INTRODUCTION

In the clinical laboratory, flow cytometry assays are critical to providing diagnostic and prognostic information to the treating clinicians. A validation or verification provides confidence that the assay will yield reliable results that can be trusted to make critical medical decisions.

Assays (or tests) may be either IVD (in vitro diagnostic) tests or LDT (laboratory developed tests). An IVD test is an assay that the manufacturer develops, optimizes, validates, and submits to a regulatory body (such as the FDA). The regulatory body approves the validation and clears it for clinical use. This allows manufacturers to sell the assay (kit) to laboratories for use in testing patient samples. The laboratory must follow the manufacturer’s instructions (standard operating procedure, SOP), and the laboratory must verify that they can reproduce the performance specifications in their lab. An LDT assay is an assay that an individual laboratory develops, optimizes, and validates using their equipment, reagents, and staff. There is currently no requirement for an outside regulatory body to approve LDTs in the United States (yet). The FDA allows for “enforced discretion” under the Clinical Laboratory Improvement (CLIA) standards. The laboratory will validate (or establish) performance specifications unique to this assay. Once validated, the laboratory can perform the assay on patient samples; however, it cannot be sold to another facility. This document will review specifications for LDT validations only [1]. A similar framework exists in the EU that regulates IVDR Compliance of LDT and is also evolving.

How the assay will be reported is critical in designing your validation plan and subsequent statistical analysis of the data. In general, assays are either quantitative, semi- (quasi-) quantitative, or qualitative. A quantitative assay uses a calibration curve and standard reference material to determine a result. An example in flow is a CD4 or CD34 enumeration assay. A semi-quantitative assay is an estimated value that cannot be directly assessed by a calibration curve. An example in flow is a minimal residual disease (MRD) or paroxysmal nocturnal hemoglobinuria (PNH) assay. A qualitative assay detects the presence or absence of a particular analyte. It is reported in a descriptive manner, rather than numeric. An example in flow cytometry is a leukemia/lymphoma immunophenotyping assay. Some assays may consist of multiple data types. An example is determining the percentage of an abnormal cell population within a leukemia/lymphoma panel. Each component must be validated appropriately.
Components of the Validation Process

It is important to define and document the experiments for your validation in a validation plan. This should consist of the method or procedure used in the validation, the reagents, equipment, and supplies used, the number and type of samples used in each validation experiment, the statistical analysis performed on each experiment as well as acceptance criteria for each parameter. The following performance specifications should be included in a validation for an LDT assay as needed: accuracy (or trueness), precision (reproducibility and repeatability), detection capability, selectivity, reference range, and sample and reagent stability. In general, the experiments are similar for quantitative, semi-quantitative and qualitative assays; however, the statistical data analysis may be different [2].

Accuracy is defined as the closeness of agreement between the average values obtained from a large series of test results when compared to an accepted reference standard. In most flow assays there is no reference standard, and these experiments should be done as a comparison to a predicate method or other lab. For quantitative or semi-quantitative assays, a correlation, Bland-Altman analysis, or t-test is used to compare results. For qualitative assays, a concordance calculation is used.

Precision is defined as the dispersion of replicate measurements using “conditions of measurement”. A condition of measurement is the variable (or lack of variable) in the experimental procedure. Types of precision are:

1. Repeatability (intra-assay precision): running the same sample multiple times in one run by the same analysts/operators
2. Intermediate (or inter-assay precision): running the same sample multiple times in different runs by the same analysts/operator
3. Inter-operator precision (reproducibility): running the same sample multiple times in different runs by the different analysts/operators.
4. Inter-instrument precision (reproducibility): running the same sample on different instruments.

Assessing precision in flow cytometry can be difficult due to sample availability, small volume, and limited stability. For this reason, a factorial design strategy can be used. Multiple factors are evaluated simultaneously to allow for a statistically robust n.

Detection capability (previously referred to as analytical sensitivity) is an umbrella term for a set of performance attributes that may be used to characterize measurement in the low-end region of the measurement interval. It can also be described as how well the assay performs with known positive samples. Here a few terms used with detection capability:
1. Limit of Blank (LOB) is defined as the highest signal obtained in the absence of the measurand. In flow cytometry, this would be the sample lacking the cells of interest (in other words a blank sample)

2. Lower Limit of Detection (LLoD) is defined as the ability to detect the measurand at a level that can reliably be distinguished from the LOB (or background). Ideally 95% of low-level samples will be detected above the LOB. LLoD is defined as presence or absence, not a quantitation.

3. Lower Limit of Quantitation (LLoQ) is defined as the lowest concentration that can be quantitated with acceptable precision.

