Opinion Paper


Recommendation for the design of stability studies on clinical specimens

https://doi.org/10.1515/cclm-2023-0221
Received February 28, 2023; accepted March 22, 2023; published online April 6, 2023

Abstract

Objectives: Knowledge of the stability of analytes in clinical specimens is a prerequisite for proper transport and preservation of samples to avoid laboratory errors. The new version of ISO 15189:2022 and the European directive 2017/746 increase the requirements on this topic for manufacturers and laboratories. Within the project to generate a stability database of European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group Preanalytical Phase (WG-PRE), the need to standardise and improve the quality of published stability studies has been detected, being a manifest deficit the absence of international guidelines for the performance of stability studies on clinical specimens.

Methods: These recommendations have been developed and summarised by consensus of the WG-PRE and are intended primarily to improve the quality of sample stability claims included in information for users provided by assay supplier companies, according to the requirements of the new European regulations and standards for accreditation.

Results: This document provides general recommendations for the performance of stability studies, oriented to the estimation of instability equations in the usual working conditions, allowing flexible adaptation of the maximum permissible error specifications to obtain stability limits adapted to the intended use.

Conclusions: We present this recommendation based on the opinions of the EFLM WG-PRE group for the standardisation and improvement of stability studies, with the intention to improve the quality of the studies and the transferability of their results to laboratories.

Keywords: instability equation; preanalytical phase; sample stability.

Introduction

The stability of an analyte in a specimen or a sample can be defined as the preservation of its physicochemical properties over time. Therefore, instability can be represented as a function of change in these properties over time, conveyed as an instability equation.

Instability is a fundamental but invisible aspect of sample quality. The ISO 15189:2012 requirements for quality and competence of laboratories [1] include the general requirement for a documented procedure to control the time and temperature during transportation as required by the requested tests (paragraph 5.4.5) and appropriate facilities to avoid deterioration during the preanalytical phase (paragraph 5.4.7). The new version in
transition, ISO 15189:2022 [2], additionally includes specific requirements for sample handling, preparation and storage:

7.1.6.2 Criteria for additional examination requests
Laboratory procedures shall include time limits for requesting additional examinations, or further examinations on the same primary sample.

7.1.6.3 Sample stability
Considering the stability of the measurand in a primary sample, the time between sample collection and performing the examination shall be specified and monitored where relevant.

These new requirements mean that the laboratory must acknowledge and monitor the time limits from sample collection to analysis and from sample collection to disposal. This time limit is usually known as the stability limit.

One of the main objectives in developing sample collection devices over the years has been to improve analyte stability. Different physical (limiting contact with air, separation of the cellular component, refrigeration, etc.) or chemical (inhibitors of enzymatic degradation systems, antiseptics, etc.) mechanisms have been implemented. The instability of analytes is affected by sample handling conditions, and it is necessary to regularly evaluate it in each health institution in the context of ongoing technological improvements.

Although sample stability is a fundamental aspect of the ability of laboratories to deliver high quality results, evaluation of the stability limit is not an explicit regulatory requirement for manufacturers of in vitro diagnostic devices. The European IVD directive 98/79/E.C. of the European Parliament [3] merely indicated, in relation to sample containers, the need to assess stability of the device as part of the requirements of the container and not the stability of the analytes. The new E.U. Directive 2017/746 [4] (which entered into force in May 2022, with a transition period until May 2027 for devices that are already on the market), in contrast, includes the following requirement in regard to stability:

Annex II, 6.1.1: Verification and validation of products to be included in technical documentation. This section shall describe the different types of samples that can be analysed, including their stability, their storage, if applicable, their transport conditions and, for time-critical methods of analysis, information on the time period between sample collection and analysis as well as their storage conditions, such as duration, temperature limits and freeze/thaw cycles.

