Research Article

Automated data analytics workflow for stability experiments based on regression analysis

Andrea Geistanger\textsuperscript{a,\,}\textsuperscript{*}, Kathrin Braese\textsuperscript{a,\,1}, Ruediger Laubender\textsuperscript{a,\,1}

\textsuperscript{a} Roche Diagnostics GmbH, Staffelsteinstr. 4, 81477 Munich, Germany

A R T I C L E  I N F O

Keywords:
Sample stability
Reagent stability
Stability studies
Data analytics workflow
CLSI EP25

A B S T R A C T

Introduction: We define a designated data analytics workflow for the evaluation of stability experiments, which takes all data situations into account. This complements the evaluation described by the CLSI EP25 [1] guideline by including a targeted exception handling algorithm and thus allows one to automatically evaluate stability data based on linear regression analysis.

Description: The evaluation of stability experiments based on regression analysis requires the calculation of the confidence interval of the regression line. The stability time is estimated at the intersection of the confidence interval with the acceptance criterion. This approach results in solving a quadratic equation, with factors that depend on the estimated intercept, slope, the measurement variability and the chosen timepoints. When defining an automated data analytics workflow for this problem, the different cases for the solutions of the quadratic equation must be considered. For some data situations there might be no solution at all, other data situations result in a negative and a positive solution and finally there might be even two positive solutions. All these cases have to be considered for the choice of the right solution to become the estimated stability time. The CLSI EP25 [1] guideline on stability evaluation of in vitro diagnostic reagents addresses this problem only superficially and might even lead to incorrect results for some specific data scenarios.

Results: We evaluate all possible data scenarios and provide examples for each. Based on the gained theoretical insights, we define a designated data analytics workflow and visualize it with a flowchart. By following this flowchart one can implement an automated analysis workflow, targeting all data scenarios with the appropriate exception handling.

Discussion: We deduce that the description for obtaining stability times according to CLSI EP25 is not fully adequate, as it addresses only best-case scenarios. However, for automated data analytics workflows all possible data situations have to be considered. With the here presented workflow one can program automated data analytics pipelines, which ensure that the right stability time is obtained, in case it exists. In addition all exceptions, where no stability times are present, are addressed in the right way and it provides hints as to the failure reason.

Background

The calculation of stability times is a mandatory component of diagnostic assay development. Stability must be proven for reagent material, for the analyte in calibrator and control samples and for the analyte in native samples taking into account respective pre-analytical handling or post-analytical storage conditions of the samples. Herein we discuss the analytical approach for stability studies, where one side of a 95% confidence interval of a linear regression line intersects the predefined acceptance criterion as proposed, for example, by the ICH Q1A(R2) guideline on stability testing of drug substances and products [2] and the CLSI EP25 guideline on the stability of in vitro diagnostic reagents [1].

For illustration purposes, we take the example data introduced in CLSI EP25 guideline [1] Appendix A. The example employs data from a shelf-life study for a new ferritin IVD reagent with study duration of 13

Abbreviations: CLSI, Clinical and Laboratory Standards Institute; ICH, International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use; IVD, In-Vitro Diagnostic; QC, Quality Control; Var(Residual), Residual Variance; EFLM, European Federation of Clinical Chemistry and Laboratory Medicine.

* Corresponding author.

E-mail addresses: andrea.geistanger@roche.com (A. Geistanger), kathrin.braese@roche.com (K. Braese), ruediger.laubender@roche.com (R. Laubender).

\textsuperscript{1} Equal contribution.

https://doi.org/10.1016/j.jmsacl.2022.01.001

Available online 8 February 2022
2667-145X/© 2022 THE AUTHORS. Publishing services by ELSEVIER B.V. on behalf of MSACL. This is an open access article under the CC BY-NC-ND license
months. Test samples include 5 calibrator samples, 2 QC samples and a native pooled sample, measured on 10 testing days, between baseline and day 402. In Fig. 1 we show the data of the native pooled sample, with the estimated regression line and the one-sided upper and lower 95% confidence limits. As the slope of the regression line is positive, the estimated stability time for this sample is given by the intersection of the upper one-sided 95% confidence limit with the acceptance criterion (grey arrow).

**Fig. 1.** Regression analysis of the native sample pool sample of the CLSI EP25 data example. The black line shows the estimated regression line, the dashed lines show the one-sided 95% upper and lower confidence interval. The acceptance criterion is at ±5% and is indicated as a black, horizontal line, since the estimated slope is positive. The estimated stability time is given at the intersection of the upper confidence interval with the acceptance criterion (grey arrow).