The calculation for each is as follows \[2, 3\]:

For the Limit of Blank (LOB)
- Run 10 normal donors
- Calculate LOB = Mean of blank + 1.645 SD

For the Limit of Detection (LLoD)
- Use LOB dataset. LLoD = Mean of Blank + 3 SD
- Or
- In clinical assays it is recommended to verify this LLoD
  - Run 5 normal donors and 5 low positive samples, multiple runs.
  - Verify that
    - No more than 5% of blank replicates exceed the low positive target
    - No more than 5% of low positives fall below the target LOD

For the LLoQ
- Performed using serial dilutions run in triplicate to expected sensitivity.
- Verify mean of triplicates recovers expected value.
- Precision of triplicates <10% (if value is ≥1%), <30% (if value is near the LLoQ, generally <1%) \[3\]

Selectivity (previously known as analytical specificity) is defined as the ability of an assay to measure the intended measurand while avoiding cross-reactivity of non-specific interference. Or how well the assay runs with known negative samples. In flow cytometry, selectivity is addressed during assay development and optimization (before the validation). It includes selection of the antigens and antibody clones through reagent titration. Also, evaluation of wash steps and buffers and the gating strategy. During validation one must document the verification that the antibodies are marking as expected (both positive and negative populations) using the validation SOP.

Stability (both reagent and sample) is defined as the lack of variability (consistency or reproducibility) in the measured analyte to time. Sample stability ensures that time between sample draw, processing, and testing does not affect results \[4\]. The same applies to reagents: if reagents are prepared (such as antibody cocktails) and there is a time lapse gap between preparation and use, it must be verified that there is no variability in results. In both instances,
you would perform experiments to compare results with aged samples or reagents and compare with results of the freshly drawn sample or freshly prepared reagents [5].

Reference intervals (reference ranges) describe the limits of distribution of the measurand within a selected normal population. For an LDT assay, your lab will either need to:

- Perform a full reference interval evaluation (outlined in CLSI EP28 using N=120 subjects with different sex and age variables) [6], or
- Find a published reference range for the assay and VERIFY that your lab can reproduce that range with a 95% confidence interval N =20 [6] (N=25 for NYS)
- Some regulatory requirements will require that you re-establish or re-verify your reference range periodically.

Documentation of the Validation Process
The results of the above experiments should be documented in a validation report and compared with acceptance criteria. A statement that the assay has passed this criterion and is acceptable for testing clinical samples is also required. Appropriate medical ad laboratory director approval is required prior to launching the test into a clinical laboratory setting [7].

After validation/verification of the respective performance specifications, ensuring these specifications can be continually met over time is necessary. This is done by ongoing quality control. Quality control (QC) is defined as a set of procedures performed by the laboratory staff for the continuous and immediate monitoring of laboratory work in order to decide whether the results are reliable enough to be released. Assay specific QC should be designed to verify critical components of the assays intended use. Validation determines if you are running the right assay while Quality Control determines if you are running the assay right.

VALIDATION EXAMPLE OF A LEUKEMIA/LYMPHOMA ASSAY
Most malignant leukocyte immunophenotyping assays have a predominant qualitative component (presence or absence of an abnormal population) as well as a minor semi-quantitative component (e.g., % abnormal cells). Both types must be validated appropriately. It is of the utmost importance to review any regulatory requirements in the design of a validation plan. In some cases, a regional or other regulatory body may have different validation criteria. When creating the validation plan, the lab should apply the most stringent criteria when there are discrepancies. It is also important to evaluate and decide on sample preparation methods before embarking on validation studies 14,15.

Accuracy
Accuracy (Diagnostic concordance; Method comparison) is determined by comparing test results from the new method with results from a current similar in-house reference method or inter-laboratory comparison or known diagnosis from an alternative method (e.g., immunohistochemistry, morphology, molecular or cytogenetic findings). A convenient common method is to use split samples with a validated assay. It is required to test at least 20 samples and to include both normal and abnormal samples. For qualitative review, compare overall diagnosis
i.e., presence or absence of hematological malignancy, as well as descriptive immunophenotype (antigen distribution and fluorescence intensity) between current and new methods. The expression/distribution of similar markers used in the panels must be comparable (e.g. B-lymphoid cells stain with CD19 in in new vs reference method) to the expected biological correlate, and the intensity of marker expression on populations of interest must be comparable (e.g., dim CD20 expression on B-cells of chronic lymphocytic leukemia). Qualitative data can be reviewed in a concordance table to calculate the degree of concordance (agreement), and the new method should show >95% concordance with the reference method [2]. For semi-quantitative review, the size of the population of interest should be comparable. This can be assessed using correlation statistics, with a target of $r^2 > 0.90$ in most situations [8].

