Scientific Working Group on DNA Analysis Methods

Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories

Table of Contents

Background.....3

Core Elements.....4

Executive Summary.....7

Section 1 – Interpretation of DNA Typing Results 8

Section 2 – Mixture Interpretation Overview and Strategies 20

Section 3 – Comparison of References and Statistical Weight of Probative Inclusions.....36

Section 4A – RMP 43

Section 4B – LR.....56

Section 4C – CPI / CPE 77

Glossary.....84

References and Suggested Readings.....91

SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories

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, subcommittees discuss topics of interest to the forensic DNA community and often develop documents to provide direction and guidance for the community. These guidelines were presented to the SWGDAM membership and approved on January 12, 2017.

This document provides guidelines for the interpretation of DNA typing results from short tandem repeats (STR).

This document contains guidelines and not minimum standards. In the event of a conflict between the FBI Quality Assurance Standards for Forensic DNA Testing Laboratories (QAS) and these

guidelines, the QAS and the QAS Audit Documents have precedence over these guidelines. The use of the term 'shall' or 'must' is not intended to transform these guidelines into standards.

These revised guidelines supersede the SWGDAM Short Tandem Repeat (STR) Interpretation Guidelines (2010). They are intended to be applied prospectively and not retroactively. With the underlying assumption that work performed prior to the issuance of these revisions was appropriate and supported by validation, revision of the applicable guidelines is not intended to invalidate or call into question the previous work. Laboratories are encouraged to review their standard operating procedures and validation data in light of these guidelines and to update their procedures as needed.

These guidelines are intended to be used for laboratories that will continue to employ binary approaches to interpret electrophoresis-based data. These guidelines may be applicable to probabilistic genotyping, next generation sequencing, and/or rapid DNA technology in a limited capacity, but are not intended for those technologies. It is anticipated that future documents will address these new technologies and methodologies. These guidelines generally address the interpretation of single-source samples and mixtures of DNA from two people. The basic concepts hold true as they relate to DNA mixtures of three or more contributors, those involving stochastic-level contributors, and with mixtures containing biologically related individuals. However, there are nuances and limitations to binary interpretation of this more complex data which will be addressed.

This document discusses strategies for low level DNA samples but is not intended to address the interpretation of analytical results from enhanced (low template) detection DNA methods. Guidance for such methods is available in the SWGDAM Guidelines for STR Enhanced Detection Methods.

Background:

The interpretation and reporting of DNA typing results for human identification purposes requires professional judgment and expertise. It is a complex process that draws upon empirical data and the overall training and experience of the scientist involved and should be reflective of the available case history and any assumptions being made. A DNA analyst should use all relevant information that is available to assist in the interpretation, such as the location from which the sample originated (e.g., intimate sample), the body fluid screening results, the quantity of DNA extracted, and the overall quality of the DNA profile.

Additionally, laboratories that analyze DNA samples for forensic casework purposes are required by the Quality Assurance Standards for Forensic DNA Testing Laboratories to establish and follow documented procedures for the interpretation of DNA typing results and reporting. Due to the multiplicity of forensic sample types and the potential complexity of DNA typing results, it is impractical and infeasible to cover every aspect of DNA interpretation by a preset rule. However, the laboratory should utilize written procedures for interpretation of analytical results with the understanding that specificity in the standard operating protocols will enable greater consistency and accuracy among analysts within a laboratory. It is recommended that standard operating procedures for the interpretation of DNA typing results be sufficiently detailed that other forensic DNA analysts can review and understand in full the laboratory's policies and practices.

Upon completion of the technical aspects of DNA analysis, DNA typing results must be verified and interpreted. The verification of the DNA typing results involves a review of peak designations and other software-generated information, as well as an evaluation of quality controls. Based on this data assessment, the DNA analyst performs interpretations, makes comparisons among samples (where appropriate) and draws conclusions. The conclusions reached as part of the interpretation process are compiled into a written report by the DNA analyst. These data and conclusions are then subjected to technical and administrative review prior to issuing a final laboratory case report.

Guidance is provided for forensic casework analyses on the identification and application of thresholds for allele detection, interpretation, and appropriate statistical approaches for autosomal STRs. This document first addresses the core elements (i.e., fundamental tenets) which must be considered when evaluating and interpreting autosomal STR typing data, determining conclusions, and reporting results. An executive summary is then provided linking the core elements with applicable validation studies. Lastly, guideline-specific sections are provided which include interpreting autosomal STR data, mixture interpretation, and statistical approaches. For clarity, examples are interspersed within each section to further illustrate specified guidelines. These examples are not meant to be all-inclusive but rather to illustrate a particular guideline. For simplicity, they are limited in display, and must be considered as representative of the entire profile, unless otherwise noted. For demonstration purposes, all examples utilize an analytical threshold of 50 RFU and a stochastic threshold of 200 RFU.

Core Elements:

Note: The numbers given in parentheses refer to the individual sections of this document that provide additional details.

I. The laboratory's interpretation guidelines and thresholds shall be based on and supported by applicable internal validation studies, publications, and scientific literature.

II. (1.1) The laboratory shall establish an analytical threshold.

III. (1.2) The laboratory shall establish criteria to evaluate internal standards and allelic ladders.

IV. (1.3) The laboratory shall establish criteria for evaluation of controls.

V. (1 Introduction) A laboratory using STR multiplexes that contain redundant loci shall establish criteria regarding the concordance of such data.

VI. (1.5) The laboratory shall establish criteria to assign allele designations to appropriate peaks.

VII. (1.6) The laboratory shall establish criteria to address the interpretation of non-allelic peaks.

VIII. (1.7) The laboratory shall establish a stochastic threshold for use with binary methods.

IX. (1.8) The laboratory shall establish peak height ratio expectations for heterozygous genotypes.

X. (1.9) For DNA mixtures, the laboratory shall establish guidelines for determination of the minimum number of contributors to a sample.

XI. (1.10) The laboratory shall establish guidelines to determine whether DNA typing results are suitable for comparisons.

XII. (2 Introduction) The primary goal of mixture interpretation shall be to determine the possible genotype combinations of the contributors.

XIII. (2.1) Any criteria (e.g., assumptions such as number of contributors and/or the presence of a known contributor) used in the interpretation of a mixed DNA sample shall be supported by the data and shall be defined and documented.

XIV. (2.3) Interpretation guidelines for mixtures must be based on mixture studies conducted using known contributors that represent the number of contributors and the range of general mixture types for which the procedure will be used in casework (e.g., mixture proportions and template quantities). The laboratory guidelines shall be sufficiently detailed to ensure confidence in the separation of the "major" versus "minor" components.

XV. (3 Introduction) The interpretation of the evidentiary profile should determine the statistical approach used. It would be inappropriate to make inclusions or exclusions based on the statistical approach without first considering the interpretation of the profile.

XVI. (3.1.1) The laboratory shall establish guidelines to ensure that, to the extent possible, DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors.

XVII. (3.1.3) The laboratory shall establish guidelines for inclusionary, exclusionary, and inconclusive determinations based on comparisons of DNA typing results from known samples to both single-source and mixed evidentiary samples.

XVIII. (3.2.1) Except for a reasonably assumed contributor, the laboratory shall perform statistical analysis in support of any inclusion (or a "cannot be excluded" conclusion), irrespective of the number of alleles detected and the quantitative value of the statistical analysis.

XIX. (3.2.2) The genetic loci and assumptions used for statistical calculations shall be documented, at a minimum, in the case notes.

XX. (3.2.3) Data that cannot be used to support inclusions shall not be used in statistical analysis at individual loci or for an entire multi-locus profile.

XXI. (3.2.4) The laboratory shall document the source of the population database(s) used in any statistical analysis.

XXII. (3.2.5) The formulae used in any statistical analysis shall be documented and must address both homozygous and heterozygous typing results, multiple locus profiles, mixtures, minimum allele frequencies, and, where appropriate, biological relationships.

XXIII. (3.2.6) Exclusionary conclusions do not require statistical analysis.

XXIV. (3.3) The three binary approaches for providing weight-of-evidence to an interpretation include the random match probability (RMP, **4A**), likelihood ratio (LR, **4B**)**,** and the combined probability of exclusion/inclusion (CPE/CPI, **4C**)**.** These methods typically assume unrelated individuals.

XXV. (3.2.5.1) Different statistical approaches (e.g., RMP, LR, CPE/CPI) shall not be combined into one calculation because they rely upon different fundamental assumptions.

XXVI. (3.4.6) If a laboratory uses source attribution statements, then criteria shall be established on which such declarations are based.

Executive Summary:

Core elements	Validation studies needed
Examine the profile and perform a quality assessment Core Elements I - VII (Section 1)	Analytical threshold Stochastic threshold Stutter thresholds Limit of linearity (off-scale data/pull-up)
Categorize allele peaks Core Elements VIII - IX (Section 1)	
Identify the minimum number of contributors to determine single source or mixture path Core Element X (Section 1)	Peak height ratio expectations
Mixture interpretation overview and strategies **The primary goal should be to determine all possible genotype combinations** Core Elements X1 - XIII (Section 2) • Assumptions • 2 person • >2 person • Major/minor • Known contributors • Potential stutter	Running applicable known mixtures (i.e., different contributor number, with different contributor proportions and template) to establish and assess protocols
Compare to reference samples Core Elements XIV - XV (Section 2)	
For inclusions, determine statistical weight of evidence with binary models Core Elements XVI - XXVI (Section 3) • RMP (Section 4a) • LR (Appendix 4b) • CPE/CPI (Section 4c)	Population allele frequencies defined and databases compared to Hardy-Weinberg Equilibrium expectations Software validation and/or performance check as needed to effectively demonstrate and confirm calculations

Section 1: Interpretation of DNA Typing Results

Introduction:

Using STR technologies for human identification, DNA typing results are derived through application of analytical software during and after electrophoresis of fluorescently-labeled amplification products. For each sample, the software translates fluorescence intensity data into electropherograms and then labels any detected peaks with such descriptors as size (in nucleotides) and peak height (in relative fluorescence units, or RFU). Using allelic ladders for reference, the software then labels peaks that meet certain criteria with allelic designations.

To ensure the accuracy of these computer-generated allele designations, the DNA analyst must verify that appropriate genotyping parameters (i.e., internal size standard and allelic ladder) were used and that the correct genotyping results were obtained for a known positive control. Additionally, if a sample is amplified using multiple kits that contain redundant loci, the DNA analyst must address the concordance of the genotyping results at the loci that are common to both kits.

The results of the analysis controls [i.e., reagent blank(s), positive amplification control(s), and negative amplification control(s)] are evaluated. If the reagent blank(s), positive amplification control(s), and negative amplification control(s) yield results that are within their prescribed specifications, the DNA analyst interprets the DNA typing results from each sample.