Although both mentioned guidelines provide formulas for the evaluation of stability times based on the mentioned approach, these formulae lack relevant details and do not address worst-case scenarios. The stability data of the native pooled sample has low variability around the regression line, compared with the acceptance criterion and there is a clearly increasing trend. However, less clear data situations might occur, especially when implementing this approach in statistical software tools. We observe that all scenarios need to be considered and clearly defined in order to avoid incorrect results for stability times.

An automated data analytics workflow for the evaluation of stability data must report the correct data-derived stability time if it exists, or, otherwise, report that it is not possible to estimate a stability time for these data. Such a workflow should also report the specific reason why no estimated stability time could be derived and indicate the next steps for the user. In terms of reproducibility and error avoidance it is unacceptable that the user has to visually check graphs, manually verify the correctness of calculated stability times or identify the root cause for the reason why no stability time could be estimated by the algorithm. Visual inspections are, of course, cumbersome and subjective.

Therefore, we present a data analytics flowchart for estimation of the stability time based on the intersection of the one-sided 95% confidence interval of a linear regression line and predefined acceptance criterion. In doing so, we follow the same logic which underlies the stability estimation calculations implemented in the statistical software SigmaPlot [3][4]. However, we did not find any clear description of the flowchart, which would allow us to reproduce the SigmaPlot approach.

In the following, we give a short overview of the statistical background of linear regression, the confidence interval approach and the quadratic equation solving. For better understanding will follow up on the example data from the CLSI EP25 guideline, given in Appendix A of this guideline together with some artificial data, which present more extreme scenarios. Although these scenarios might occur infrequently, it is important to handle them properly in an automated data analytics workflow.

**Applied Methods**

Let $Y_{it}$ denote the result of a measurement of an analyte for sample $i \in \{1, 2, \ldots, n_i\}$ at time point $t \in \{1, 2, \ldots, n_t\}$ with replicate $r \in \{1, 2, \ldots, n_r\}$.

Denote by $Y_i$ the mean of the replicates of sample $i$ at time point $t$, that is

$$Y_i = \frac{1}{n_r} \sum_{r=1}^{n_r} Y_{it}.$$  \hfill (1)

Let $X_t$ be the time of measurement, with $X_1 = 0$ the first time of measurement, called also baseline measurement, and with $X_n = T$ the last time of measurement, with $T \in \mathbb{R}^+$, with $n \geq 3$. The linear regression analysis for each sample can either be based on the measured concentration values or on the percentage deviation from the different time points to the baseline measurement. The choice here depends on the concentration of the sample and the absolute value of the measurements. For samples with low analyte concentration an analysis based on recoveries for higher concentrated samples might be more comparable.

We denote the percentage deviation of the mean of replicates at time point $t$ from the mean of the replicates at $t_1$ with

$$Z_t = \left( \frac{Y_t}{Y_{t_1}} - 1 \right) \times 100.$$  \hfill (2)

We present our results for percentage deviation values. However, they can be easily replaced by measured concentrations. For each sample $i$, a linear regression between the result $Z_{it}$ and the time $X_t$ is assumed, that is

$$Z_{it} = a_i + b_i \times X_t + \varepsilon_{it},$$  \hfill (3)

where $a_i$ denotes the intercept, $b_i$ denotes the slope and $\varepsilon_{it}$ denotes the random error. It is assumed that the random errors are identically and independently normally distributed with mean 0 and variance $\sigma^2$.

**Linear regression modelling**

For this bivariate linear regression model, estimators of the coefficients $a_i$ and $b_i$ are given by

$$\hat{b}_i = \frac{\sum_{t=1}^{n_t} (Z_{it} - \bar{Z}_i) \times (X_t - \bar{X})}{\sum_{t=1}^{n_t} (X_t - \bar{X})^2}$$  \hfill (4)

with

$$Z_i = \frac{1}{n_t} \sum_{t=1}^{n_t} Z_{it}$$  \hfill (5)

and

$$X_i = \frac{1}{n_t} \sum_{t=1}^{n_t} X_t.$$  \hfill (6)

The intercept is estimated by

$$\hat{a}_i = Z_i - \hat{b}_i \times \bar{X}.$$  \hfill (7)
Predicted values can be calculated as
\[ \hat{Z}_i(X) = \hat{a}_i + \hat{b}_i \times X. \]

where \( X \in \mathbb{R}^+ \) denotes any time point not necessarily used in the design.

The difference between a percent deviation at time \( t \) and its predicted value \( Z_a - \hat{Z}_a \) is called the residual and the estimator for the standard deviation of the residuals is given by
\[ \hat{\sigma}_i = \sqrt{\frac{\sum_{i=1}^{n} (Z_a - \hat{Z}_a)^2}{n_t - 2}} \]  \hspace{1cm} (9)

The squared standard deviation of the residuals is called the variance of the residuals and is denoted by \( \hat{\sigma}_i^2 \).