![Table](image)

Figure 1. Example of an abbreviated method comparison table assessing immunophenotype, final diagnosis, and size of the abnormal population. Diagnostic and immunophenotype concordance are 100%.
Figure 2. Example of graph assessing the correlation between abnormal population sizes using the old vs new method, showing $r^2$ of 0.99. Graph created in Excel.

**Precision**

*Intra-assay precision* (within-run precision; repeatability; instrument precision) is the closeness of agreement between results of measurements obtained under identical conditions and is measured with replicate measurements from the same run. Precision is assessed quantitatively using standard deviation (SD) and coefficient of variation (CV). One stained sample should be analyzed with 10 replicates in a single run, on each instrument being evaluated. The samples should have 100% categorical concordance for qualitative variables. For semi-quantitative variables, the mean, standard deviation, and %CV should be determined, with a target of CV<10%, though higher variability may be acceptable for infrequent populations as seen in MRD detection, fetomaternal hemorrhage (FMH), and PNH assays $^{[2, 8, 9]}$. 
Figure 3. Intra-assay precision example using an abnormal sample. Size of abnormal population and final diagnosis were recorded. Statistics were calculated using Excel.

Inter-assay precision (between-run precision; reproducibility) is the closeness of agreement between results of successive measurements obtained under changed conditions and is measured with replicate measurements from multiple runs. Three to five samples (combination of controls and fresh patient samples, to include at least 1 abnormal patient sample) can be analyzed in triplicate and assayed at time 0 and at maximum stability time as determined by stability studies. A desirable target for assay imprecision is a CV of less than 10% and 100% diagnostic concordance \[^2,8\]. Inter-operator precision (staff procedural precision) can be assessed by following the inter-assay precision protocol per each performing employee using the same sample \[^8\].
Figure 4. Inter-assay precision experiment example. Three samples run in triplicate on two different flow cytometers. Average CV<2.0% for semi-quantitative variable (size of abnormal population) and both instruments showed concordance of qualitative result (diagnosis).

An alternative to the separate traditional approaches above is to create a factorial design. This allows evaluation of multiple variables within a single experiment. For example, one may design a study that assesses different instruments and/or different operators simultaneously. Factorial designs, when planned appropriately, can save significant time and laboratory resources [2, 7, 8].

**Detection capability**

The *detection capability* (analytical sensitivity) in a qualitative assay is the ability to recognize a finding above background [8]. Leukemia and lymphoma immunophenotyping is a qualitative assay and data is reported in descriptive terms in nominal (e.g. yes/no) or ordinal (e.g. +, ++, ++++) formats [2] that is not designed to report stringent qualitative data such those reported in an MRD assay; however, it is recommended to verify the LOD for a given neoplastic population in this assay.

Verification of the LOD for a qualitative assay involves the recovery of the desired number of target (neoplastic) populations in a stained sample after the acquisition of the required number of total events. The recommended optimal LOD values for different cell lineages are B-cells, 0.1%; T-cells, 1%; Myelomonocytic cells, 0.5%; and Plasma cells, 0.1%[2]. The LOD can be verified by performing a spiking experiment of known abnormal samples using a sample containing an appropriate normal background population [2].
Figure 5. Example of LLoD study for B-cells in CLL. A peripheral blood sample was prepared such that the spiked sample contained 0.1% of the abnormal population. The acquisition protocol on the flow cytometer was set to collect a minimum of 100,000 total leukocytes (CD45+ cells) and a minimum of 100 target events to achieve a CV of <10%. The assay was run in triplicate with all 3 stained samples using the same settings. In this example, the mean of the B-cells was 0.1%, with an acceptable precision of the target population (CV of 7%).

Diagnostic sensitivity/Sensitivity concordance
The diagnostic sensitivity or sensitivity concordance, depending on the type of method used for concordance, should also be assessed. These are calculated similarly, but the terms differ depending on the type of comparative standard used in the study. Diagnostic sensitivity applies when comparing results to specimens that have a confirmed clinical diagnosis, while sensitivity concordance compares results to results of a comparative method. The percentage of samples correctly identified by the assay as containing a population of interest/abnormal population can be determined. Data from a 2x2 table from accuracy/method comparison can easily be used to calculate sensitivity TP/ (TP+FN), where TP=true positives and FN=false negatives. A target is >95% sensitivity [2].