1.1 Analytical Threshold

The analytical threshold should be based on signal-to-noise analyses of internally derived empirical data. An analytical threshold defines the minimum height requirement at and above which detected peaks can be reliably distinguished from background noise. Because the analytical threshold is based upon a distribution of noise values, it is expected that occasional, non-reproducible noise peaks may be detected above the analytical threshold. Usage of an exceedingly high analytical threshold increases the risk of allelic data loss.

1.2 The laboratory must develop criteria to evaluate internal standards and allelic ladders.

1.3 Controls are required to assess analytical procedures.

1.3.1 The laboratory must establish criteria for evaluation of the following controls, including but not limited to: reagent blank and positive and negative amplification controls.

1.3.2 The laboratory must develop criteria for the interpretation and documentation of results in the event that the controls do not perform as expected.

1.4 Locus Designation: The laboratory must have criteria to address locus assignment for alleles. The criteria should address alleles that fall above the largest or below the smallest allele (or virtual bin) of the allelic ladder.

1.5 Allele Designation: The laboratory must designate alleles as numerical values in accordance with recommendations of the International Society of Forensic Genetics.

1.5.1 Allele designation is based operationally on the number of repeat sequences contained within the allele and by comparison to an allelic ladder.

1.5.2 The laboratory must establish guidelines for the designation of alleles containing an incomplete repeat motif (i.e., an off-ladder allele falling within the range spanned by the ladder alleles). This designation includes the number of complete repeats and, separated by a decimal point, the number of nucleotides in the incomplete repeat (e.g., FGA 18.2 allele).

1.5.3 Extrapolation of an above/below ladder allele to a specific designation should also be supported by precision studies, validation and determination of measurement variance. Above/below ladder alleles should be designated as either greater than (>) or less than (<) the respective ladder allele (or virtual bin), or designated numerically when appropriate extrapolation can be used. When the ">" or "<" designation is used, the laboratory should establish criteria, based on relative sizes, for the comparison of such alleles among samples.

1.6 Non-allelic peaks

Because forensic DNA typing characterizes STR loci using PCR and electrophoretic technologies, some data that result from this analytical scheme may not represent actual alleles that originate in the sample. It is therefore necessary, before the STR typing results can be used for comparison purposes, to identify any potential non-allelic peaks. Non-allelic peaks may be PCR products (e.g., stutter, non-template dependent nucleotide addition, and non-specific amplification product) or instrumental artifacts (e.g., spikes, raised baseline, and incomplete spectral separation resulting in pull-up or bleed-through).

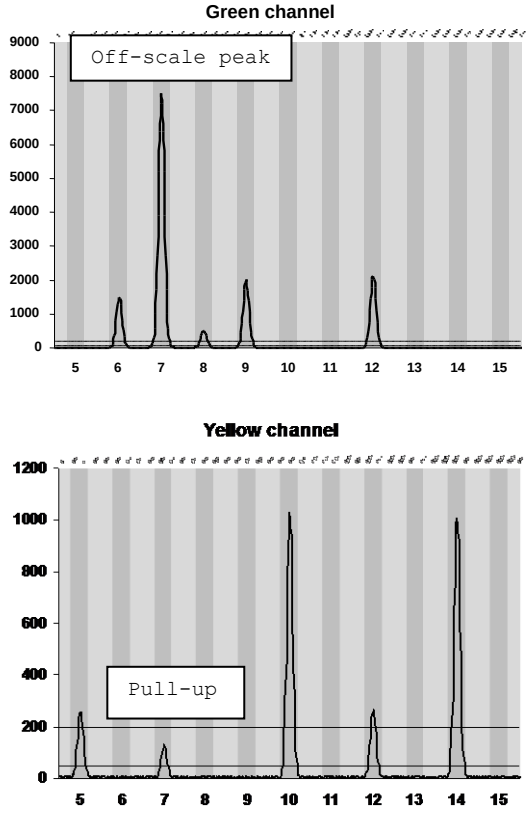
1.6.1 The laboratory must establish criteria based on empirical data (obtained internally or externally), and specific to the amplification and detection systems used, to address the interpretation of non-allelic peaks. These guidelines should address identification of non-allelic peaks and the uniform application, across all loci of a DNA profile, of the criteria to identify non-allelic peaks.

1.6.1.1 In general, the empirical criteria are based on qualitative and/or quantitative characteristics of peaks. As an example, dye artifacts and spikes may be distinguished from allelic peaks based on morphology and/or reproducibility. Stutter and non-template dependent nucleotide addition peaks may be characterized based on size and amplitude relative to an allelic peak.

1.6.1.2 The laboratory must establish guidelines addressing off-scale data. Fluorescence detection instruments have a limited linear range of detection, and signal saturation can result in off-scale peaks. Following peak detection, such

peaks in the analyzed data are assigned an artificial height value which is not representative of the true amplitude. Peak height values for off-scale peaks should be used with caution in quantitative aspects of interpretation (e.g., stutter and peak height ratio assessments).

Example: *This example illustrates how signal saturation from off-scale data can result in an increase in relative fluorescence unit (RFU) signal for artifacts, making interpretation of the minor contributor in a mixture more difficult.*



This example is a mixture of two contributors at an approximate 3:1 ratio. The locus in the green channel contains an off-scale major peak (bin 7) resulting in an artificial peak height for allele 7. This artificial peak height leads to an apparent elevated amount of stutter product, which is detected in both bins 6 (n-4 stutter) and 8 (n+4 stutter), above the laboratory's stutter expectation and becomes designated as alleles by the genotyping software. In addition, this same off-scale peak is causing spectral pull-up in an adjacent dye channel (yellow) in another locus of the mixture. This pull-up peak is in an actual bin (bin 7) in the yellow channel and thus is being designated as an allele by the software. These artifacts are indistinguishable from actual alleles (i.e., above the analytical threshold, good peak morphology, in an allele sizing bin, etc). If the peaks in bins 6 and 8 of the green channel and bin 7 in the yellow channel are not recognized as artifacts and

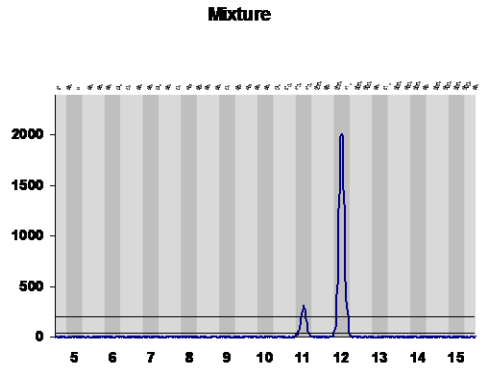
are used in the interpretation of the profile, this would result in over-estimating the number of contributors and complicating the interpretation for the true minor contributor in the mixture.

Empirical data from internal validation studies may be used (e.g., dilution series including high concentrations) by laboratories as a strategy to establish an upper RFU threshold (i.e., limit of linearity) to help determine when artifacts from off-scale data are most likely to occur. This threshold will then caution laboratories in interpreting this type of data.

1.7 Stochastic threshold

Amplification of samples containing low-level DNA may be subject to stochastic effects, where two alleles at a heterozygous locus exhibit considerably different peak heights (i.e., peak height ratio generally <60%) or an allele fails to amplify to a detectable level (i.e., allelic dropout). Stochastic effects within an amplification reaction may affect one or more loci irrespective of locus or allele size. A threshold can be applied to alert the DNA analyst that all of the DNA typing information may not have been detected for a given sample. This threshold, referred to as a stochastic threshold, is defined as the peak height value below which it is reasonable to assume that allelic dropout may have occurred within a single-source sample. This definition of a stochastic threshold may not be appropriate to the interpretation of mixtures when allele sharing is possible, including the effects of sharing amongst stutter and minor allele peaks; however, a stochastic threshold may still be useful in providing an indication that data may be missing (i.e., when data is present below that threshold).

Example: *This example illustrates how stutter can result in an increase in RFU signal for a minor contributor allele, bringing it above the stochastic threshold, making interpretation of the minor contributor in a mixture more difficult.*



Based upon the examination of the entire DNA profile, this mixture is assumed to be from two contributors, and is further interpreted based on peak heights (i.e., major/minor). In this example, while the peak in bin 11 is in stutter position from the major allele 12, it is above the laboratory's established stutter threshold and thus considered as an actual allele of the minor contributor. While it is above the laboratory's stochastic threshold, the possibility that stutter from allele 12 may be artificially inflating the detected RFU value of allele 11 must be considered. Dependent upon the RFU value of allele 11 and the stutter expectations at this locus, the interpretation of the minor contributor may need to include the possibility of an undetected allele. As such, a CPI or CPE calculation may be inappropriate, and the RMP and LR calculations for the minor contributor may need to account for the possibility of an undetected allele (e.g., $2P_{11}$).

1.7.1 A stochastic threshold must be based on empirical data derived within the laboratory and specific to the quantitation and amplification systems (e.g., kits) and the detection instrumentation used. The RFU value below which it is reasonable to assume that, at a given locus, allelic dropout of a sister allele may have occurred constitutes a stochastic threshold. It is noted that a stochastic threshold may be established by assessing allelic dropout and peak height ratios across multiple loci in a dilution series of DNA amplified in replicate.

1.7.1.1 If measures are used to enhance detection sensitivity (i.e., allelic height), the laboratory must perform additional studies to establish independent criteria for application of a separate stochastic threshold(s). Such measures may include but not be limited to increased amplification cycle number, increased injection time, and post-amplification purification/concentration of amplified products. Refer to the SWGDAM STR Guidelines for Enhanced Detection Methods for further

guidance. Instances where the laboratory decreases detection sensitivity, such as decreased injection time because of off-scale RFU values, may not require a separate stochastic threshold.

1.7.1.2 For samples where an assumption can be made as to the number of contributors, the laboratory should establish criteria for comparison of allelic peaks which fall below the stochastic threshold. As an example, if a locus in an assumed single-source sample exhibits two alleles, one or both of which are below the stochastic threshold, the laboratory may use that locus for comparison purposes. Also, the presence of male DNA may be established based on a Y-allele at amelogenin or a result at other male-specific loci that is below the stochastic threshold.

1.7.1.3 For samples exhibiting low-level data where an assumption as to the number of contributors cannot be confidently made, relying on the stochastic threshold alone may be insufficient to determine that all alleles have been detected. As an example, if a locus in an assumed mixture is exhibiting two alleles, both of which are above the stochastic threshold, the interpretation must consider the possibility that this locus is not fully representative of the genotypes of all of the contributors.

1.8 Peak Height Ratios

Intra-locus peak height ratios (PHR) can be calculated for a given locus by dividing the peak height of an allele with a lower RFU value by the peak height of an allele with a higher RFU value, and may be expressed as a percentage. PHR evaluation may be used to establish potential expectations for allele pairing to define genotypes for mixed samples.

