Scientific Working Group on DNA Analysis Methods

Validation Guidelines for the Use of an Expert System with Forensic Samples

Table of Contents

SWGDAM Validation Guidelines for the Use of an Expert System with Forensic Samples

The Scientific Working Group on DNA Analysis Methods, better known by its acronym of SWGDAM, is a group of scientists representing Federal, State, and Local forensic DNA laboratories in the United States and Canada. During meetings, which are held twice a year, committees discuss topics of interest to the forensic DNA community and develop documents to provide direction and guidance. These guidelines, drafted by the SWGDAM Casework Expert System Ad Hoc Working Group, were presented to the SWGDAM membership, and approved on XXXX.

This document provides guidelines for the validation of an Expert System for use with forensic samples. In the event of a conflict between the FBI's *Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS)* or the National DNA Index System (NDIS) Operational

Procedures and these guidelines, the FBI's *Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS)* and/or NDIS Operational Procedures have precedence over these guidelines. Absent any other directive, the use of the term shall or must is not intended to transform these guidelines into standards.

An Expert System is a software program or set of software programs designed to interpret single source DNA data in accordance with laboratory defined quality assurance rules and identify DNA data not satisfying laboratory defined quality assurance rules, without human intervention. All other samples continue to require analyst interpretation and review. This document gives guidance for laboratories to use already available Expert Systems to analyze single-source forensic samples. Expert Systems are not intended to replace manual evaluation of mixed DNA samples or manual review of CODIS eligibility. Additional research is necessary to expand the scope beyond single-source forensic samples. The validation and use of Expert Systems for reference samples is not applicable to these recommendations, nor are these guidelines intended for use with a Rapid DNA System.

1. Introduction

A validated Expert System may be used to complete the data review of single-source forensic samples with complete data present at all tested loci. If a validated Expert System is used and a sample is determined to be acceptable based on the validated parameters, manual review of the sample by an analyst and/or technical reviewer is not required. Use of an Expert System shall be approved by NDIS prior to uploading eligible samples to CODIS as described in the NDIS Operational Procedures Manual. NDIS approval is not required for Expert System review of forensic samples that will not be uploaded to CODIS

Laboratories should be aware of the limitations of Expert Systems. Expert Systems must be implemented in accordance with laboratory defined quality assurance rules and be able to accurately identify data that does and does not satisfy such rules. Appropriate validation studies and ongoing quality control testing shall occur.

2. Validation Criteria

Use of an Expert System for review of single source casework samples shall be developmentally validated as defined in the FBI's *Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS)*. Profiles reviewed by an Expert System that will be entered into CODIS shall also be developmentally validated in accordance with applicable NDIS Operational Procedures. For the purpose of CODIS entry, if attempting to validate and Expert System not currently approved by NDIS, contact the FBI's CODIS Unit.

With the exception of legally protected information, underlying scientific principle(s) utilized by software which impact the analytical process or interpretation shall be publicly available for review or published in a peer-reviewed scientific journal.

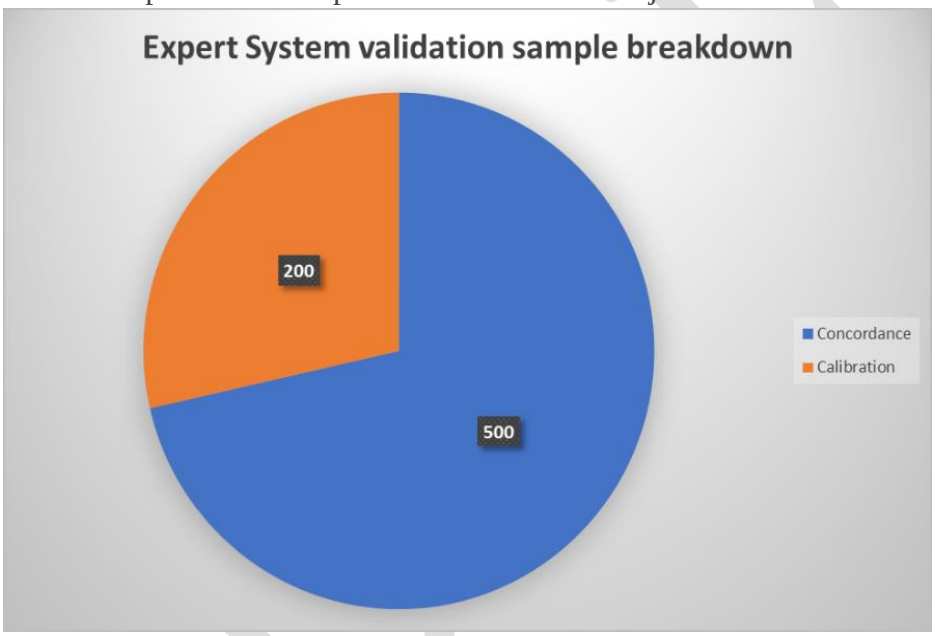

Figure 1: Expert System validations are comprised of two major components: calibration set and concordance set, each requiring a minimum number of sample evaluations.