Diagnostic specificity/Specificity concordance
Similarly, diagnostic specificity applies when comparing results to specimens that have a confirmed clinical diagnosis, while sensitivity concordance compares results to results of a comparative method. These can also be determined using the accuracy/method comparison data set. Calculate specificity as TN/ (TN+FP); where TN=true negatives and FP= false positives. A target is specificity >95% [2].
### Concordance Table and Calculations for Reference

<table>
<thead>
<tr>
<th>Candidate Method</th>
<th>Comparative Method</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>A</td>
<td>B</td>
<td>A+B</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>C</td>
<td>D</td>
<td>C+D</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>A+C</td>
<td>B+D</td>
<td>A+B+C+D</td>
<td></td>
</tr>
</tbody>
</table>

Specificity concordance = D/(B+D)*100%
Sensitivity concordance = A/(A+C)*100%

### Concordance Table for Presence or Absence of Abnormal Population

<table>
<thead>
<tr>
<th>Candidate Method</th>
<th>Comparative Method</th>
<th>Positive</th>
<th>Negative</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>56</td>
<td>0</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>0</td>
<td>55</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>55</td>
<td>111</td>
<td></td>
</tr>
</tbody>
</table>

Specificity = 100%
Sensitivity = 100%

Figure 6. Example of 2x2 comparison table from a study of 111 leukemia/lymphoma cases using the comparative method approach.

**Selectivity**

*Selectivity* must demonstrate that the antibodies and panels are used correctly and must specifically identify the targeted populations of interest. It should include a summary of how and why the final panels were chosen. Thoughtful panel design is critical and must include careful consideration of desired target populations, normal and abnormal immunophenotypes of cell populations, fluorochrome properties, clone choice, combinations of markers, and possible staining intensities. This process must include consultation with a variety of resources, including published peer-reviewed literature sources, specialized society recommendations, laboratories with well-established panels at other institutions, experts in the field, and vendor suggestions [10].

Once panels have been constructed, pre-analytic sample preparation steps must be refined, including lyse and wash steps, buffers, and antibody titration [5]. Panel templates must be designed to accurately and completely display the relevant information about all cell populations of interest. Gating strategies must then be determined for each panel, highlighting target populations without selectively excluding other populations that may be of interest.
In addition, antibody specificity must be verified for each antibody to ensure that each marks target populations as expected. This can be accomplished by selecting validation cases that show expected positive and negative cell populations.

Figure 8. Example of antibody specificity verification with CD19.
Specimen Stability

Specimen stability testing is performed on a range of sample types and disease states, with a minimum of 3-5 samples, stained at multiple time points. Stability is established for qualitative testing as 100% categorical concordance across time points. Changes in the fluorescent intensity of antigen expression levels should be <0.5 log decade difference from the baseline. For quantitative testing, stability is established as the latest time point where the clonal population shows <20% change from baseline in 80% of cases \(^2\). In such instances where the population is rare, higher variation may be acceptable.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Baseline</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
<th>Time 5</th>
<th>Time 6</th>
<th>Concordance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Abnormal</td>
<td>Abnormal</td>
<td>Abnormal</td>
<td>Abnormal</td>
<td>Abnormal</td>
<td>Abnormal</td>
<td>Abnormal</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>Myeloma cK</td>
<td>Myeloma cK</td>
<td>Myeloma cK</td>
<td>Myeloma cK</td>
<td>Myeloma cK</td>
<td>Myeloma cK</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>AML</td>
<td>AML</td>
<td>AML</td>
<td>AML</td>
<td>AML</td>
<td>AML</td>
<td>AML</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>ALL (T)</td>
<td>ALL (T)</td>
<td>ALL (T)</td>
<td>ALL (T)</td>
<td>ALL (T)</td>
<td>ALL (T)</td>
<td>ALL (T)</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 9. Example of qualitative component of a specimen stability study \(^2\).

Cocktail Stability

Any antibody cocktails need to be assessed for stability. This is accomplished by comparing results of a cocktail prepared on Day 0 with results of a freshly prepared cocktail or individual pipetted antibodies and repeated at regular time intervals \(^5\). At each interval, compare the signal to noise (S/N) ratio or the MFI of each marker as well as the percentage of the positive markers between the two tubes. Cocktails should be evaluated at least every week for the duration of the cocktail stability study. Ideally, both non-diseased and diseased samples (if applicable) should be used for the stability study. Ensure that tandem dyes remain stable in the cocktail (for example, if there is a breakdown in the tandem dye (e.g., PE-CY7), the signal from the tandem might appear in the parent channel (PE)). An ideal target for acceptability is that the MFI of each positive marker should be within 0.5-1.0 log difference or signal to noise ratio should be within 10-15% difference, and the percentage of the abnormal population should be within 10-15% \(^5\).
Figure 10. Example of cocktail stability assessment comparing cocktail stability to individually pipetted antibodies.