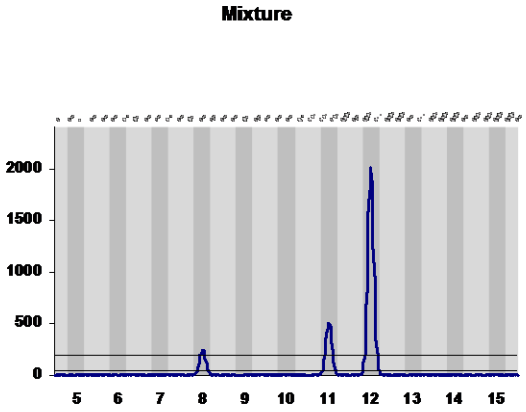
1.8.1 The laboratory shall establish PHR expectations based on empirical data derived from DNA typing results from single-source samples. Different PHR expectations can be applied to individual loci; alternatively, a single PHR expectation can be applied to all loci.

1.8.1.1 The laboratory should evaluate PHRs at various DNA template levels (e.g., a dilution series of DNA). It is noted that different PHR expectations at different peak height ranges may be established. As DNA template levels decrease, uncertainty in the ability to reliably define genotypes (allele pairs) increases, especially at or below the stochastic threshold.

1.8.2 With mixtures, stutter from a major allele(s) may cause an artificial imbalance in the PHR calculation results of minor genotypes.

1.8.3 With mixtures, contribution from a minor allele(s) may cause an artificial imbalance in the PHR calculation results of major genotypes due to allele stacking. As the number of contributors to a mixture increases, this effect may become more pronounced.

Example: *This example illustrates how stutter can result in an increase in RFU signal for a minor contributor allele making interpretation of the minor contributor in a mixture more difficult.*

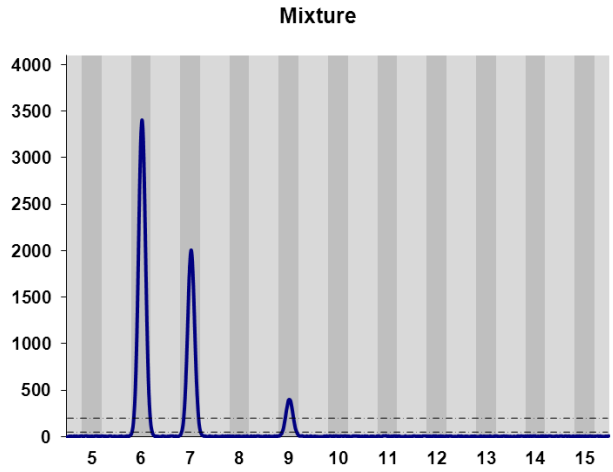


Bin	RFU
8	210
11	500
12	2000

Genotype	Peak height ratio
8,11 (no stutter considerations)	210 / 500 = 42%
8,11 (stutter consideration in bin 11 from peak in bin 12)	210 / [500 – (2000*10% expected stutter contribution)] = 70%

Example: *This example illustrates how alleles shared between contributors can result in an increase in RFU signal for a major contributor allele making interpretation of the major contributor in a mixture more difficult.*

Based upon the examination of the entire DNA profile, this mixture is assumed to be from three contributors, and is further interpreted based on peak heights (i.e., major/minor).



Bin	RFU
6	3400
7	2000
9	400

The two most significant alleles (bins 6 and 7) do not meet the laboratory's established 60% peak height ratio expectations for heterozygous genotypes (2000 / 3400 = 59%). However, this should not immediately discount the possibility of the most significant contributor to the mixture having genotype 6,7.

One way the mixture could be explained is with the following genotypes, using the assumption of perfectly balanced alleles in each genotype:

Genotype	RFU	
6,7	2,000	2,000
6,6	500	500
6,9	400	400

The laboratory's validation studies of mixtures containing three or more contributors should provide guidance towards such interpretations.

1.9 Number of contributors to a DNA Profile

In order to properly interpret a DNA profile, assumptions are necessary. An important assumption relates to the number of contributors.

For example, a sample is assumed to have originated from a single individual if one or two alleles are present at all loci for which typing results were obtained (although tri-allelic loci may occur), and the peak height ratios for all heterozygous loci are within the empirically determined values. It is noted that peak height imbalances may be seen in the typing results from, for example, a primer binding site variant that results in attenuated amplification of one allele of a heterozygous pair. Likewise, degraded, inhibited, and/or low level single-source DNA samples may exhibit poor peak height balance with heterozygous alleles.

A sample is assumed to have originated from more than one individual if three or more alleles are present at one or more loci (excepting tri-allelic loci) and/or the peak height ratios between a single pair of allelic peaks for one or more loci are below the empirically determined heterozygous peak height ratio expectation. Generally, the minimum number of contributors to a mixed sample can be determined based on the locus that exhibits the greatest number of allelic peaks. As an example, if at most five alleles are detected per locus, then the DNA typing results are consistent with having arisen from at least three individuals. Additional information such as peak heights, the total number of alleles detected across all autosomal loci (Perez et al. 2011), the known heterozygosity of loci (Buckleton et al. 2007), the assumed presence of a known contributor, and/or Y-STR data may further support the assumption of the number of contributors.

1.9.1 The laboratory should define the number of alleles per locus and the relative intra-locus peak height requirements for assessing whether a DNA typing result is consistent with originating from one or more sources. The minimum number of loci should be defined for determination of whether a sample is a mixture.

1.9.1.1 If the evidence is assumed to consist of two or more close biological relatives, utilizing allele counts may underestimate the true number of contributors to the evidence due to the high degree of allele sharing.

1.9.2 For DNA mixtures, the laboratory must establish guidelines for when a finite number of contributors can be assumed. The number of contributors should be supported by the data. All alleles above analytical threshold should be used in this assessment.

1.9.3 For DNA mixtures when a finite number of contributors cannot be assumed, the laboratory must establish guidelines for determination of the minimum number of contributors to a sample. All alleles above analytical threshold should be used in this assessment.

1.10 The laboratory shall establish guidelines to determine whether DNA typing results are suitable for comparisons. Results deemed unsuitable for comparisons include those that are uninterpretable (e.g., data of limited or poor quality) as well as those that do not meet quality assurance guidelines as defined by the laboratory.

1.11 Evaluating Replicate Typing Results

1.11.1 Where multiple amplifications and/or injections are generated for a given sample extract, the laboratory shall establish guidelines for determining which results are used for comparisons and statistical calculations. If composite profiles are used, the laboratory must establish guidelines for the generation of the composite result.

Section 2: Mixture Interpretation Overview and Strategies

Introduction:

The primary goal of mixture interpretation shall be to determine the possible genotype combinations of the contributors. The determination of contributing genotypes will allow for inclusions or exclusions based upon those genotypes and not merely the presence or absence of alleles in a mixture. Evaluating a mixture for the presence or absence of a potential contributor's alleles, although unavoidable in certain circumstances (e.g., indistinguishable mixtures), is an incomplete interpretation of the data.

The evaluation of a mixture as a whole will provide additional information, beyond an assessment of the potential number of contributors to the mixture (see Section 1), that may guide the interpretation. Additional factors that may be evaluated include the presence of stochastic level contributor(s), evidence of degradation or potential inhibition, whether the profile indicates major/minor contributor(s) or alternatively, indistinguishable contribution levels among contributors. Each of these factors may have an effect on the interpretation of the results.

An individual's contribution to a mixed biological sample is generally proportional to their quantitative representation within the DNA typing results. Accordingly, it may be possible to further resolve a mixture using the relative proportions of the various contributors (e.g., mixture ratios). If the amounts of biological material from multiple donors are similar, or if there are many contributors to a mixture, it may not be possible to further separate the mixture profile into genotypes of the contributing individuals. Intensity differences between alleles may be due to allele sharing. When contributor genotypes cannot be distinguished because of similar contribution levels, the sample is an indistinguishable mixture.

Some evidence items may be expected to yield DNA from the individual from whom the sample was taken. If another source of DNA is present in sufficient quantity in such a sample, a mixture of DNA is likely to be detected. Based on this expectation, any DNA typing results that match an expected contributor may be assumed and considered in the interpretation of the profile.

Interpreting DNA Typing Results for Mixed Samples

2.1 Criteria used for mixture interpretation: Any criteria used in the interpretation of a mixed DNA sample shall be supported by the data and shall be defined and documented. The laboratory shall establish guidelines regarding the application of assumptions to the data.

2.1.1 If assumptions are made as to the number of contributors, additional information such as the number of alleles at a given locus and their relative peak heights can be used to distinguish major and minor contributors.

2.1.2 If assumptions are made as to the presence of known contributor(s), the genotypes used in the deconvolution shall be documented. See section 2.5 for additional information.

2.1.3 A laboratory may define other quantitative characteristics of mixtures to aid in further refining the contributors. Validation studies shall be used to inform decisions regarding the relative contribution of each donor (e.g., mixture ratios).

2.1.3.1 Mixture ratios can vary from locus to locus, and laboratories using this approach should establish expectations from examination of validation data.

2.1.3.2 Based on the quality of the profile, differential degradation of the contributors to a mixture may impact the mixture ratio across the entire profile.

2.2 Two-person mixtures: The laboratory should establish guidelines based on peak height ratio assessments for evaluating potential sharing of allelic peaks between contributors and for determining whether the genotypes of the two contributors to a mixed DNA typing result are distinguishable. When assessing peak height ratios, pair-wise comparison of all potential genotypic combinations should be evaluated.

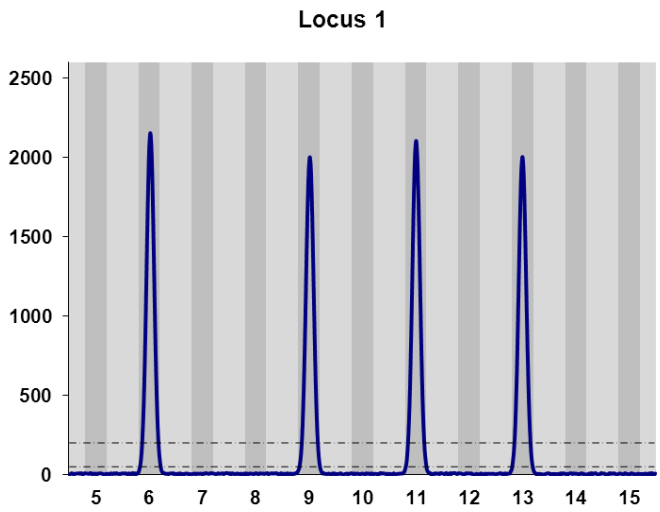
2.2.1 Distinguishable two-person mixtures: A distinguishable DNA mixture is one in which relative peak height ratios allow for the interpretation of the genotype of the major contributor and possibly the minor contributor. A distinguishable mixture contains a

distinct contrast in signal intensities (e.g., peak heights) between the different contributors' alleles. Discernment of the STR typing results for the major or minor contributor to a mixture may effectively constitute deduced single-source profiles, or in other instances may be limited to single genotypes at only some loci with the remaining loci yielding multiple potential genotypes for the major or minor contributor.

