PART 2 Hemostasis Test Validation, Reference Intervals, and Performance

Silmara Montalvão

TOPICS COVERED

- ✓ Method for Preparation of PNP
- ✓ How to Validate Coagulation Equipment and Tests
- ✓ Samples for Equipment/Reagent Validation Processes
- ✓ Validation or Verification
- ✓ Precision and Accuracy
- ✓ Establishing a Reference Interval
- ✓ Validation Versus Verification of Reference Interval
- ✓ Statistical Analysis of Reference Interval
- ✓ Internal Quality Control and External Quality Assessment

Pooled normal plasma (PNP) is an essential component for the hemostasis laboratory, as it is used in different testing protocols, from evaluating a prolonged APTT to evaluating specific and non-specific inhibitors. It can also be used as reference material for calibration and normal control if ideal conditions are observed for this purpose. The following is a protocol for how PNP should be prepared.

Table 4. Requirements for the preparation of PNP

Donors	Minimum 20 normal healthy individuals not taking medications that interfere with clotting factors and coagulation reaction. It is acceptable to include women taking oral contraceptives. An approximately equal number of males and females is desirable. The age range should be 20 to 50 years.
Anticoagulant	0.109 M (3.2%) trisodium citrate dihydrate buffered with N-2-hydroxyethylpiperazine. N-2-ethanesulphonic acid (HEPES) at 5 g per 100 ml trisodium citrate.
Collection	Donors are bled between 9:00 a.m. and 11:00 a.m. using 60 ml plastic disposable syringes and 21-gauge butterfly needles.

Method for Preparation of PNP:

- ✓ Collect 54 ml blood and mix with 6 ml anticoagulant in plastic containers.
- ✓ Store sample on melting ice during preparation of pool.
- ✓ Centrifuge at 4°C for 15 minutes at 2500g.
- ✓ Pool plasma in plastic non-contact container.
- ✓ Aliquot in 1.5 ml plastic vials in 0.5 ml aliquots.
- ✓ Snap freeze on dry ice/solid CO_2 if available. Alternatively, place immediately on an open shelf at -70°C.
- ✓ Complete the above procedure within four hours.
- ✓ Stable at -70°C for > six months.

A PNP prepared in this way will have levels of factor II (FII), factor V (FV), factor VII (FVII), factor IX (FIX), factor X (FX), factor XI (FXI), factor XII (FXII), high-molecular-weight kininogen (HMWK), and prekallikrien (PKK) of around 1 U/ ml or 100 U/dl, although the levels of FVIII and von Willebrand factor (VWF) vary widely in different pools of PNP. A local PNP should be calibrated in International Units (IU), since international standards are now available for all the above-mentioned clotting factors, except for FXII. The pool can be used uncalibrated with an assumed potency of 100 U/dl or 1 U/ml for FXII. To calibrate in IU, it is

necessary either to obtain calibrated World Health Organisation (WHO) reference preparations (which are held at the National Institute for Biological Standards and Control, South Mimms, Potters Bar, Herts, U.K.) or to purchase suitable commercial reference plasma that has been calibrated in IU by the manufacturer. Consideration should be given to replacing such a plasma pool every 12 to 18 months, unless there is evidence from internal quality control (IQC) results that stability has been maintained.

Method for calibration of local PNP:

- ✓ Obtain a calibrated standard, such as WHO international standard (IS) (minimum two vials).
- ✓ On two different days, use one vial of IS and four aliquots of local PNP.
- ✓ On day one test IS, local, local, local, local, IS, and repeat this using fresh dilutions of each plasma.
- ✓ On day two test local, local, IS, IS, local, local, local, and repeat this using fresh dilutions of each plasma.

Calculate potency of each aliquot of local standard against average of results with the two IS.

- ✓ The mean result of 4 aliquots x 2 dilutions x 2 days (n = 16) is assigned to the local standard as its potency.

