

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/346470854>

International Council for Standardization in Haematology (ICSH) laboratory guidance for the evaluation of haemostasis analyser-reagent test systems. Part 1: Instrument-specific iss...

Article in *International Journal of Laboratory Hematology* · November 2020

DOI: 10.1111/ijlh.13411

CITATIONS

5

READS

296

8 authors, including:



Chris Gardiner

Leeds Teaching Hospitals NHS Trust

124 PUBLICATIONS 16,197 CITATIONS

[SEE PROFILE](#)



Robyn Coleman

Sullivan Nicolaides Pathology

7 PUBLICATIONS 189 CITATIONS

[SEE PROFILE](#)



Moniek de Maat

Erasmus MC

421 PUBLICATIONS 13,693 CITATIONS

[SEE PROFILE](#)



Akbar Dorgalaleh

185 PUBLICATIONS 1,450 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Neutrophils, NETs, fibrin and wound healing [View project](#)



department of Hematology, Iran Unuversity of Medical Science [View project](#)



REVIEW

International Council for Standardization in Haematology (ICSH) laboratory guidance for the evaluation of haemostasis analyser-reagent test systems. Part 1: Instrument-specific issues and commonly used coagulation screening tests

Chris Gardiner¹ | Robyn Coleman² | Moniek P. M. de Maat³ | Akbar Dorgalaleh⁴ | Marión Echenagucia⁵ | Robert C. Gosselin⁶ | Masahiro Ieko⁷ | Steve Kitchen⁸

¹Haemostasis Research Unit, University College London, London, UK

²Department of Haematology, Sullivan Nicolaides Pathology, Bowen Hills, QLD, Australia

³Department of Haematology, Erasmus Medical Centre, Rotterdam, The Netherlands

⁴Department of Hematology and Blood Transfusion, School of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran

⁵Centro Nacional de Hemofilia, Banco Municipal de Sangre del Distrito Capital, Caracas, Venezuela

⁶Thrombosis and Hemostasis Center, University of California, Davis Health System, Sacramento, CA, USA

⁷Health Sciences University of Hokkaido, Ishikari-Tobetsu, Japan

⁸Sheffield Haemophilia and Thrombosis Centre, Royal Hallamshire Hospital, Sheffield, UK

Correspondence

Chris Gardiner, Haemostasis Research Unit, University College Hospital, London, UK. Email: chris.gardiner@uclmail.net

Abstract

Before a new method is used for clinical testing, it is essential that it is evaluated for suitability for its intended purpose. This document gives guidance for the performance of verification, validation and implementation processes required by regulatory and accreditation bodies. It covers the planning and execution of an evaluation of the commonly performed screening tests (prothrombin time, activated partial thromboplastin time, thrombin time and fibrinogen assay), and instrument-specific issues. Advice on selecting an appropriate haemostasis analyser, planning the evaluation, and assessing the reference, interval, precision, accuracy, and comparability of a haemostasis test system are also given. A second companion document will cover specialist haemostasis testing.

KEYWORDS

haemostasis, laboratory automation, precision

1 | BACKGROUND

Before a new method is used for clinical testing, it is essential that it is evaluated for suitability for its intended purpose. Since the Recommendations for Evaluation of Coagulation Analyzers¹ was published in 2006 a great deal has changed in the clinical haemostasis laboratory.

Laboratory accreditation is now more widespread and rigorous. In many countries, this accreditation requires stringent validation

and verification of diagnostic tests. Examples of the additional testing challenges for clinical haemostasis laboratories are the wide use of automated sample volume and preanalytical screening for haemolysis, icterus and lipaemia (HIL), direct oral anticoagulants, novel treatments for haemophilia and an increasing repertoire of haemostasis assays. While regulatory authorities may approve isolated reagent platforms (eg dilute thrombin time), approved instruments are typically associated with specific reagent applications. Consequently, it may no longer be relevant to consider an instrument evaluation in isolation.

TABLE 1 Glossary of terminology

Term	Definition	Alternative terminology
Accuracy	Closeness of agreement between the result of a measurement and a true value of the measurand	Trueness
Analyte	The chemical substance that is the subject of chemical analysis	
Analytical measurement range	The range of test values that a test system can directly measure without prior dilution or concentration.	
Bias	The difference between the expected (true) test result and the measured test result.	
Analytical sensitivity	The lowest concentration of an analyte that may be reliably measured.	Often synonymous with lower limit of detection
Analytical specificity	The ability of a test system to measure solely the measurand	Specificity
Calibration	A set of operations that establish the relationship between measurement response and the value of the quantity being measured	
Comparability	Closeness of agreement between the results of test system under evaluation with an established method (predicate device).	
Diagnostic sensitivity	The probability of getting a positive test result in subjects with the disease.	
Diagnostic specificity	The proportion of people who are free of a specific disease and are so identified by the test.	
Imprecision	Dispersion of independent results of measurements.	Expressed as coefficient of variation.
Laboratory Developed Test (LDT)	A test developed in a single laboratory, or an approved test which has been modified or used outside its intended scope.	Laboratory defined method
Linearity	The ability of a test system to obtain results which are directly proportional to the concentration of an analyte within a given range	
Lower limit of detection	The minimum detectable concentration of an analyte that can be distinguished from a blank.	Detection limit
Lower limit of quantitation	The minimum concentration at which an analyte can reliably detected with acceptable precision and accuracy.	Quantitation limit
Measurand	The quantity intended to be measured.	CLIA uses "analyte"
Precision	The closeness of agreement between results of successive measurements	Repeatability
Reference interval	The set of values 95% of the normal population falls within	Normal range, reference range, biological RI
Reportable range	The range test result values over which the laboratory can establish or verify the accuracy of the test system.	Clinically reportable range, measuring interval, reportable interval.
Sensitivity	The change in response of a measurement system divided by the corresponding change in a value of the quantity being measured.	
Test system	A laboratory test performed by a defined procedure with specified reagent(s) and equipment.	Examination procedures, measurement procedure, analytical method, assay, procedure, laboratory test, test method, test, or device.
Validation	Provision of objective evidence through a defined process that a test system meets requirements for an intended use.	
Verification	An abbreviated process to demonstrate that a test system performs in substantial compliance to previously established claims.	

Metrological terminology has evolved differently in Europe, the United States and elsewhere, and this is reflected in documents from different regulatory authorities (Food and Drug Administration [FDA] USA, Conformité Européenne [CE marking] Europe, National Association of Testing Authorities [NATA] Australia, Pharmaceuticals and Medical Devices Agency [MDA] Japan, and the Medical Device Single Audit Program). The terms used in this guideline (and the alternative terminology) are listed in Table 1.

Clinical Laboratory Improvement Amendments (CLIA²) refers to test systems and, while ISO 15189 discusses equipment acceptance testing, this is secondary to verification and validation of "examination procedures."³ Other accreditation bodies/regulatory authorities refer to the *analytical method, assay, procedure, laboratory test, test method, device*, or simply the *test*.⁴⁻⁷ For the purposes of this document, *test system*: (ie a laboratory test performed by a defined procedure with specified reagent(s) and equipment, is the most appropriate term.

While existing standards/regulations/guidance dictate the laboratory requirements for method validation and verification, detail is lacking with the onus very much on the laboratory to provide the missing detail. For example, ISO 15189 states that the "validation shall be as extensive as is necessary and confirm, through the provision of objective evidence, that the specific requirements for the intended use of the examination have been fulfilled."³ Most guidance documents are better suited to clinical chemistry, in which well-defined analytes with well-documented reference materials or standards are assayed and a range of suitable matrices is available. Performance characteristics listed in ISO 15189 include consideration of measurement trueness, detection limit and quantitation limit, diagnostic specificity, and diagnostic sensitivity. These characteristics are not applicable to several coagulation tests such as prothrombin time (PT) and activated partial thromboplastin time (APTT). Similarly, while many clinical chemistry systems can be assessed using recovery studies in contrived samples to produce abnormal or results with a wide range of concentrations, this approach is not always applicable to haemostasis test systems.

