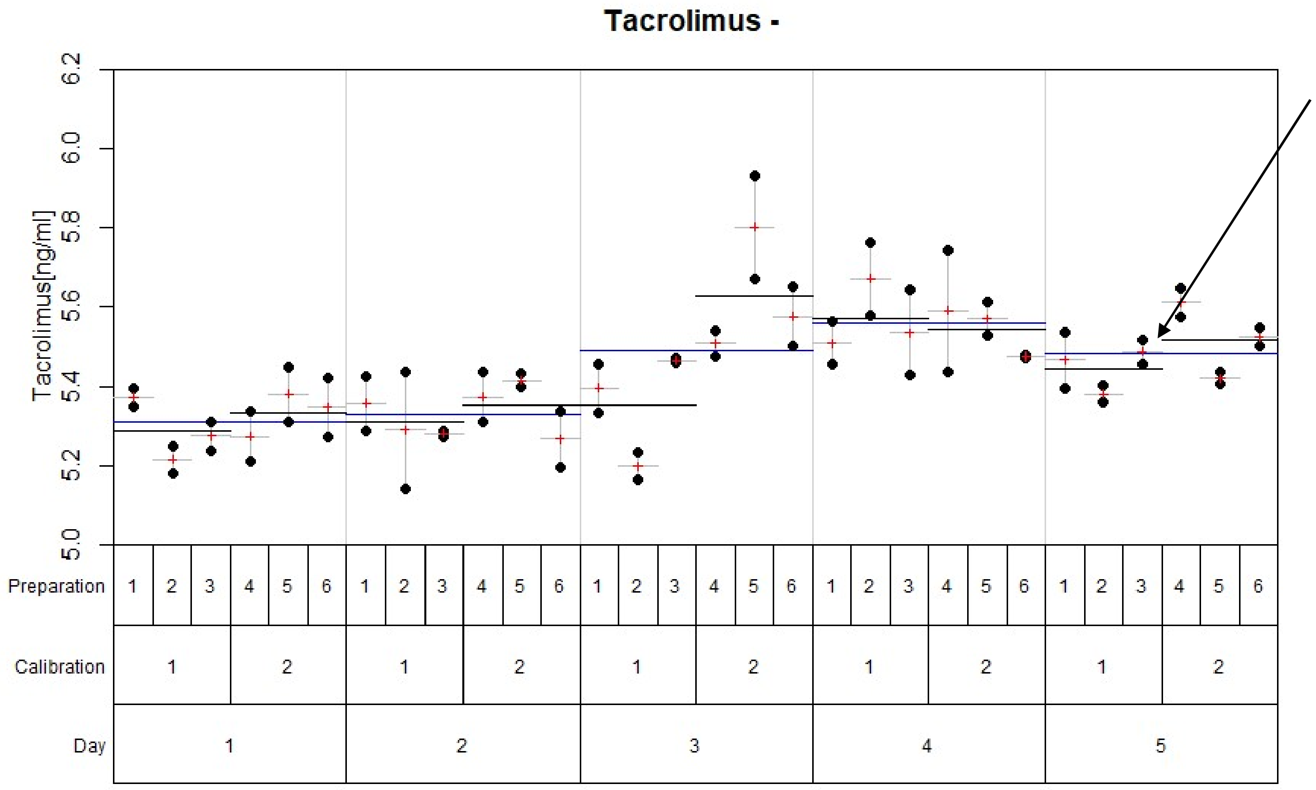


10 mean values from calibration events

The between-calibration variability is the variability of the 10 mean values, corrected by the estimated between-sample preparation variability.

Correction ensures that only the variability part due to calibration is taken into account.