The standards mentioned above clearly focus on risk management for both laboratories and industry, and the extent or comprehensiveness of the assessment of stability should reflect this. It is therefore essential to document the impact of the loss of sample stability on patient care, which can manifest itself either as a single finding in a patient with a specific undetected problem (e.g. transport delay) or as a general effect between laboratories with different sample handling procedures (e.g. glucose determination in a tube with a targeted stabilizer vs. determination in serum without stabilizer, alongside other biochemical analytes) [5, 6].

Multiple experimental stability studies have been published for most of the analytes used in daily practice. However, many of these have been carried out with obsolete devices, such as sample collection systems without vacuum or separation systems. There are bibliographic compilations of stability studies with recommended stability limits produced by different professional groups, such as German Society for Clinical Chemistry and Laboratory Medicine (DGKL) [7], American Association for Clinical Chemistry (AACC) [8] and Clinical and Laboratory Standards Institute (CLSI) [9–11]. Although these studies were carried out with outdated methods or have poor experimental quality, they continue to significantly influence laboratory practice and stability recommendations.

A fundamental problem in achieving transferability between studies is that only few define and report instability equations or share raw data. Usually, the results are expressed in the form of stability limits, obtained by applying widely varying maximum permissible error (MPE) specifications. Significant differences are observed even when it is possible to retrospectively calculate instability equations from the published study data [12]. Knowledge about the specific conditions under which studies were conducted is essential to allow comparability of results. However, many studies do not adequately provide this information, e.g. sample collection, processing, storage or transportation or the instruments and methods used for analysis, etc. Given these shortcomings, in 2019, the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group Preanalytical Phase (WG-PRE) published a “Checklist for REporting Stability Studies” (CRESS) [13]. This recommendation was intended to encourage the authors of such studies to transparently provide uniform information for interpretation and meta-analysis.

The CRESS does not, however, include guidance on how to design a stability study or assess the results. There are only few national [14, 15] and no international recommendations covering this topic. Therefore, the aim of this recommendation is to fill this void. This paper focuses on how to design studies for those specimens most frequently analysed in clinical laboratories (i.e. blood and urine). For other specimens, this guidance might also be applicable, but researchers conducting studies in matrices not covered in this recommendation would need to consider the characteristics of the sample type of interest. When taking long-term analytical variability due to changing batches of material or equipment into account, the proposed study design can also be applied to long-term storage of biological samples (Biobanking).
Although clinical laboratories may conduct stability studies to verify the manufacturers’ stability claims, the primary responsibility for this task lies with the assay supplier company. All stability data, including the stability conditions, time periods studied and statistical confidence of the provided instability equations must be provided alongside the reagents, according to the recommendations in this document. As sampling devices are a fundamental determinant of stability, collaboration with collection device manufacturers is essential.

**Study design**

**Stability conditions**

Each manufacturer or laboratory should determine what conditions must be tested, depending on a risk assessment, taking into account the intended sample collection and handling procedures.

For a given analyte, it is necessary to start from an understanding of the causes of its degradation or increase. The number of studies to be carried out will depend on the number of identified variables except for time, which will be used as the only independent variable in the design of studies, as discussed below.

Common variables which contribute to analyte instability in clinical samples include:

- **Cellular metabolism and cell lysis.** Contact with the cells produces an exchange of substances, even without cell destruction [16–18]. Cell lysis results in the release of enzymatic material that can lead to altered metabolism of constituents. Blood samples are centrifuged and serum or plasma is separated from the cells (with a physical barrier or by transfer into a different receptacle) to halt the impact of cellular metabolism on the composition of serum or plasma. After centrifugation, some cells (particularly platelets) and cell debris, usually remain in the serum and in a greater degree in plasma, depending on the method and conditions used for separation. Urine samples can contain varying amounts of inflammatory or epithelial cells and microorganisms, influencing stability [19, 20].

- **Contact with air; diffusion and evaporation.** Uncapping the tube during analysis causes evaporation of the solvent, generating an increase in the concentration of most analytes [21] and diffusion of dissolved gases, with loss of carbon dioxide and a subsequent increase in pH [8, 22, 23]. Even if the tube is kept closed, these processes must be taken into account as some gases can diffuse through the walls of the container, especially in plastic tubes [22–25]. Collection of urine takes place in contact with air, although it can be subsequently transferred to vacuum tubes.