How to Validate Coagulation Equipment and Tests: Before a new method is used, it is essential that it be evaluated for its suitability for the intended purpose. This chapter provides a general recommendation on how to plan and carry out the processes required for the selection and evaluation of hemostasis analyzers/testing systems. These recommendations are not intended to replace regulations or standards, but rather to provide guidance on the steps necessary to meet good laboratory practice.

The extent of the evaluation of a test system 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 bodies or statutes, (c) the available resources of the laboratory. When choosing a test system, research should be carried out to determine which hemostasis analyzers are available. A list of requirements should be drawn up to identify the best system for your laboratory, detailing the physical characteristics of the laboratory as well as the required performance of the equipment. Validation planning is an important stage and is related to the results obtained since good planning can generate a technically adequate validation. To this end, it is important to estimate a realistic time frame for the evaluation, based on the resources available and the scope of the evaluation. It is necessary to define the details of the assessment process (in stages), which must be reviewed and approved by the person in charge of the sector. The validation plan should detail the evaluation parameter (e.g. imprecision), the test(s) to be performed and the desired result (e.g. statistical limits). As some test systems may require unusual samples or those covering a wide range of values, it may be advisable to start collecting and freezing samples weeks or even months in advance. The quantities of consumables (including reagents) required for the evaluation of the test system should be estimated with room for contingency plans in case additional work is required. The laboratory must document each stage of the evaluation process including preventive instrument maintenance, temperature evaluations, and data generated in the validation or verification stages. The laboratory manager must analyze the data and record the result of the analysis. It is advisable to have a dedicated and labelled digital record folder for the validation and evaluation of the system. Regardless, any form of test system evaluation should be readily available for any need to verify laboratory suitability.

Samples for Equipment/Reagent Validation Processes: The sample processing used in the validation process should be the same as for patient test samples. For the PT test, studies show that the stability of processed plasma is 24 hours when kept at room temperature, however, the faster it is processed, the better the quality assurance. Plasma samples for other tests must be tested within 4 hours of collection. If samples cannot be tested within acceptable stability limits, platelet-poor plasma (PPP) samples should be produced, aliquoted, and stored frozen at -70°C (Favaloro et al, 2008). Before analysis, frozen samples

must be thawed at 37°C (3 to 5 minutes for aliquots of up to 1 ml) and mixed immediately before testing (Kitchen et al, 2021).

Validation or Verification: In general, validation is a process that must be carried out on completely new test systems or a test developed in the laboratory. Verification, on the other hand, is a process that can be applied to already validated test systems that have recently been introduced to the diagnostics market. Verification can also be applied to evaluation after equipment has been relocated. If a test system has been cleared by the regional regulatory authority, only the verification process of the test system can be carried out locally. In this case, verification can be defined as providing objective evidence that a given test system locally meets the specifications set by the manufacturer. Deviation from the manufacturer's instructions for use for a test system will require follow up for system validation (Castellone, 2017).

Precision and Accuracy: After installing the equipment, it is important to assess the degree of intra-run imprecision of the tests to be evaluated. This practice is useful for identifying outliers and variability in the measurement system. The number of tests required for an evaluation will depend on whether the test system(s) require validation or verification (Gardiner et al, 2021a). For accuracy studies, the samples tested will depend on the type of patient samples expected to be tested in the lab. Accuracy is an important assessment, defined as the agreement between a measurement and the true value. Accuracy is usually carried out by comparing new instruments or reagent systems with an existing or predetermined method (Eusebi, 2013). The comparison between systems should be carried out using statistical analysis. Comparability between systems can be analyzed using linear regression (normal, weighted, Deming, or Passing-Bablok, as appropriate), Bland Altman trend graphs, and paired t-tests (or Mann Whitney U-tests if the data is not normally distributed) (Jensen and Kjelgaard-Hansen, 2006). The acceptability criteria will be specific to the test. For the APTT test, for example, two different test systems using different reagents could generate results with differences that could be clinically significant (Montalvão et al, 2020). Assessments of the sensitivity, specificity, and negative and positive predictive value of the test are also relevant and fundamental information for some tests. Calibrated assays for which there is an international standard (e.g. FVIII) and standardized methodology, should produce a regression line with a slope close to 0.90–1.10, with a strong correlation ($r > 0.95$) and no clinically significant bias (Gardiner et al, 2021b). The exact number of samples for the statistical evaluation of the verification/validation process will depend on the acceptability criteria for each test. Participating in an EQA program can be useful for establishing the accuracy of the system before putting it into a patient routine (Montalvão et al, 2022).