For stability analysis, a one-sided \( 1 - \alpha \%- \)confidence interval for a predicted value has to be calculated. In most cases the analysis of stability studies is based on \( \alpha = 0.05 \) or \( \alpha = 0.01 \), however the calculations of stability times can be performed for all values with \( \alpha < 0.5 \). Without loss of generality we regard the case of \( \alpha = 0.05 \). The calculation of the confidence interval depends on the direction of the slope:

- If the estimated slope \( \hat{b}_i \) is positive, then the confidence interval limit is given by
  \[ \hat{L}_i(X) = \hat{Z}_i(X) + t_{1(1-\alpha df)} \times \hat{\sigma}_i \times \sqrt{\frac{1}{n_t} \times \left( \frac{(X - \bar{X})^2}{TSS_X} \right)} \]  \hspace{1cm} (10)

- If the estimated slope \( \hat{b}_i \) is negative, then the confidence interval is given by
  \[ \hat{U}_i(X) = \hat{Z}_i(X) - t_{1(1-\alpha df)} \times \hat{\sigma}_i \times \sqrt{\frac{1}{n_t} \times \left( \frac{(X - \bar{X})^2}{TSS_X} \right)} \]  \hspace{1cm} (11)

where \( t_{1(1-\alpha df)} \) is the critical value of Student’s central t-distribution usually with \( \alpha = 0.05 \) as significance level and with \( df = n_t - 2 \) degrees of freedom and
\[ TSS_X = \sum_{i=1}^{n} (X_i - \bar{X})^2 \]  \hspace{1cm} (12)
as total sum of squares for the variable \( X \).

From (10), (11) we see that a confidence interval can only be estimated for the given stability design if \( \hat{\sigma}_i > 0 \) holds. This implies that at least three observations are present, and that these observations do not fall on a single line. Such a “perfect fit” situation occurs in experiments where the resolution of the measurements is lower than the ongoing degradation. We discuss the calculation steps for this case in more detail below.

The acceptance criterion is defined as the maximum limit of acceptance of the one-sided confidence interval of the regression line. If the regression analysis is based on absolute concentration values, an absolute criterion will be applied. If, however, the percentage deviation from the baseline mean is used as data to calculate the regression line, a relative criterion will be used. The relative criterion is defined as percentage deviation from \( Y_{RL} \) and denoted by \( \delta_i \in \mathbb{R} \). In case of a negative slope \( \hat{b}_i \) the acceptance criterion \( \delta_i \) is negative and in case of a positive slope \( \hat{b}_i \) the acceptance criterion \( \delta_i \) is positive.

When an absolute criterion is given, but the analysis shall base on relative deviations, the absolute criterion has to be converted to a relative criterion: Denote the absolute criterion by \( \eta_i \). The corresponding relative criterion is then given by \( \delta_i = \frac{\eta_i}{Y_{RL}} \).

For the estimation of the stability time \( X_{st} \), we first consider the case that \( \delta_i > 0 \) holds. Based on the confidence interval, the stability time \( X_{st} \) is calculated as intersection between the acceptance criterion \( \delta_i \) and the one-sided 95% confidence interval.

Solving \( \hat{L}^+_i = \delta_i \) or \( \hat{U}^-_i = \delta_i \) for \( X \) leads to the following quadratic equation [1]:
\[ C_1 \times X^2 + C_2 \times X + C_3 = 0, \]  \hspace{1cm} (13)

where the coefficients of the quadratic equation are
\[ C_1 = 1 - A \times \hat{b}_i^2, \]  \hspace{1cm} (14)
\[ C_2 = 2 \times A \times \hat{b}_i - (\delta_i - \hat{a}_i) - 2 \times \bar{X}, \]  \hspace{1cm} (15)
\[ C_3 = \bar{X}^2 + \frac{TSS_X}{n_t} - A \times (\delta_i - \hat{a}_i)^2 \]  \hspace{1cm} (16)

with
\[ A = \frac{TSS_X}{t_{1(1-\alpha df)} \times \hat{\sigma}_i}. \]  \hspace{1cm} (17)

It is well known that solving the quadratic equation in (13) results in two roots which are given by [1]:
\[ X_{st}^1 = -\frac{C_2 + \sqrt{C_2^2 - 4 \times C_1 \times C_3}}{2 \times C_1}, \]  \hspace{1cm} (18)
\[ X_{st}^2 = -\frac{C_2 - \sqrt{C_2^2 - 4 \times C_1 \times C_3}}{2 \times C_1}. \]  \hspace{1cm} (19)

In order to obtain numerically stable solutions, we change the equations (18) and (19) into equivalent formulas, dependent on the sign of \( C_2 \). This protects against catastrophic cancellation when doing subtraction [5].