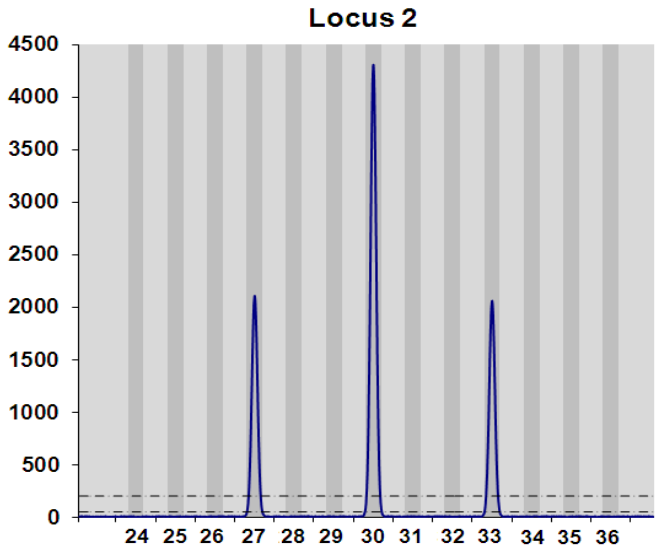
2.2.2 Indistinguishable two-person mixtures: If the amounts of biological material from the two donors are similar, it may not be possible to further refine the mixture profile into individual genotypes. However, an assumption as to the number of contributors can be used to limit genotype possibilities in some instances (e.g., a four-allele locus in a two-person mixture is not reasonably expected to have any homozygous genotypes contributing). Mixture profiles deemed indistinguishable may still be interpretable. Individuals may still be included or excluded as possible contributors to an indistinguishable mixture (see following example). The laboratory should establish guidelines for identifying mixtures for which no major or minor contributor can be discerned.

Example: *This example illustrates a method, using peak height information, which can be utilized to infer the genotypes of a mixture assumed to be from only two contributors.*

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors.



Bin	RFU
6	2150
9	2000
11	2100
13	2000



Bin	RFU
27	2100
30	4300
33	2050

Given the close peak height values of the four alleles detected at locus 1, without utilizing the assumption of a known contributor it is not possible to confidently define the unique genotypes of the two contributors.

However, using the assumption that only two contributors are present, the assumption that tri-allelic individuals are rare, and the interpretation that the mixture is an approximate 1:1 ratio of the two contributors, it is reasonable to conclude that anyone without one of the below listed genotypes for locus 1 should be excluded as a potential contributor to the mixture.

Possible genotype of contributor 1	Peak height ratio	Possible genotype of contributor 2	Peak height ratio
6,9	93%	11,13	95%
6,11	98%	9,13	100%
6,13	93%	9,11	95%
9,11	95%	6,13	93%
9,13	100%	6,11	98%
11,13	95%	6,9	93%

Furthermore, assuming all alleles are detected, if an individual is included in the interpretation results from locus 1, they must also be included in locus 2 in order to be included as a contributor to the mixture as a whole. Using a 1:1 ratio of contributors to determine the proportion of the RFU value of any shared allele to assign to each genotype, the following genotype combinations are available for consideration:

Possible genotype of contributor 1	Peak height ratio	Possible genotype of contributor 2	Peak height ratio	Peak height ratio expectations met?
27,27	n/a	30,33	48%	No
27,30	24%	27,33	51%	No
27,30	98%	30,33	95%	Yes
27,30	49%	33,33	n/a	No
27,33	51%	27,30	24%	No
27,33	98%	30,30	n/a	Yes
27,33	49%	30,33	24%	No
30,30	n/a	27,33	98%	Yes
30,33	48%	27,27	n/a	No
30,33	95%	27,30	98%	Yes
30,33	24%	27,33	49%	No
33,33	n/a	27,30	49%	No

If the laboratory's validated minimum peak height ratio expectations are 60% for heterozygous genotypes, then several of the locus 2 combinations of potential genotypes would be excluded as contributing to the mixture (listed as No in column 5).

Therefore, combining the information from locus 1 with locus 2, many possible genotype combinations exist for each of the contributors. After a thorough interpretation has been completed, there are several genotype combinations that are encompassed by the simple presence of the detected alleles (such as a homozygote 11,11 or 6,6), that are not reasonable to assume for either contributor.

Possible genotype of contributor 1		Possible genotype of contributor 2	
Locus 1	Locus 2	Locus 1	Locus 2
6,9	27,30	11,13	30,33
6,9	27,33	11,13	30,30
6,9	30,30	11,13	27,33
6,9	30,33	11,13	27,30
6,11	27,30	9,13	30,33
6,11	27,33	9,13	30,30
6,11	30,30	9,13	27,33
6,11	30,33	9,13	27,30
6,13	27,30	9,11	30,33
6,13	27,33	9,11	30,30
6,13	30,30	9,11	27,33
6,13	30,33	9,11	27,30
9,11	27,30	6,13	30,33
9,11	27,33	6,13	30,30
9,11	30,30	6,13	27,33
9,11	30,33	6,13	27,30
9,13	27,30	6,11	30,33
9,13	27,33	6,11	30,30
9,13	30,30	6,11	27,33
9,13	30,33	6,11	27,30
11,13	27,30	6,9	30,33
11,13	27,33	6,9	30,30
11,13	30,30	6,9	27,33
11,13	30,33	6,9	27,30

2.3 Greater than two-person mixtures: Interpretation guidelines for mixtures must be based on mixture studies conducted using known contributors that represent the number of contributors and the range of general mixture types (e.g., mixture proportions and template quantities) for which the procedure will be used in casework. If a laboratory will be interpreting mixtures containing stochastic level data, the validation studies on which the interpretation guidelines are based should contain mixtures with stochastic level data.

2.3.1 Laboratories may establish criteria for determining which peaks are of such greater amplitude that it is reasonable they represent or contain the genotype of a major contributor to the mixture.

2.3.2 Where a clear major contributor does not exist in mixtures of greater than two people, further determination of definitive contributor genotypes may not be possible. The laboratory should establish guidelines based on peak height ratio assessments for identifying mixtures for which no major or minor contributors can be discerned. Probabilistic genotyping may be helpful in these instances.

2.3.3 When multiple major contributors are present in a mixture of more than two people, the laboratory may establish guidelines to resolve the portion of the mixture that contains the DNA types from the multiple major contributors. The laboratory should establish guidelines based on peak height ratio assessments, number of contributors, and/or mixture ratios for determining whether multiple major contributors are present and distinct from low level data in a mixed sample. The laboratory guidelines should be sufficiently detailed to ensure confidence in the separation of the "major" versus "minor" components. The interpretation of the "major mixture" may be performed regardless of the interpretation of the "minor portion" of the mixture (i.e., the "minor portion" may be uninterpretable).

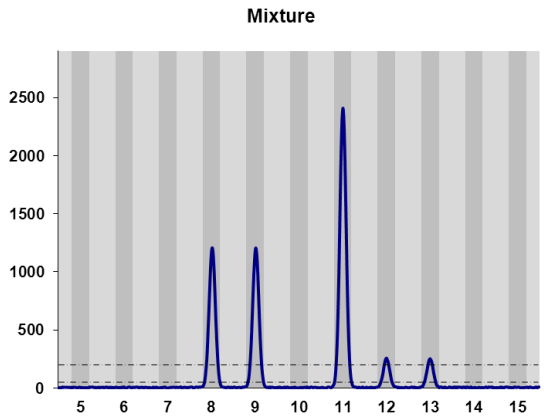
Example: *This example illustrates a method, using peak height information, that can be utilized to distinguish a major mixture of two or more contributors, separate and distinct from minor contributor(s). This is just one example of how this separation may be determined and other mechanisms may be used if supported by validation data.*

Based upon the examination of the entire DNA profile, the mixture example below is assumed to have originated from only three contributors. While there are 210 possible genotype combinations of three contributors that could produce a five-allele mixture, given the peak height information, many of these combinations can easily be discarded as unreasonable. For instance, assuming one contributor to be genotype 8,9, the second contributor 11,12, and the third contributor homozygous 13,13 would only be possible if the second contributor had a heterozygous peak height ratio of 10.4% (250/2400).

The only four reasonable genotype combinations would be:

Contributor 1	Contributor 2	Contributor 3
8,9	11,11	12,13
11,11	8,9	12,13
8,11	9,11	12,13
9,11	8,11	12,13

The ratio of the three contributors would therefore be 4.8 : 4.8 : 1



Bin	RFU
8	1200
9	1200
11	2400
12	250
13	250

Based upon this interpretation, a mixture calculation statistic (e.g., Likelihood Ratio, modified RMP, or CPE) can be applied to the two contributors in the major mixture, and a separate statistic (e.g., RMP) can be applied to the single minor contributor.

It must be noted however that the more contributors that are present in the mixture, the more difficult the interpretation becomes to confidently parse out alleles and possible genotypes of the major contributors as opposed to the additive effects of multiple contributors.

If the mixture were actually comprised of five contributors, one possible combination of genotypes would be:

	Contributor 1	Contributor 2	Contributor 3	Contributor 4	Contributor 5
Genotype	8,11	11,11	9,9	9,12	9,13
RFU	1200, 1200	1200	600	300, 250	300, 250
PHR	100%	N/A	N/A	83%	83%

The ratio of the five contributors would therefore be 4 : 2 : 1.1 : 1 : 1

This is meant to demonstrate that the number of contributors that are interpreted to be present may have an effect on the assignment of genotypes. Defining the alleles 8, 9, and 11 as the major mixture would therefore generate a statistic that would not be reflective of the true composition of the mixture.

2.4 Mixtures with a Single Major Contributor and One or More Minor Contributors

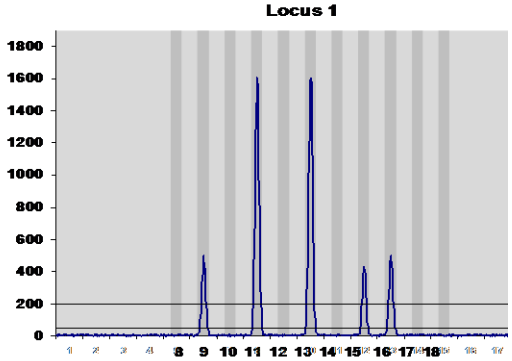
2.4.1 In general, heterozygous alleles attributed to a major contributor should meet the laboratory's established peak height ratio expectations for single-source samples. Minor contributor alleles may be shared with one or more heterozygous alleles from the major contributor at various loci throughout a mixed DNA profile. As a consequence of this allele stacking, heterozygous peaks from a major contributor may be outside established peak height ratio expectations. Laboratories may use other quantitative means of determining possible major and minor genotypes (e.g., mixture ratios).

