Opinion Paper

Michael Cornes*, Pieter Vermeersch, Ana-Maria Šimundić, Alexander Von Meyer, Tomáš Šálek, Brendan Meyer, Sean Costelloe, Vincent De Guire, Ruben Gomez-Rioja and Janne Cadamuro, on behalf of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group for the Preanalytical Phase (WG-PRE)

The final part of the CRESS trilogy – how to evaluate the quality of stability studies

https://doi.org/10.1515/cclm-2024-0527
Received April 26, 2024; accepted April 26, 2024;
published online May 16, 2024

Abstract: High quality laboratory results are critical for patient management. However, poor sample quality can impact these results and patient safety. To ensure reliable and accurate results laboratories must be aware of each analyte's stability under various storage conditions and matrices to guarantee correct and dependable outcomes. This knowledge allows laboratories to define the allowable delay between sample collection and centrifugation/analysis for all analytes to guarantee appropriate results quality and interpretation. The EFLM Working Group for the Preanalytical Phase (WG-PRE) therefore established a 4-step plan to tackle this issue, aiming to standardize and harmonize stability studies for improved comparison and meta-analysis. The plan included the development of checklists and how-to guides for performing and reporting stability studies as well as a central resource of stability data. This manuscript deals with the issue of evaluating publications and incorporating them into a central resource. To evaluate stability studies, the CRESS checklist was used to structure 20 sections used to judge the quality of studies. Each section has 4 levels of quality, with scores converted to numerical values and weighted based on expert opinion. Based on this, a final score ranging from A to D was determined. The procedure was then tested on six manuscripts and checked for agreement between expert judgements. The results demonstrated that the proposed evaluation process is a useful tool to distinguish between best in class manuscripts and those of lower quality. The EFLM WG-PRE strongly believes that the provided recommendations and checklists will help improving stability studies both in quality and standardisation.

Keywords: stability; preanalytical; evaluation; quality

***Corresponding author: Michael Cornes**, Clinical Chemistry Department, Worcester Acute Hospitals NHS Trust, Worcester, UK, Phone: +44 1905 760843, E-mail: Michael.cornes@nhs.net
Pieter Vermeersch, Clinical Department of Laboratory Medicine, KU Leuven University Hospitals Leuven, Leuven, Belgium. https://orcid.org/0000-0001-7076-061X
Ana-Maria Šimundić, Department of Global Medical & Clinical Affairs, Business Unit for Preanalytics, Greiner Bio-One GmbH, Kremsmünster, Austria; and Faculty of Pharmacy and Medical Biochemistry, Zagreb University, Zagreb, Croatia. https://orcid.org/0000-0002-2391-5241
Alexander Von Meyer, Institute for Laboratory Medicine and Microbiology, Munich Municipal Clinic Group, Munich, Germany
Tomáš Šálek, Institute of Clinical Biochemistry and Diagnostics, Medical Faculty in Hradec Králové, Charles University, Prague, Czechia; and Department of Clinical Biochemistry and Pharmacology, The Tomas Bata Hospital in Zlín, Zlín, Czechia. https://orcid.org/0000-0002-8392-5003
Brendan Meyer, BD Life Sciences – Winnersh, Wokingham, UK
Sean Costelloe, Department of Clinical Biochemistry, Cork University Hospital Group, Cork, England. https://orcid.org/0000-0002-2464-3735
Vincent De Guire, Department of Biochemistry, Maisonneuve-Rosemont Hospital Research Centre, University of Montreal, Montreal, Quebec, Canada
Ruben Gomez-Rioja, Department of Laboratory Medicine, La Paz-Carlos III-Cantoblanco University Hospital, Madrid, Spain. https://orcid.org/0000-0002-3429-0427
Janne Cadamuro, Department of Laboratory Medicine, Paracelsus Medical University Salzburg, Salzburg, Austria. https://orcid.org/0000-0002-6200-9831

Aims

We describe a tool to evaluate the quality of stability studies, aiming to standardize and harmonize the grading of such studies for improved comparison and meta-analysis as a starting point, contributing to the development of a database of graded stability studies.