\[ C_2 < 0 : X_{st}^1 = \frac{2 \times C_1}{-\sqrt{C_2^2 - 4 \times C_1 \times C_3}}; C_2 > 0 : X_{st}^1 = \frac{-C_2 - \sqrt{C_2^2 - 4 \times C_1 \times C_3}}{2 \times C_1}. \]  \hspace{1cm} (20)

\[ C_2 < 0 : X_{st}^2 = \frac{2 \times C_1}{C_2 + \sqrt{C_2^2 - 4 \times C_1 \times C_3}}; C_2 > 0 : X_{st}^2 = \frac{-C_2 + \sqrt{C_2^2 - 4 \times C_1 \times C_3}}{2 \times C_1}. \]  \hspace{1cm} (21)
In its mathematical essence, the problem of estimating stability times reduces to solving a quadratic equation and obtaining at least one real-valued, positive solution. The solutions for the quadratic equation in Eq. (13) are presented in [1,6]. However, none of the aforementioned publications provides guidance on how to verify if the mathematically valid solution for (13) also corresponds to the solution of the underlying problem which is finding the correct stability time. This becomes evident if both roots (18) and (19) have positive solutions or if the discriminant, i.e. the formula under the square root in (18) and (19), is negative so that there is no solution for (13) at all.

In addition, we have to ensure that the estimated intercept is lower than the acceptance criterion in case of a positive slope and higher than the acceptance criterion in case of a negative slope. If this is not the case, it means that already at baseline the stability criterion is not met. From a mathematical point of view, the aforementioned case may lead to (13) having one (or two) positive solution(s) which, however, would describe the intersection of the acceptance criterion with the opposite confidence interval limit at baseline. If (13) is not satisfied, then the upper confidence interval limit is smaller than the specification for a negative slope and the upper confidence interval limit at baseline. If (13) is satisfied, then the upper confidence interval limit is larger than the specification for a positive slope.

Comparison of the statistical t-test for the slope (see eg. [1]) with the acceptance criterion in case of a negative slope and higher than the specification for a positive slope and higher than the acceptance criterion in case of a positive slope and higher than the acceptance criterion in case of a negative slope is crucial.

For the definition of an adequate data analysis workflow and a software implementation which provides the user with the correct statistical t-test for the slope is \( p < 0.05 \). This means that the slope is statistically different from 0. Note that this is a major criterion within [1], for the estimation of the stability time. However, this criterion does not guarantee the estimation of a correct stability time.

For (13) we can derive the following important property: If \( C_3 < 0 \) holds, then \( \tilde{L}_i(0) < \tilde{\delta}_i \) or \( \tilde{L}_i(0) > \tilde{\delta}_i \) holds. The coefficient \( C_3 \) can be interpreted as the difference between the specification \( \tilde{\delta}_i \) and the upper confidence interval limit at baseline. If \( C_3 < 0 \) holds, then the upper confidence interval limit is below the specification for a positive slope and above the specification for a negative slope.

Finally, it is important to mention, that the discriminant of Eq. (13) can only become negative if \( C_1 > 0 \) and \( C_3 > 0 \), in case that the absolute value of the intercept is lower than the absolute value of the acceptance criterion.

Within the Results section we show how the interplay between the two factors \( C_1 \) and \( C_3 \) impact the solutions of the stability time calculation.

Note that both solutions (20) and (21) can result in stability times, which are greater than the maximal tested time range. The CLSI EP25 guidelines forbids extrapolation further than the maximal tested range, whereas other guidelines [2] allow for extrapolation, with additional measures of care. Certainly, one has to be very careful when extrapolating beyond the tested range, as it is not known whether the degradation model stays the same. Extrapolation always assumes that the change pattern will continue to apply beyond the tested time. However, it might make sense in some situations to allow for some extrapolation, i.e. 10% beyond the maximal time. Therefore, we do not include a extrapolation/no extrapolation rule within our examples or within the data analytics workflow. If there is a defined extrapolation rule, it should be added to the workflow in the last calculation step.

**Perfect linear fit**

For the sake of completeness, we also consider the case of a perfect linear fit implying that all data points fall on one perfect line such that \( \tilde{\delta}_i = 0 \). This situation typically arises when working with data whose resolution (in terms of the number of reported digits) is so low that all data points of the measurement series have the same value. It can be seen from (10) and (11) that such a case does not allow for the calculation of a confidence interval. Hence, there is only an intersection of the acceptance criterion with the regression line. Furthermore, the following cases have to be distinguished:

If the estimated slope is zero so that \( \tilde{b}_i = 0 \) holds, then the stability time \( X_{\text{est}} = X_{\text{at}} \), and all observations per timepoint.