The laboratory shall perform and complete the appropriate components of a validation in accordance with the FBI's *Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS)* and applicable NDIS Operational Procedures. All Expert System settings used during the internal validation shall be documented in the validation summary. The remainder of this document discusses the requirements for internal validation.

At least 200 unique samples shall be analyzed to establish the rules and thresholds for the software. This set of 200 samples is referred to as the calibration set. A "sample" is defined as a profile resulting from the analysis of one DNA specimen where the DNA profile is known. The

calibration portion of the validation establishes rules and thresholds while configuring the Expert System software to detect quality issues typically seen in the laboratory's data. At a minimum the calibration dataset shall contain all the challenges listed in Table 1. The calibration data set is a targeted and focused number of samples with known anomalies to stress test the system configurations. To effectively test software setting configurations, a dataset containing samples that will intentionally trigger quality flags shall be created. Adding problematic, as well as high quality, data allows the laboratory to ensure the Expert System responds appropriately to the spectrum of profile quality produced by the laboratory. The quality issues in the dataset shall be recorded prior to testing the Expert System and the performance of the Expert System shall be measured against the known issues for each sample. The more challenges the Expert System is introduced to during calibration, and the more closely the dataset mimics data produced at the laboratory, the more capable the Expert System will be in its evaluations. This process also allows the laboratory to understand the limitations of the software and what situations require human review. The Expert System shall detect quality issues with either the same or more stringent requirements used by the current system in the laboratory. For example, if the current system requires two peaks within one locus to have a peak height ratio of 60%, the Expert System must detect peaks that fall outside of the established percentage. The set should also include high quality samples to measure review efficiency. If high quality data is often flagged as needing human review, the system may be over-calibrated. The review of the calibration set should answer the question "can the system identify and alert the user to all known and commonly observed quality issues?"

A concordance study shall be performed to demonstrate that the system performs as well as, or better than, the current system used by the laboratory. The concordance study shall consist of a minimum of 500 unique samples. The concordance data set is a more generalized data set comprised of data produced and reviewed over time in the laboratory. Samples included in this concordance test set should be representative of samples analyzed in the laboratory. Evaluation of non-concordant data is conducted to determine if the Expert System performs as well as the currently validated allele calling procedure in the laboratory. The 200 unique samples from the calibration set shall not be included as part of the concordance study. To assess the concordance of allele designations between the Expert System and the validated system currently in place, the

samples in this dataset require a documented review under the currently validated process. These results will be the standard by which the Expert System is compared. Through the course of the concordance review, the laboratory may discover the Expert System performs more consistently and accurately than the process currently in place. This should be detailed in the validation summary. The review of the concordance set should answer the question "does the system produce the same (or better/more consistent) review conclusions as the current review process?"

Using data from the concordance and/or calibration study, the laboratory should demonstrate that the Expert System does not incorrectly accept alleles or profiles that should require human review. Expert Systems use a system of flags and/or scores to measure specific quality metrics, depending on the software program(s) being used. Within this document, the term "flag" is used to describe the signaling of the Expert System software that a locus or profile has not met the required quality metric. Flags and/or scores may be used in combination to meet any specific requirements depending on the software being used. A properly calibrated Expert System will sometimes flag samples that will pass review by an analyst, with or without edits. The laboratory shall have policies and procedures in place for the interpretation of samples that are flagged by the Expert System. The policies and procedures shall cover the manual review or reanalysis of the samples as appropriate.

The FBI's *Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS)* requires that new software or new modules of existing software that are used as a component of analysis and/or interpretation of DNA data shall be subject to internal validation specific to the laboratory's intended use prior to implementation in analysis. The impact of the Expert System on DNA data interpretation and technical review should be considered when designing the appropriate validation studies. The internal validation of an Expert System shall be specific for each of the following: Expert System software, instrument and associated data collection software, and DNA typing kit.

The laboratory shall evaluate the sample number and type as outlined below to demonstrate the potential limitations and reliability of the Expert System. Internal validation should establish the limits of the Expert System. The internal validation shall be conducted in accordance with applicable sections of the FBI's *Quality Assurance Standards for Forensic Testing Laboratories*.