Background

Laboratory medicine plays a critical role in a significant proportion of the clinical decisions regarding patient

management [1]. It is therefore of the utmost importance that published laboratory results are reliable and accurate. However, as surmised by the phrase 'Garbage in, garbage out', the quality of the results can only be as good as the quality the samples delivered to the laboratory. Laboratory medicine has established excellent procedures to ensure quality in the analytical phase during the last century. In addition, over the last few decades, significant progress has been made in establishing awareness of quality in the extra analytical steps factoring in all aspects of the total testing process, from the clinical decision to take a sample to the interpretation of laboratory results [2–4]. In 2012 the laboratory standard ISO15189 [5], introduced a requirement for laboratories to improve and maintain the quality in the preanalytical phase which has led to an increase in interest and in publications on this topic. To avoid poor quality samples with subsequent poor quality results, it is essential that laboratories are able to determine the quality of the sample and the associated quality of the analytes which are to be tested. In order to do this, laboratories need to know the stability limits for every analyte in a variety of different storage conditions and matrices. This knowledge allows laboratories to define the analyte-specific acceptable delay between sample collection and centrifugation/analysis necessary to guarantee results and interpretation of sufficient quality to answer the question asked by the requesting clinician.

When reviewing published stability studies, a certain level of redundancy of tested analytes as well as a heterogeneity in applied methodology and results can be found, making it difficult to transfer stability criteria from such studies onto the local setting. Our Working Group for the Preanalytical Phase (WG-PRE) of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) therefore established a 4-step plan, aiming to standardize and harmonize stabilities data for better comparison and meta-analysis (Figure 1). The first step was to produce a checklist to guide reporting of stability studies which was published in 2020 [6] and the second step, published in 2023, was to produce a how to guide for performing stability studies [7]. The final 2 steps involve a mechanism to evaluate published stability data and to then incorporate this into a central free-to-use resource.

Figure 1: The EFLM WG-PRE 4-step plan, aiming to standardize and harmonize stabilities data for better comparison and meta-analysis.

What is stability?

Defining stability in the context of a biomarker can be difficult. The international Vocabulary in Metrology (VIM) [8] defines stability only for analytical instruments as a 'property of a measuring instrument, whereby its metrological properties remain constant in time'. This statement is applicable to analytes by slight rephrasing to: 'a property of a measurand, whereby its metrological properties remain constant in time'. Furthermore, this definition could be extended, stating analyte stability as the timeframe during which the measurand is being stable under defined conditions, or by stating that a particular analyte changes less than a defined criteria (percentage) over a defined period under defined conditions [6]. In other words, is it defining a limit past which the sample should not be used for the respective analyte testing. Ideally, the result deterioration is expressed by a regression equation, making it possible to calculate individual acceptance limits for the local setting. Either way the main goal should be to ensure that laboratory professionals can add the highest value possible to a laboratory report to ensure laboratory results of sufficient quality for patient's safety.

Variables affecting stability

There are many contributing variables that may affect the stability of a sample and it is important that these are all

understood and accounted for when determining whether the sample is of sufficient quality for analysis. This range of variables needs to be controlled and documented when performing stability studies.

Apart from the time between collection and analyte measurement, variables include the sample type (e.g. blood, serum, plasma, cerebrospinal fluid, saliva or other bodily fluid) or the collection tube type (e.g. EDTA, Citrate, Heparin, etc.) which can vary among manufacturers for the same tube. Sample mixing with the additive may also have an impact on the quality, as inadequate mixing can lead to poor stabilisation of the sample by additives and over vigorous mixing can cause cell damage. Additionally, the tube filling volume may contribute to poor sample quality in terms of potential dilution effects or inadequate additive effect. Another major variable affecting sample stability is the temperatures the sample is exposed to during transport, centrifugation or storage. The centrifugal force during plasma/serum separation is another important factor that can influence the sample quality, e.g. if separation is incomplete. Additional variables potentially affecting sample stability and quality include light exposure if the analyte is light labile, sample evaporation and specifics of the laboratory instrumentation and reagent reaction kinetics [9].

Sample stability may vary between individuals as some have cells which leak more readily than others *in vitro* (e.g. potassium), and others have different enzyme activities, cell counts, protein concentration, all of which potentially influences the analyte's stability. The metabolism in the sample may also lead to the production of the analyte of interest and increase its concentration.