If the estimated slope is non-zero, one has to check whether the absolute value of the intercept is lower than the absolute value of the acceptance criterion or not. If this is not the case, a positive stability time \( X_{\text{est}} = X_{\text{at}} \) does not exist, since the acceptance criterion is not met even at...
baseline. However, if the absolute value of the intercept is lower than the absolute value of the acceptance criterion, a positive stability time $X_{\text{est}}$ can be estimated by

$$X_{\text{est}} = \frac{\delta - \hat{a}}{\hat{b}}.$$  \hspace{1cm} (22)

**Results**

For better understanding and visualization of the aforementioned considerations on the calculation of stability times, we provide a calculation example including interpretation, based on the example data of the CLSI EP25 guideline [1] Appendix A. In addition, we present data on artificial samples to point out the limitations of the analysis approach provided in CLSI EP25 and to visualize the data scenarios which must be covered by a reliable workflow for the calculation of stability times.

In order to show the interplay between the slope of the regression line and the standard deviation of the residuals, we calculated the factors $C_1$ and $C_3$ along with the discriminant of the quadratic equation for different combinations of slopes and residual standard deviations. In doing so, we used the distribution of the timepoints according to the data example data presented in CLSI EP25 and set $\delta$ to 5\%, $\alpha = 0.05$ and the intercept to 0.

Fig. 2 shows the results of the above calculations dividing them into five different color-coded areas: The lightest grey area describes combinations of the residual variance $\hat{\sigma}^2_i$ and slope $\hat{b}_i$ where the discriminant is negative. In the bottom left grey area $C_3 \leq 0$, whereas for the
upper three light grey areas it holds that $C_3 > 0$. In the black and top right grey areas it holds that $C_1 < 0$, whereas for the left grey areas $C_1 > 0$. The white and grey lines indicate where $C_1 = 0$. With exception of the white line, the area below the black curve describes combinations of slope $\hat{b}_i$ and residual variance $\hat{\sigma}^2_i$, for which the estimated stability time $X_{\text{est}}$ exists.

Data scenarios with solutions for the stability time

In Fig. 2, the black area ($C_1 < 0$ and $C_3 < 0$) reflects situations where the recovery of the measured samples, as compared to baseline, shifts over time so that the related lower (for negative slope) or upper (for positive slope) one-sided confidence interval intersects with the acceptance criterion at exactly one positive time point. Note that in this case the correct stability time results from Eq. (20), while the solution given by Eq. (21) corresponds to the intersection of the acceptance criterion with the opposite and thus undesired confidence interval (i.e., lower (upper) confidence interval for positive (negative) slope). CLSI EP25 lacks the above case discrimination and might thus lead to incorrect stability times.

We see that the data of the native sample pool presented in the CLSI example falls within this area. For this sample the estimated regression line together with the acceptance criterion and confidence intervals is shown in Fig. 3A. The dashed vertical line displays the estimated stability time at 393 days, whereas the dotted line exemplifies the intersection of the lower confidence interval with the acceptance criterion.
which is at day 555.

For the bottom left grey area, we obtain a positive stability time from (20) and a negative one for (21). In this case, the relevant confidence interval intersects with the acceptance criterion, once in the positive range and once in the negative range. Here, $C_1 > 0$ and $C_3 < 0$ holds, which means that the slope is not statistically different from zero, but at baseline the confidence interval is within the acceptance criterion. For a positive slope the confidence interval intersects with the acceptance criterion only once in the positive range and the lower confidence interval never intersects with the acceptance criterion.

If the definition of the stability time does not require the statistical significance of the slope, (20) gives the calculated stability time.

Fig. 5. Flowchart for calculating the stability time $X_{est}$. 

A. Geistanger et al.
However, there are different opinions, whether the statistical significance of the slope should be taken into account or not. [7] argues for the statistical significance for the estimation of shelf life in drug or reagent stability studies. We value this argument, especially if the substance is subjected to conditions that cause degradation. However, for sample stability studies where the stability of the analyte should be shown, we do not necessarily see the need for statistical significance. Nevertheless, we will include in the definition of the data analysis workflow both possibilities.

Fig. 3B shows the estimated regression line, the one-sided 95% upper and lower confidence intervals and the acceptance criterion for calibrator sample L2. The dashed line displays the estimated stability time at 583 days given by (20). In the case that extrapolation beyond the maximal measured time is not allowed, this value would be trimmed to 402 days. The second solution as given by (21) is negative and describes the intersection of the upper confidence interval with the acceptance criterion, which is in the negative range. The lower confidence interval never intersects with the upper acceptance criterion.