Establishing a Reference Interval: Establishing a normal reference interval is one of the most important tasks carried out in the laboratory, as most medical decisions are made based on laboratory results. Coagulation tests present a unique set of challenges. The reagents used in routine tests can have different sensitivities to coagulation based on the concentration and type of phospholipid and activators. A classic example of this is reagents used for the APTT test, which use particles (e.g. kaolin, cellite, silica) or chemicals (e.g. ellagic acid) that directly affect the sensitivity and specificity of the test. In this context, the different particles, as well as the phospholipid class and fatty acid composition, are not standardized. Therefore, reagents from different suppliers may have different compositions. All these properties of the reagents need to be considered when establishing a normal reference interval. Therefore, it is essential for the laboratory to carry out a local determination of the normal reference range so that the set of reagents, equipment, and procedures can be taken into account when assessing the patient. Health is not a well-defined condition and is often a relative term. The ideal group in some cases could be closely matched with the population under investigation with respect to age and sex. However, such careful selection is not essential for many coagulation tests. In practice, the selection of healthy normal subjects for establishment of a normal range will be influenced by practical considerations. Healthy hospital employees who do not receive any medication and healthy blood donors can be successfully used. There are important considerations in relation to normal ranges, which are given below. The condition of the normal subjects when blood is collected can influence the results obtained. This includes a review of the evidence for the effects of physical stress (e.g. up to 10-hour persistence of a 2.5-fold increase in FVIII/VWF), mental stress (e.g. increase in FVIII

and VWF after acute mental stress), hormone effects, circadian variations, and the effects of posture and diet. Some general recommendations were made, which were not restricted to investigation of female patients. These were as follows:

- ✓ Abstain from intense physical exercise for 24 hours prior to venipuncture.
- ✓ Use an environment where physical and mental stress are reduced.
- ✓ Abstain from fatty foods and smoking on the morning of venipuncture.
- ✓ Obtain samples early in the morning (7 a.m. to 9 a.m.), after the subject has been seated in a relaxed position for 20 to 30 minutes.

Validation Versus Verification of Reference Interval: The results of the reference interval are statistically evaluated, and the type of statistical evaluation is based on the number of individuals used. The process can include validation or full verification, where the reference interval is previously established. Validation requires a study of a minimum of 120 individuals, while verification of a reference range requires only 20 individuals to demonstrate that a test performs as previously established. The reference interval must be verified with any change of reagent, lot number, and instrument or collection system. The mean and standard deviation (SD) can be calculated (Gardiner et al, 2021a).

Statistical Analysis of Reference Interval: The SD is the dispersion of the data around the mean. The more dispersed the data, the greater the deviation. The confidence interval measures the level of uncertainty. If the 95% confidence level is chosen, the intervals will be estimated at the 2.5th and 97.5th percentiles of the distribution of results. This ensures that your confidence interval of values contains the true average of 95% of the population. Higher confidence levels will have wider reference intervals, while lower confidence intervals are narrower (Henny et al, 2016).

Different statistical methods can be used to evaluate data:

(1) Parametric method: is used when the population distribution is normal or Gaussian.

(2) Non-parametric method: does not require probability laws due to careful selection of subjects and a sufficient number of (≥120) individuals tested.

(3) Robust method: for use on a limited number of individuals without requiring the distribution to be Gaussian and measures position (location) and dispersion (spread) instead of the mean and standard deviation.