Between-Calibration Preparation Variability Estimation

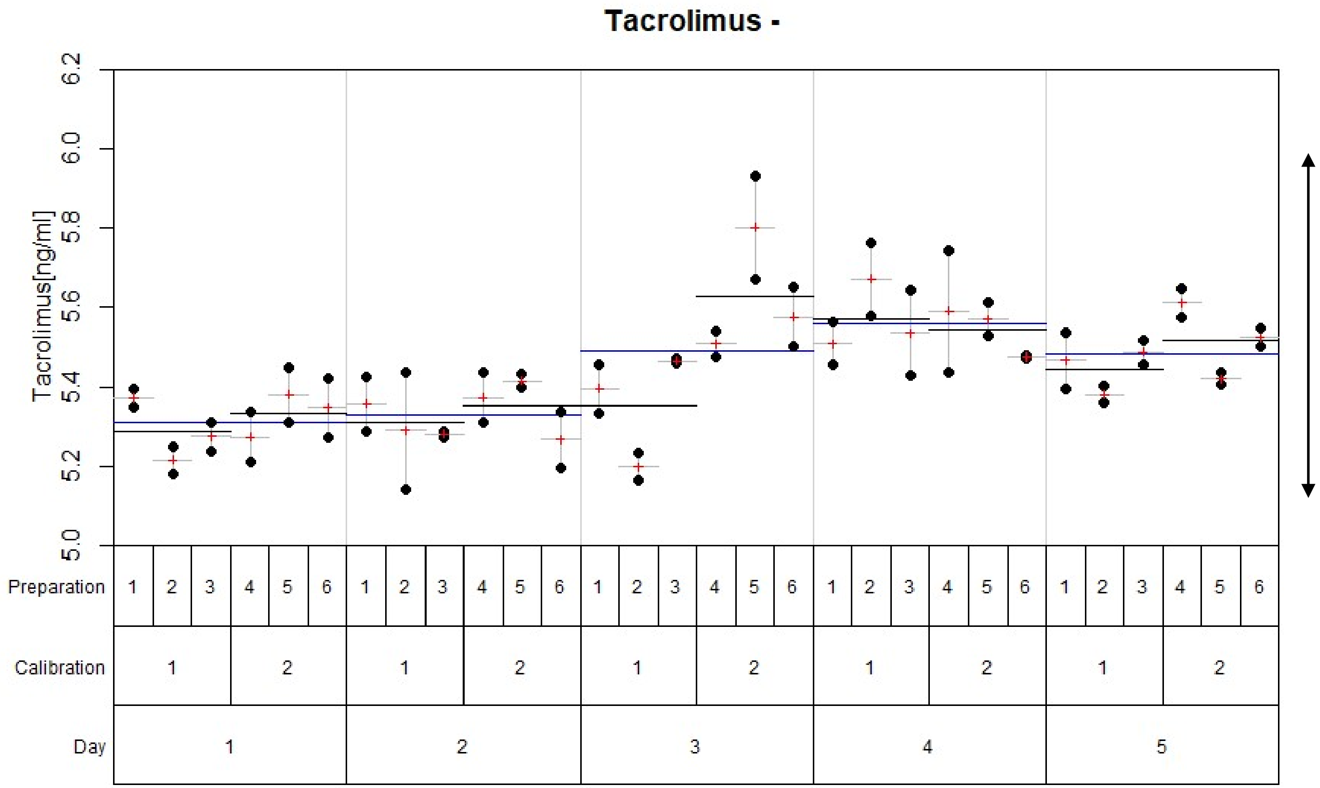


5 mean values from calibrator preparation events

The between-calibration preparation variability is the variability of the 5 mean values, corrected by the estimated between-calibration variability.

Correction ensures that only the variability part due to calibrator preparation is taken into account.

Total Variability – Intermediate Precision Estimation



Total variability (SD) is estimated as the square root of the sum of the squared variance-components SDs.

$$\text{SD(Total)} = \sqrt{SD^2(CPrep) + SD^2(Calibration) + SD^2(SPrep) + SD^2(Repeat)}$$

Result output table

Variance components for Day, Calibration and Preparation

Concentration ~ Day/Calibration/Preparation

	Degrees of freedom	Sum of Squares and Mean Squares		Variance	% of Total Variability	Standard Deviation	Coefficient of Variation in %
	DF	SS	MS	VC	%Total	SD	CV[%]
total	17.04	NA	NA	0.03	100.00	0.16	2.98
Day	4.00	0.56	0.14	0.01	28.34	0.09	1.59
Day:Calibration	5.00	0.26	0.05	0.01	22.20	0.08	1.40
Day:Calibration:Preparation	20.00	0.33	0.02	0.00	14.05	0.06	1.12
error	30.00	0.28	0.01	0.01	35.42	0.10	1.77

Because of the correction of the higher variance components, by the lower ones, % of Total Variability for each variance component can be calculated.

Rownames define the individual variance components, based on the given formula

total – Sum of all variance-components, overall variability

Day – Variability between the days or calibrator preparations

Day:Calibration - Variability between calibrations within a day

Day:Calibration:Preparation – Variability between sample preparations, within a calibration

error – Repeatability, precision on the lowest level

Confidence intervals of variance-components

Confidence intervals for variance components can be calculated based on Chi-Square distributions. **Degrees-of-freedom (DF)** are the main parameter of the Chi-Square distribution.

The higher the DF, the narrower the distribution. → High DFs, narrow confidence intervals.

DFs for the variance-components are determined by the

- numbers of levels of the variance-component factor

- % of total variability of the factor: Higher the % of total variability, more levels are necessary. → makes it difficult to get the optimal sample size of the individual factors before the experiment.

(Henn-egg-problem)

Name	DF	LCL	CV[%]	UCL
total	17.04	2.24	2.98	4.46
Day	1.45	0.77	1.59	18.27
Day:Calibration	2.23	0.75	1.40	7.48
Day:Calibration:Preparation	3.24	0.64	1.12	3.85
error	30.00	1.42	1.77	2.37

Asymmetric CIs occur, with small DFs.

Individual Variance-components can have low DFs, although total precision can be estimated good enough.

CIs for CVs are based on the CIs of SDs. No further adjustment for the uncertainty of the mean.

References

- CLSI EP9c – Measurement Procedure Comparison and Bias Estimation Using Patient Samples, 3rd Edition, 2018
- Bland JM, Altman DG, Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-310
- Linnet K. Evaluation of regression procedures for method comparison studies, Clin Chem 1993; 39(3):424-432
- 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 J. Clin Chem Clin Biochem. 1983; 21(11):709-720

References

- CLSI EP05-A3 – Evaluation of Precision of Quantitative Measurement Procedures, 3rd Edition, 2014
- Searle, S.R, Casella, G., McCulloch, C.E. (1992), Variance Components, Wiley New York
- Satterthwaite, F.E. (1946), An Approximate Distribution of Estimates of Variance Components., Biometrics Bulletin 2, 110-114 Gaylor,D.W., Lucas,H.L., Anderson
- mcr: Method Comparison Regression; Version 1.2.1, E. Manuilova et al, https://cran.r-project.org/, last visited Mai 2020
- J. Taibon et al, An isotope dilution LC-MS/MS based candidate reference method for the quantification of cyclosporine A, tacrolimus, sirolimus and everolimus in human whole blood, Clinical Biochemistry, Available online 23 November 2019
- VCA: Variance Component Analysis; Version 1.4.2, A. Schuetzenmeister et al, https://cran.r-project.org/, last visited Mai 2020