2 | SCOPE OF THIS GUIDELINE

This is the first of two ICSH guidance documents. Recommendations are based on information from peer-review publications, the authors' personal experience and expert opinion, as well as good clinical laboratory practice principles. This document will address the evaluation of commonly used screening tests, that is, prothrombin time (PT) with International Normalized Ratio (INR), activated partial thromboplastin time (APTT), thrombin time (TT) and fibrinogen assays, and instrument-specific issues. The second document will cover the evaluation of other haemostasis assays used on haemostasis analysers. Guidance or recommendations for the evaluation of point-of-care devices (eg PT/INR for patient use) or near-care systems (eg thromboelastography) are outside the scope of this document.

This guidance document provides recommendations on how to plan and execute the processes required for the selection and

evaluation of haemostasis analysers/test systems. It is not intended to replace regional requirements, regulations, or standards, but rather to give guidance on the steps necessary to meet the requirements of the regulatory authorities.

The extent of an test system evaluation will depend on several factors including: (a) the intended use of the test system, (b) whether the test system has been deemed approved for clinical use by regional regulatory bodies or statutes, (c) the resources available to the laboratory.

3 | SELECTION OF A HAEMOSTASIS ANALYSER

Market research should be performed to establish which haemostasis analysers are available. Publications such as CAP Today (<https://www.captodayonline.com/coagulation>) produce a listing of instrumentation with operational features, approved in the US. A specification of requirement (SOR) should be drafted, detailing the physical laboratory characteristics (eg floor/tabletop space, electrical service, etc) limitations with the laboratory's required and desirable instrument features and characteristics. An example of this SOR that may be requested by the laboratory and/or instrument manufacturer are shown in Table 2.

3.1 | Recommendations for SOR

- Describe the desired equipment in terms of its intended use (instrument test menu) and the required level of performance, rather than by a generic description or brand name.
- Be concise, but sufficiently detailed to enable manufacturers to address all relevant costs.
- Detail the criteria for system consideration, including any health and safety requirements, any international and national quality requirements, and comply with relevant national and international laws.
- Describe any physical requirements or restrictions, including but not limited to footprint (floor of countertop space), electrical, waste, or other environmental needs.

The weighting given to acceptance may depend on additional laboratory settings. For example, a laboratory with a large paediatric workload will place a great deal of importance to minimum sample size, whereas options for clotting factor and von Willebrand factor assays will be greater importance to a haemophilia treatment centre.

4 | BRIEF ASSESSMENT OF LOAN INSTRUMENT

If the laboratory has the option of loaning the instrument prior to acquisition, a limited assessment of performance may be useful

TABLE 2 Information required prior to analyser selection

Information required by the laboratory	Information required by the manufacturer
Analyser requirements	Operating environment
Safe operating temperature	Range of temperature
Range of acceptable humidity	Range of humidity
Restrictions regarding direct sunlight	Bench/floorspace available
Sound level produced by device dBA	Sources of vibration
Heat dissipation	Exposure to dust
Physical dimensions (height, width, length and weight)	Proximity of instrument to power source/number of power points
Water/drainage requirements	Provision of uninterruptible power supply
Power requirements (voltage, uninterruptible power source)	Proximity of instrument to water/drainage
Compatibility with laboratory automation system	Width of doorways/elevators into the laboratory
Analyser capabilities	Degree of laboratory automation/robotics
Throughput (tests per hour)	Laboratory workload
Sample volume/microsampling capacity	Daily workload (samples/tests per day)
Acceptable primary tube sizes/types	Number of stat tests per day
Test menu and capacity	Expected turnaround time for routine and stat samples
Sample and reagent dead volumes	Breakdown by hour of how samples are received by the laboratory
Closed tube sampling	Peak hourly workload
Stat sample turnaround time	QC (number of levels, daily frequency) and calibration regimes required
Reflex testing/redilution capabilities	Sample tubes used
On-board reagent capacity (stirring/refrigeration)	Test menu required
User-defined test programming	Expected changes in future workload
Ability to use reagents from other manufacturers	Experience and skill level/mix of laboratory personnel
Maintenance requirements	Number of people requiring training
Personnel requirements	Expected training period required
Analytical performance	Storage capacity for reagents/consumables
Precision	Information technology
Comparability	Current laboratory/hospital information systems
Linearity	Anticipated changes in laboratory/hospital information systems
Interfering substances	Reliability of laboratory/hospital information systems
Carryover	Connectivity options to the internet for remote support by the manufacturer
Limit of detection/quantitation	IT security and any antivirus policies
Reportable range	
On-board reagent stability	
Information management	
Interface capabilities	
Compatibility with laboratory/hospital information systems	
On-board quality control capabilities	
Standalone capability/data storage	
Bar code reading options	
Ability to run concurrent lots of reagents	
Calibration curves storage capability	
Recalculation of results	
Customer support	
Technical support and service agreement	
Response times	
Ongoing training and education	
Peer group data analysis	

in making the purchasing decision. The scope of this assessment will be dependent on the length of the loan, and the type of work performed by the laboratory. As a minimum, the PT, APTT and TT and/or Clauss fibrinogen assays should be assessed for within-run imprecision (repeatability) by testing at least two levels of QC

10-20 times in a single run, and (if time permits) between-run (intermediate or within-laboratory) imprecision by testing the same QC samples in triplicate over 3-5 days.^{8,9} A limited comparability assessment (eg 10-50 samples over days with a wide range of values), comparing the results of the new method with results of samples tested by the current routine test system (predicate device) may also be performed.

Whereas a full and thorough evaluation of the test system in accordance with this guidance is the preference this is not always possible and for many reasons. Where it is not possible sites should request from the supplier a full list of sites using the test system and set up either a visit to see the system in use or at least talk to the appropriate scientist at one or more sites by way of reference. This is best executed independent of the supplier.

Depending on the type of laboratory work to be performed, a broader range of tests may be assessed. For example, a one-stage PT-based factor assay, a one-stage APTT based factor assay, a chromogenic assay and an immunoturbidometric assay (and aggregation-based assay if available) should give a good indication of the instrument's capabilities.

Note: Data collected from loaned equipment does not mitigate the laboratory responsibility for fulfilling instrument evaluation recommendations for newly installed equipment used for clinical testing and should not be used in the final evaluation.

5 | PLANNING THE EVALUATION

It is important to estimate a realistic timescale for the evaluation, based on the available resources and the scope of the evaluation. We recommend the development of an evaluation plan (see Table 3) that details the stepwise evaluation process, to be reviewed and approved by the section director or designate. The plan should detail the evaluation parameter (eg imprecision), the test(s) to be performed, and the desired outcome (eg statistical limits).

As some test systems may require unusual samples or those that cover a wide range of values, it may be advisable to start collecting and freezing samples weeks, or even months, in advance (see below for recommended freezer temperatures). The quantities of consumables (including reagents) required for the test system evaluation should be estimated with allowance for contingency plans in case additional work is required (eg additional reagents of the same lot numbers with an expiry data beyond the end of the planned evaluation date should be held in reserve).

The laboratory must document each stage of the evaluation process, including (but not limited to) any instrument preventative maintenance, temperature assessments, and data generated from validation or verification steps. The laboratory director or designate must review the data and record the outcome of the review. It is advisable to have a dedicated, labelled binder or digital record folder for the system evaluation. Regardless, any form of test system evaluation documentation must be readily available for regulatory or accrediting agency inspections.

TABLE 3 Evaluation plan

Item	Action required
Establish need for new analyser	Write specification of requirement (SOR) and request tenders according to local regulations.
Establish short list	Compare manufacturer's specifications with SOR.
Select analyser	Consider preacquisition assessment.
Preacquisition assessment	Review data against manufacturer's specifications.
Establish scope of evaluation	Verification or validation? Establish number of samples/test required
Plan evaluation	Estimate time and resources required. Identify and allocate resources. Collect Samples in advance.
Installation	Conduct preliminary precision testing to verify instrument integrity. Review results before proceeding.
Calibration and precision testing	Review results and compare with manufacturer's specification before proceeding
Comparability testing	Review results at regular intervals. Identify problems and perform additional testing if required.
Reference intervals	Check results against manufacturer's data and/or published RI. Perform additional testing if required.
Additional testing	If validation rather than verification is required, perform additional checks, for example, on-board reagent stability
Final review	Thoroughly review all data. Identify necessary changes to reporting (eg RI and therapeutic ranges) and inform clinicians. Review SOPs and training requirements required before implementation.
Set timetable for implementation	Establish a realistic schedule for implementation to ensure minimal disruption.
Implementation	Perform debrief after implementation. Review workflow and perform surveillance.