Example: *This example illustrates how caution is required when interpreting a major contributor versus the additive effects of multiple contributors sharing alleles.*

While one or more loci in a mixture may appear to originate from a single major contributor and one or more minor contributors, the entire electropherogram should be considered prior to making this interpretation.

Based upon the examination of the entire DNA profile, the mixture in this example is determined to have originated from at least three contributors.

Locus 1 of this mixture has alleles 9, 11, 13, 15, and 16 all above the stochastic threshold.

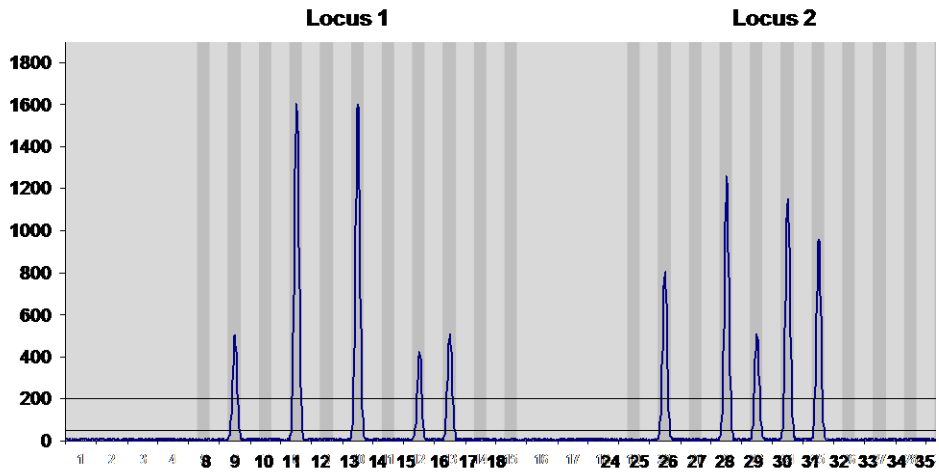


Bin	RFU
9	500
11	1600
13	1600
15	420
16	500

While the 11 and 13 alleles demonstrate the highest peaks and meet the peak height ratio expectation for a heterozygous genotype, potential additive effects of the contributors must be considered.

It is possible that results from this locus represent a mixture of three individuals, comprised of a major contributor (11,13) and two minor contributors (who possess the 9, 15 and 16 alleles at a minimum). However, it is also possible that the mixture is comprised of four individuals with genotypes as follows: 11,11 and 9,13 and 13,15 and 13,16 at a close ratio of contributors.

The information at all loci across the profile must be used in evaluating the number of contributors and in determining their relative proportions.



Bin	RFU
26	800
28	1250
29	500
30	1150
31	950

Locus 2 of this mixture has alleles 26, 28, 29, 30, and 31 all above the stochastic threshold.

Using the information from Locus 2, it appears as though there is no distinct major contributor to the mixture.

Since the interpretation of the mixture as a whole does not support a clear major genotype, interpreting the presence of a major contributor of genotype 11,13 at Locus 1 would not be appropriate.

2.4.2 After deconvolution, the DNA typing results attributed to a single minor contributor should also meet PHR expectations. The PHR expectations of a minor contributor may be reduced due to stochastic peak height variation and the additive effects of peak sharing (e.g., minor peak and stutter peaks).

2.4.3 Determination of a single genotype for a minor contributor may be possible at only some loci because multiple possible genotypic combinations, potential allelic dropout, and/or masking of the minor contributor's alleles by those of the major contributor or by stutter from the major contributor's alleles preclude such determination at other loci. Probabilistic genotyping may be helpful in these situations.

2.5 Mixtures with an Assumed Contributor(s): The laboratory should establish guidelines for determining whether separation of an assumed contributor's profile is applicable (e.g., based on the types of evidentiary items). The obligate foreign alleles may effectively constitute a single-source profile, if there is one DNA contributor in addition to the individual from whom the sample was taken, or a mixture profile if there are multiple additional DNA contributors.

2.5.1 At a minimum, when there is no indication of sharing of the assumed and foreign alleles, the laboratory should designate obligate alleles of the foreign contributor.

2.5.2 To further determine the obligate alleles in a profile, the laboratory may establish guidelines (e.g., peak height ratio thresholds or mixture ratios) for addressing potential sharing of alleles among the individual assumed to have contributed to a sample and the additional contributor(s).

2.6 Interpretation of Potential Stutter Peaks in a Mixed Sample

2.6.1 For mixtures in which minor contributors are determined to be present, a peak in stutter position may be determined to be 1) a stutter peak, 2) an allelic peak, or 3) indistinguishable as being either an allelic and/or stutter peak. This determination is based principally on the height of the detected peak that is in the stutter position, its height relative to other peak heights in the mixture, and its relationship to the stutter percentage expectations established by the laboratory.

2.6.2 Generally, when the height of a peak in the stutter position exceeds the laboratory's stutter expectation for a given locus, that peak is consistent with being of allelic origin and should be designated as an allele.

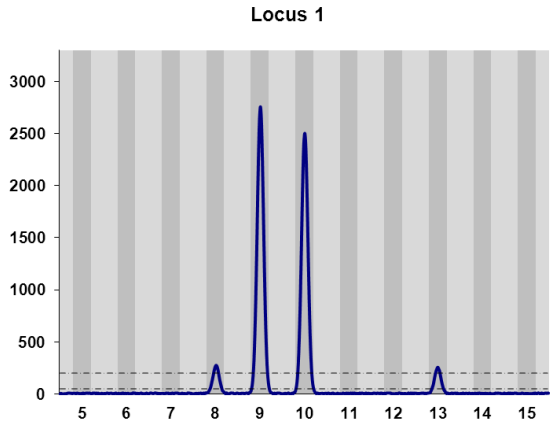
2.6.3 If a peak is at or below the stutter threshold there are two possibilities: either the peak is completely stutter or it is a composite of stutter product and an allele from another contributor. The peak should be considered as potentially a composite of stutter and allele if the peak height of the potential stutter peak(s) is consistent with (or greater than) the heights observed for any allelic peaks that are conclusively attributed (i.e., peaks in non-stutter positions) to the minor contributor(s). The designation of stutter peaks as possible allelic peaks is unnecessary at a locus if the consideration of stutter peaks would violate the documented assumption as to the number of contributors to the mixture (e.g., a two-person distinguishable mixture with four alleles).

Example: *This example illustrates how peaks in stutter positions but below the stutter percentage expectation might need to be considered as true alleles of the minor contributor.*

Based upon the examination of the entire DNA profile, this mixture is assumed to be from only two contributors. The person of interest (POI) is considered as a possible minor contributor.

Locus 1 demonstrates a clear separation of a major and minor contributor. The major contributor can readily be assumed to be genotype 9,10. The minor contributor must have allele 13 within his or her genotype. Since the allele 13 is above the stochastic threshold and not in a stutter position bin it is reasonable to

assume the minor contributor is not experiencing allelic dropout at this locus. The peak in bin 8 was filtered by the analysis software as being below the laboratory's 10% stutter filter for this locus; this peak meets the expectations of being simply stutter.



Bin	RFU
8	270 (below the stutter threshold)
9	2750
10	2500
13	250

Assuming 100% peak height ratio balance and that the peak in bin 8 is only stutter, there are three possible genotypes for the minor contributor.

	Contributor 1	Contributor 2
Genotype	9,10	13,13
RFU	2750, 2500	250
PHR	91%	N/A
Ratio of contributors	21 : 1	
Genotype	9,10	9,13
RFU	(2750-250), 2500	250, 250
PHR	100%	100%
Ratio of contributors	10 : 1	
Genotype	9,10	10,13
RFU	2750, (2500-250)	250, 250
PHR	82%	100%
Ratio of contributors	10 : 1	

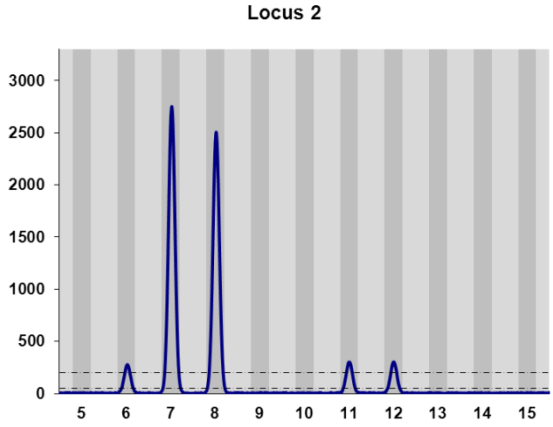
Since it is impossible to know exactly how much stutter occurred in this amplification from bin 9 into bin 8, it is possible that the peak in bin 8 may be comprised of a combination of stutter and of detected signal from the minor contributor. In this example, that decision may be informed by the PHR between the 8 and the unambiguous 13.

Because the peak in bin 8 may not be stutter, there is one additional possible genotype for the minor contributor.

	Contributor 1	Contributor 2
Genotype	9,10	8,13
RFU	2750, 2500	270, 250
PHR	91%	93%
Ratio of contributors	10 : 1	

While the possibility of peaks of the minor contributor being masked by stutter is a legitimate concern, not every peak in every stutter position should ultimately be designated as indistinguishable from stutter.

Locus 2, of the same mixture as locus 1 in this example, again demonstrates a clear separation of a major and minor contributor. The major contributor can readily be assumed to be genotype 7,8. The minor contributor can readily be assumed to be genotype 11,12. Since two unambiguous alleles are designated for the genotype of the minor contributor it is unnecessary to assume (given the assumption that tri-allelic genotypes are rare) that the minor contributor is masked by the major contributor or by stutter peaks of the major contributor, at this locus. The peak in bin 6 was filtered by the analysis software as being below the laboratory's 10% stutter filter for this locus; this peak meets the expectations of being simply stutter.

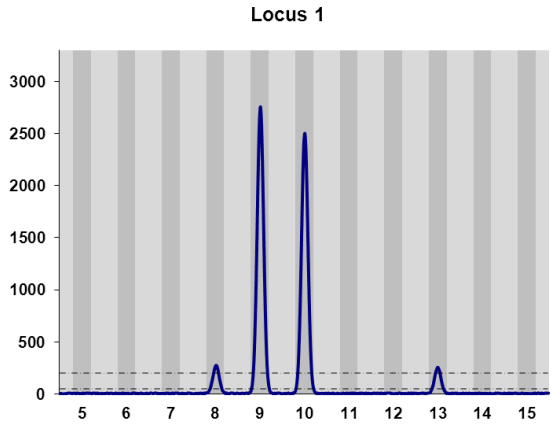


Bin	RFU
6	270 (below the stutter threshold)
7	2750
8	2500
11	250
12	250

Example: *This example illustrates how peaks in stutter positions, below the stutter threshold, might not need to be considered as true alleles of the minor contributor, based upon assumptions used in the interpretation.*

Other assumptions used in the interpretation of the evidence may also be utilized in the interpretation of peaks in stutter position as being defined as stutter or as indistinguishable from stutter.