Methods

In order to evaluate sample stability studies, we used the EFLM Checklist for Reporting Stability Studies (CRESS) as a foundation to structure 20 sections against which stability publications would be assessed [6]. Each section then had 4 levels of quality grading, ranging from A for "Best in Class" to D for a fail in that section (Table 1). Initial requirements for the categories A to D were produced following expert discussion among EFLM WG-PRE members. Six published stability studies from the last 10 years were then circulated among these members to test the methodology [10–15]. The manuscripts were sent to members of the EFLM WG-PRE alongside the scoring criteria. Category grading was then refined following these pilot results and the feedback from members of the group.

The scores of A to D for each criteria was then converted to a numerical value (4 for an A down to 1 for a D) and each of the 20 sections was weighted ranging from 0.5 to 3, depending on the importance assigned to each section based on the expert opinion of the EFLM WG-PRE. For example, a low weight was applied to questions on funding and ethics as these factors play a minor role in evaluating sample stability, compared to details about the study population, samples used and the analytical method which are some of the highest weighted criteria (Table 2). Unweighted results were also calculated to demonstrate the merit of using weighted scores.

For weighted scores the maximum score possible was calculated by multiplying the maximum score for each category from the CRESS checklist by the criteria weighting and adding them all together. A final score of A to D was then calculated, based on a final score being a set percentage of the maximum achievable. Percentages used were 80, 60 and 40 % following expert discussion.

For the test stability studies a weighted score for each category from the CRESS checklist stability study was calculated by multiplying the average score from the experts for each category from the CRESS checklist by the weighting and adding them together. For both weighted and unweighted a final score of A to D was then calculated, based on the final score being a set percentage of the maximum achievable as above.

Results

Table 1 presents a standardized model utilizing the CRESS checklist strategy for evaluating the quality of stability manuscripts. This framework covers typical sections and important parameters commonly found in stability manuscripts, encompassing 20 distinct evaluation criteria. Defined by consensus, each item includes a specific question and the definition of four levels of quality, ranging from best in class (A) to Fail (D), thereby providing a score from 4 to 1. Additionally, weighted scores based on the importance of each criterion were also determined and are detailed in Table 2. To validate this strategy, six stability manuscripts were evaluated by members of the WG-PRE.

There were varying levels of heterogeneity between both assessors and manuscripts but this would be expected to some extent due to the remaining subjective nature of the process and the piloting phase of the evaluation process. However, even at this stage, the majority of scores differed only by a single category. The observed grading variations were lower in the higher scoring studies, compared to the lower scoring manuscripts.

The final scoring of the manuscripts ranged from A to C and the ranking itself for the chosen manuscripts didn't display any difference. Table 3 displays the scores for weighted and unweighted results alongside the percentage scores. However, the percentage scores did differ ranging from 52 to 81 % for the non-weighted scores and 51–84 % for the weighted scores. This emphasises that weighting is of importance to demonstrate the differences between stability studies.

Table 1: Working group for the preanalytical phase.