3. Settings or Parameters to Define and/or Evaluate during Validation

The Expert System validation is twofold. First, software parameters are set by the user. These settings "pass" data falling within the parameters and force quality flags to fire when data outside of the parameters are encountered by the Expert System. Second, specific data quality issues (e.g., pull up, mixtures) are passed through the Expert System in two phases utilizing the calibration and the concordance set in order to challenge the parameters and ensure that data quality issues requiring a human review are correctly flagged. Parameter settings and challenge testing that are not applicable to the Expert System software being validated should be addressed in the validation summary.

If no data exists to challenge the parameter, the quality flags or parameters may be adjusted to cause the quality flag to fire as needed. Laboratories may not have profiles which exhibit all the challenges listed below. To ensure the Expert System software functions properly, laboratories may elect to temporarily alter the settings to a degree which is less or more stringent than the value intended for implementation. This will cause quality data to "trigger" the software flags and consequently demonstrate its capabilities to detect issues. Temporarily adjusting the stringency of the settings can ensure a more robust testing of the software occurs, as well as provide users with comprehensive experience regarding how the system communicates quality related alerts. If the parameters are adjusted for testing purposes, it should be conducted for targeted parameters or challenge testing (e.g., stutter) and must be addressed in the validation summary.

4. Software Parameters

4.1 Allele number or Ploidy settings define the number of allowable peaks at a locus. A quality flag will indicate when more than the maximum number is

encountered or when a locus at which peaks are expected has no peaks above the detection threshold.

4.1.1 The Expert System shall be configured to permit only single-source samples to pass without human intervention; therefore, the maximum allele number for autosomal loci shall be set to two (diploid). The maximum allele number for Y-chromosome loci shall be set to one; additionally, an imbalance flag may be set to fire for this locus if more than one allelic peak is present. The allele number setting is a required setting and is not laboratory dependent.

4.1.2 Under the above settings, this parameter shall indicate all mixed samples and samples having total allelic dropout at a locus. In addition, this flag will catch tri-allelic genotypes at autosomal loci, and duplicated genotypes at Y-chromosome loci. This flag will also detect incidents of elevated stutter, spectral pull up, and other amplification or electrophoresis-related artifacts.

4.1.3 This parameter shall be challenged as follows:

4.1.3.1 Five mixed samples to include at least one mixture demonstrating more than one allele at a Y-STR locus.

4.1.3.2 Five single source samples having amplification- or electrophoresis-related artifacts.

4.1.3.3 Five samples with complete locus dropout.

4.1.4 If available in the laboratory, the parameter shall also be challenged by one sample having a tri-allelic locus.

4.2 **The detection threshold** is the minimum height at which the Expert System will label peaks. Depending on the software used, this may be the same as or different than the **analytical threshold** which is the minimum height requirement, determined through validation testing, at or above which detected peaks/signal can be reliably distinguished from background noise. Peaks/signal at or above

this threshold are generally not considered noise and are either artifacts or true alleles.

4.2.1 The detection threshold may be set to the same height as the analytical threshold. Laboratories may choose to implement a detection threshold lower than their analytical threshold to assess sub-analytical threshold data.

4.2.2 The analytical threshold will be the same threshold determined during validation of the DNA typing kit. It should not be raised to avoid the detection of mixtures and/or artifacts.

4.2.3 If no quality flag for detection threshold exists in the Expert System, challenges and quality flags will be covered with the settings of the homozygous minimum peak height, heterozygous minimum peak height, and maximum expected alleles.

4.2.4 The laboratory's validation of the Expert System shall illustrate that the allele calls are appropriately assigned based on the laboratory's established analytical threshold and may not require additional verification or testing.

4.3 **A broad peak** is defined as when the width of a peak exceeds a maximum set peak width.

4.3.1 The broad peak parameter shall be evaluated and adjusted according to validation data, as necessary.

4.3.2 The broad peak quality flag and/or peak width threshold shall be able to detect excessively wide peaks, to avoid missing heterozygotes and minor components separated by one base pair.

4.3.3 The broad peak quality flag and/or peak width threshold shall be able to detect loss of resolution from poor injection or migration of the DNA profile or internal size standard when this interferes with profile interpretation.

Figure 2: A configuration which inaccurately accepts broad peak width can result in mischaracterization of heterozygous loci.

4.3.4 The broad peak quality flag and/or peak width threshold shall be adjusted if broad peaks are incorrectly passed as allelic or if alleles separated by one base pair are not properly resolved.

4.3.5 A broad peak configuration setting which is too restrictive can lower efficiency and result in the unnecessary manual review of acceptable data.

4.3.6 A minimum of five injection resolution challenges shall be appropriately flagged and/or fall outside of the laboratory's set threshold.

4.3.7 A minimum of five base pair resolution challenges at three different loci shall be appropriately flagged and/or fall outside of the laboratory's set threshold.