As an example, if the evidence sample is a vaginal swab from the victim of a sexual assault it may be reasonable to assume the presence of her DNA on the item. If the assumed known contributor is the minor contributor to the two-person mixture, comparing their known reference profile to the mixture may discount a peak in a stutter bin as being indistinguishable from stutter.



Bin	RFU
8	270 (below the stutter threshold)
9	2750
10	2500
13	250

Knowing the victim's DNA profile at this locus is 9,13 it would be unreasonable to assume that the peak in bin 8 is a true allele of a minor contributor as this would violate the assumptions of the mixture being only two contributors with the victim being the minor contributor. As such, the peak in bin 8 could reasonably be reported as meeting the expectations of being simply stutter.

Section 3: Comparison of References and Statistical Weight of Probative Inclusions

Introduction:

Once the data has been evaluated and potential genotype combinations for contributors have been determined, comparison of reference samples can be made. With any relevant inclusions, statistical calculations are performed following proper interpretation on evidentiary DNA profiles to provide an assessment of the significance of an inclusion. The simple presence or absence of alleles may be insufficient to draw a proper conclusion due to the possibility of allelic drop-out. The interpretation of the evidentiary profile should determine the statistical approach used. The choice of a statistical approach should not drive the interpretation. It would be inappropriate to make inclusions or exclusions based on the statistical approach without first considering the interpretation of the profile.

There are two general approaches for determining the statistical weight of probative inclusions: (1) binary (threshold-based concept where alleles and genotypes for a contributor are either present or absent) and (2) probabilistic genotyping (PG). Both approaches (binary and PG), when properly validated and implemented, are scientifically sound, and appropriately convey the weight of the evidence. The choice of approach is influenced by the laboratory's current protocols, validation, and available resources. Laboratories are encouraged to evaluate their current casework and the complexity of mixtures encountered to determine the best approach to employ when evaluating the weight for any inclusion.

For any included individual, all of his/her genotypes should be considered both in the interpretation and in the statistical calculation. A statistical calculation more accurately reflects the interpretation if it does not include extraneous genotype combinations as is done in unrestricted computational models. For example, if an interpretation excludes particular genotype combinations, the most appropriate statistical calculation would not include those genotypes.

3.1 Comparison of DNA Typing Results

Using the binary model, the following determinations can be made upon comparison of evidentiary and known DNA typing results (and between evidentiary samples):

- The known individual cannot be excluded (i.e., is included) as a possible contributor to the DNA obtained from an evidentiary item.
- The known individual is excluded as a possible contributor.
- The comparison is inconclusive.
- The DNA typing results from multiple evidentiary items are consistent or inconsistent with originating from a common source(s).

3.1.1 The laboratory must establish guidelines to ensure that, to the extent possible, DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors.

3.1.2 DNA typing results may not be obtained at all loci for a given evidentiary sample (e.g., due to DNA degradation, inhibition of amplification and/or low-template quantity); a partial profile thus results.

3.1.2.1 For partial profiles, the determination of which alleles/loci are suitable for comparison and statistical analysis should be made prior to comparison to the known profiles.

3.1.2.2 The laboratory should establish guidelines for inclusions and exclusions when a known individual's DNA profile is not fully observed in the evidentiary profile.

3.1.3 The laboratory must establish guidelines for inclusionary, exclusionary, and inconclusive determinations.

3.1.3.1 Comparisons (e.g., inclusions/exclusions) should not simply be based on the presence or absence of alleles for a particular contributor at a given locus, but instead be based on the possible genotype(s) that are being evaluated for each

contributor at each locus while considering the overall quality of the profile (e.g., degradation, preferential amplification, inhibition, drop-out) and any assumptions being made with the interpretation.

3.1.4 For mixtures for which two or more individuals cannot be excluded as potential contributors, the laboratory may establish guidelines for assessing whether all of the DNA typing results obtained from the mixed sample are accounted for by the multiple known samples.

3.1.5 Because assumptions regarding the evidence (e.g., intimate sample, number of contributors) can impact comparisons, the laboratory should establish guidelines for documenting any assumptions that are made when formulating conclusions.

3.1.6 A verbal scale or predicate may be used by a laboratory to give context to the weight of the evidence based on the associated number. This association of words with numbers may serve as a guide to further assist the court with evaluating the weight of the evidence.

3.2 Requirements for reporting statistics

3.2.1 Except for a reasonably assumed contributor, the laboratory shall perform statistical analysis in support of any inclusion (or a "cannot be excluded" conclusion) irrespective of the number of alleles detected and the quantitative value of the statistical analysis.

3.2.1.1 The laboratory should establish reporting guidelines where multiple stains from the same or separate items have provided genetic information that is consistent with originating from a common source(s) but having various levels of discrimination.

3.2.2 The genetic loci and assumptions used for statistical calculations must be documented, at a minimum, in the case notes.

3.2.3 Data that cannot be used to support inclusions shall not be used in statistical analysis at individual loci or for an entire multi-locus profile.

3.2.3.1 For a distinguishable mixture, a major contributor(s) profile may be suitable for statistical analysis even in the presence of minor contributor results that are unsuitable for comparison.

3.2.4 The laboratory must document the source of the population database(s) used in any statistical analysis.

3.2.5 The formulae used in any statistical analysis must be documented and must address both homozygous and heterozygous typing results, multiple locus profiles, mixtures, minimum allele frequencies, and, where appropriate, biological relationships.

3.2.5.1 Given a profile for which multiple statistical methods may be applicable (e.g., CPI, RMP, or LR), the laboratory should have guidelines for the selection of the method and suitable formula(e) to be used for statistical application. These multiple statistical methods shall not be combined into one calculation for the given profile. For example, the LR and CPI cannot be multiplied across loci in the statistical analysis of an individual DNA profile; the CPI is a probability, while the LR is a ratio of two conditional probabilities. Additionally, the LR and CPI utilize the interpreted number of contributors differently.

3.2.6 Exclusionary conclusions do not require statistical analysis. Likewise, statistics on the evidentiary sample are not required when no comparison is made to a known sample.

3.2.7 Statistical models commonly used for estimating DNA profile rarity typically involve an assumption of unrelated individuals. Based on possible case scenarios and the question posed, it may be necessary to consider and compute statistical calculations where specific biological relationships are assumed.

3.3 Binary Statistical Models

Binary models are useful when a clear contributor can be deconvoluted, thereby eliminating unreasonable genotype combinations based on quantitative information (i.e., peak height ratio expectations and/or mixture ratios). These approaches, when established using appropriate empirical data, may provide consistent and objective interpretations. They typically assume unrelated individuals, and include the random match probability (RMP), likelihood ratio (LR), and the combined probability of exclusion/inclusion (CPE/CPI) for providing weight to the interpretation.

3.4 Statistical formulae

3.4.1 There are two statistical calculations commonly used in forensics for evaluating the weight of evidence when results from an evidentiary DNA profile are consistent with the DNA profile from an individual in question: the profile probability and the match probability.

3.4.2 The profile probability determines the rarity of the observed profile in a particular population group. This method is described in National Research Council (NRC) II recommendation 4.1, and corrects for any two alleles that may be identical by descent within an individual using an F value (commonly referred to in practice as θ) for homozygous genotypes.

3.4.2.1 When the interpretation is based upon the assumption of a single contributor (or a single major or single minor contributor to a mixture), the RMP formulae are those described in NRC II recommendations 4.1, 4.3, and 4.4.

3.4.2.2 For heterozygote genotypes, the formula is 2pq. This formula is NRC II 4.1b.

3.4.2.3 For homozygote genotypes, the formula is $p^2 + p(1-p)\theta$, where typically $\theta = 0.01$ (for most U.S. groups) or 0.03 (for some isolated populations) as recommended by NRC II. This formula is NRC II 4.4a.

3.4.2.4 For single-allele loci where the zygosity is in question (e.g., it falls below the stochastic threshold), the formula 2p, as described in recommendation 4.1 of NRC II, may be applied to this result. Alternatively, the formulae $2p - p^2$ or $p^2 + 2p(1-p)$ may be used to address this situation without double-counting the proportion of homozygotes in the population.

3.4.3 The match probability, described by the sampling formulae, conditions the probability for the observed profile in a particular population group given that it has already been observed at least once (i.e., matches the individual in question). This method is described in NRC II recommendation 4.2, and corrects for any two alleles that may be identical by descent between two individuals using an F value (commonly referred to in practice as θ). These conditional probability formulae are listed in NRC II 4.10a and 4.10b and shown below:

$$\Pr(aa \mid aa) = \frac{(2\theta + (1-\theta)P_a)(3\theta + (1-\theta)P_a)}{(1+\theta)(1+2\theta)}$$

$$\Pr(ab \mid ab) = \frac{2(\theta + (1-\theta)P_a)(\theta + (1-\theta)P_b)}{(1+\theta)(1+2\theta)}$$

3.4.4 Laboratories should determine which statistical calculations are appropriate to use for calculating genotype frequencies based on the population groups they are reporting and any sub-population effects that may exist.

3.4.5 Any locus deemed unsuitable for comparison or inconclusive shall not be used for statistical weight, which effectively assigns the locus a value of 1.

3.4.6 If a laboratory uses source attribution statements, then it must establish guidelines for the criteria on which such a declaration is based. This typically involves establishing a statistical threshold within a specific population group of size N, based on a confidence level (1-α). The general formula for random match probability calculations is $(1 - \text{RMP})^N \geq 1 - \alpha$.

If these calculations are equal to or greater than the confidence level, then the contributor who is included may be concluded, with a high degree of confidence, to be the source of that profile. Uncertainty in the RMP can be accommodated by using a value that has been adjusted by 10 fold as supported by NRC II.

Section 4: Statistics

Three approaches are outlined with explanatory text and examples. These approaches are random match probability (RMP – section 4A), likelihood ratio (LR – section 4B), and the combined probability of inclusion/exclusion (CPI/CPE – section 4C).