In Fig. 2, the top left light grey area below the black curve shows combinations of slope and residual variation where the right confidence interval intersects with the acceptance criterion twice in the positive range. Here, (21) gives the first intersection, (20) the second intersection. Similar to SigmaPlot [3], we define this second intersection as obtained stability time. Note that CLSI EP25 does not mention this data situation at all. According to the guideline in this situation the stability time would be set to the maximum time of the stability experiment, as the slope of the regression line is not statistically different from 0. This situation is visualized by the artificial sample No. 1 in Fig. 3C. The estimated slope is positive, but not statistically significant, and at baseline the upper confidence interval is broader than the acceptance criterion. As the upper confidence interval is a convex function, it becomes more narrow as it progresses towards the regression line and intersects the acceptance criterion once at 24 days and the second time at 320 days. We define the second intersection at 320 days as the obtained stability time.

Data scenarios with no solutions for the stability time

For the areas above the black curve in Fig. 2, we do not obtain adequate stability times. This also holds for the white separation line between the black and bottom left grey area. In these cases, further evaluations of (a) the observed measurement error within the experiment, (b) the defined acceptance criterion and (c) the measurements at baseline have to be made.

The light grey, parabolic area in Fig. 2, defines where the discriminant is negative and there exist no solution of the Eq. (13). This is the case if none of the confidence intervals intersect with the acceptance criterion. Fig. 4A shows this case at the example of the regression analysis for the second artificial sample. Here, the measurement error of the experiment needs to be checked carefully. If, however, the measurement error is within the expected range of the measurement procedure, one can show that the design and measurement procedure is not good enough for the defined acceptance criterion.

Within Fig. 2, the top right grey area defines combinations of slope and residual variation, resulting in \( C_1 < 0 \) and \( C_2 > 0 \), which means that the estimated slope value is statistically significant, but at baseline the confidence interval of interest is broader than the acceptance criterion. In this case, none of the solutions for (13) gives a stability time. (20) results in negative values and the positive value obtained by (21) describes the intersection of the opposite confidence interval with the acceptance criterion. Fig. 4B shows this case at the example of the regression analysis for artificial sample No. 3. The slope of the sample is already quite steep and also shows a higher residual variation, such that already at baseline the upper confidence interval is out of the acceptance criterion. Only the lower confidence interval intersects with the acceptance criterion at a positive time point, on day 80. However, this is not the correct value for the stability time. Note that this case is not at all considered in CLSI EP25, although the slope is statistically significant. Samples which fall within the top right grey area have a relatively steep slope, compared to the acceptance criterion and a higher measurement error. In addition to checking the measurement error within the experiment, it is important to reconsider the validity of the baseline measurements.

The reconsideration of the validity of the baseline measurements is especially important for results where the estimated intercept is already above/below the acceptance criterion, as shown by the regression analysis of the artificial sample No. 4 in Fig. 4. Also in this case, there does not exist a stability time, as the regression analysis tells us that requirements are not even met at baseline.

For the sake of completeness, we briefly discuss the problematic case of the white separation line between the black and bottom left grey area. For combinations of slope and residual standard deviation falling exactly on this line, we have \( C_1 = 0 \), which is equivalent to a p-value for the significance test of the slope of exactly 0.05. In this case, the lower and upper confidence interval asymptotically approach the acceptance criterion, but none intersect with it. Therefore, the stability time does not technically exist in this case, which we consider extremely unlikely.

Exception Handling

For the implementation of an automated data analytics workflow, all process steps have to be defined narrowly, based on the considerations of the theoretical calculations. Before starting the analysis the following exception handling needs to be implemented:

![Fig. 6. Three sampling schemes for the CLSI stability example, with 10 sampling time points, between 0 and 402 days.](image-url)
• The number of time points is at least three, that is \( n \geq 3 \).
• There are no negative times \( X_i \).
• There are no negative results \( Y_{mr} \), since we consider concentrations that are always positive.

In addition, it might be useful to warn the user in the case of the following data inputs:

• There are results \( Y_{mr} \) below limit of quantitation (LoQ) or above the upper limit of the measuring interval.
• There are missing data, in particular data are missing for any time point and the data are unbalanced with respect to the desired number of replicates \( n_i \) across all time points.
• The design is different for samples within the same experiment, in particular concerning the number of replicates \( n_i \) and the number of time points \( n_t \) and the location of the times \( X_i \).