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
1	Title/ keywords	Does the title clearly indicate that the content relates to a stability study and contains the component(s)/ analyte(s) and sample material tested?	Title clearly indicates that this is a stability study and mentions the tested analytes (or type of analytes if a large number) the sample matrix, the storage temperatures and time period over which it was conducted Keywords also reflect these facts	Title clearly indicates that this is a stability study and mentions the tested or type of analytes Some or all additional information regarding the study is missing	Title states that the paper is a stability study but lacks additional information	No clarity that the manuscript is a stability study	Title must reflect that the publication was designed to primarily study stability, the analytes covered, the matrix they are in and the time period covered to ensure it is found in any searches
2	Abstract	Does the abstract state that this is a stability study and include the aim of the study? Does it include a short description of the study design including the analyte(s) tested, sample matrix, container type and manufacturer, the number of samples tested, the duration of time and any other relevant conditions tested. Finally, are the major results and a conclusion included?	The abstract states that the study is a stability study and clearly defines what the aims of the study are It includes an overview of the study design and any protocols followed. It defines which analytes are being examined (or type of analytes if a large number), in which matrices, in which tube types and manufacturer, under which condition and over what period of time. Headline results and the key findings of the study are highlighted and appropriate conclusions are drawn	The abstract states that the study is a stability study and defines what the aims of the study are It includes a limited overview of the study design and defines which analytes (or type of analytes if a large number) are being examined. Headline results and the key findings of the study are highlighted and appropriate conclusions are drawn Some information is missing	The abstract states that the study is a stability study An overview of the study design is included Headline results and the key findings of the study are highlighted Some or all other information is missing	Unclear description of the aims and methodology	The abstract must accurately reflect the aims, methods/ components, results and conclusions of the paper
3	Introduction	Are the rationale and importance of knowledge about stability of sample material for different analytes pointed out and emphasised? Are the different conditions and factors contributing to stability mentioned? Does it detail what the current knowledge about the stability of the analytes included in the study is and what the background to	Discusses the importance of sample stability Provides a thorough overview of the position using relevant literature and existing evidence in the area and clearly outlines what the knowledge gap being filled by the study in question is and why this is needed to benefit the profession and patients	Discusses the importance of sample stability Provides a limited overview of the position and relevant studies in the area but is lacking in some literature sources. Outlines what question is being answered by the study but does not discuss the knowledge gap Does not fully discuss benefits for patients or the profession	Discusses sample stability Outlines what question is being answered by the study	Incomplete background to the study and its purpose	A well written introduction is important to draw people's attention to the study, and to explain why the subject is important with impact on patients

Table 1: (continued)

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
		the study is. This should include why it is important and what evidence gap it is filling					
4	Aim	Is the aim of the study clearly described and is it clear how the study will address this?	Clearly defines the aims of the study, including the analytes (or type), matrix, tube type and manufacturer, analyser and manufacturer, temperature and duration involved. It is clear how the current study will address the aims	Defines the aims of the study and mentions the matrix, time period, storage condition and analytes investigated. Some additional information is missing	Defines the aims of the study and mentions the matrix, time period, storage condition and analytes investigated	Unclear definition of the study aims	A precise description of the aim is necessary to perform the study and make use of its results
5	Materials and Methods	Are the materials and methods described in enough detail to allow other healthcare settings to consider applying the data to their own population and if not is there sufficient detail to replicate the study?	States where study was done and when it started and finished Describes materials and methods used in sufficient detail for other laboratories to replicate the study. Provide details of reagents and assay performance There is sufficient detail to consider transfer of the study to different healthcare settings	Describes materials and methods used in sufficient detail for other laboratories to replicate the study. Provide details of reagents and assay performance. Some information is missing preventing transferability of the study to different healthcare settings	Describes the equipment used but lacks detail of reagents and lot numbers	Very limited overview of the methodology insufficient to allow the study to be reproduced	One must clearly understand the underlying methodologies and assumptions of any scientific study in order to judge whether or not the conclusions can be generalised or the results applied to other settings. Inadequate description, or an adequate description of materials/ methods that cannot be utilised elsewhere, seriously reduce the usefulness of published conclusions
6	Measurand	Are the measurands clearly documented?	All measurands/analytes clearly defined using unambiguous standardised international terminology. Sufficient detail provided to allow the study to be replicated by other laboratories	All measurands/ analytes clearly defined. Sufficient detail provided to allow the study to be replicated by other laboratories	All measurands/ analytes partially defined	Analytes are not well defined and/or referred to as a collection e.g. standard biochemical tests	Adequate description of the measurand
7	Samples	Does the manuscript detail the sample collection procedure? This should include details of the matrix, sample volume,	Defines the tube type, sample matrix, tube type/additive(s) and the manufacturer The volume of sample collected is included, whether they were specifically collected	Defines the tube type, manufacturer, matrix and additive(s) The volume of sample collected is included, whether they were	Defines the tube type, manufacturer, matrix and additive(s)	Sample collection unclear perhaps specifying only serum/plasma	Detailed description of samples and their collection conditions is important to be able to validly evaluate and compare stability studies

Table 1: (continued)