5.1 | Selection of samples for evaluation

Residual plasma samples that were collected as part of a patient's routine care are typically used for test system evaluations. Blood samples collected from normal donors for reference interval (RI) determination, or from patients, for the sole purposes of test system evaluation may require informed consent. Therefore, each site must determine whether local Institutional Review Board (IRB) or equivalent approval is required prior to blood collection.

Only citrated plasma is suitable for most haemostasis testing. The sodium citrate concentration for the evaluation samples must be the same as that used for patient testing, with recommended concentration at 3.2% (0.109 mol/L).¹⁰ The citrated blood should be collected according to local recommendations and processed to obtain platelet poor plasma (PPP) defined as having a residual plasma platelet count of $<10 \times 10^9/L$.

For PT testing, the stability of processed plasma is 24 hours, when maintained at room temperature. Plasma samples for other tests should be tested within 4 hours of collection, although there are other literature supporting longer stability when maintained at room temperature.¹¹ Samples collected for unfractionated heparin (UFH) testing should be centrifuged within 1 hour of collection and processed within 4 hours.^{12,13}

If samples cannot be tested within acceptable stability limits, the PPP samples should be aliquoted into labelled cryovials (or equivalent) and stored frozen at -70°C or below until required.¹⁴

To obtain the wide range of samples required for a thorough evaluation, it may be necessary to use some frozen samples. If -70°C freezers are not available, storage of processed PPP plasma samples at -20°C for up to 2 weeks is acceptable provided that freezers with automatic defrost cycles are not used where partial thaw of samples may occur.¹²

Prior to analysis, frozen samples should be thawed at 37°C (3-5 minutes for aliquots of up to 1 mL) and mixed thoroughly immediately prior to testing. The use of frozen samples is preferable to lyophilized plasma, as lyophilization may introduce artefactual changes as compared to freezing plasma. However, for laboratories with limited resources, lyophilized plasmas may be used if there are no alternatives, provided they are tested within their period of post-reconstitution stability. Results from a fresh plasma may differ from results obtained from the same plasma after freezing and thawing.¹⁵ As such, once frozen plasma that has been appropriately thawed, it should be tested within 4 hours on both predicate and new test systems.

It is not always possible to collect all the samples required for test system evaluations, for example, samples from patients with a range of factor deficiencies. In this situation, contrived samples may be used, for example, serial dilution of known reference material with factor deficient plasmas to produce a range of factor deficient plasmas across the reportable range. It is important to realize that contrived plasmas may not behave in the same manner as plasma samples from patients with factor deficiencies.¹⁶

6 | INSTALLATION AND TRAINING

Installation will normally be performed by the manufacturer/distributor. There may be institutional requirements or departments (eg plant operations) that verify the safety of the instrument prior to use (eg assure no electrical discharge, proper grounding requirements). In all cases, a copy of the manufacturer's electrical safety certification should be provided by the supplier (eg IEC/EN 61010-1). There should be adequate logistics, such as space (for instrumentation), storage (eg refrigeration of evaluation reagents) and personnel to assure proper evaluation. Processes should be implemented to assure that evaluation instrument(s) and reagent(s) are labelled accordingly to assure they are not used for clinical testing.

Instrument training may be undertaken on-site or at the manufacturer's premises but should be completed before the performance evaluation, especially for new test systems. For on-site programmes, the selected laboratory personnel should be available for training without interruption. The training course should cover instrument theory and operation, maintenance and basic trouble shooting. The training should have documentation attesting to operator competence. The instrument operator's manual should be supplementary to, not instead of, thorough training and readily available to system users. The operator's manual should be available in a language understood in the laboratory.

Where reagents or wash/rinse materials are to be provided by the supplier, which may be specific to the test equipment, Instructions for Use (IFU) and COSHH/MSDS sheets should be provided and evaluation staff familiarized with them.

6.1 | Recommendation

- All local instrument validation and/or verification work should be performed by the laboratory staff in the location in which the instrument will be used.
- All staff involved in the evaluation should undergo an instrument familiarization period, and a competency assessment prior to evaluation process.
- Documentation of staff instrumentation competency must be maintained in the evaluation binder.

7 | VALIDATION OR VERIFICATION?

As it is important to understand when a test system requires validation or verification (this is discussed in detail by Antonelli et al¹⁷), this guidance document covers several different approaches:

- Validation of wholly new test systems or a Laboratory Developed Test (LDT – see below)
- Verification of newly introduced validated test systems
- Verification of test systems already in use in a laboratory for accreditation purpose

- Limited verification of test systems following the relocation of equipment (required by CLIA).

This list is not exhaustive, and each laboratory will need to determine the scope of any validation or verification. ICSH recognizes that there is no universal definition of an LDT and it is therefore the responsibility of the individual laboratory/institution to decide whether a test system constitutes a LDT and to implement a validation procedure that meets their local regulatory requirements.

If the validation of a test system is intended to provide data for an application for regulatory approval, it will be more comprehensive than a local validation of an LDT. Similarly, transferring reagents already in routine use to a new instrument, if both instruments use the same measurement system and interpretative criteria, would require a more limited verification than if both instrument and reagent were changed.

7.1 | Recommendation

- Wherever possible, medical laboratories should select procedures which have been validated for their intended use.
- If a test system has not received the relevant regulatory approval, it will require validation locally, which is defined as the provision of objective evidence through a defined process that a test system meets requirements for an intended use.

ISO 15189 describes four situations in which the local laboratory should validate a procedure which it defines as: Nonstandard methods; laboratory designed or developed methods; standard methods used outside their intended scope; validated methods subsequently modified. Defining what constitutes sufficient degree of local modification to require new validation may not be straightforward and may be interpreted differently by different regulatory or accreditation bodies. Validation is required for LDT. An LDT is defined by the FDA as an in vitro diagnostic test that is manufactured by and used within a single laboratory.¹⁸ In this context, standard methods used outside their intended scope and validated methods which have been subsequently modified are also considered LDTs. This is particularly relevant for laboratories testing paediatric samples with test systems which have not received regulatory approval for use in paediatric samples. The use of a reagent/kit on another manufacturer's analyser, when that particular reagent/analyser combination does not have regulatory approval, is also considered an LDT by most regulatory authorities.

Generally, if a test system has been cleared by the relevant regional regulatory authority, it is only necessary to verify the test system locally. Verification may be defined as the provision of objective evidence that a given test system locally fulfils the manufacturer established specifications. In most cases, the established specifications will be provided by the manufacturer and are dependent on following the manufacturer's IFU for the instrument/reagent combination described. Deviation from the manufacturer's IFU for a test system will require validation as an LDT (Table 4).

TABLE 4 LDT for coagulometer: Guidance only, consult regional or regulatory rules and requirements for the identification of an LDT

LDT Element	Examples	Coagulometer considerations ^a	Validation considerations
Modification of an existing measurand protocol	<ul style="list-style-type: none"> • Modifying dilutions for factor assay testing • Modifying calibration curve parameters • Modifying diluent for factor assays • Prolonging incubation times for low levels of factor activity 	<ul style="list-style-type: none"> • Instrument protocol name change • Instrument reagent name change(s) • Instrument calibrator setup change • Instrument calibration acceptance change 	P, MC, Lin, RI, AMR
Introduction of new reagent into existing instrument programme	<ul style="list-style-type: none"> • Modifying defined PT reagent for factor assay • Modifying defined APTT reagent for factor assay • Modifying defined calibrator source • Modifying defined factor deficient plasma for factor assays • Modifying Heparin Anti-Xa to FX DOAC testing 	<ul style="list-style-type: none"> • Instrument protocol name change • Instrument reagent name change(s) • Instrument calibrator setup and potential sample acceptance change • Potential interfacing, reporting, billing • Potential clinician notification 	P; MC; RI; Lin; AMR; OBS; CO
Introduction of new measurand instrument protocol	<ul style="list-style-type: none"> • Creating new chromogenic factor protocol (eg VII, IX) • Creating ecarin based assay • Creating textarin time • Creating new platform for protein S activity • Analysing animal samples^b 	<ul style="list-style-type: none"> • New instrument protocol • New instrument reagent changes • New instrument calibrator requirements • New Interfacing, reporting, billing • Clinician notification 	P; MC; RI; Lin; AMR; OBS; RCO (SBO required if sample type outside manufacturer IFU or single pipet coagulometer system)

Abbreviations: AMR, analytical measurement range; Lin, linearity, including lower limit of quantitation; MC, Method Comparison; OBS, on-board reagent stability; P, Precision; RCO, Reagent carryover; RI, Reference Interval; SCO, Sample carryover.