Section 4A: The Random Match Probability (RMP)

Introduction:

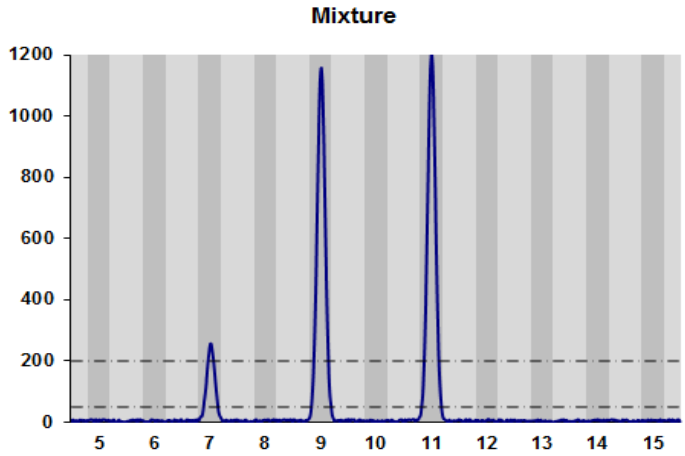
The Random Match Probability (RMP) is the probability that the DNA of a randomly chosen person has the same profile as the DNA of an evidentiary sample. While the RMP is commonly thought of in terms of single-source profiles, this formula also applies to mixture calculations where the number of contributors is assumed. At a locus, the RMP calculation (also referred to as a "modified RMP" or "mRMP") is the sum of the probabilities for all of the genotypes that represent possible contributors to a DNA mixture under the assumption of a defined number of contributors. The RMP is distinguished from the CPI that does not utilize an assumption as to the number of contributors.

A limited set of examples is provided below which includes explanations of:

- the fundamentals of the RMP statistic as it applies to mixtures
- the difference between unrestricted and restricted calculations
- the inclusion of the possibility of undetected data (allelic dropout)
- the inclusion of peaks indistinguishable from stutter

4A.1 When the interpretation of a mixture is conditioned upon the assumption of a particular number of contributors, the RMP is the sum of the individual probabilities for the genotypes included following a mixture deconvolution.

Example: *This example illustrates a fundamental concept of the RMP approach - the mixture is interpreted under the assumption of a specified number of contributors, and the RMP is the sum of the individual probabilities for the included genotypes. In this case, the number of contributors is assumed to be two, one of whom is assumed, and the RMP for the second contributor is the sum of the individual probabilities of the possible genotypes for that contributor.*

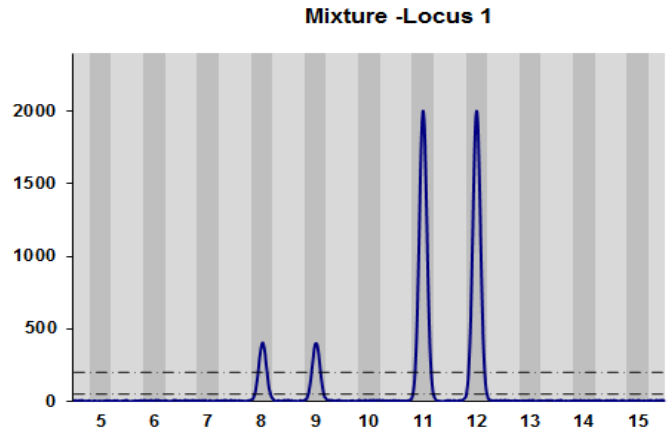


Bin	RFU
7	250
9	1150
11	1200

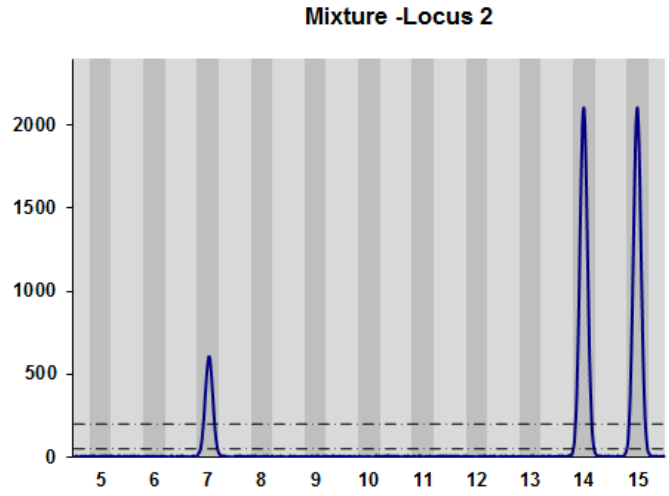
In a sperm fraction mixture (at a locus having alleles 7, 9, and 11) assumed to be from two contributors, one of whom is the victim (having genotype 9,11), the sperm contributor (minor) genotypes included post-deconvolution might be 7,7, 7,9, and 7,11. In this case, the RMP for the sperm DNA contributor could be calculated as $[P_7^2 + P_7(1-P_7)\theta] + 2P_7P_9 + 2P_7P_{11}$.

4A.2 The RMP may be used to combine multiple possible genotypes for a contributor at each locus, and the sums are then multiplied across loci.

Example: *This example illustrates a fundamental concept of the RMP approach - the mixture is interpreted under the assumption of a specified number of contributors, and the RMP is the sum of the individual probabilities for the included genotypes at each locus. In this case, the number of contributors is assumed to be two, one of whom is assumed, and the RMP for the second contributor is the product of the sums of the individual probabilities of the possible genotypes at each locus for that contributor.*



Locus 1	
Bin	RFU
8	400
9	400
11	2000
12	2000

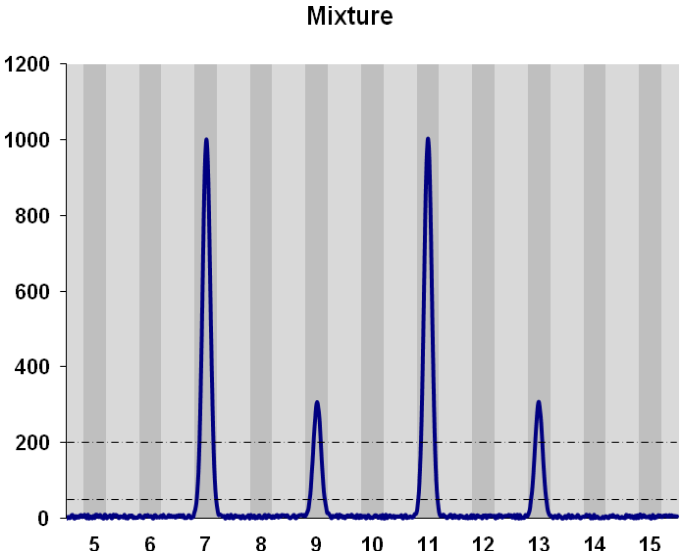


Locus 2	
Bin	RFU
7	600
14	2100
15	2100

In a sperm fraction assumed to be from two contributors, one of whom is the victim, the sperm contributor (minor) genotypes included post-deconvolution might include only a single genotype (8,9) at Locus 1, but multiple possible genotypes (7,7 or 7,14 or 7,15) at Locus 2. In this case, the two-locus RMP for the sperm DNA contributor could be calculated as $2P_8P_9 * [P_7^2 + P_7(1-P_7)\theta + 2P_7P_{14} + 2P_7P_{15}]$.

4A.3 In addition to assumptions of the number of contributors, quantitative peak height information and mixture ratio/proportion assessments may be included in the interpretation and statistical support for an inclusionary statement of an individual in an evidentiary profile. Calculations performed using interpretations incorporating this information are termed "restricted." When this quantitative peak height information is not included, the resultant calculation is termed "unrestricted."

Example: *This example illustrates the difference between the unrestricted and the restricted RMP, under the assumption of two unknown donors.*



Bin	RFU
7	1000
9	300
11	1000
13	300

For the unrestricted RMP, relative peak height differences are not utilized, so all two-allele combinations using two of the four alleles at this locus are included as possible genotypes. The possible genotype pairs are:

First Contributor	Second Contributor
7,9	11,13
7,11	9,13
7,13	9,11
9,11	7,13
9,13	7,11
11,13	7,9

For this unrestricted interpretation, the six possible genotypes are 7,9; 7,11; 7,13; 9,11; 9,13 and 11,13. The RMP would be calculated as:

$2P_7P_9 + 2P_7P_{11} + 2P_7P_{13} + 2P_9P_{11} + 2P_9P_{13} + 2P_{11}P_{13}$

For the restricted RMP, relative peak height differences are utilized. Given the peak heights displayed, the only reasonable pairings are the 7 allele with the 11 allele, and the 9 allele with the 13 allele. Therefore, the genotypes of the contributors are:

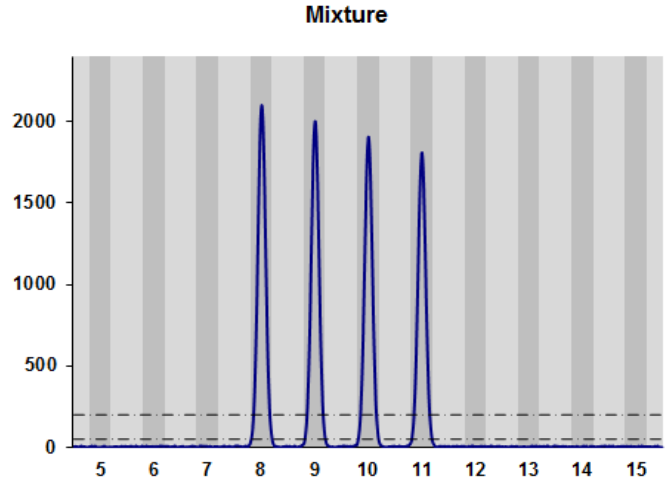
Major Contributor	Minor Contributor
7,11	9,13

For this restricted interpretation, the RMP for the major contributor would be calculated as $2P_7P_{11}$, and the RMP for the minor contributor would be calculated as $2P_9P_{13}$.

This example clearly indicates the value of incorporating peak height information into the determination of the RMP – the list of genotypes included in the statistic is narrowed down, while the list of genotypes excluded from the statistic is increased.

4A.4 The unrestricted RMP might be calculated for mixtures that display no indications of allelic dropout. The formulae include an assumption of the number of contributors, but relative peak height information is not utilized.

Example: *This example illustrates that for two-person mixtures, for loci displaying four alleles, homozygous genotype frequencies would not typically be included in the unrestricted RMP calculation. This is in contrast to the CPI calculation, which would include the homozygous genotype frequencies.*



Bin	RFU
8	2100
9	2000
10	1900
11	1800

While the CPI calculation for this locus would be $(P_8 + P_9 + P_{10} + P_{11})^2$, the unrestricted RMP calculation, under the assumption of two contributors, would require the subtraction of the homozygous genotype probabilities i.e., $(P_8 + P_9 + P_{10} + P_{11})^2 – P_8^2 – P_9^2 – P_{10}^2 – P_{11}^2$.

Note that for two-person mixtures, the formulae for loci displaying one, two, or three alleles are identical to the CPI calculations discussed in section 4C.