4.4 **Global filters** will filter out peaks with a peak height ratio less than or equal to a set threshold. Peaks with a peak height ratio greater than the threshold are labeled. Global filters shall not be used with an Expert System to analyze forensic samples.

4.5 **The homozygote threshold** is defined as the peak height determined through internal validation studies, below which it is reasonable to assume that, at a given locus, allelic dropout of a sister allele in a heterozygous pair may have occurred.

4.5.1 The homozygote threshold shall be based on empirical data. The homozygote threshold may be determined during internal validation of the DNA typing kit or during validation of the Expert System (i.e., when a stochastic threshold was not determined during validation of the typing kit). The homozygote threshold (possibly referred to as the "homozygous minimum peak height") is used to ensure that all heterozygous loci are properly interpreted by the Expert System. This setting must be equal to or more stringent than a stochastic threshold determined through the kit validation (as applicable). If the laboratory has validated a stochastic threshold, the homozygote threshold should be equal to or more stringent than the stochastic threshold determined during the typing kit validation (i.e., the low homozygous setting should be higher than the stochastic

threshold). The Expert System shall always indicate when peaks do not meet or exceed the stochastic threshold.

4.5.2 This threshold shall be able to identify a heterozygous locus where drop-out of a sister allele may be occurring.

4.5.3 This flag should be challenged via low-level single-source data. A minimum of ten samples having known heterozygote genotypes where only one peak is called by the Expert System shall be correctly flagged by the software.

4.6 **The heterozygote threshold** is defined as the peak height, above which two peaks are assumed to be from a heterozygote pair, presuming they meet other requirements (e.g., peak height ratio). The heterozygous threshold may provide an additional quality flag for reviewing low quality data.

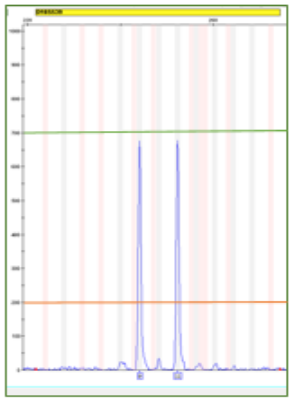

Figure 3: Orange line is the set analytical threshold (e.g., 200RFUs). Green line is the set heterozygote threshold (e.g., 700RFUs). If heterozygote peaks fall between the lines that locus will be flagged.

4.6.1 The heterozygote threshold will be greater than or equal to the analytical threshold.

4.6.2 This parameter should be challenged in conjunction with the analytical/detection threshold parameter, the peak height ratio (imbalance) parameter and homozygote threshold.

4.7 **Stutter parameters** describe the ratio of acceptable stutter.

4.7.1 The stutter parameters shall be based on empirical data. The stutter parameters determined during developmental or internal validation of the typing kit may be applied. This threshold shall be able to filter stutter peaks. The stutter parameters shall not be set too high as to intentionally filter out minor alleles and/or have a mixture profile appear as single source.

4.7.2 Global stutter filters will filter a set stutter ratio across all loci. Global stutter filters are not to be used in an Expert System for the evaluation of forensic samples.

4.7.3 The locus stutter ratio is the ratio of acceptable stutter at a given locus. This threshold is to be applied at the individual locus level depending on the software setting. This threshold shall include reverse stutter and should, as applicable, include forward and partial repeat stutter.

4.7.4 The Expert System may not have a specific quality flag for elevated stutter; however, the Expert System shall be able to identify instances of elevated stutter in a single source profile. Quality flags to identify elevated stutter may include but are not limited to the allele number (ploidy) and peak height ratio (imbalance) flags. This parameter shall be challenged via samples having a peak(s) in a stutter position that exceeds the locus- or allele-specific stutter threshold. This challenge requires a minimum of five observations. These observations shall occur at a minimum of five different loci.

4.8 **Peak Height Ratio Threshold** is defined as the greatest imbalance two heterozygote sister alleles can exhibit at one locus and still reasonably originate from the same contributor. It is typically denoted as a ratio of intensity of one peak over another at one locus.

4.8.1 The Peak Height Ratio or Imbalance flag indicates if the peak height ratio between the lowest and the highest peak at a locus is less than the minimum peak height ratio defined in the analysis method.

4.8.2 The laboratory shall establish peak height ratio expectations based on empirical data derived from DNA typing results from single-source samples. Different peak height ratio expectations may be applied to individual loci; alternatively, a single peak height ratio may be utilized if that value is sufficient to detect suspicious or uncommon imbalance for all loci to which it is applied. Peak balance consistent with empirically determined ratios reflect high quality data. A setting which accepts

minimum imbalance ensures only high-quality data are passed by the system.

4.8.3 This parameter shall be challenged via single-source samples with known heterozygote peak height ratio imbalances. This challenge requires a minimum of five observations of heterozygote peak imbalance.