Other validation components, such as reference ranges and linearity/analytical measurement range (AMR), are not applicable for qualitative methods.

**Ongoing QC and monitors**

Ongoing QC monitors ensure that the assay performs as expected throughout the life of the assay. This should include but is not limited to:

- Instrument QC
- Reagent and cocktail QC must meet predetermined acceptability criteria (as determined by the laboratory)
- Verification that internal controls are marking as expected for each case
- Verification of positive controls at least monthly, with rare antigens at least semiannually
- Verification of compensation setting on each day of testing
- Competency assessment of analyst

**VALIDATION EXAMPLE OF A PNH ASSAY**

Since the PNH assay does share both a qualitative component (presence or absence of a PNH clone) as well as semi-quantitative component (size of the PNH clone), both components need to be validated. If a laboratory decided to either start a PNH assay or modify a PNH assay, they will need to perform a full validation of this assay. This should include the following:

- Optimize instrument settings for both assays (previously qualified) [12]
- Panel design and reagent/fluorochrome selection [12]
- Titration of the relevant antibodies [12,14]
- Developing an acquisition/analysis template including gating plots, diagnostic plots and control plots. [13]
- Determination of accuracy (agreement) [9]
- Determination of precision (reproducibility) [9]
- Determination of selectivity (analytical specificity): Qualitative and semi-quantitative [9]
- Determination of detection capability (analytical sensitivity) [9]
- Stability (sample stability and reagent/cocktail stability) [9]
Our laboratory already had a validated antibody panel for both the RBC and the WBC assay based on current PNH Guidelines [see 2018 ICCS/ESCCA PNH Guidelines for PNH Testing] [12, 13]:

- **RBC panel**: CD235a-FITC/CD59-PE

The 2 color RBC panel and the 6 color WBC panel had been previously validated on a 10 color platform (BC Navios) and since the new 10 color Navios EX did not require modification of the assay, the panel and reagents did not need a validation. Their performance just needed to be verified.

Specimen preparation followed the Stain first/then Lyse procedure for the WBC assay and a 1:100 dilution was used for the RBC assay. Please refer to the 2018 ICCS/ESCCA PNH Guidelines for PNH Testing [12, 13] and the ICCS Q&S Module #1 for more detail. [15]

![Example of a 6C PNH WBC panel with gating and diagnostic plots (left) and control plots using lymphocytes (right)](image)

Figure 11. These 2-color (2C) RBC and 6-colour (6C) WBC panels were previously performed and validated on a 10-color (10C) platform (BC Navios). Since the new 10C Navios EX did not require modification of the assay, the panel and reagents did not need a validation, so that their performance just needed to be verified.
All of the validation requirements had already been performed but the following still needed to be verified.

**Accuracy (Trueness, Agreement)**

20+ normal (PNH negative) and abnormal (PNH+) cases were run on both the reference instrument and the new instrument

a. **Qualitative component**: The goal was 100% concordance for all PNH testing cases

b. **Semi-quantitative component**: PNH clone sizes needed to match with $R^2$ value of greater than 0.95

**WBC (Monocytes and Neutrophils)**
RBC (PNH Type III, Type II and Total PNH Clone)

Figure 13. Regression analysis on the scatter plots shows PNH monocytes, granulocytes, RBC type II, type III and total RBC % shows excellent correlation with $R^2$ value > 0.99

Precision
Acceptability criteria: results within 2 SD
a. Repeatability (intra assay precision): running the same sample multiple times in one run by the same analysts/operators

RBC type III, II and total PNH clone

<table>
<thead>
<tr>
<th>Precision - run one case 5x</th>
<th>% PNH Type III</th>
<th>% PNH Type II</th>
<th>% Total RBC PNH clone</th>
</tr>
</thead>
<tbody>
<tr>
<td>R19-2830 (#1) 12/9/2019 MB/AD</td>
<td>55.6</td>
<td>14.4</td>
<td>70</td>
</tr>
<tr>
<td>#2 12/9/2019 MB/AD</td>
<td>55.5</td>
<td>14.4</td>
<td>70</td>
</tr>
<tr>
<td>#3 12/9/2019 MB/AD</td>
<td>55.5</td>
<td>14.4</td>
<td>70</td>
</tr>
<tr>
<td>#4 12/9/2019 MB/AD</td>
<td>55.5</td>
<td>14.4</td>
<td>70</td>
</tr>
<tr>
<td>#5 12/9/2019 MB/AD</td>
<td>55.5</td>
<td>14.4</td>
<td>69.9</td>
</tr>
</tbody>
</table>

| Mean | 55.58 | 14.4 | 69.98 |
| SD   | 0.044721 | 0 | 0.045 |
| CV   | 0.080 | 0 | 0.064 |