- **Exposure to light.** It is not common to use opaque materials to limit exposure of the sample to light irradiation of varying degrees during the whole process. Some analytes are unstable depending on the intensity, the duration and type of light exposure encountered [26].

- **Tube orientation and mixing.** Mixing is a factor that influences the speed of biochemical reactions. It is an important aspect for samples that must be transported over long distances. The vertical orientation of tubes of serum allows complete coagulation and reduces mixing of its components, also avoiding the adherence of fibrin to the inner wall of the tube cap [7].

- **Adsorption.** In plastic containers or sample tubes with a separating gel, there may be a transfer of components from the specimen to the material of the tube/gel or vice versa. It is especially relevant for the detection of some trace elements, drugs or hormones [27–29].

- **Temperature** is an important catalyst for chemical reactions causing reduced stability [30–34].

- **Preservatives.** In some collection devices preservatives are added to samples to alter their behaviour (anticoagulants or procoagulants), inactivate or prevent the growth of microorganisms [35–37] or improve the stability of a particular analyte (i.e. tubes with a glucose stabilizer, tubes with EDTA and protease inhibitors for nucleic acids or peptides analysis) [38, 39]. These substances may alter the stability of other analytes.

- **Variability due to the collection process:** Such variables include possible deterioration by hemolysis or incorrect filling of the tube, which may affect the stability of the sample’s constituents. These effects should be limited by adhering to according sampling guidelines [40].

- **Analytical method:** The analytical method determines the analyte’s property (e.g. enzyme activity vs. mass) or the part of the detected molecule (e.g. immunoassays against intact molecule or fragments). Even with very similar methodologies, differences in observed stability have been reported [41].

Many of these potentially influencing factors are subject to wide inter-individual variability due to different cell concentrations in the sample or genetic variability in the degradation mechanisms [42]. This inter-subject variability leads to the common observation among published studies that different slopes for the individual instability equations for each subject are reported, which causes a proportional increase in the dispersion of results over time, making predictions of stability for a particular subject more inaccurate the longer the time elapses.

The presence of medications, supplements or microorganisms in the sample is a possible cause of variation in
the expected stability that has been poorly studied and should also be taken into account.

Although it is possible to design stability studies by modifying several conditions simultaneously, which would allow the development of multivariate predictive models, it must be taken into account that many of the stability conditions mentioned are not continuous variables and those that are, such as temperature, do not affect all analytes uniformly. Moreover, the a posteriori distinction between the tested conditions may be difficult, contributing to inaccuracy of instability predictions. Hence, it seems more appropriate to define a set of standard stability conditions and to perform “one-at-a-time” studies, where all stability conditions are fixed and only the effect with respect to time is evaluated.

When designing the study, the most common local methods of storage should be tested for. If samples are stored primarily at the clinics or the doctor’s office, investigating analyte stabilities in whole blood may be of greater importance than in plasma/serum.

The two “basic” studies would evaluate stability of specimen/samples at the common handling temperatures, which would mean that, applying local settings (collection tubes and instruments, analytical methods, etc.) each provider/laboratory would have to perform studies on samples stored at ambient temperatures or refrigerated specimen (reflecting extra-laboratory pre-analytical processes) and refrigerated or frozen sample (reflecting intra-laboratory pre-analytical processes). Other temperatures can be sensible to check also, especially if an important influence of temperature is demonstrated for the analyte. Studies should strictly control, record and document the temperature at each stage of the study, as this is a major influencing variable.

The investigated storage durations will depend on the common sample transport and handling practices. Consideration should be given to the locally common time frame for pre-analytical sample handling, including the need for transport to distant laboratories and the time needed for sample processing (centrifugation or addition of preservatives). Additionally, the duration and manner of analytical processing has to be considered, potentially taking up considerable processing time at room temperature, as well as the sample retention time in case of subsequently added requests.