Calibration challenge category	suggested minimum number of challenges
Allele Number	**15**
Broad Peak	**10**
Stutter	**5**
Peak Height Ratio	**5**
Degradation	**5**
Inhibition	**5**
Primer Peak (where applicable)	**5**
Contamination	**5**
Low Homozygote	**10**
Micro-variant	**5**
Mixture	**30**
Off-scale	**5**
Out of Marker Range	**5**
Spikes	**1**
Upstream or downstream of the allelic ladder	**5**
Pull-up	**5**
(-) A	**5**
Size Standard	**5**
Positive control	**5**
Single Source-evidence	24
Partials	10
Known references	10
Positive amplification controls	5
Negative controls-extraction	5
Negative controls-amplification	5
Allelic Ladders	5
Total	200

Concordance category	suggested minimum number of comparisons
Single Source-evidence	150
Single Source-known reference	20
Mixture	100
Partial	100
Low quant	30
Difficult samples	10
Positive control-amplification	30
Negative control-amplification	30
Negative control-extraction	30
Total	500

Table 1: Calibration and Concordance data will consist of samples exhibiting multiple quality issues across multiple case types. **Bolded** categories are required challenges and shall be evaluated in the validation. Non-bolded categories are those that should be considered when finalizing the samples for the calibration and concordance sets, but do not constitute samples required for validation. Low quant samples, for example, would be those samples which require the sample to be amplified at full volume. Difficult samples would be described as unique and/or peculiar samples observed in the laboratory which are a challenge for manual interpretation and should be evaluated to ensure the Expert System identifies the sample for human review.

5. Additional Software Parameters

The following parameters may not be available in all Expert System software. If software settings are available and implemented by the laboratory, they shall be assessed as described below. If no software setting is available, these data quality challenges should be properly flagged by other Expert System parameters, but no specific testing of the parameter is required.

5.1 **Degradation** occurs mainly when a sample is exposed to certain environmental factors and may lead to profiles having allelic and/or locus dropout, typically at the larger loci.

5.1.1 Some Expert System software can allow for a degradation quality flag to be set. If the laboratory chooses to implement this feature, the laboratory should define the point at which degradation interferes with the interpretation of single-source samples and internally validate an acceptable degradation slope. Samples meeting this laboratory definition shall not be passed by the Expert System.

5.1.2 If a specific quality flag does not exist in the software, laboratories can utilize the quality flags for allele number, peak height ratio (imbalance) threshold, homozygote threshold, heterozygote threshold, etc., to identify degraded samples.

5.1.3 If a laboratory is specifically implementing a degradation setting, degradation shall be challenged with at least 5 samples that have a degradation pattern. If not implementing a setting specifically for degradation, the laboratory shall challenge the Expert System with at least 5 degraded samples and describe in the validation summary how the system will respond to those samples.

5.2 **Inhibition** occurs when a sample is exposed to factors that interfere with the PCR reaction.

5.2.1 Some Expert System software can allow for the detection of inhibition occurring because of environmental insults, the addition of known

inhibitors to the PCR reaction, or amplification of unpurified lysis products (e.g., dirty extracts). Inhibition may be specifically detected via dye balance or inter-locus peak height ratio flags. If the laboratory chooses to implement this feature, the laboratory shall define the point at which inhibition interferes with the interpretation of single-source samples and internally validate an acceptable inhibition threshold. Samples meeting this laboratory definition shall not be passed by the Expert System.

5.2.2 If a specific quality flag does not exist in the software, laboratories can utilize the quality flags for allele number, peak height ratio (imbalance) threshold, homozygote threshold, heterozygote threshold, etc., to identify inhibited samples.

5.2.3 If a laboratory is specifically implementing a setting to detect inhibition, this setting shall be challenged with at least 5 samples that have an inhibition pattern. If not implementing a setting specifically for inhibition, the laboratory shall challenge the Expert System with at least 5 inhibited samples and describe in the validation summary how the system will respond to those samples.

5.3 P**rimer peak detector** is a quality flag that identifies the presence or absence of primers in negative control samples indicating that the appropriate reagents and amplified products were added.

5.3.1 If configurable within the software, the Expert System shall identify samples where primer peaks are not detected or present.

5.3.2 If a laboratory is specifically implementing a setting to detect primer peaks, this setting shall be challenged with at least five samples that do not contain a primer peak. Data may be generated by creating samples that contain only formamide and internal size standard and defining the sample as a negative control, or by adding less than the required amount of amplified product to the electrophoresis plate.