Figure 14. Intermediate precision on type III, type II and total PNH RBCs show results are within 2SD and CV of <1%
WBC (Neutrophils and Monocyte clone)

<table>
<thead>
<tr>
<th></th>
<th>PNH Monos</th>
<th>PNH Neuts</th>
</tr>
</thead>
<tbody>
<tr>
<td>F19-2411 run 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F19-2411 run 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F19-2411 run 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F19-2411 run 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F19-2411 run 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>99.46</td>
<td>98.88</td>
</tr>
<tr>
<td>SD</td>
<td>0.152</td>
<td>0.045</td>
</tr>
<tr>
<td>CV</td>
<td>0.152</td>
<td>0.045</td>
</tr>
</tbody>
</table>

Figure 15. Intermediate precision on PNH Monocytes and Neutrophils shows results are within 2SD and CV of <1%

b. Intermediate (or inter assay precision): running the same sample multiple times in different runs by the same analysts/operator: Performed but not shown

c. Inter-operator precision (reproducibility): running the same sample multiple times in different runs by the different analysts/operators: Performed but not shown

d. Inter-instrument precision (reproducibility): running the same sample on different instruments. Performed but not shown

Selectivity
This describes how well the assay performs with known negative specimens (how many false positives do we have?). In this experiment we used 5 PNH specimens with a confirmed PNH diagnosis and 15 specimens from normal donors and compared the results to another laboratory [9].

<table>
<thead>
<tr>
<th>Candidate Method</th>
<th>True Diagnosis</th>
<th>Total</th>
<th>Specificity = (TN+TN+FP) x100</th>
<th>Estimated Clinical Sensitivity = [A+(A+C)]x100</th>
<th>Estimated Clinical Specificity = [D+(B+D)]x100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>5 (A=true positive)</td>
<td>0 (B=False Positive)</td>
<td>5 (A+B)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>0 (C=False Negative)</td>
<td>15 (D=True Negative)</td>
<td>15 (C+D)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5 (A+C)</td>
<td>15 (B+D)</td>
<td>20</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Figure 16. Example of estimating Sensitivity and Specificity using 2 x 2 contingency table

Detection Capability (Analytical Sensitivity)
This describes how well the assay performs with known positive specimens (how many false negatives do we have). It is the ability of an assay to distinguish signal from the background and precisely measure low amounts of the measurand [9, 3]. This includes:

a. Analytical Sensitivity:
   - Limit of Blank (LOB): The highest signal expected in the absence of the measurand (PNH cells); LOB = Mean + 1.645 SD
• Limit of Detection (LOD): the ability to detect the measurand (PNH cells) at a level which can be distinguished from the LOB (95%). LOD = Mean + 3SD

<table>
<thead>
<tr>
<th>Acc#</th>
<th>Date</th>
<th>Tech</th>
<th>#RBCs</th>
<th>#PNH RBC</th>
<th>#PNH cells/million</th>
<th>%PNH RBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>F19-2311</td>
<td>12/6/2019</td>
<td>AD</td>
<td>721,340</td>
<td>2</td>
<td>3</td>
<td>0.00%</td>
</tr>
<tr>
<td>F19-2336</td>
<td>12/10/2019</td>
<td>AD</td>
<td>1,101,027</td>
<td>7</td>
<td>6</td>
<td>0.00%</td>
</tr>
<tr>
<td>F19-2362</td>
<td>12/12/2019</td>
<td>AD</td>
<td>760,985</td>
<td>1</td>
<td>1</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table: Example of LOB and LLoD evaluation

b. Functional Sensitivity: Lower Limit of Quantification (LLOQ): The lowest concentration that can be reliably detected with acceptable accuracy and precision. The determination for this can be done through a spiking experiment. Blood from a normal donor and a PNH+ sample with a clone size of at least 50% in the WBCs and 20% clearly delineated Type III RBCs should be used. The sample should be processed within 24 hours (48 hours is also acceptable if the sample was kept at 4°C). For the RBC part, it is best to use a PNH sample that has no or very few PNH Type II phenotypes as the separation between PNH type II and normal RBCs may be difficult to standardize (see example below). Follow ICCS Quality and Standards Module - "Verification of PNH assay sensitivity through spiking experiment" which includes an Excel spreadsheet for statistical assessment of LOB, LOD and LLOQ that each laboratory can use for the specific validation [3].
Figure 18. Schematic illustration of spiking experiment showing neat (PNH blood), dilutions and control (normal blood).