^aThe coagulometer may require these steps be performed in a specific sequence.

^bPotentially, depending on animal source, viscosity, anticipated clotting times, etc if not validated by manufacturer.

Some regulatory authorities require a limited verification of test systems following the relocation of equipment. At the very least, this will require a reassessment of imprecision at least two levels for each test affected. Verification should also be done for any loan analyser which is installed as a temporary replacement for an analyser removed from service for repair or upgrade.

7.2 | Recommendations

- Verification is sufficient for test systems with regulatory approval.
- Test systems which have not been validated, including LDTs, will require evidence that the test system meets requirements for an intended use. This may include but may not be limited to establishing a new RI, extended comparability testing, on-board reagent stability, carryover, and establishment of therapeutic ranges appropriate to the laboratory. Test-specific linearity, LoD and LoQ should be listed as both critical to evaluation/use where expected by accreditation bodies.
- If an approved reagent is used on an instrument from another manufacturer and there is no IFU for this reagent/instrument test system, this constitutes an LDT.

8 | PERFORMANCE EVALUATION

At the very start of any evaluation (immediately after installation), we recommend performing the within-run imprecision of the measurands to be evaluated (see below). This ensures that all the instrument operations (eg pipetting, detection channels) are working and appropriately calibrated. This procedure is useful for identifying “fliers” (ie sporadic statistical outliers), and variability in detector channels. Identifying problems early during the validation/verification can save valuable resources such as laboratory staffing time and instrument consumables.

The amount of testing necessary for an evaluation will depend upon whether the test system(s) require validation or verification. For method comparison or accuracy studies, the samples tested will depend on the type of patient samples expected to be tested in the laboratory. In most situations, it is only necessary to validate or verify the performance of a test system with respect to the types of samples tested in your laboratory. For example, while it is important to understand the sensitivity of screening tests to anticoagulant effect (pharmacodynamics) used in your patient population, if a given anticoagulant is never (or rarely) encountered in your laboratory, testing samples from patients receiving that drug may not be considered a wise use of resources.

8.1 | Recommendation

- If the coefficient of variation of the initial within-run imprecision is statistically higher than the manufacturer's specification, or otherwise higher than expected, contact the manufacturer before proceeding with the evaluation.
- Quality control ranges should be established during the performance evaluation.

8.2 | Precision testing

Precision refers to the closeness of agreement between independent test results. In practice, we measure imprecision, that is, the degree of variation about the mean. This may be reported as the standard deviation (SD), or the coefficient of variation (CV, mean divided by SD) which is usually expressed as a percentage (%CV).

There are two measurements of precision in common use: Within-run and between-run, with total precision a cumulation of both data sets. Within-run imprecision (repeatability) is assessed by testing the same sample multiple times in a single run of measurements. Each test should be run independently (eg run PT multiple times, then run APTT multiple times). Between-day/within-laboratory/intermediate imprecision is typically evaluated by testing two or more levels of QC over several days. As most haemostatic analytes are not stable beyond a period of a few hours, precision is best assessed using frozen and aliquoted samples or freshly reconstituted lyophilized normal and abnormal control plasmas. In the absence of resources to obtain multi-level controls, pooling of samples to obtain adequate volumes for precision testing would be acceptable. If pooling of samples is to be used for between-run precision, then the sample(s) should be appropriately aliquoted and frozen, with each day of precision testing performed on thawed samples (see comments above on acceptable freeze and thaw conditions).

To establish the imprecision of a method (for validation purposes), CLSI recommends that two replicates of each QC are performed twice a day on each of 20 days.⁸ To verify the imprecision of a method, CLSI recommends five replicates over 5 days with one daily series.⁹ The manufacturer's precision claims should be verified by comparing the measured imprecision to the manufacturer's claim.

To validate an assay, it is necessary to perform repeat calibrations over five separate days to check calibration curve stability. This may be ascertained by calculating the CV% of the slope values and is essential for assays which use stored calibration curves.

8.2.1 | Recommendations for precision testing

- For within-run precision:
 - A minimum of 10 replicates of a single sample should be tested within a single run. The mean and CV% should be calculated and compared to IFU limits

- Ideally there should be ≥ 20 replicates of each sample, with 2-3 samples covering the measurement range of the measurand (eg for PT would be normal, prolonged, and markedly prolonged). The mean and CV% should be calculated and compared to IFU limits.
- Pooling of plasma samples to determine precision of abnormal samples is acceptable if proper mixing is assured.
- For between-run precision
 - Minimum should be duplicate analysis of quality control material once daily for at least 5 days. The mean and CV% should be calculated and compared to IFU limits.
 - Ideally, replicate analysis of quality control material on each shift for at least 5 days to be performed by multiple testing personnel. The mean and CV% should be calculated and compared to IFU limits.
 - Pooling of plasma samples to determine precision of normal or abnormal samples is acceptable if each day of precision testing is performed on thawed material.

8.3 | Reference intervals

Validation of a new method requires establishing a reference interval (RI) in a minimum of 120 samples from normal healthy individuals.¹⁹⁻²¹ The most important consideration in developing reliable RIs is the proper selection of reference subjects, testing and adequate number of subjects, and avoidance of preanalytical sources of error. Reference individuals should not be hospital or clinic patients unless absolutely necessary, as might be in the case when establishing RIs for children or elderly patients.²¹ Although the selection of suitable donors and consideration of preanalytical variables is beyond the scope of this document, important considerations in collecting samples include; ensuring that the normal samples are collected, processed in the same way as patient samples, ensuring that the donor has not undergone intense physical exercise in the previous 24 hours, and excluding donors who are pregnant or receiving hormonal therapy. Establishing a new RI is covered in detail elsewhere.^{19,21,22} We do not recommend using patient samples that are within the normal limits reported by the laboratory, as these samples may not reflect normal levels of factor activity, especially in hospitalized patients where acute phase reactants (fibrinogen and factor VIII) are typically elevated.

Transference of a manufacturer's RI (if provided in the reagent IFU) is acceptable if verified. A minimum of 20-40 ostensibly healthy donors should be tested. If 18/20 (90%) of these donors fall within the manufacturer's RI, then transference may be used. If <90% of the donors are within the manufacturer's RI, then additional donor samples are required, increasing the number of donors to 30 or 40. In some situations, age-specific RIs (eg neonatal or paediatric populations) or clinical threshold cut-off values (eg D-dimer for venous thromboembolism exclusion) may be required, which are usually provided by reagent IFU or reference literature.

Note: Literature for paediatric and neonatal reference intervals for screening tests are typically generated from one reagent type and should be interpreted and used with caution.

Laboratories reporting PT/INR results calculated from a locally derived geometric mean PT (GMNPT) will need to test at least 20 normal plasmas.²³ Donors for GMNPT determination or RI verification should be ostensibly healthy individuals who are not on anti-coagulants, oral contraceptives and have not had vigorous exercise in the previous 24 hours.²⁴ Many laboratories without access to phlebotomy services and/or normal donors may purchase normal donor plasma panels from approved commercial sources. It should be noted that the sample collection and processing can affect results of PT and APTT¹² and commercial panels may not be collected and processed in the same way as the patient samples arriving in the local laboratory.

Reference interval data should be checked for normality and, if the data have a non-normal distribution, a nonparametric RI (eg 95% confidence intervals) should be used.

8.3.1 | Recommendations for RI determination

- The use of transference of manufacturer's RI (if provided in the manufacturer IFU) for verifying test system performance.
- That transference of manufacturer RI cannot be used for LDTs
- In the absence of appropriate donors, we recommend that institutions utilize RI for neonates, infants and other paediatric populations derived from appropriate literature.^{25,26}
 - The RI noted on any report should indicate these are from publications and were not locally determined.
- Samples collected from patients within the normal range of the laboratory measurand should not be used for RI determination or verification, except for paediatric or geriatric RIs.