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
		manufacturer and tube type, any additives, whether samples were pooled and whether or not the samples were surplus or taken for the study	for the study or if surplus samples have been used and whether they were pooled is stated Details of how the samples were collected and any protocols followed is included	specifically collected for the study or if surplus samples have been used Some additional information is missing			
8	Origin of samples	Does the manuscript clearly define the patient population or source that the samples came from sufficiently that the results can be applied to similar populations	The population from which samples were taken is defined to include as a minimum geographical location, any relevant comorbidities, age, sex, any factor that could impact the analyte Note these may have been included deliberately but should be documented. States how this information was obtained and verified	The population from which samples were taken include some l information additional to basic demographics if relevant, potentially impacting the analytes but is incomplete Does not mention how this information was obtained or verified	Basic patient population detail e.g. age/gender provided but lacking further details	Patient population not defined	Knowledge of the population from which samples are taken is important to aid other laboratories in applying the stability data to their own population
9	Preanalytical conditions	Are all preanalytical aspects of the study described in sufficient detail to allow all aspects of the study to be replicated and applied to different healthcare settings and/or to allow a more detailed data evaluation, taking all potential biases into account?	States details of how the sample was collected including patient preparation and sample mixing Details sample transport conditions e.g. time, season, temperature, forces etc. Details all times and conditions the sample was exposed to from collection to analysis including the overall length of time to first analysis Indicates any interferences in the sample(s) and any other relevant preanalytical factors such as freeze thaw cycles Sufficient detail provided to allow the study to be replicated and to consider transfer of the study to different healthcare settings	States details of how the sample was collected including sample mixing Details sample transport conditions e.g. time, temperature, forces etc. Some detail missing but sufficient detail provided to allow the study to be replicated however insufficient detail to consider transfer of the study to different healthcare settings	Details relevant preanalytical conditions from the time zero point onwards including storage conditions, time temperature, centrifugation etc. Insufficient detail to replicate study without making assumptions	No specifications on preanalytical biases	It is important to include details of the whole preanalytical journey to enable other laboratories to understand any potential factors that could have influenced the results

Table 1: (continued)

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
10	Analytical procedure	Is the method(s) used to measure the analytes of interest described in sufficient detail	States the method of analysis, the reagent used including lot number, its traceability, the analytical platform, relevant details of the reaction including any deviations from manufacturers recommendations or references to a source where the method is detailed. Includes within and between batch variation (CVA%) data and confirms that IQC was monitored during the study States the number of replicate analyses, defines whether analysis was done as a single batch or multiple batches and states any between batch controls as relevant Sufficient detail provided to allow the study to be replicated and to consider transfer of the study to different healthcare settings	States the method of analysis, the analytical platform or references a source where the method is detailed Stated the number of replicate analyses and confirms that IQC was monitored during the study Define whether analysis was done as a single batch or multiple batches and state any between batch controls as relevant Sufficient detail provided to allow the study to be replicated but insufficient to consider transfer of the study to different healthcare settings	States the method of analysis and the analytical platform. Insufficient to fully enable study replication without making assumptions	Limited information on the analytical procedure or the method is obsolete and no longer valid	Adequate description of the analytical method is necessary to ensure transferability of data
11	Spiking studies (if applicable)	Was spiking necessary as part of the study and if so how was it performed	Justifies why (if any) spiking studies were necessary and performed. Details what protocol was followed including the material used to allow replicability	Justifies why (if any) spiking studies were necessary and performed. Insufficient protocol detail to allow replicability	States spiking was performed and material used but does not detail protocol	No details on spiking provided	To understand the transferability of a study it is important to understand all aspects and any sources of variation
12	Duration of study	Is there a clear description of the study duration and frequency of analysis	States and justifies the duration over which stability of the analytes in the study will be studied and all the time points analysed through the course of the study	States the duration over which stability of the analytes in the study will be studied and all the time points analysed through the course of the study. Does not justify the time points or duration	States the duration over which stability of the analytes in the study will be studied	Unclear or no mention of the duration and or frequency of time points	It is important this information is readily available to allow other healthcare professionals to understand the transferability of the paper
13	Storage conditions	Were the storage conditions during	Defines sample storage conditions clearly	Defines storage conditions clearly		Unclear storage details included	It is important this information is

Table 1: (continued)