Data Analytics Workflow Flowchart

Fig. 5, 6 shows a flowchart that describes the conditions where it is possible to estimate the stability time \( X_{mr} \). This flowchart is the backbone of a software implementation for estimating the stability time \( X_{mr} \). Following this chart, a straightforward implementation of a data analytics workflow for all possible data scenarios is achievable. In order to describe the area below the black curve, as shown in Fig. 2, the following notations are introduced:

\[
G = \frac{(\delta_i - \hat{a}_i)^2 \times n \times TSS_X}{\hat{b}^2_{[1\times d]} \times (TSS_X + n \times \bar{X}^2)} \quad (23)
\]

\[
H = \frac{(\delta_i - \hat{a}_i) \times n \times \bar{X}}{TSS_X + n \times \bar{X}^2} \quad (24)
\]

\[
I = \frac{\hat{b}^2_{1\times d} \times TSS_X + n \times (\hat{a}_i + \hat{b}_1 \times \bar{X} - \delta_i)^2}{\hat{b}^2_{[1\times d]}} \quad (25)
\]

The workflow starts with the evaluation with regard to whether the residual standard deviation is positive and different from zero. If this is not the case, all measured values fall on a line and the workflow enters the decision tree for the perfect linear fit. If the residual standard deviation is different from zero the next level checks, whether the estimated intercept of the regression line is larger than the acceptance criterion. However, this approach results in solving a quadratic equation, which mathematically has either no solution or two solutions. Although guidelines and publications show these results, they do not clarify which of the two solutions of this equation should be used and for which data situation. However, a fully automated data analytics workflow can only be put in place, if this is done in a rigorous manner.

We show that the existence of an intersection of the respective confidence interval with the acceptance criterion mainly depends on the combinations of the coefficients \( C_1 \) and \( C_3 \) of the quadratic equation and the relation between the intercept and the acceptance criterion. For different data situations, one has to define the solution to obtain the correct stability time. In addition, we give for each scenario a clear example to exemplify the different intersection situations of the confidence interval with the acceptance criterion. Also extreme data scenarios have to be considered for the definition of a fully automated data analytics workflow. One important scenario is that all data points fall on a line, i.e. the residual variation is 0. This might happen if the resolution of the measurements is lower than the change of the analyte over time. In addition to the step-by-step data analytics workflow definition, we give guidance for further evaluations in case no solutions for the stability time exist.

The gained insights of the different data scenarios can be used to deduce comparison measures for different sampling schemes and their suitability for the defined acceptance criterion. The acceptance criterion for stability should be suitable for a result to be used in a medical decision and ideally based on total allowable error requirements. Total allowable error requirements should be based on models defined by the 2014 European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) consensus conference [8] including clinical outcome studies, a selected proportion of biological variation, or, when information derived from the first two models are lacking, state-of-the-art of the measurement performance. If the acceptance criterion is based on total error requirements, an appropriate sampling scheme must be defined for stability time estimation. Comparison of different schemes is already helpful in the planning phase of a stability study. They allow one to deduce the maximal allowed measurement variability for the given sampling scheme and acceptance criterion within the planning phase. This maximum allowed measurement variability can then be compared to the actual measurement variability of the measurement system to check whether the planned stability study can be performed in this manner. More details are provided in the supplemental material.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All authors are employees at Roche Diagnostics GmbH, Germany.

Comparison of different sampling schemes

The CLSI EP25 guideline [1] assumes a minimum of three timepoints, equally distributed across the study's duration. Additional timepoints should be added for addressing specific questions. As laid down in Chapter 3, there are some data scenarios where the analysis of stability data might not have a solution for the stability time. This is most often the case if too much variability is present within the data, resulting in too wide confidence intervals depending on whether we ask for the statistical significance of the slope or not. In the first case, the entire time range becomes the stability time, in the second, the solution is again given by (20), in case that \( C_1 \) is not equal to 0.

Discussion

We carefully examined different data situations for estimation of the stability time according to the analysis approach of CLSI EP25 [1] guideline. This approach defines the stability time as the intersection of a respective confidence interval of the regression line with the predefined acceptance criterion. However, this approach results in solving a quadratic equation, which mathematically has either no solution or two solutions. Although guidelines and publications show these results, they do not clarify which of the two solutions of this equation should be used and for which data situation. However, a fully automated data analytics workflow can only be put in place, if this is done in a rigorous manner.

We show that the existence of an intersection of the respective confidence interval with the acceptance criterion mainly depends on the combinations of the coefficients \( C_1 \) and \( C_3 \) of the quadratic equation and the relation between the intercept and the acceptance criterion. For different data situations, one has to define the solution to obtain the correct stability time. In addition, we give for each scenario a clear example to exemplify the different intersection situations of the confidence interval with the acceptance criterion. Also extreme data scenarios have to be considered for the definition of a fully automated data analytics workflow. One important scenario is that all data points fall on a line, i.e. the residual variation is 0. This might happen if the resolution of the measurements is lower than the change of the analyte over time. In addition to the step-by-step data analytics workflow definition, we give guidance for further evaluations in case no solutions for the stability time exist.