6. Challenging the Parameter Settings

As the laboratory validates an Expert System for use in casework, the laboratory shall configure the system's settings to address a core set of quality issues typically observed in casework analysis. The core quality issues that shall be assessed are as follows:

6.1 **Contamination** in a reagent blank or negative control is defined as any allelic peaks above the detection threshold. Although not required, laboratories may choose to set the detection threshold used by the Expert System for reagent blanks and other negative controls lower than the one determined during validation, in order to ensure that sub-threshold allelic peaks are flagged by the Expert System, if that is part of the laboratory's routine assessment of these controls. The detection threshold used for negative controls shall not be higher than that determined during validation.

6.1.1 The Expert System's ability to detect contamination shall be challenged by marking a sample with allelic data as a negative control, or by using a negative control that has been designated as contaminated.

6.1.2 This challenge requires a minimum of five observations.

6.2 **Drop-in** is defined as a non-reproducible allele in a profile or control that does not originate from the principal DNA donor(s). The laboratory should have validation studies that demonstrate whether their amplification and electrophoresis processes are affected by allelic drop-in. Drop-in can affect both controls and casework samples. Samples that could be considered as having allelic drop-in shall not be passed by the Expert System.

6.2.1 The laboratory shall set a minimum peak height threshold (for both heterozygote and homozygote peaks) at or above any drop-in cap or threshold determined by the laboratory during amplification kit and/or electrophoresis system validation.

6.2.2 These quality flags will already be challenged and no samples containing drop-in specifically will need to be challenged. If the laboratory's typing

kit validation studies demonstrated drop-in, those samples should be included in the calibration set.

6.3 **Drop-out (Missing Allele and Missing Locus)** is defined as failure to detect an allele within a sample or failure to amplify an allele during PCR. Samples having allelic dropout shall not be passed by the Expert System.

6.3.1 There is not a specific quality flag for detecting drop-out, however, utilizing the quality flags for homozygous minimum peak height to detect sister allele drop-out and allele number, or other similar indicator(s), for total locus drop-out allows for drop-out to be detected.

6.3.2 The listed quality flags will already be challenged and no additional samples with drop-out specifically will need to be challenged.

6.4 **Micro-variant** is defined as a peak that falls outside one of the defined bins in the allelic ladder or Expert System. Quality flags such as Off Ladder Allele, Off Bin, or BIN indicate if a sample includes a micro-variant allele.

6.4.1 The Off Ladder Allele/Off Bin/BIN quality flags are typically hard-coded into the Expert System to flag any peak/allele that falls outside one of the defined bins within the marker range. The laboratory needs to review the defined bins and whether there is any bin overlap in the allelic ladder. During validation, the laboratory may choose to add bins for commonly observed micro-variants.

6.4.2 For a micro-variant, the Off Ladder Allele/Off Bin/BIN flag shall indicate the detection of off ladder alleles within the allelic ladder due to the presence of alleles or artifacts having a base pair size different from the canonical allele size.

6.4.3 The challenge set shall include samples with off ladder/off bin alleles within the marker range.

6.4.4 This challenge requires a minimum of five observations.

6.5 **Mixture** is defined as a sample having more than one contributor. A combination of flags may indicate a mixture.

6.5.1 Some Expert System software can allow for a mixture quality flag or indicator to be configured by the user. If the laboratory chooses to implement this feature, the laboratory shall define the point at which allele number and peak height ratio interferes with the interpretation of single-source samples. In other systems, the mixture (MIX) quality flag is hard coded into the Expert System to flag samples based on allele number and peak height ratio settings. No additional user settings are required.

6.5.2 The Expert System shall be able to indicate a potential mixed source sample.

6.5.3 Mixed samples shall be tested by analyzing a series of samples containing DNA from two or more individuals. Mixture samples selected shall vary in contributor ratio and contributor number similar to those encountered in the laboratory.

6.5.4 This challenge requires analyzing a mixture dilution series for different contributor ratios.

6.5.4.1 Challenge samples shall include the upper and lower limit of the contributor ratios encountered and identified as mixtures in the laboratory. The goal is to determine the limit of detection for minor contributors above the detection threshold that the Expert System will flag.

6.5.4.2 If a contributor ratio threshold has been validated and is used for profile interpretation, the Expert System shall be challenged against a contributor ratio series that is one ratio beyond the established threshold (e.g., if the established threshold is 1:10, then the system could be tested at 1:5, 1:10 and 1:15).

6.5.4.3 If a contributor ratio threshold has not been established, the Expert System shall be challenged against a mixture series that ranges from 1:1 to a ratio where the amplification kit can no longer detect a mixture.

6.5.4.4 Documentation of the contributor ratios used to challenge the Expert System shall be included in the validation summary.