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
		the study clearly defined, documented and controlled	including how they were monitored. Details regarding storage until time zero also documented Details of storage monitoring included, and details of thawing processes (if applicable) included Sufficient detail provided to allow the study to be replicated and to consider transfer of the study to different healthcare settings	including how they were monitored. Details regarding storage until time zero also documented Sufficient detail provided to allow the study to be replicated but insufficient to consider transfer of the study to different healthcare settings	Defines storage conditions and duration		readily available to allow other healthcare professionals to understand any variables in the study and therefore the transferability of it
14	Statistical data analysis	Are details of all statistics used presented with a justification as to why they were selected	Includes justification for the number of samples tested (*a-priori* power calculation) and the number of replicates detailing how this minimises analytical imprecision and whether mean or medians were used for replicates Includes a defined instability equation Defines and justifies what statistical tools were used	Describes the number of samples tested and the number of replicates and whether mean or medians were used for replicates. Some justification provided Includes a defined instability equation Defines statistical tools were used	Describes the number of samples tested and the number of replicates and whether mean or medians were used for replicates Defines what statistical tools were used	Minimal or inappropriate statistics used	To understand the data and any potential statistical anomalies it is important that those performing the study describe and justify the statistical methodologies used. This should include work to ascertain that sufficient samples have been processed
15	Outliers	Has testing for outliers of (a) replicates (b) samples per subject (c) between subjects been performed?	States how outlier testing was performed and how many outliers were identified and define what tools were used to remove outliers and why	Limited outlier testing was performed states and how many outliers were identified. Limited details of mechanism to define outliers	Limited outlier testing was performed states and how many outliers were identified Insufficient details provided around outlier process	Outlier analysis not performed or not stated to have been performed	The presence of outliers modifies the estimates and confidence intervals. In the case of comparison between subjects, the possible existence of individual factors in a specific sample that modify the stability of a quantity must be considered
16	Acceptability criteria	Has a definition of what the maximum permissible difference (MPD) been stated and justified RCV values of all	There is a definition for the MPD and a justification for why that level was chosen e.g. RCV Follows the Milan hierarchy to defining	There is a definition for the MPD and a justification for why that level was chosen e.g. RCV Doesn't follow Milan hierarchy to	There is a definition for the MPD but no justification	No definition of inappropriate MPD used	Each stability study performed will have a different reason for the study to be performed RCV values of all analytes should be calculated. It is

Table 1: (continued)

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
		analytes should be calculated	analytical performance specifications	defining analytical performance specifications			important to understand why the level was chosen with justification as to why this was the case in the studying institute
17	Results	Is there a good clear portrayal of the results in a variety of formats? Is raw data available? Estimates of instability; adequately calculated and presented?	Presents data in at least two of textual, graphical and tabulated format Presents PD% for each experiment and each subject under study and for each sampling time The average of the PD% and its CI in each sampling time is used to compare against the MPD specification Presents data using consistent terminology throughout the manuscript with the use of SI units A stability equation is presented Raw data is available as a supplemental file	Presents data in a tabulated format plus, graphical and/or textual Data is presented using consistent terminology throughout the manuscript with the use of SI units Some additional information is missing. Raw data are missing	Presents data in a tabulated or textual format	No data provided, just headline results	The presentation of data in a variety of ways is important to ensure the full picture is painted. The inclusion of an instability equation and the raw data is critical to allow other laboratory professionals to understand the data and apply it to their own healthcare setting and requirements
18	Discussion	A final discussion of the data is included stating how the study has or has not addressed the original aims	States how the study addressed or otherwise the original aims of the study Discusses findings relative to similar studies and any similarities or differences identified and discussed The implications of the results for the profession are highlighted with a view to the transferability of the results Discusses any limitations identified in the study	Through discussion, not all aspects are covered in sufficient detail	States how the study addressed or otherwise the original aims of the study. Compare the results to those of other studies. Limitations are discussed	Poor or no discussion just a restatement of results	The discussion is critical in summarising the findings in the context of not only the problem it set out to address but also in analysing how the results can applied to other healthcare settings and crucially be further developed in the future
19	Funding (if applicable)	Were there any funding sources as part of the study	State any funding sources or not and states whether financial or in the form of consumables. Full	State any funding sources or not and states whether financial or in the form of consumables. What	Source of funding stated but no additional details	No statement	It is important to identify funding sources