Position meters sort the data from smallest to largest in equal parts, while dispersion analyses the distance between the values of the distribution are from the center. To check for outliers, the data can be visually inspected and evaluated in a method proposed by Dixon (Henny et al, 2016). With a sample size of 20 (reference interval check), two outliers are allowed. If there are more than two outliers, an additional 20 samples should be tested. If there are another two outliers, other sources of error, such as reagents, analyzer problems, or biological variation, should be investigated (Henny et al, 2016). It may be necessary to carry out a full reference interval.

Internal Quality Control and External Quality Assessment: Quality assurance (QA) is an overall term that may be used to describe all measures taken to ensure the reliability of laboratory testing and reporting. This includes the choice of test, the collection of a valid sample from the patient, analysis of the specimen and the recording of results in a timely and accurate manner, through to interpretation of the results, where appropriate, and communication of these results to the referring clinicians. IQC and external quality assessment (EQA) (sometimes referred to as proficiency testing) are two distinct, yet complementary, components of a laboratory quality assurance program. IQC is used to establish whether a series of techniques and procedures are performing consistently over a period. It is therefore deployed to ensure day-to-day laboratory consistency. EQA is used to identify the degree of agreement between one laboratory's results and those obtained by other centers.

Internal quality control: IQC is used to establish whether a series of techniques and procedures perform consistently over time. The expression "quality control" is commonly used to describe the set of procedures used to check that the results of laboratory investigations are reliable enough to be released to assist clinical decision making, monitoring of therapy, and diagnosis of hemostatic abnormalities. Quality control procedures should be applied in a way that ensures immediate and constant control of results generation.

Within a laboratory setting, the quality of results obtained is influenced by many factors, including:

- Appropriate sample collection and handling
- Selection of suitable techniques and maintenance of an up-to-date manual of standard operational procedures
- Use of reliable reagents and reference materials
- Selection of suitable automation and adequate maintenance
- Adequate records
- Reporting system for results

In addition, the quality of results obtained in routine practice is highly dependent on the selection, training, and motivation of an appropriate complement of suitable personnel.

IQC is particularly useful to identify the degree of precision of a particular technique—precision being the degree of agreement among repeat measurements on one sample. To ensure that the results obtained in the laboratory are reliable, they must be precise in their analyses. IQC guarantees the daily consistency of an analytical process and thus helps determine whether patient results are reliable enough to be released. An IQC program should refer to institutions that support laboratory quality guidelines, such as International Organisation for Standardisation (ISO) or Clinical and Laboratory Standards Institute (CLSI). Laboratory tests must be able to identify physiological and pathological results regardless of when the test is carried out. In the ISO 15189:2022 recommendations, the laboratory shall have an IQC procedure for monitoring the ongoing validity of examination results, according to defined criteria, that verifies the attainment of the intended quality and ensures consistent validity pertinent to clinical decision making. IQC shall be performed at a frequency that is based on the stability and robustness of the examination method and the risk of harm to the patient from an erroneous result. In CLSI H47-A2 Vol 28 No 207.8, for all non-manual coagulation test systems, minimally, the laboratory must include at least two levels of control material for every 8 hours of operation and each time a reagent is changed.

Internal quality control materials: To assess the precision of a particular method, it is necessary to perform repeated analyses of aliquots of the same sample. It is important to include quality control (QC) samples with normal and abnormal values to ensure that a method is under control at different levels of a particular analyte, since relatively minor changes in an analytical process may be more apparent when testing an abnormal control. The control material should be similar in properties to test samples and be analyzed concurrently. Quality control materials of human origin are more likely to closely resemble human test samples. All vials or aliquots of the control material should be practically identical, so that any variation in test results is not a consequence of vial-to-vial variation. The QC material should also be stable for its intended period of use. With respect to hemostatic tests and assays, plasma samples must be deep frozen (preferably at -35°C or lower) or lyophilized in order ensure adequate stability for use as QC material. For reconstitution of lyophilized samples, it is important to use distilled water with pH 6.8–7.2 and to allow at least five minutes for reconstitution. If commercial QC material is used, this should be reconstituted according to manufacturer's instructions using an accurate pipetting system. If deep frozen QC material is used, this should be thawed rapidly at 37°C for five minutes. In the selection of QC material, the risk of transmission of blood-borne viruses should be considered. High-risk material should not be used. At least one QC material should be included with each group of screening tests or assays. For screening tests, it may be most appropriate to include a normal QC in this way and to test abnormal QC materials once per day or shift, or when doubt exists about whether a method is under control. A QC material with a reduced level should be included with tests used for the diagnosis and monitoring of congenital