6.6 Off Scale (OS) or Saturated (SD) data is flagged when a sample and/or individual locus within the sample has fluorescence that exceeds the limited linear range of the detection instrument and results in signal saturation. Off scale data is not an issue in itself but serves as a potential indicator of other artifacts which could be mistakenly interpreted as allelic data.

6.6.1 The laboratory shall establish guidelines for addressing off scale data.

6.6.1.1 Laboratories shall enable features which scan for off scale or saturated data and review loci, with special attention given loci appearing as heterozygous which are most vulnerable to mischaracterization from off scale data.

6.6.1.2 The weighting of the off-scale or saturated flag shall be such that it would require the manual review of the sample.

6.6.2 The challenge set shall include samples with data that exceeds the saturation point of the detection instrument and samples from a dilution series that include high concentrations of DNA.

6.6.3 This challenge requires a minimum of five saturated samples.

6.7 Outside Marker Range is a quality flag that detects when one or more peaks are between two marker size ranges.

6.7.1 The laboratory's Expert System shall be set so that all peaks between the predefined smallest/largest marker range (interlocus space) are assessed. The smallest and largest markers in each dye shall be set to capture routine alleles at the edges of these ranges. It is not possible to configure the Expert System to capture or flag peaks present outside of the ranges for each dye. The outside marker range flag shall indicate if labeled peaks are detected between two marker size ranges.

6.7.2 The challenge set shall include samples with labeled peaks between two marker size ranges.

6.7.3 This challenge requires a minimum of five observations. The laboratory may adjust their marker ranges as needed to ensure this flag is challenged.

6.8 **Positive control** is a DNA sample or known profile that the laboratory uses to monitor or assess the quality of the DNA typing or interpretation process (e.g., amplification positive control, extraction positive control).

6.8.1 The Expert System shall assess control concordance for both amplification and extraction positive controls, as applicable to the laboratory's workflow. These positive controls shall not only have the correct alleles called but should also meet other Expert System parameters set by the user (e.g., PHR, minimum peak height).

6.8.2 The Expert System's ability to assess the positive control may be challenged by marking a non-positive control sample as a positive control, or by using a known positive control that has been failed by a human evaluation.

6.8.3 This challenge requires a minimum of five observations of a flagged positive control. For laboratories using an extraction positive control in addition to a positive amplification control, at least one of each control shall be flagged for a total of five.

6.9 **Shouldering (Minus-A)** is an artifact which is the result of all amplified fragments not being adenylated following the completed PCR reaction (non-template dependent nucleotide addition). This artifact presents as a second peak in close proximity, one base pair shorter than the allele. It has been referred to as a split peak and shouldering.

6.9.1 The Expert System shall identify those samples affected by incomplete adenylation. This may be done individually or in combination with other flags (e.g., off ladder, peak height ratio, allele number).

6.9.2 Global Minus-A filters will filter a set minus-A ratio across all loci. Global Minus-A filters shall not be used in an Expert System for the evaluation of forensic samples.

6.9.3 This challenge requires a minimum of five observations.

6.10 **Size Standard** is a quality flag that indicates when there is a problem with the internal size standard.

6.10.1 The Expert System shall assess internal size standard base pair designation and peak morphology in all samples and controls.

6.10.2 This challenge requires a minimum of five observations of a flagged internal size standard.

6.11 **Spikes** are cross channel artifacts caused by a voltage change during capillary electrophoresis.

6.11.1 While rarely observed, the Expert System shall identify those samples containing a spike.

6.11.2 This challenge requires a minimum of one observation of spike flagged by the Expert System.

6.12 **Upstream or downstream allele** are alleles that are larger or smaller than the range of alleles in the allelic ladder.

6.12.1 This issue is distinct from peaks outside the marker range in that the peaks are within a defined marker range. These peaks are those alleles that would require a "<" or ">" designation for CODIS entry and searching.

6.12.2 The laboratory shall review the defined bins used by the Expert System and determine whether alleles requiring a ">" or "<" designation will be flagged by the Expert System. The laboratory shall not create bins with the "<" or ">" designation for use on casework samples.

6.12.3 This challenge requires a minimum of five observations.

6.13 **Pull-Up** is defined as signal at a location and base pair range that is not attributed to allelic data and is the result of incomplete spectral dye separation.

6.13.1 The Expert System shall correctly flag samples containing pull-up peaks. Pull up may be flagged via the allele number, peak height, and/or peak height ratio (imbalance) parameters.

584 **6.13.2** The challenge set shall include samples containing peaks identified as
585 pull-up by a qualified analyst.

586 **6.13.3** This challenge requires a minimum of five observations.