Table 1: (continued)

Item number	Section/ parameters	Question	A (Best in class)	B Intermediate	C Minimum	D Fail	Rationale
			details provided of what aspects were funded	aspects of the study were funded is undefined			
20	Ethics	Was there any ethical approval required and if so was it granted. Was patient consent required and if so was it obtained?	Includes a statement regarding ethical approval or stated to adheres to national regulations on the use of human samples for research use Statement about patient consent included if applicable. A statement on national requirement should be included	Includes a statement regarding ethical approval or stated to adheres to national regulations on the use of human samples for research use	Includes a statement regarding ethical approval. Including stating if not required	No statement	It is important to state that the study has followed all appropriate ethics

CVA, coefficient of analytical variation; IQC, internal quality controls; MPD, maximum permissible difference; RCV, reference change value.

Table 2: Weighting of scores based on importance of quality criteria applied for evaluating stability studies.

Item number	Section	Weighting
1	Title/keywords	1
2	Abstract	1
3	Introduction	2
4	Aim	1
5	Materials and methods	3
6	Measurand	3
7	Samples	3
8	Origin of samples	3
9	Preanalytical conditions	3
10	Analytical procedure	3
11	Spiking studies (if applicable)	0.5
12	Duration of study	2
13	Storage conditions	3
14	Statistical data analysis	3
15	Outliers	2
16	Acceptability criteria	3
17	Results	3
18	Discussion	3
19	Funding (if applicable)	0.5
20	Ethics	0.5

Discussion

Patients have the expectation that results of the ordered tests are accurate and reflect the true state of their health. Laboratory medicine professionals are aware that there are various factors that can affect the accuracy of results, of which most are to be found in the preanalytical phase, affecting sample stability, among others. It is critical that laboratories are aware of the impact of any delays and conditions samples are exposed to during this delay. For many laboratory specialists, consulting the literature is the first place to search for stability data. The current problem is that the measurands evaluated in the published studies overlap and the methodologies differ, making it hard or impossible to retrieve the desired information or to apply the results onto the local settings. To that end, as discussed above, the EFLM WG-PRE have put together a package to guide and standardize the conduct and evaluation of stability studies. The final part of this package, as detailed in this manuscript, was to produce a standardized evaluation process to assess the quality of stability studies, following the sections detailed in the CRESS checklist. This guide will

Table 3: Results of scoring of quality of sample stability studies.

Type	Tanner et al. [14]	Oddoze et al. [12]	Cuhadar et al. [10]	Henriksen et al. [11]	Kift et al. [13]	van Balvaren et al. [15]
Weighted	B	B	B	B	C	B
Non weighted	B	B	B	B	C	B
Percent weighted	B	A	B	A	C	B

allow laboratorians to identify studies that have followed a standard methodology and will contain all the information they require to make an informed decision. Of the six manuscripts evaluated, 2, 3 and 2 scored an A, a B and a C, respectively. There were no scores of D (fail) which perhaps reflects the quality of the peer review process in eliminating the poorest quality manuscripts. It is also worth noting that although this process identifies high quality studies, manuscripts that score low are not necessarily of poor result quality, they just do not adhere to all of the CRESS checklist and therefore will not have all the information now stated to be required in a good quality manuscript.

The quality ranking (A–D) did not vary whether weighted or non-weighted scoring was used, but there was a difference in the percentage scores. This indicates that for a larger volume of manuscripts a difference in categorisation would occur and would differentiate between best in class papers and lower quality manuscripts.

The aim of providing this package of guidance on sample stability is to encourage people to not only perform high quality stability studies but to report their results in high quality manuscripts and to share their raw data. As a fourth step, the EFLM WG-PRE is currently working on a database which will contain published stability data in a structured way and to which local (unpublished) stability data can be uploaded. The vision is that the relevant manuscript is linked to this data and subsequently representatives of the EFLM WG-PRE will be able to perform meta-analysis and assess the manuscript to assign a quality score of A–D to it. That way a laboratory medicine professionals can search for stability information on a particular analyte and filter for the storage conditions in question and be able to apply the data to their own setting.