deficiency states associated with bleeding. In all cases, the control material must be treated exactly like test samples, if possible

Acceptable limits of variation: For commercial IQC, samples manufacturers often provide a target range of acceptable values. In the case of screening tests and occasional assays, the results obtained will depend on the reagents and endpoint detection system used to perform the tests. The target range must take these effects into account. Where a target range is not available for a particular technique, this can be established locally. The IQC material is tested repeatedly (minimum 10 times) on different days when the method is known to be under control (as indicated, for example, by within target results on an alternative QC material). The mean and SD of these results are then calculated. The SD is the square root of the sum of d2 divided by n-1, where d is the difference of individual results from the mean and n is the number of determinations. The SD is a measure of the spread of results; the larger the SD, the greater the spread of results. Another important parameter is the coefficient of variation (CV), which is the SD expressed as a percentage of the mean (CV = SD divided by mean, multiplied by 100%). The CV of results on different days for PT and APTT of a QC sample should always be less than 8%, and preferably lower. For assays such as FVIII:C and FIX, CVs of less than 10% should be attainable for tests performed over several days. In most cases, results obtained for an IQC sample will show a normal (Gaussian) distribution. It is common practice to set the target range for IQC results as the mean +/- 2 SD, since this should include 95% of values. Individual results should be recorded on a chart that identifies the target range. Results outside this range indicate that the QC material has deteriorated or been handled incorrectly, or that the method is not properly controlled. Repeat testing of further QC material will then differentiate between these two possibilities, with further out-of-limit results confirming that the test system is out of control.

External quality assessment: In large EQA schemes, retrospective analysis of results obtained by participating laboratories permits the identification, not only of poor individual laboratory performance, but also of reagents and methods that produce unreliable or misleading results. The primary function of EQA is proficiency testing of individual laboratory testing. The World Federation of Hemophilia (WFH) International External Quality Assessment Scheme (IEQAS) includes analyses of particular relevance to the diagnosis and management of bleeding disorders (for further information, contact the WFH). Data from this scheme have been published in the following references:

- Jennings I, Kitchen S, Woods TAL, Preston FE. Development of a World Federation of Hemophilia External Quality Assessment Scheme: results of a pilot study. *Haemophilia* 1996; 2: 4–46.
- Jennings I, Kitchen S, Woods TAL, Preston FE. Laboratory performance of haemophilia centres in developing countries: 3 years' experience of the World Federation of Hemophilia External Quality Assessment Scheme. *Haemophilia* 1998; 4: 739-746.
- Jennings I, Kitchen DP, Woods TA, Kitchen S, Walker ID, Preston FE. Laboratory performance in the WFH EQA programme 2003-2008. *Haemophilia*. 2009; 15:571-7.
- Silmara Montalvão, Ian Jennings, Christopher Reilly-Stitt, Dianne Kitchen, Steve Kitchen. Quality of diagnosis and lab monitoring of people with hemophilia and other bleeding disorders across the continents: WFH IEQAS programme 2016-2023.