Since sample collection, transport and analysis processes take usually up hours rather than days, the recommended duration unit is the hour, so that the slope of the instability equation will express the hourly percentage change, reflecting routine processing times more accurately.

If there are changes in the diagnostic devices or a significant variation in any indicator of their performance, their impact on stability should be rechecked.

**Recommendation:** Use a “Check one set of variables at a time” design, based on studying the effect of time on the stability of analytes in samples collected, processed, and stored under a fixed set of stability conditions.

**Recommendation:** The number of stability conditions and the time periods to be studied should be defined in accordance with locally common procedures for sample collection, processing, transport, and storage.

**Patient/sample selection**

The number and type of subjects included in a stability study is important to ensure statistical validity of the model and transferability to a population with similar characteristics.

The use of samples from patients is recommended instead of healthy volunteers. In this way, the study samples are representing greater inter-individual variability and include clinically relevant analyte concentrations, avoiding the need to carry out spiking studies.

Since the initial analyte concentration may affect its stability [25], stability studies should be performed in samples with concentrations close to the level of clinical decision limits, considering the type of laboratory and the population of interest. However, in most cases the concentration of the sample before analysis is unknown, except in very controlled cases. Therefore, it is difficult to choose patients well and often study samples will have concentrations within the biological reference interval.

In order for the results of the stability study to be transferable, and to avoid misinterpretation of identified variations, sample collection and pre-analytical handling should be standardized before any other actions relevant to the analysis results. Mishandling of the sample during the pre-analytical phase may lead to phenomena affecting stability, such as hemolysis, bacterial contamination or contact with air. Therefore, it should be ensured that sample collection and processing prior to the experimental study is performed in a time frame at least equal to that of best laboratory practice.

It is imperative for stability studies to simultaneously collect multiple equivalent samples from the same patient. This usually involves obtaining a larger number of samples. In various situations, such as when collecting blood, it is important to keep in mind that there may be variations in the sample’s quality as a result of the collection process, mostly because the sample must be obtained by inducing venous stasis. Additionally, the emotional stress of the patient associated with the venepuncture itself may result in the release of hormones or other substances into the blood.
Significant differences have been reported between the results from different tubes, which were obtained in the same extraction. Hence, in order to ensure the equivalence of all tubes, care should be taken that phlebotomy is conducted in a standardized fashion according to current recommendations [40]. Ideally, to avoid collection induced systematic biases, the order of samples collected for the different storage conditions/durations should be randomized.

Stability studies should simulate erroneous delays in sample processing. For simulation of sample storage between collection and processing (e.g. centrifugation), we advise that each specimen associated with each storage period be obtained in a separate dedicated primary tube. This is especially important when using vacuum tubes, as aliquoting the specimen will allow contact of the specimen with air and possible contamination. In order to ensure that the study results reflect the routine real-world working procedures, it is imperative to use primary tubes for each storage condition simulation. These samples should be randomized after collection and handled in accordance with the supplier’s instructions.

In case the processed samples are to be conserved, the usual working conditions should be simulated. When using vacuum tubes with built-in separation systems (e.g. gel), direct analysis in the primary tube itself is the recommended method of operation for the majority of common analytes. In this case, it is also recommended to obtain distinct specimens for each storage duration, since in event of a delay in the analysis, the sample would remain centrifuged and capped in the primary tube.

If the local process includes aliquoting, the stability study could be carried out in this manner. It should be remembered that the use of leftover samples, pooling or spiking samples will always introduce new potential biases. If for ethical, technical or financial reasons no other study design is feasible, or if this is the standard method of sample handling in the local laboratory, this should be taken into account when attributing the instability effect to the primary tube.

The CRESS guide includes an extensive list of aspects to be considered, which are a fundamental part of the sample collection guidelines [9–11].

**Recommendation:** Samples for stability studies should be obtained and stored under ideal routine working conditions prior to analysis.