8.4 | Accuracy

Accuracy is defined as closeness of agreement between a measurement and the true value. No standard exists for noncalibrated coagulation screening tests (PT, APTT and TT), so there is no true value. Accuracy is typically performed comparing new instrumentation or reagent systems to an existing or predicate method.

A degree of standardization for INR has been achieved, so verification of the accuracy of a PT/INR may be best achieved by testing suitable EQA materials or through evaluation comparability with an established method. If the PT/INR test system is not validated, this will require assigning an International Sensitivity Index (ISI)/calibrating the INR which is beyond the scope of this document and readers should refer to WHO guidance.²³ INR calibrants are commercially available for certain test systems and there are other commercial sources that provide frozen plasma of varying assigned INRs, which are designed to assist local verification of ISI and MNPT.

8.4.1 | Recommendation

Early provisional registration in an established external quality assurance scheme (EQA) may be useful to establish accuracy of the test system ahead of patient testing.

8.5 | Comparability

In the absence of true value for PT, APTT and TT, comparability with an established method may serve as a substitute for accuracy determination. It is essential that the samples used for the assessment of comparability span the reportable range of each tests system. The number and type of samples tested will depend upon whether the new test system has been previously validated, the expected types of patient samples to be tested in the laboratory and the resources available.

Of note, the PT and APTT have multiple clinical uses. Each can be used for: (a) for the initial investigation of factor deficiencies, (b) monitoring anticoagulant therapy (eg warfarin or UFH), or (c) assessing efficacy of treatment (eg replacement therapy such as fresh frozen plasma, prothrombin concentrates, etc). In addition, many APTT reagents are influenced by lupus anticoagulant. As such, comparability studies should use samples that reflect the clinical purpose of each test at a given laboratory. If the laboratory test samples do not include the use of the APTT for UFH monitoring, then comparing those samples to a predicate method may not be as relevant as other clinical uses for this test. It is important for the laboratory to understand the clinical utility of their provided test menu and be able to address differences between test systems once transition occurs to a new system.

Comparability should be analysed by linear regression (ordinary, weighted, Deming, or Passing-Bablok, as appropriate), Bland Altman bias plots, and paired-*t* test (or Mann Whitney *U* test if the data are not normally distributed). The acceptability criteria will be test-specific, for example, for tests such as PT or APTT for which there is no standard, two different test systems using different reagents will usually generate results with differences which may be both clinically (suggesting a change in patient diagnosis or management) and statistically significantly different, especially if the RIs are different between reagent platforms. If a new test system represents an improvement on the existing method, accepting changes in results (and RI) using the new test system would be justified. Additional considerations for comparability examination could include sensitivity, specificity, the negative and positive predictive value of the test. If however, two test systems use the same reagent on different analysers, the correlation is generally very strong ($r > .98$) and this can give rise to a small bias which may be statistically highly significant but far too small to be of clinical relevance. Calibrated assays for which there is an international standard (eg fibrinogen) and standardized methodology, should produce a regression line with a slope close to unity (0.90-1.10), with strong correlation ($r > .95$) and no clinically significant bias (eg 90% of samples within 15% difference).

The regulatory agencies do not specify the number of samples required to establish comparability but recommend that the samples should cover the full reportable range. CAP suggest 20-40 samples may be sufficient for method verification. When a *t* test is used to compare the results of a measurand in two test systems for a one-SD difference, with a preset alpha of 0.05, 16 samples are required to obtain a statistical power of 0.80, while 26 samples would be required to obtain a power of 0.95.^{7,27}

Example: If after testing 20 samples, the paired-*t* test yields a probability (*P*) value of .80 and correlation has a *r* value of .99, this would be adequate for test verification and running additional tests is unlikely to change the outcome. On the contrary, if after 20 tests *P* < .05, *r* < .90, and/or Bland Altman analysis shows a significant bias (eg <90% of samples within 15% difference) or growing bias with increasing values, additional testing would be indicated. Where agreement around clinical decision limits may be of particular importance (eg D-dimer results close to cut-off incorporated into protocols for exclusion of venous thromboembolism) then analysis of a sub-group of data points around such decision limits is useful.

Validation of a previously unvalidated test will require a minimum of 40 samples but perhaps >100 samples may be required. The precise number of samples will depend upon the acceptability criteria. A large laboratory may run many PT tests over several days to assess the comparability of a new method but is unlikely to test a similar number of samples using expensive specialized assays.

8.5.1 | Recommendations

- For test system verification. The absolute minimum number of samples required for comparability testing is 20. Whether further samples will be required depends upon the test concerned and whether acceptability criteria have been met using the minimum number of tests.
 - Minimum limits of statistical acceptability may include *t* test of *P* < .05, regression *r* > .90, and/or lack of significant bias (>90% of samples within 15% difference) or within total allowable error
- For accuracy validation of an LDT or other nonvalidated test system, a minimum of 40 samples but more than 100 samples may be required.
- No more than two samples should be tested from the same patient (if patient ID is known)

8.6 | PT/INR

Due to differing phospholipid compositions and sources of tissue factor, significant differences in results from different PT test systems are commonplace. Consequently, the method comparison may not reach the desirable statistical limits as used for calibrated assays.

This does not preclude the use of the new test system, as comparability testing is meant as an informative process so the laboratory or interpretive/end users are aware of the areas of bias and sensitivities of the new test system compared to that currently used. These may be reduced by expressing the result as a PT ratio but this is not widely used in clinical practice.

Converting the PT to INR should in theory eliminate clinically significant biases in patients receiving vitamin K antagonists (VKA). Validation of a previously unvalidated INR test system will require ISI assignment and comparability testing with at least 20 normal plasmas and 60 samples from patients receiving VKA.²³ The relative difference between the mean INR of the two methods for samples within the therapeutic range should be within 0.2 INR units and the line describing the relationship between logarithms of patient prothrombin times should ideally pass through the mean of the logarithms of normal prothrombin times.²³ If these conditions are not met, this indicates that either the ISI assignment or the MNPT is incorrect.

To verify an established PT/INR test system (ie with a method-specific ISI) it is necessary to test 20-40 samples from stably anticoagulated patients receiving VKA and 20 normal plasmas.⁷ Although two INR methods may only be truly linear up to an INR of 4.5 it is still important that samples used span the full reportable range, as clinicians use these results to guide clinical decision making. Acceptability has previously been defined as >85% of samples within the therapeutic range resulting in a difference of <0.5 INR units.^{7,28}

The PT/INR is not reliable for the measurement of DOAC, and should not be used to quantify DOAC concentrations.²⁹ Furthermore, the sensitivity of thromboplastins to DOACs is highly variable, being dependent on both the drug and reagent. However, it may be important for staff to understand the sensitivity of PT/INR test system to DOACs regularly encountered in their laboratory, for example, in a patient with known DOAC exposure, a prolonged PT may indicate the presence of DOAC.²⁹ Some PT reagents are totally insensitive to Apixaban.^{30,31} If the manufacturer has not provided information on DOAC sensitivity, the laboratory may wish to establish the sensitivity of a new PT test system to all DOACs regularly encountered in its patient population. This will require testing 20-30 plasmas covering the "on-therapy" drug concentration for each DOAC.³² Verification of manufacturer's stated sensitivities would require a smaller number of plasmas from patients receiving each relevant DOAC covering a wide range of DOAC concentrations. If parenteral thrombin inhibitors (DTI, for example, argatroban and bivalirudin) are used in the hospital population, the sensitivity of the PT/INR test system to these should also be determined.

Large reference laboratories and those serving haemophilia treatment centres (HTC) may be required to verify or validate the sensitivity of a new PT tests system to deficiencies of tissue factor pathway factor (factors II, V, VII and X). As congenital deficiencies of these factors are very rare, this can be achieved by serial dilution of known reference material with factor deficient plasmas.^{24,31,33}

Plasmas from patients with congenital deficiencies should also be used if available.