The WFH IEQAS was launched in 2004 to monitor and improve laboratory performance in hemophilia treatment centers (HTCs) worldwide. Laboratories can participate in this scheme to assess their quality assurance systems and the reliability of their test results. IEQAS improves and standardizes laboratory diagnosis by auditing the effectiveness of the internal quality assurance systems in place and providing a measure of the laboratory's competence. The United Kingdom's National External Quality Assessment Scheme (UK NEQAS) operates the scheme for Blood Coagulation, based in Sheffield, which has been inspected by the United Kingdom Accreditation Service Ltd (UKAS) and granted full accreditation to ISO 17043 for all listed tests. The mandate of the WFH IEQAS is to provide an EQA for tests of blood coagulation and to promote high standards of performance and practice. EQA, together with IQC procedures, are vital components of overall laboratory quality assurance. In addition, the WFH IEQAS provides an advisory service to participants through exchanges on lab diagnosis, including a participants' meeting during the

biennial WFH World Congress and onsite/virtual visits to provide training, as needed. The WFH IEQAS Committee is responsible for overseeing the IEQAS program. The committee is comprised of an independent chair appointed by the WFH, the Scheme program director, IEQAS program staff in the Sheffield Teaching Hospitals (host institution), and WFH staff and volunteers. The IEQAS Committee oversees all operational aspects of the program, reviews participation in the scheme, analyzes results, monitors global laboratory performance, and provides advisory support for centers registered in the scheme. The WFH IEQAS surveys are distributed three times per year, typically in March, July, and November. All surveys usually include PT, APTT, FVIII assay, and FIX assay. Two of the three surveys include VWF antigen assay and ristocetin cofactor/VWF activity assay. One out of the three surveys include two other factor assays so that FII, FV, FVII, FX, and FXI assays are all assessed at some stage, along with fibrinogen. For details, contact neqas@coageqa.org.uk.

References

Castellone DD. Establishing reference intervals in the coagulation laboratory. *Int J Lab Hematol* 2017; 39 Suppl 1: 121–127.

Eusebi P. Diagnostic accuracy measures. *Cerebrovasc Dis.* 2013; 36(4): 267–272.

Favaloro EJ, Lippi G, Adcock DM. Preanalytical and postanalytical variables: the leading causes of diagnostic error in hemostasis? *Semin Thromb Hemost* 2008; 34(7): 612–634.

Gardiner C, Coleman R, de Maat MPM, Dorgalaleh A, Echenagucia M, Gosselin RC, Ieko M, Kitchen S. 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* 2021a; 43(2): 169–183.

Gardiner C, Coleman R, de Maat MPM, Dorgalaleh A, Echenagucia M, Gosselin RC, Ieko M, Kitchen S. International Council for Standardization in Haematology (ICSH) laboratory guidance for the verification of haemostasis analyser-reagent test systems. Part 2: Specialist tests and calibrated assays. *Int J Lab Hematol* 2021b; 43(5): 907–916.

Henny J, Vassault A, Boursier G, Vukasovic I, Mesko Brguljan P, Lohmander M et al. Recommendation for the review of biological reference intervals in medical laboratories. *Clin Chem Lab Med* 2016; 54(12): 1893–1900.

Jensen AL, Kjelgaard-Hansen M. Method comparison in the clinical laboratory. *Vet Clin Pathol* 2006; 35(3): 276–286.

Kitchen S, Adcock DM, Dauer R, Kristoffersen AH, Lippi G, Mackie I, Marlar RA, Nair S. International Council for Standardization in Haematology (ICSH) recommendations for processing of blood samples for coagulation testing. *Int J Lab Hematol* 2021; 43(6): 1272–1283.

Montalvão SAL, Francisco AP, da Silva BLQ, Huber SC, Aguiari HJ, Fernandes MCGL et al. From hemophilia to deep venous thrombosis patient samples: How to perform an easy coagulometer validation process according to available guidelines. *Clin Appl Thromb Hemost* 2020; 26: 1076029620915512.

Montalvão SAL, Lowe A, Kitchen S. Advantages of external quality assessment-EQA programs. *Haemophilia* 2022; 28(4): 679–686.