Table 1. RBC spiking experiment 6-16-2020

<table>
<thead>
<tr>
<th>Dilution</th>
<th>Run 1</th>
<th>Run2</th>
<th>Run3</th>
<th>Avg</th>
<th>SD</th>
<th>%CV</th>
<th>%CV*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neat</td>
<td>98,928</td>
<td>99,37</td>
<td>98,306</td>
<td>99,029</td>
<td>99,48</td>
<td>99,43</td>
<td>98,682</td>
</tr>
<tr>
<td>1:10</td>
<td>108,004</td>
<td>6.33</td>
<td>6,834</td>
<td>95,136</td>
<td>6.30</td>
<td>5,996</td>
<td>95,408</td>
</tr>
<tr>
<td>1:100</td>
<td>170,711</td>
<td>0.69</td>
<td>1,173</td>
<td>103,819</td>
<td>0.63</td>
<td>652</td>
<td>217,388</td>
</tr>
<tr>
<td>1:1,000</td>
<td>204,752</td>
<td>0.07</td>
<td>220</td>
<td>204,615</td>
<td>0.07</td>
<td>138</td>
<td>189,439</td>
</tr>
<tr>
<td>1:10,000</td>
<td>963,901</td>
<td>0.007</td>
<td>75</td>
<td>1,159,725</td>
<td>0.007</td>
<td>104</td>
<td>771,769</td>
</tr>
<tr>
<td>1:100,000</td>
<td>6,479,788</td>
<td>0.0017</td>
<td>24</td>
<td>1,085,137</td>
<td>0.0005</td>
<td>24</td>
<td>1,020,399</td>
</tr>
<tr>
<td>Normal</td>
<td>1,613,568</td>
<td>0.0004</td>
<td>71</td>
<td>1,089,489</td>
<td>0.0004</td>
<td>71</td>
<td>985,577</td>
</tr>
</tbody>
</table>

*Acceptable precision between the 3 replicates should be; if PNH clone is > 0.1%, a %CV should be <90% if PNH clone is <0.1%, a %CV should be <100%. LLOQ below.

Table 2. PNH Neutrophils 6-25-2020

<table>
<thead>
<tr>
<th>Dilution</th>
<th>Run 1</th>
<th>Run2</th>
<th>Run3</th>
<th>Avg</th>
<th>SD</th>
<th>%CV</th>
<th>%CV*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neat</td>
<td>45,255</td>
<td>67.8</td>
<td>30,702</td>
<td>47,225</td>
<td>67.8</td>
<td>32,009</td>
<td>46,272</td>
</tr>
<tr>
<td>1:10</td>
<td>46,734</td>
<td>6.3</td>
<td>2,942</td>
<td>48,612</td>
<td>6.3</td>
<td>3,048</td>
<td>52,430</td>
</tr>
<tr>
<td>1:100</td>
<td>65,692</td>
<td>0.7</td>
<td>447</td>
<td>64,829</td>
<td>0.7</td>
<td>465</td>
<td>145,951</td>
</tr>
<tr>
<td>1:1,000</td>
<td>166,698</td>
<td>0.1</td>
<td>199</td>
<td>166,539</td>
<td>0.1</td>
<td>120</td>
<td>103,123</td>
</tr>
<tr>
<td>1:10,000</td>
<td>355,239</td>
<td>0.06</td>
<td>34</td>
<td>519,660</td>
<td>0.06</td>
<td>41</td>
<td>484,370</td>
</tr>
<tr>
<td>Normal</td>
<td>569,886</td>
<td>0.07</td>
<td>7</td>
<td>603,014</td>
<td>0.07</td>
<td>5</td>
<td>537,398</td>
</tr>
</tbody>
</table>

*Acceptable precision between the 3 replicates should be; if PNH clone is > 0.1%, a %CV should be <90% if PNH clone is <0.1%, a %CV should be <100%. LLOQ below.