8.7 | APTT

Comparability testing of APTT should include normal plasmas and abnormal plasmas with APTTs covering the full reportable range. The type of samples will depend upon the patient population and the type of testing performed in the laboratory. If the APTT is used to monitor UFH therapy, at least 20 samples for verification and at least 30 for validation,³⁴ from patient's receiving therapeutic UFH doses (assayed for anti-Xa activity) must be tested to determine the APTT heparin therapeutic range (HTR)^{24,35} (see supplemental methods). In vitro spiking of UFH will overestimate reagent sensitivity and thus is not recommended.³⁴ In hospitals where low molecular weight heparin or DOACs have fully replaced the use of UFH, it is not necessary to establish a heparin HTR.

As with the PT/INR, the sensitivity of APTT test reagents to DOACs and parental thrombin inhibitors is highly variable, and the same advice applies. Where these anticoagulants are used, the laboratory may wish to verify or establish sensitivity to DOACs or parenteral DTIs on the hospital formulary.³⁶ For parenteral DTIs, it may be necessary to inform the pharmacy or clinical staff if large variations of clotting times are observed between predicate and investigational APTT method are observed.

The discussion regarding lupus anticoagulant (LA)-sensitivity in the selection of APTT reagents is beyond the scope of this guideline but is covered in detail by Fritsma et al.³⁷ Although for general APTT use, it is recommended to avoid using LA sensitive reagents,³⁷ nevertheless, it may be important for the laboratory to understand the LA effect of the APTT test systems. If the manufacturer has provided data regarding the sensitivity of the test system to lupus anticoagulant, verification may be achieved by testing a small number of plasma samples from patients with known LA and a similar number of patients known to be negative for antiphospholipid antibodies (10 in each group should suffice). Review of EQA results may also assist in reagent selection in terms of lupus sensitivity but should always be locally verified.

Understanding the sensitivity and specificity of APTT reagents to factor deficiencies is important. Most reagents are designed to detect deficiencies in FVIII, FIX and FXI at a clinically significant level (0.3 IU/mL in North America, whereas in Europe 0.40–0.45 IU/mL is the normal). In contrast, in the context of screening for increased risk of bleeding, high sensitivity to FXII and/or lupus anticoagulant is considered undesirable.³⁷ If the manufacturer has provided data on sensitivity to factor deficiency, verification of factor sensitivity may be unnecessary in most small laboratories if the other comparability data are acceptable. This will not be the case for laboratories associated with HTCs where establishment or verification of factor sensitivity may be required (see supplementary methods).

8.8 | Thrombin time

Comparability testing of thrombin time test systems should include plasmas from patients with disseminated intravascular coagulation (DIC) or critically ill patients with raised D-dimer levels, patients receiving LMWH, patients receiving UFH (if used locally), severe liver disease, patients with low fibrinogen (hypofibrinogenemia, <2.0 g/L) and high fibrinogen (hyperfibrinogenemia >5.0 g/L). Additionally, samples from patients with abnormally functioning fibrinogen (dysfibrinogenemia), and patients receiving Dabigatran and/or parenteral thrombin inhibitors may be tested if available and relevant.

Thrombin time test systems vary considerably in thrombin concentration, calcium ion concentration and sample dilution schemes between manufacturers, and are sometimes performed as LDTs. It is unlikely that thrombin time test performance evaluations would satisfy statistical criteria, especially if the RIs are different.

8.9 | Fibrinogen assays

Fibrinogen assays based on the Clauss method differ from the other screening tests in that they are calibrated against a reference material. In addition to comparability testing in samples with fibrinogen concentration spanning the reportable range, it is also necessary to verify the lower limit of detection (LOD), lower limit of quantitation (LOQ) and linearity (see Table 1 for definitions). This is outside the scope of this document and will be covered by the companion ICSH guidance document for evaluation of specialized haemostasis test systems.

Test systems which have not previously been validated may also require investigation of on-board reagent stability and standard curve stability over several days or weeks.

Several manufacturers also offer a PT-derived fibrinogen value, based on the magnitude of the change in plasma optical density during the clotting process during a PT test. Reporting this value is usually restricted to the RI but, even within this range, the method can give apparently normal results in patients with an abnormal Clauss fibrinogen assay.

8.9.1 | Recommendation

- The use of the PT-derived fibrinogen value is not recommended in the diagnostic laboratory.³⁸

9 | EFFICIENCY ASSESSMENT AND INSTRUMENT-SPECIFIC ISSUES

The efficiency assessment should include, but may not be limited to, ease of use, throughput using a representative laboratory workload, barcode reader error rates, compatibility with the laboratory

information system, time required for maintenance, reliability (down time and manufacturer response).

Manufacturers are required to submit validated instrument specifications during the regulatory approval process. This will include assessment of reagent carryover, sample carryover, on-board reagent stability and interference. Assessment of these specifications is not necessary for method verification but may be required for validation of LDTs (eg when using a reagent on another manufacturer's analyser) or other methods without regulatory approval). These processes are described in the supplementary methods. Other components of method evaluation including sample carryover and interference may be considering necessary for LDTs in some regulatory environments. Laboratories should follow relevant local and regional guidance on this issue.

10 | IMPLEMENTATION

The implementation of the chosen analyser will be influenced by many factors. The major determinant will be the scope of change to be implemented. An update in similar technology and offered tests might cause no significant disruption, whereas a change in technology may require significant change management process, internally and externally to the laboratory. The nomination of a project manager to oversee and lead the change process is a key step in a successful project. Clear documentation, including the use of checklists to assure completion of all necessary implementation components would be favourable. Requirements outside the laboratory environment, including changes in Information Technology (IT), hospital coding or billing requirements (especially if new tests are going to be implemented) should be considered. A timeline for installation (setting achievable targets), regulatory requirements, training, supply chain, LIS interactions, engineering support, the effect on end users, reports, laboratory publications also need to be addressed prior to implementation. Additionally, the number of sites that may be affected, the function of those sites (core/reference or satellite laboratory) and the population they service (hospital or community) must be considered. The potential disruption to service provision must be assessed and contingency plans implemented where required, especially if drug monitoring or therapeutic targets are changed, as this may require institutional approval and processes that take weeks to implement, such as UFH dosing protocols. Whenever such changes are made, the users of the service must receive advance notice of the planned date/time of changeover with confirmation that the change occurred as planned immediately after it occurs.

An extensive test of the LIS interface end to end, from sample presentation to result reporting is critical.

A test plan should be set that covers all likely scenarios for the laboratory. Each step needed to ensure that the expected outcome is achieved must be documented as follows:

- Most instruments have bi-directional interfaces. Ensure the correct test is ordered and that results are filed into the correct

report field.

- Check if the HIL and error flagging, if available, transmits in a useable format to allow remote authorization, otherwise access may be required to the instrument work screen to check error messages and sample integrity checks.
- Check all reporting units are correct. Include in the review, the final report(s) to the clinician, in all available formats, paper and electronic. A hospital setting may have one uniform data management system but referrers in the community may use a variety of practice management software.
- If installing at more than one instrument or site, check all LDT are programmed correctly.
- Repeat the LIS Interface testing for each instrument installation.
- Where space is available, running a new instrument up besides the old one allows a smoother transition.
- If replacing like-for-like, then there may be a gradual switch over of assays. Where the change is distinct then a nominated date will be required.
- If reference ranges/tests/report units are affected, then consideration should be given as to the handling of archived or cumulative reports.
- If replacing more than one instrument, a rolling replacement programme, with sufficient time between each placement for review of any issues that arise, is a wise approach.
- After implementation, make time for a process debrief as it can provide a unique learning opportunity.
- Review laboratory workflow to ensure the benefits of any additional features of the new instrumentation are maximized.
- Assure proper (laboratory approved) procedures (manual or electronic) are readily available for each laboratory staff member that will perform system testing (for both instrumentation and measurand).
- Assure departmental procedure for designation of critical values and critical tests, especially if new test system(s) are implemented that have been identified as a critical test or have been assigned a critical value threshold.³⁹
- Assure sufficient laboratory personnel are properly trained (create a competency plan and document each operator competence) to operate the instruments during laboratory hours of operation.
- After implementation of new system(s), consider a surveillance after a sufficient period (eg 1 month) of all implemented tests and reporting methods to assess client satisfaction.