**Recommendation:** To avoid manipulation, the use of distinct primary tubes for each storage duration that is to be investigated, is recommended.

**Recommendation:** Samples from patients are preferable, as they reflect the concentrations of interest of the analytes better than those from healthy subjects.

### Experimental design

As stated in ISO Guide 35:2017 [43]: The existence of instability for a specific set of conditions is evaluated by simultaneously obtaining several samples from a patient and maintaining them under the same controlled conditions of stability for a variable time. Once stressed, the samples should be analysed in a manner that minimises analytical error. This can be achieved by tightly controlling random error by replicate determination, and systematic error, by two strategies:

- If the analysis is carried out in several series, as each study time point is reached (real-time studies), strict internal quality control (and recalibration when necessary) should be performed before each analytical batch to minimise the bias between series. There should be no changes in reagent lot between series.
- Analyse all samples in the same analytical series, eliminating bias between series (isochronous studies). This is only possible if the samples can be stored under reference conditions where there is an established and documented long-term stabilisation method e.g. cryopreservation of samples in deep freeze. If samples are kept in deep freeze but not analysed in a single run, it is advisable to analyse at least each complete subject in a single run.

In either model, strict internal quality control processes must accompany the stability study, and lot changes in reagents, controls, or interventions on the measuring equipment during sample analysis should be avoided.

The existence of a reference long-term storage method may be based on previous studies or a preliminary study, conducted to demonstrate its validity, covering the intended study period. This information should include the number of acceptable freeze/thaw cycles and should provide information on the manner of thawing and homogenisation. This information is especially important if the sample is to be stored in the primary tube.

The use of replicate measurements to reduce imprecision should be used for detecting outliers. Significant laboratory errors will be detected as a variation between replicates much higher than the usual laboratory imprecision. For samples with adequate precision, the mean of the results of replicate measurements is considered the best estimate of the true value. Duplicate analysis is sufficient unless the analytical method is known to be prone to high laboratory imprecision, in which case it may be necessary to increase the number of replicates.

A key issue of the study design is the number of samples required. The sample size is connected to the likelihood of
detecting the instability effect (=effect size). In the proposed regression model, the magnitude of the effect is related to the slope of the fitting line. A slope that is barely different from zero indicates the absence of instability, meaning that at any storage duration studied, the difference from baseline sample is not different from zero. Very discrete effects, with slopes of the regression equation close to zero, will require a very large sample size, prolonged storage duration and minimal analytical error (see validation section).

It is possible to estimate the needed sample size if data about the stability equation and the typical dispersion of the results are known a priori, although there is disagreement about the calculation method and reliability [44]. As an example, a study with a slope of −0.2%/h (=5% decrease in 24 h), considering an alpha and beta error of 0.05 and 0.2, respectively, would need a sample size of 100, which is equivalent to 20 subjects with 5 observation time points.

Storage time points are the moments during the total time of the study at which the loss of stability is assessed. Ideally, in the regression model the time points are random and evenly distributed for any subject over the total study time. This is often technically complicated, so fixed time points for all the subjects may be appropriate as long as the uniform distribution is maintained and at least five time points are included. At least five time points are necessary to verify the fit of the instability equation to a linear model [45].

Each manufacturer/laboratory must select the appropriate study design for the type of sample and analytes to be measured. Suggestions on study designs for some common specimens is provided in Table 1.

Recommendation: If there is evidence of a reference long-term stability method, like deep freezing, it is possible to choose an isochronous design (measuring all samples at the end of the study in one batch) to limit the effect of analytical variation between runs.

Recommendation: Stability studies must be accompanied by strict internal quality control and bias correction to minimize the potential impact of analytical error.

Recommendation: Measurements should be performed at least in duplicate to limit the influence of analytical random error. The mean of the results of replicate measurements is considered the best estimate of the true value.

Recommendation: Fixed time points for all subjects could be used. At least five uniform distributed storage time points are required to establish a valid instability equation.