587 **7. Implementation**

588 Once validation has been completed, the laboratory should look toward how the Expert
589 System will be implemented. Considerations for implementation include but are not
590 limited to, documented interpretation guidelines for samples rejected by the Expert
591 System, how quarterly recertification will be achieved, requirements following repair or
592 service to the Expert System, and the control of the Expert System.
593

Validation Checklist (Guidance on how to find sample sets, define parameters, or evaluate parameter settings)

1-Getting Started	
Is the validation being performed using employees of the laboratory? (Individuals who are not provided by or affiliated with the Expert System vendor)	
Have the appropriate number of unique samples been compiled, reviewed, and classified according to quality issues?	
2-Configuration and Calibration	3-Concordance Study
Have at least 200 samples been designated for the calibration study?	Have at least 500 samples been designated for the concordance study? (for internal validations)
Does the calibration data set contain both problematic and high-quality data?	Have these samples undergone an initial review?
Is the data set a mix of samples, ± controls, and ladders?	Were the results of the initial review compared to the Expert System review?
Is the sample set representative of data routinely produced by the laboratory?	Did the initial review process identify quality issues not documented in the Expert System review?
Does the sample set contain, at a minimum, the required quality issues listed in Table 1?	If so, what steps were taken to ensure the Expert System review produces results at least as accurate as the initial review process?
Were the software settings configured to detect quality issues in a manner at least as sensitive as the initial review[1] procedure?	Did the Expert System review identify quality issues not documented in initial review?
Has a minimum homozygote threshold been configured to a degree such that it is at least as stringent as the manual review process?	If so, what new issues were identified?
Does the threshold demonstrate that profiles exhibiting partial dropout have not been designated as "Accept"?	Was the issue documented in the validation summary?
Is the Expert System detection threshold sufficiently configured to detect the appropriate RFU range?	What, if any, steps were taken to ensure the quality of the data produced under the initial review process?
Was Peak Height Ratio a consideration in the establishment of this threshold setting?	
Did the calibration portion of the validation demonstrate background signal filter configurations do not mischaracterize data? (i.e., ensure samples with low level additional contributors are properly classified as mixtures)	
Did the validation demonstrate the Expert System's ability to consistently detect issues as required by the NDIS Operational Procedures?	
Did the validation demonstrate that the Expert System settings evaluate controls to a degree at least as stringent as those used to evaluate samples?	
Did the validation demonstrate the Expert System does not make incorrect allele calls in cases where the results are classified as "Accept"?	
4-Implementation	
Has the laboratory developed interpretation guidelines and procedures to resolve quality challenged samples detected by the Expert System? (samples which the ES classifies as Edit or Reject)	
Has the laboratory created a dataset to complete its quarterly recertification?	
Prior to recertification, was the dataset supplemented to contain samples from recent analyses?	
Does the laboratory have procedures to recertify the Expert System following repair, service, or calibration?	

[1] **Initial review refers to the laboratory's manual review procedure {*Labs may not use an Expert System calibrated for offenders for a casework comparison*}**

Roadmaps for Calibration & Concordance

Calibration Roadmap which describes the general steps required to ensure System settings perform with accuracy.

- ensure settings are consistent with amplification kit validation, current review protocols, and interpretation guidelines
- select appropriate number of samples for each challenge area which will stress the System settings- depends heavily on typical case type throughput for each laboratory
- record all applicable issues associated with each sample (typically identified through the current review process)
- verify the System identified all of the anticipated issues
- determine if the Expert System identified issues not identified in the current review process
- evaluate if the Expert System review is as stringent (or more stringent) as the current review process
- calibration is complete when the System is competent to indicate a problem exists for all quality issues identified through the current review process

Concordance Roadmap which describes the general steps required to ensure the results of the Expert System review are consistent with the current review process.

- select an appropriate number of samples ensuring each sample has undergone a documented manual review
- record all issues associated with each sample (typically identified through the current review process)
- collect any discrepancies between the manual review and Expert System review for each sample using a spreadsheet, database, or other tool
- evaluate if the Expert System review is as stringent (or more stringent) as the current review process
- concordance is complete when the Expert System demonstrates fidelity or superiority as it relates to discovering all quality issues identified through the current review process
- concordance must demonstrate the Expert System produces results of the same (or better) quality as the current review process

References and Suggested Readings

Federal Bureau of Investigation (2020) *NDIS Operational Procedures Manual*, available at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-procedures-manual.

Federal Bureau of Investigation (2020) *Quality Assurance Standards for Forensic DNA Testing Laboratories*, available at https://www.fbi.gov/services/laboratory/biometric-analysis/codis.

Scientific Working Group on DNA Analysis Methods (SWGDAM) (2017) *SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*, available at http://www.swgdam.org.

Scientific Working Group on DNA Analysis Methods (SWGDAM) (2016) *SWGDAM Validation Guidelines for DNA Analysis Methods*, available at http://www.swgdam.org.