10.1 | Measurement of uncertainty

Estimation of measurement uncertainty (MU) has become a common requirement in some regulatory environments. The purpose of MU is to ensure results are fit for purpose and provide information that laboratories can use to improve their test methods and make informed decisions. There are a number of different approaches to

calculate MU and in the selection of the method of calculation the laboratory should consider the nature and purpose of the test, the ease of access of the information available to make the calculation of MU and how the test results will be used as well as a consideration of risk associated with decisions based on the test result.⁴⁰⁻⁴² The most practical method for most laboratories is to use internal QC data over an extended period (minimum 6 months) to calculate mean and SD. The SD is then multiplied by 2 to reflect 95% confidence limits. Some of the commercial QC programs provide for the calculation of MU. Result bias may be assessed using external QAP over several cycles. MU calculations should be re-assessed at intervals to ensure ongoing relevance to the test results reported.

11 | SUPPLEMENTARY METHODS

11.1 | Factor sensitivity

The factor sensitivity of a PT or APTT test system is defined as the maximum level of factor activity which produces a clotting time above the upper limit of the reference interval. For the APTT, establishment of sensitivity to FVIII, FIX, FXI and FXII by serial dilution of known reference material with factor deficient plasmas will be the minimum requirement.²⁴ The reference material should have a factor concentration close to 100 IU/dL/100% of normal for all relevant factors, while the deficient plasma should contain <1 IU/dL (<1%) of factor to be adjusted.

However, serial dilution with a single source of factor deficient plasma and normal plasma may not be sufficient.¹⁶ In this situation, it is normally desirable to test samples from patients with a wide range of factor deficiencies, if the samples are available. Some laboratories may also wish to simulate multiple factor deficiencies by diluting normal plasmas. In addition, the sensitivity of the APTT test system to novel treatments such as emicizumab⁴³ and extended half-life FVIII and FIX products may be useful in HTC's that routinely use these products.⁴⁴

11.2 | Heparin sensitivity

Where the APTT is used to monitoring UFH therapy, the APTT test system should be adequately responsive to UFH. If the manufacturer has not provided an APTT range corresponding to the recommended therapeutic UFH concentration (typically 0.3 to 0.7 anti-Xa heparin units/mL), the laboratory will need to establish a therapeutic reference interval. This is achieved by testing at least 20 (at least 30 for validation) plasma samples from patients receiving UFH therapy, covering the entire therapeutic range. In vitro spiked samples are not suitable.²⁴ A regression graph of APTT in seconds (y-axis) against heparin anti-Xa international units (x-axis) is plotted and the time in seconds corresponding to the therapeutic range is recorded.³⁴ Data scatter may be very variable as factor VIII levels can affect response of APTT to heparin.

11.2.1 | Recommendations

- At least 20 samples from patients receiving therapeutic doses of UFH (ex vivo) should be tested for verification of manufacturer's specifications and at least 30 samples for validation of an unvalidated method
- In vitro spiking is not suitable for determining HTR.

11.3 | Carryover testing

Carryover testing is not required for test systems with regulatory approval, nor is it required for analysers with dedicated delivery systems for thrombin containing reagents and PT reagents. However, if a single probe is used for all reagents or used for both reagents and samples, it may be necessary to assess reagent carryover in previously unvalidated test systems. This is achieved by repeatedly testing several aliquots of a single plasma with a thrombin sensitive method (eg APTT or one-stage FVIII clotting assay) and a method using reagents with a high concentration of thrombin (eg Clauss fibrinogen or antithrombin assay) and looking for a progressive shortening of the clotting time in the thrombin sensitive test. Detailed instructions are given in CLSI H-57A.⁷ Carryover of heparin neutralizers present in PR reagents may also cause carryover problems and this should be investigated when a PT reagent is used on another manufacturer's analyser.

Assessment of sample carryover is not normally required for modern haemostasis analysers while testing plasma samples.

11.4 | Assessing the influence of interfering substances

This is rarely necessary for previously validated test systems. If it is required, the clot detection method will have a major bearing on the evaluation of HIL testing.

Optical interference due to HIL is test system specific and tends to cause problems at lower wavelengths. Haemolysis that occurs in vitro in blood samples prior to testing interferes in mechanical as well as optical systems.⁴⁵ To assess the ability of an instrument to deal adequately with HIL, a large number of samples with HIL concentrations spanning the range encountered in the laboratory need to be tested on each unvalidated test system/LDT (typically 30 for each interfering substance). Plasma haemoglobin, triglyceride, cholesterol, and bilirubin concentrations should be analysed using specific assays. The maximum HIL level at which a result may be reliably obtained should be recorded. If desired, the relationship between measured HIL and HIL instrument flags may be determined but this is beyond the scope of this document. It should be noted that samples created by addition of haemolysate to plasma may not replicate all the relevant changes associated with haemolysis that has occurred during sample collection and processing.⁴⁵

If lipoglycopeptide antibiotics are widely used, the influence of these drugs on the test system may need to be assessed if not already done by the manufacturer.

11.5 | Assessing reagent on-board stability

For LDTs and other unvalidated test systems, it may be necessary to assess the on-board stability of reagents. This will also apply to test systems used outside the manufacturer's recommended operating conditions. Freshly reconstituted, or opened, reagents should be placed on the analyser and, after the recommended equilibration time, at least two levels of freshly prepared QC material should be tested in triplicate, every 24 hours until the manufacturer's stated stability time is reached or the QC values have drifted outside acceptable limits, for example, outside the between-day precision limits for the test system.

ACKNOWLEDGEMENTS

All authors were part of the ICSH coagulometer evaluation Working Group chaired by CG and all contributed to the design, drafting and editing of this document. ICSH is a "not for profit" organization.









CONFLICT OF INTEREST

CG serves as a consultant for Sysmex UK. MPMd.M. has received speaker fees from Siemens, Roche, Werfen. RCG has provided expert testimony for dabigatran and rivaroxaban testing, has received honoraria from Siemens Healthcare Diagnostics, Machaon Laboratories and Diagnostica Stago, and serves as a consultant for Diagnostic Grifols and UniQure, and advisory board member for BioMarin. SK has received speaker/consultancy fees from Roche, Siemens, Werfen, and performed funded studies for Grifols, TCoag, Roche, Diagnostica Stago and Sysmex. RC, AD, ME, and MI report no conflicts.

DATA AVAILABILITY STATEMENT

There are no primary data associated with this publication.

ORCID

Chris Gardiner  <https://orcid.org/0000-0002-2318-0062>
 Robyn Coleman  <https://orcid.org/0000-0002-1488-2644>
 Moniek P. M. de Maat  <https://orcid.org/0000-0001-7749-334X>
 Akbar Dorgalaleh  <https://orcid.org/0000-0002-0125-9319>
 Marión Echenagucia  <https://orcid.org/0000-0002-3777-7185>
 Robert C. Gosselin  <https://orcid.org/0000-0002-5669-8722>
 Masahiro Ieko  <https://orcid.org/0000-0003-1854-4543>
 Steve Kitchen  <https://orcid.org/0000-0002-6826-8519>