### Study evaluation

Once the samples have been analysed, the data should be checked for major errors. First, errors in replicate measurements could be assessed by comparing with the usual imprecision of the laboratory. The coefficient of variation (CV%) of the replicates is compared with the customary coefficient of variation of the instrument used (CVa%). Replicates with a CV% higher than three times the CVa% should be reviewed and, in the case of suspected error, eliminated. If the number of outliers in the experiment is higher than 5% of the samples, analytical performance should be reviewed in depth.

The calculated mean of the replicate results will thereafter be used as the best estimator of the true value for further statistical evaluation.

The loss of analyte stability is defined as the difference between the baseline sample value (t₀) and each successive test sample value (tᵳ). In order to compare the difference observed with permissible error specifications and to facilitate comparison between quantities, the relative difference is expressed as percentage (PD%) according to the formula

\[
PD\% = \left( \frac{t_\text{x} - t_0}{t_0} \right) \times 100
\]

- \(t_\text{x}\): Average concentration of test sample n in time x.
- \(t_0\): Average concentration of basal sample n.

Adjustment of the global instability equation can be made by regression analysis using least-squares adjustment. The instability equation does not have an interception coefficient, as the design implies that the baseline sample has no instability itself. Hence, the regression equation has to be forced to pass through the origin of coordinates, also known as “Regression Through Origin” (RTO). The goodness-of-fit is calculated in this case using the “simple \(R^2\)” [46], which calculates residuals against zero instead of against the mean of the data. The interpretation thereof is equivalent to the one for the coefficient of determination (\(R^2\)). Caution should be taken, because even when allowing to force the regression through the origin, not all statistical packages indeed compute the simple \(R^2\).

RTO fitting with a linear model, observed in most publications, yields a first-degree equation, with a single coefficient (slope; \(\beta\)) and without intercept term, of the type:

\[
PD\% = \beta \times \text{time (h)}
\]

This model facilitates understanding and transferability of results but may not be the best possible fit. From the graphical review of the linear model, it is advisable to evaluate different fitting models (polynomial, exponential), comparing the increase in the determination coefficient.
Table 1: Design selection suggestion (basic strategy in bold).

<table>
<thead>
<tr>
<th>Primary specimen</th>
<th>Laboratory habitual procedure</th>
<th>Stability conditions of concern</th>
<th>Design recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>Serum</td>
<td>Glucose</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Serum separator</td>
<td></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Blood</td>
<td>Plasma</td>
<td>Ammonium</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Plasma separator</td>
<td></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Blood</td>
<td>Plasma</td>
<td>Citrate, no separation</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Plasma separator</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Blood</td>
<td>Blood</td>
<td>Plastic syringe</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>High</td>
</tr>
<tr>
<td>Blood</td>
<td>Blood</td>
<td>EDTA</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>EDTA</td>
<td>Mean corpuscular volume</td>
<td>External</td>
</tr>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>Blood</td>
<td>Blood</td>
<td>EDTA</td>
<td>HbA₁c</td>
</tr>
<tr>
<td></td>
<td>EDTA</td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>Urine, random</td>
<td>Centrifuged urine</td>
<td>Vacuum secondary tube</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Low</td>
</tr>
</tbody>
</table>

(Instability equation: PD% = 0.79*time (h))

MPE, maximum permissible error; SL, stability limit.

Figure 1: Graphical representation of the percentage difference (PD%) with respect to storage time (hours). Estimation of the instability equation by regression through the origin. Linear model with confidence interval for the slope. MPE, maximum permissible error; SL, stability limit.

(simple R²) representing the goodness of fit in the regression model. For comprehension of the study results, a graphical representation may be very helpful, plotting PD% against time on a scattergram (Figure 1).

Once the fit of the instability equation has been determined, it has to be proven that the equation could not reasonably be due to chance by using a hypothesis test, demonstrating that the observed slope is non-zero at a given significance level. The common method to do so is the t-test, using the standard error of the slope. This also allows for the calculation of the confidence interval of the slope, which is convenient for the transferability of the instability equation and for assessing the quality of the study. Small instability effects will be difficult to demonstrate, needing a greater sample size. If the hypothesis test is non-significant, it is challenging to distinguish whether there is no loss of stability or whether the design is too underpowered to demonstrate it.