The gained insights of the different data scenarios can be used to deduce comparison measures for different sampling schemes and their suitability for the defined acceptance criterion. The acceptance criterion for stability should be suitable for a result to be used in a medical decision and ideally based on total allowable error requirements. Total allowable error requirements should be based on models defined by the 2014 European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) consensus conference [8] including clinical outcome studies, a selected proportion of biological variation, or, when information derived from the first two models are lacking, state-of-the-art of the measurement performance. If the acceptance criterion is based on total error requirements, an appropriate sampling scheme must be defined for stability time estimation. Comparison of different schemes is already helpful in the planning phase of a stability study. They allow one to deduce the maximal allowed measurement variability for the given sampling scheme and acceptance criterion within the planning phase. This maximum allowed measurement variability can then be compared to the actual measurement variability of the measurement system to check whether the planned stability study can be performed in this manner. More details are provided in the supplemental material.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: All authors are employees at Roche Diagnostics GmbH, Germany.
as compared to the acceptance criterion. However, the insights we gained from the extensive analysis of the two factors $C_1$ and $C_3$ can be used to compare different sampling schemes for a stability experiment. Using these additional insights, we are able to avoid data scenarios where there doesn’t exist a solution for the estimated stability time.

The definition of the sampling scheme does not only depend on statistical considerations, but also on already acquired knowledge on the expected stability, laboratory logistics and handling possibilities. However, each proposed sampling scheme can be evaluated based on the formula for $C_3 \leq 0$, in order to obtain the maximal measurement variability, for which $C_3 \leq 0$ holds for different assumed values of the estimated intercept. In case the maximal measurement variability is lower or equal to the known variability of the measurement procedure, the sampling scheme and measurement procedure are not appropriate for the stability analysis. Here, the sampling scheme needs to be adapted, by adding more sampling points, or either by shifting the sampling points more to extreme time points. [9] already criticized the application of equally spaced sampling schemes and emphasized that sampling schemes with more measurements at the extremes leads to better statistical properties and more reliable stability time estimates. Another possibility to adapt the sampling scheme consists in adding replicate measurements at each timepoint and using the mean of those replicates as data point for the analysis. Note that the variability of the mean then reduces by the factor $\sqrt{n}$. As an example, we compare the sampling scheme example, outlined in CLSI EP25, with two others, one with more data points at the extremes and the other with more data points in the middle of the considered time period.

In Fig. 7, the maximum allowed standard deviation of the residuals is plotted on the y-axis against an assumed intercept on the x-axis. The three different lines correspond to the three assumed measurement schemes. For the CLSI example scheme, the maximum allowed standard deviation of the residuals should stay below 5%, in case that the intercept is zero and below 2%, if the estimated intercept is 3%. Only below these values of the standard deviation, the results of slope and residual standard deviation will fall in the black area and hence we can obtain an estimate of the stability time.

The alternative scheme No. 1 allows for higher standard deviations of the residuals, as the spread of the measured time points is wider than those of the CLSI example. On the other side we observe for the alternative scheme No. 2, how drastically the range of the allowed standard deviation of the residuals is reduced, only because more time points are present in the center of the time range. The maximum standard deviation of the residuals is directly comparable to either the standard deviation of the measurements, in case of a stability analysis, based on the measured concentrations. In case of an analysis based on the percentage deviation this value is comparable to the coefficient of variation of the measurement method.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmsacl.2022.01.001.

References

[1] CLSI EP25-A. Evaluation of Stability of In Vitro Diagnostic Reagents. Approved Guideline. Wayne, PA: Clinical; Laboratory Standards Institute; 2009.
[2] ICH. Stability testing of new drug substances and products. ICH guideline Q1A(R2). International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. 2003.
[3] SigmaPlot. Computing shelf life time with sigmaplot 2021.
[4] SigmaPlot. Validation of the shelf life macro 2021.
[5] Cook JD. The quadratic formula and low-precision arithmetic 2018.
[6] J. Pum, Evaluating sample stability in the clinical laboratory with the help of linear and non-linear regression analysis, Clinical Chem. Lab. Med. 58 (2) (2020) 188–196, https://doi.org/10.1515/cclm-2019-0596.
[7] S.-C. Chow, Statistical design and analysis of stability studies, Chapman & Hall, New York, 2007, https://doi.org/10.1201/9781584889069.
[8] S. Sandberg, C.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, et al., Defining analytical performance specifications: Consensus statement from the 1st strategic conference of the european federation of clinical chemistry and laboratory medicine, Clinical Chem. Lab. Med. 53 (2015) 835–835, https://doi.org/10.1515/ cclm-2015-0007.
[9] K. DeVore, Have more confidence in your stability data: Two points to consider, J. Pharmaceutical Biomed. Anal. 41 (2005) 293–298, https://doi.org/10.1016/j. jpha.2005.10.038.