Table 3. PNH Monocytes 6-25-2020

<table>
<thead>
<tr>
<th>Dilution</th>
<th>Run 1</th>
<th>Run2</th>
<th>Run3</th>
<th>Avg</th>
<th>SD</th>
<th>%CV</th>
<th>%CV*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neat</td>
<td>4,059</td>
<td>62.8</td>
<td>2,549</td>
<td>4,255</td>
<td>63.1</td>
<td>2,685</td>
<td>4,333</td>
</tr>
<tr>
<td>1:10</td>
<td>2,918</td>
<td>9.8</td>
<td>286</td>
<td>3,087</td>
<td>9.3</td>
<td>287</td>
<td>3,365</td>
</tr>
<tr>
<td>1:100</td>
<td>3,961</td>
<td>1.1</td>
<td>44</td>
<td>4,068</td>
<td>1.1</td>
<td>44</td>
<td>8,833</td>
</tr>
<tr>
<td>1:1,000</td>
<td>9,906</td>
<td>0.1</td>
<td>14</td>
<td>10,050</td>
<td>0.1</td>
<td>19</td>
<td>6,190</td>
</tr>
<tr>
<td>1:10,000</td>
<td>22,105</td>
<td>0.06</td>
<td>2</td>
<td>32,026</td>
<td>0.06</td>
<td>4</td>
<td>25,317</td>
</tr>
<tr>
<td>Normal</td>
<td>35,735</td>
<td>0.06</td>
<td>0</td>
<td>38,138</td>
<td>0.06</td>
<td>2</td>
<td>33,163</td>
</tr>
</tbody>
</table>

*Acceptable precision between the 3 replicates should be; if PNH clone is > 0.1%, a %CV should be <90% if PNH clone is <0.1%, a %CV should be <100%. LLOQ below.

Figure 19: Table 1 (RBC), Table 2 (Neutrophils) and Table 3 (Monocytes) show the results of the spiking experiment and suggested LLOQ for each lineage. For the LLOQ determination, typically at least 50PNH cells should be acquired for reliable quantification. Based on the results above, the LLOQ for the RBCs is 0.01%, the LLOQ for the neutrophils between 0.01% and 0.02% depending on the number of neutrophils acquired and the LLOQ for the monocytes is variable (0.1%-1%) depending on the number of monocytes available.
Stability
- All reagents, antibodies and cocktails were tested for stability.
- Sample stability was also tested over the course of 7 days for RBCs and 4 days for WBCs (refrigerated vs room temperature)

Ongoing QC and monitors
Implementing proper QC monitoring is essential for ensuring that the assay performs as expected. This should include but is not limited to the points below.

1. Instrument QC
2. Reagent and cocktail QC must meet predetermined acceptability criteria
3. Monthly assay control to verify assay performance. The suggested assay acceptability criteria are as follows:
   a. A normal donor should show 0.00% PNH Neutrophils, 0.00% PNH Monocytes and 0.00% Type III PNH cells
   b. The CD64++ Monocytes should fall within the 3rd decade of the plot
   c. The CD15++ Neutrophils should fall within the 3rd decade
   d. The CD45+ WBC should fall within 3rd decade
   e. The MFI should be +/- 2 in 2nd decade, +/- 20 in 3rd decade, +/- 100 in 4th decade
   f. PNH-positive samples should show similar clone sizes in all flow cytometers used
4. Verification of sensitivity (suggested yearly)
5. Verification of LOB (suggested yearly)
6. Evaluation of carryover for rare event assays (suggested yearly)
7. Competency assessment of analyst

SUMMARY
There is to date no "one fits all" way to validate/verify various flow cytometry assays. The most recently published CLSI guideline H62[2] is a welcome and much-needed addition to the repertoire and has added significant value in addressing the crucial components and steps to consider. It is important to realize that although this document provides a comprehensive overview, each flow cytometry assay has a different purpose (research vs clinical, semiquantitative vs qualitative, etc.) and types of results. All of this should be taken into consideration when designing and performing an assay-specific validation or verification, or related assay-specific QC monitors, to ensure that all pre-analytic, analytic and post-analytic components are adequately assessed and monitored throughout the life of the assay. We hope that by providing concrete real-life examples of studies performed for each necessary component, this document will bring value to the readership and clarity to the often-overwhelming task of validation in flow cytometry.
References:


11. CAP Accreditation Program 09.22.2021, Flow cytometry checklist; FLO 23737.


Reviewed and approved by:
- Silvia Bunting, Director, Hematopathology and hematology, Cleveland Clinic, Weston, Florida at Cleveland Clinic
- Nina Rolf, BC Children’s Hospital Research Institute, University of British Columbia, Vancouver, Canada
- George Deeb, hematopathologist, Emory University Hospital in Atlanta, Georgia.
For any questions on this module or any other suggestions, please email info@cytometry.org

The documents posted on ICCS website may contain product or vendor names which are provided for platform specific guidance. Any reference within the ICCS Quality and Standards modules to any vendor, product or educational material by trade name, trademark or manufacturer does not constitute or imply the endorsement or recommendation by ICCS.