The big limitation of the proposed evaluation process is the subjectivity of some aspects of the process. We saw differences in the ranking of sections of the different manuscripts assessed and while this was useful in allowing us to eliminate areas that could be misclassified or where the degree of overlap was too great, it will always be a point of variation in this type of process. That said it is very unlikely to impact on the final outcome of any classification due to the wide ranges. It could also be mitigated by having 2 verifiers for all studies.

It is the hope of the EFLM WG-PRE that this final paper in the trilogy in combination with the other manuscripts will lead to an improvement in both the way that stability studies are performed, as well as in the way that they are reported and evaluated. We believe that adhering to the proposed guidance will result in improvements in both quality and standardisation of stability studies.

Research ethics: Not applicable.
Informed consent: Not applicable.
Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission. All authors contributed to the conception, discussion and testing of methodology and all edited manuscript. MC drafted manuscript.
Competing interests: The authors state no conflict of interest.
Research funding: None declared.
Data availability: Not applicable.

References

1. Hallworth MJ. The “70% claim”: what is the evidence base? Ann Clin Biochem 2011;48:487–8.
2. Plebani M, Sciacovelli L, Aita A. Quality indicators for the total testing process. Clin Lab Med 2017;37:187–205.
3. Sciacovelli L, Padoan A, Aita A, Basso D, Plebani M. Quality indicators in laboratory medicine: state-of-the-art, quality specifications and future strategies. Clin Chem Lab Med 2023;61:688–95.
4. Lundberg GD. Adding outcome as the 10th step in the brain-to-brain laboratory test loop. Am J Clin Pathol 2014;141:767–9.
5. ISO 15189:2022. Medical laboratories – requirements for quality and competence. Available from: https://www.iso.org/standard/76677.html [Accessed 17 May 2020].
6. Cornes M, Simundic AM, Cadamuro J, Costelloe SJ, Baird G, Kristensen GBB, et al. The CRESS checklist for reporting stability studies: on behalf of the European federation of clinical chemistry and laboratory medicine (EFLM) working group for the preanalytical phase (WG-PRE). Clin Chem Lab Med 2020;59:59–69.
7. Gomez-Rioja R, Von Meyer A, Cornes M, Costelloe S, Vermeersch P, Simundic AM, et al. Recommendation for the design of stability studies on clinical specimens. Clin Chem Lab Med 2023;61:1–11.
8. BIPM. International vocabulary of metrology (VIM). Available from: https://www.bipm.org/en/publications/guides/vim.html [Accessed 17 May 2020].
9. Gómez Rioja R, Martínez Espartosa D, Segovia M, Ibarz M, Llopis MA, Bauça JM, et al. Laboratory sample stability. Is it possible to define a consensus stability function? An example of five blood magnitudes. Clin Chem Lab Med 2018;56:1806–18.
10. Cuhadar S, Atay A, Koseoglu M, Dirican A, Hur A. Stability studies of common biochemical analytes in serum separator tubes with or without gel barrier subjected to various storage conditions. Biochem Med 2012;22:202–14.
11. Henriksen LO, Faber NR, Moller MF, Nexo E, Hansen AB. Stability of 35 biochemical and immunological routine tests after 10 hours storage

and transport of human whole blood at 21 °C. Scand J Clin Lab Invest 2014;74:603–10.
12. Oddoze C, Lombard E, Portugal H. Stability study of 81 analytes in human whole blood, in serum and in plasma. Clin Biochem 2012;45: 464–9.
13. Kift RL, Byrne C, Liversidge R, Babbington F, Knox C, Binns J, et al. The effect of storage conditions on sample stability in the routine clinical laboratory. Ann Clin Biochem 2015;52:675–9.
14. Tanner M, Kent N, Smith B, Fletcher S, Lewer M. Stability of common biochemical analytes in serum gel tubes subjected to various storage temperatures and times pre-centrifugation. Ann Clin Biochem 2008;45: 375–9.
15. van Balveren JA, Huijskens MJAJ, Gemen EFA, Péquériaux NCV, Kusters R. Effects of time and temperature on 48 routine chemistry, haematology and coagulation analytes in whole blood samples. Ann Clin Biochem 2017;54:448–62.