REFERENCES

- Gardiner C, Kitchen S, Dauer RJ, Kottke-Marchant K, Adcock DM. Recommendations for evaluation of coagulation analyzers. *Lab Hematol*. 2006;12:32-38.
- Clinical Laboratory Improvement Amendments. State operations manual appendix C - survey procedures and interpretive guidelines for laboratories and laboratory services, Vol 3. 2017:5-21.
- International Organization for Standardization. ISO 15189:2012 - Medical laboratories-Requirements for quality and competence. Int Organ Stand ISO. 2012.
- PMDA. Current situation of regulations and premarket review in future of companion diagnostics in Japan approval reviews of IVD products: devices and reagents. 2019. <https://www.pmda.go.jp/files/000153694.pdf>. Accessed July 1, 2020.
- National Association of Testing Authorities Australia (NATA). Validation and verification of quantitative and qualitative test methods. 2018.
- Australian Government Department of Health. National Pathology Accreditation Advisory Council (NPAAC). NPAAC Secretariat.
- Clinical Laboratory Standards Institute. *H57-A Protocol for the Evaluation, Validation, and Implementation of Coagulometers*, Vol 28. Wayne, PA: Clinical and Laboratory Standards Institute; 2008:1-33.
- Clinical Laboratory Standards Institute. *EP5-A2 Evaluation of Precision Performance of Quantitative Measurement Methods; Approved Guideline – Second Edition*, Vol 24, 25th edn. 2004:1.
- Clinical Laboratory Standards Institute. *EP15-A3 User Verification of Precision and Estimation of Bias; Approved Guideline—Third Edition*, Vol 34, 12th edn. Wayne, PA. 2014:1.
- Clinical Laboratory Standards Institute. In: Adcock DM, ed. *H21 A5 Collection, Transport, and Processing of Blood Specimens for Common Laboratory Tests*, Vol 24, 25th edn. Wayne, PA: Clinical and Laboratory Standards Institute; 2008.
- Favaloro EJ, Lippi G, Adcock DM. Preanalytical and postanalytical variables: the leading causes of diagnostic error in hemostasis? *Semin Thromb Hemost*. 2008;34:612-634.
- Adcock Funk DM, Lippi G, Favaloro EJ. Quality standards for sample processing, transportation, and storage in hemostasis testing. *Semin Thromb Hemost*. 2012;38:576-585.
- Adcock D, Kressin D, Marlar RA. The effect of time and temperature variables on routine coagulation tests. *Blood Coagul Fibrinolysis*. 1998;9:463-470.
- Gosselin RC, Honeychurch K, Kang HJ, Dwyre DM. Effects of storage and thawing conditions on coagulation testing. *Int J Lab Hematol*. 2015;7:551-559.
- Gosselin RC, Dwyre DW. Determining the effect of freezing on coagulation testing. *Blood Coagul Fibrinolysis*. 2015;26(1):69-74.
- Lawrie AS, Kitchen S, Efthymiou M, Mackie IJ, Machin SJ. Determination of APTT factor sensitivity – the misleading guideline. *Int J Lab Hematol*. 2013;35:652-657.
- Antonelli G, Padoan A, Aita A, Sciacovelli L, Plebani M. Verification or validation, that is the question. *J Lab Precis Med*. 2017;2:58.
- FDA/CDRH. The Public Health Evidence for FDA oversight of laboratory developed tests: 20 case studies Office of Public Health Strategy and Analysis. FDA. 2015.
- Castellone DD. Establishing reference intervals in the coagulation laboratory. *Int J Lab Hematol*. 2017;39(Suppl 1):121-127.
- Mackie I, Cooper P, Lawrie A, Kitchen S, Gray E, Laffan M. Guidelines on the laboratory aspects of assays used in haemostasis and thrombosis. *Int J Lab Hematol*. 2013;35:1-13.
- Horowitz GL, Altaie S, Boyd JC, et al. *EP28-A3c: Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline—Third Edition*, Vol 28. 2010.
- Marlar RA. Hemostasis test validation, performance and reference intervals. In: Kitchen S, Preston FE, Olson JD, eds. *Quality in Laboratory Hemostasis and Thrombosis*, 28, 5th edn. Hoboken, NJ: John Wiley & Sons, Ltd; 2009:9-18.
- WHO Expert Committee on Biological Standardisation. WHO Expert Committee on Biological Standardization. *World Health Organ Tech Rep Ser*. 2013;Annex 6:271-316.
- Clinical Laboratory Standards Institute. *H47-A2 One-Stage Prothrombin Time (PT) Test and Activated Partial Thromboplastin Time (APTT) Test; Approved Guideline*, 2nd edn, Vol 28. Wayne, PA. 2009.

25. Andrew M, Paes B, Johnston M. Development of the hemostatic system in the neonate and young infant. *J Pediatr Hematol Oncol.* 1990;12:95-104.
26. Toulon P, Berruyer M, Brionne-François M, et al. Age dependency for coagulation parameters in paediatric populations: results of a multicentre study aimed at defining the age-specific reference ranges. *Thromb Haemost.* 2016;116:11-16.
27. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J.* 2003;20:453-458.
28. Fitzmaurice DA, Gardiner C, Kitchen S, Mackie I, Murray ET, Machin SJ. An evidence-based review and guidelines for patient self-testing and management of oral anticoagulation. *Br J Haematol.* 2005;131:156-165.
29. Gosselin RC, Adcock DM, Bates SM, et al. International Council for Standardization in Haematology (ICSH) recommendations for laboratory measurement of direct oral anticoagulants. *Thromb Haemost.* 2018;118:437-450.
30. Gosselin R, Grant RP, Adcock DM. Comparison of the effect of the anti-Xa direct oral anticoagulants apixaban, edoxaban, and rivaroxaban on coagulation assays. *Int J Lab Hematol.* 2016;38:505-513.
31. Gardiner C, Kohama K, Patel I, et al. A performance evaluation of a novel human recombinant tissue factor prothrombin time reagent (Revohem™ PT). *Int J Lab Hematol.* 2017;39:532-538.
32. Douxfils J, Gosselin RC. Laboratory assessment of direct oral anticoagulants. *Semin Thromb Hemost.* 2017;43:277-290.
33. Dumoulin EN, Fiers L, Devreese KM. Investigation of sensitivity for coagulation factor deficiency in APTT and PT: how to perform it? *Clin Chem Lab Med.* 2016;54:e169-e172.
34. Marlar RA, Clement B, Gausman J. Activated partial thromboplastin time monitoring of unfractionated heparin therapy: issues and recommendations. *Semin Thromb Hemost.* 2017;43:253-260.
35. Kitchen S, Jennings I, Woods TA, Preston FE. Wide variability in the sensitivity of APTT reagents for monitoring of heparin dosage. *J Clin Pathol.* 1996;49:10-14.
36. Van Cott EM, Roberts AJ, Dager WE. Laboratory monitoring of parenteral direct thrombin inhibitors. *Semin Thromb Hemost.* 2017;43:270-276.
37. Fritsma GA, Dembitzer FR, Randhawa A, et al. Recommendations for appropriate activated partial thromboplastin time reagent selection and utilization. *Am J Clin Pathol.* 2012;137:904-908.
38. Mackie IJ, Kitchen S, Machin SJ, Lowe GDO. Guidelines on fibrinogen assays. *Br J Haematol.* 2003;121:396-404.
39. Gosselin RC, Adcock D, Dorgalaleh A, et al. International council for standardization in haematology recommendations for hemostasis critical values, tests, and reporting. *Semin Thromb Hemost.* 2019;46:398-409.
40. Lim YK, Kweon OJ, Lee M-K, Kim B, Kim HR. Top-down and bottom-up approaches for the estimation of measurement uncertainty in coagulation assays. *Clin Chem Lab Med.* 2020;58:1525-1533.
41. National Association of Testing Authorities Australia (NATA). General accreditation guidance estimating and reporting measurement uncertainty of chemical test results. 2018. <https://www.nata.com.au/phocadownload/gen-accreditation-guidance/Estimating-and-reporting-measurement-uncertainty-of-chemical-test-results.pdf>. Accessed July 1, 2020.
42. Clinical and Laboratory Standards Institute. In: Kallner A, ed. *EP29-A Expression of Measurement Uncertainty in Laboratory Medicine*, Vol 32, 1st edn. Wayne, PA: Clinical and Laboratory Standards Institute; 2012:1.
43. Adamkewicz JI, Chen DC, Paz-Priel I. Effects and interferences of emicizumab, a humanised bispecific antibody mimicking activated factor VIII cofactor function, on coagulation assays. *Thromb Haemost.* 2019;119:1084-1093.
44. Young GA, Perry DJ. Laboratory assay measurement of modified clotting factor concentrates: a review of the literature and recommendations for practice. *J Thromb Haemost.* 2019;17:567-573.
45. Woolley A, Golmard J-L, Kitchen S. Effects of haemolysis, icterus and lipaemia on coagulation tests as performed on Stago STA-Compact-Max analyser. *Int J Lab Hematol.* 2016;38:375-388.

How to cite this article: Gardiner C, Coleman R, de Maat MPM, et al. International Council for Standardization in Haematology (ICSH) laboratory guidance for the evaluation of haemostasis analyser-reagent test systems. Part 1: Instrument-specific issues and commonly used coagulation screening tests. *Int J Lab Hematol.* 2020;00:1-15. <https://doi.org/10.1111/ijlh.13411>