It should be noted that the computed instability equation serves as an explanatory rather than a predictive model. Extrapolation of the obtained instability equation outside the studied time frame is not recommended, both in the case of intending to demonstrate that there is no loss of stability beyond the maximum studied storage time and in the case of

<table>
<thead>
<tr>
<th>Time (h)</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>14</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>PD%</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>16</td>
</tr>
</tbody>
</table>

![Graphical representation of the percentage difference (PD%) with respect to storage time (hours). Estimation of the instability equation by regression through the origin. Linear model with confidence interval for the slope. MPE, maximum permissible error; SL, stability limit.](image-url)
estimation of the future instability. If additional storage times are required, a fresh experiment must be conducted.

Once the hypothesis test has proven the model to indicate a loss of stability, it is important to look at the goodness-of-fit, indicated by the simple $R^2$. A simple $R^2 > 0.7$ is desirable.

In order to identify inter-individual variability, we recommend providing a graphical analysis of the results, including a separate visualisation of every subject (Figure 2). The results should be plotted on a scattergram with the PD% on the ordinate (y-axis) and the storage time on the abscissa (x-axis), representing each patient studied with a different marker. Individual regression equations typically show some tendency to scatter over time, as inter-individual variability manifests itself as small variations in the slope. Any subject or patient with clear variations from an obvious regression trend should prompt further research into specific conditioning elements (e.g. abnormal cellularity, use of medications or substances, etc.). These cases may be considered outliers only when extremely rare conditions are demonstrated in the subject in comparison to the rest of the study population. If a significant proportion of subjects with similar behaviour is observed, the presence of an uncontrolled instability variable should be suspected.

The result of the analyte-specific stability study should indicate the significance level of the instability equation and the strength of the relationship between storage time and PD%. It is recommended to express the results in the following manner:

- If the instability equation is statistically significant ($p<0.05$) and has sufficient strength (simple $R^2 > 0.7$), after applying a linear model, indicate: “Under these conditions, an analyte increase/decrease of n% per hour is expected up to x hours (maximum studied storage time point)”.

- If the instability equation is statistically significant ($p<0.05$) and has sufficient strength (simple $R^2 > 0.7$), after applying a non-linear model, indicate: “Under these conditions, a non-linear increase/decrease in analyte concentration is observed over time” and specify degree of deterioration at selected times as an example.

- If the slope obtained is not different from zero ($p>0.05$), state: “Under these conditions no significant loss of stability of analyte is expected up x hours (maximum studied storage time point)”.

- If the slope obtained is significantly different from zero, but the simple $R^2$ obtained is <0.7, indicate: “Under these conditions, a possible analyte increase/decrease of n% per hour is expected up to x hours (maximum studied storage time point), and needs to be confirmed”.

In order to obtain a stability limit (SL), derived from the instability equation, a maximum permissible error (MPE) must be defined. The SL corresponds to calculated storage
Concluding remarks

The provided recommendations are intended to assist diagnostic assay suppliers and laboratories in designing and interpreting stability studies in compliance with manufacturing and clinical laboratory accreditation regulations. This manuscript is freely available to the medical laboratory community and especially to scientific societies that can greatly assist in its development and dissemination.

The EFLM WG-PRE would like to encourage providers and laboratories to perform and publish new standardized stability studies under different habitual working and storage conditions. The presentation of the entire data in the form of the mean instability equation is essential to ensure their quality and transferability. We are confident that this recommendation will help guiding this process.

Acknowledgments: This document, on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group Preanalytical Phase (WG-PRE), has been presented, discussed, reviewed and approved at the plenary meetings of the group, whose current members are listed at the EFLM site: https://www.eflm.eu/site/page/a/1156.

Research funding: None declared.

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

Competing interests: Authors state no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

References