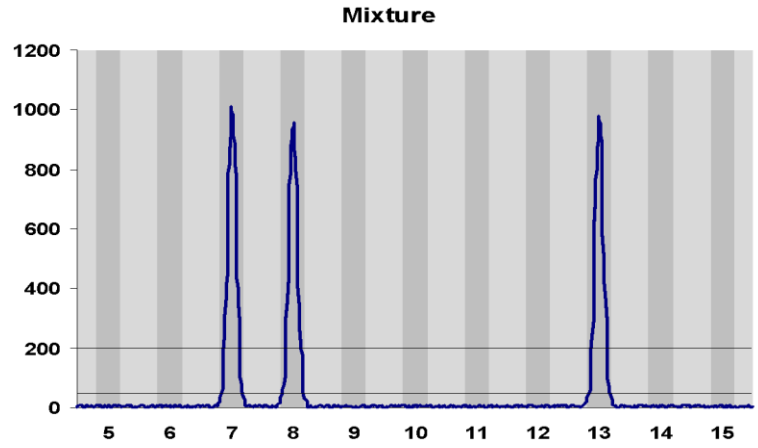
4A.5 The restricted RMP might be calculated for mixtures that display no indications of allelic dropout. The formulae include an assumption of the number of contributors and relative peak height information is utilized.

Example: *This example illustrates the use of the restricted RMP at a locus displaying three alleles, with no reasonable assumption of allelic dropout.*

Given the presence of alleles 7, 8, and 13 well above the stochastic threshold in a mixture assumed to be comprised of only two contributors, the interpretation could be a heterozygote-homozygote pairing or a heterozygote-heterozygote pairing where one allele is shared. Therefore, without consideration of the RFU values, the following genotype combinations are possible:

First Contributor	Second Contributor
7,7	8,13
7,8	7,13
7,8	8,13
7,8	13,13
7,13	7,8
7,13	8,8
7,13	8,13
8,8	7,13
8,13	7,7
8,13	7,8
8,13	7,13
13,13	7,8

However, upon examination of the electropherogram, it is seen that the RFU values of the three alleles are generally equivalent.



Bin	RFU
7	1000
8	950
13	975

Therefore, it is unreasonable to assume that the two contributors share a common allele and each has an unshared allele. In such mixtures, the shared allele would be expected to be detected at an RFU value equivalent to the sum of the RFU values of the unshared alleles. For example, if this mixture were comprised of genotypes 7,8 and 8,13, the peak in bin 8 would be expected to be approximately 2,000 RFU.

Using these assumptions, the list of potential genotypes can be reduced as such:

First Contributor	Second Contributor
7,7	8,13
7,8	13,13
7,13	8,8
8,8	7,13
8,13	7,7
13,13	7,8

Applying the RFU values to the list of potential genotypes shows that the two contributors are present in an approximately 2:1 ratio of contributors:

First Contributor	Second Contributor
7,7 (1,000 RFU)	8,13 (1,925 RFU)
7,8 (1,950 RFU)	13,13 (975 RFU)
7,13 (1,975 RFU)	8,8 (950 RFU)
8,8 (950 RFU)	7,13 (1,975 RFU)
8,13 (1,925 RFU)	7,7 (1,000 RFU)
13,13 (975 RFU)	7,8 (1,950 RFU)

Defining the First Contributor as the major contributor, the list of potential genotypes is further reduced:

Major Contributor	Minor Contributor
7,8 (1,950 RFU)	13,13 (975 RFU)
7,13 (1,975 RFU)	8,8 (950 RFU)
8,13 (1,925 RFU)	7,7 (1,000 RFU)

Only individuals who have genotype 7,8 or 7,13 or 8,13 would be included as the major contributor. All others would be excluded, even if their alleles are present within the mixture.

The RMP for this interpretation of the major contributor must account for all heterozygotes represented by alleles 7, 8, and 13, and would be calculated as:
$2P_7P_8 + 2P_7P_{13} + 2P_8P_{13}$

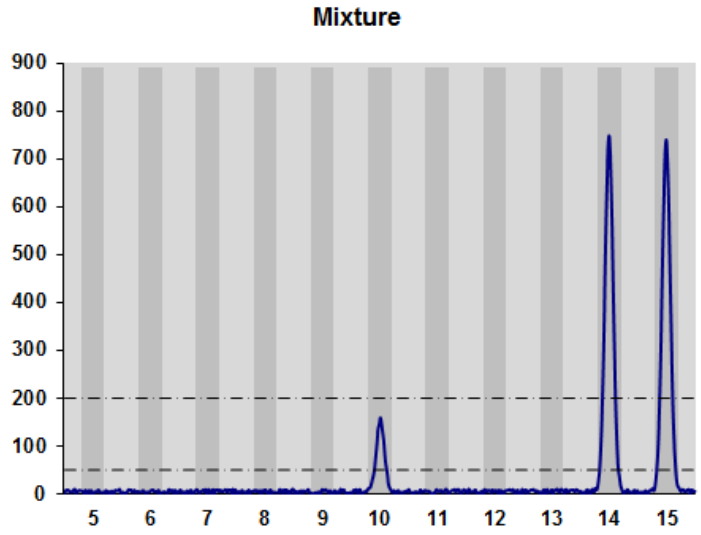
Only individuals who have genotype 7,7 or 8,8 or 13,13 would be included as the minor contributor. All others would be excluded, even if their alleles are present within the mixture.

The RMP for this interpretation of the minor contributor must account for all homozygotes represented by alleles 7, 8, and 13, and would be calculated as:

$[P_7^2 + P_7(1-P_7)\theta] + [P_8^2 + P_8(1-P_8)\theta] + [P_{13}^2 + P_{13}(1-P_{13})\theta]$

4A.6 If the interpretation of the evidence is that a contributor's genotype may be experiencing allelic dropout, utilizing a RMP of 2p (or $2p-p^2$) would incorporate all genotypes that include allele p.

Example: *This example illustrates how a locus may be used in the RMP calculation when only a single minor allele is detected below the stochastic threshold.*



Bin	RFU
10	150
14	740
15	730

This is a mixture (at a locus having alleles 10, 14 and 15) assumed to be from two contributors, where the major contributor is interpreted as being genotype 14,15, based on peak height ratios. The RMP for the major contributor would be calculated as $2P_{14}P_{15}$.

There remains an obligate minor contributor allele 10 below the stochastic threshold. There is a reasonable expectation that the sister allele to the allele 10 may be undetected. The RMP for the minor contributor must account for all genotypes at this locus that include the allele 10, since the minor contributor's genotype may not be fully detected.

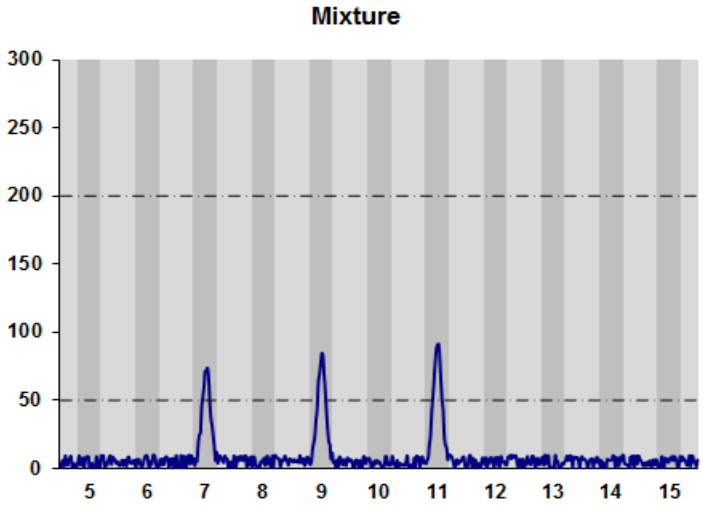
For the minor contributor, using an allele frequency for allele 10 of 10%:
Using section 3.4.2.4 (RMP = 2p), the RMP is 2*0.10 = 0.20 = 1 in 5.00
Using section 3.4.2.4 (RMP = $2p - p^2$), the RMP is $(2*0.10) - (0.10)^2 = 0.19$ = 1 in 5.26
Using theta (0.01) adjustment, the RMP = $p^2 + p(1- p)\theta + 2p(1- p)$, the RMP is $(0.10)^2 + (0.10)(1- 0.10)(0.01) + 2(0.10)(1- 0.10) = 0.1909$ = 1 in 5.24

If the locus is not utilized, the RMP for the minor contributor = 1 in 1 using section 3.4.5. This approach does not account for the available information regarding the minor contributor as described above.

This example highlights a limitation of the statistics within the binary approach. In a probabilistic approach, a lower probability of the evidence may be assigned if the POI had a genotype that included the allele 10 with an undetected second allele.

4A.7 In a low-level mixture assumed to be comprised of only two contributors, with the possibility of allelic dropout, the RMP must account for possible heterozygous genotype pairings in which one of the alleles may have dropped out.

Example: *This example illustrates how the RMP is used to account for a heterozygote-heterozygote pairing at a locus where one of the four alleles may have dropped out.*



Bin	RFU
7	70
9	80
11	90

In a low-level mixture assumed to be from two contributors and having alleles 7, 9, and 11, with all three alleles below the stochastic threshold, the interpretation may be that the two contributors could be a homozygote-heterozygote pairing where all alleles were detected, a heterozygote-heterozygote pairing where all alleles were detected, or a heterozygote-heterozygote pairing where a fourth allele might have dropped out:

First Contributor	Second Contributor
7,7	9,11
9,9	7,11
11,11	7,9
9,11	7,7
7,11	9,9
7,9	11,11
7,9	7,11
7,9	9,11
7,11	9,11
7,11	7,9
9,11	7,9
9,11	7,11
7,9	11,F
7,11	9,F
9,11	7,F
11,F	7,9
9,F	7,11
7,F	9,11

In this case, the RMP must account for all heterozygotes and homozygotes represented by these three alleles, but also all heterozygotes that include one of the detected alleles, paired with an unknown allele, F, that has dropped out. The possible genotypes for the detected alleles are (7,7), (9,9), (11,11), (7,9), (7,11), and (9,11), while the possible genotypes where an undetected allele is paired with one of the detected alleles are (7,F), (9,F), and (11,F). Genotype F,F is not permitted since this would violate the assumption of only two contributors.

The RMP for this interpretation could be calculated as $(2P_7 – P_7^2) + (2P_9 – P_9^2) + (2P_{11} – P_{11}^2) – 2P_7P_9 – 2P_7P_{11} – 2P_9P_{11}$.* Since $2P_7$ includes $2P_7P_9$ and $2P_7P_{11}$, $2P_9$ includes $2P_7P_9$ and $2P_9P_{11}$, and $2P_{11}$ includes $2P_7P_{11}$ and $2P_9P_{11}$, the formula subtracts $2P_7P_9$, $2P_7P_{11}$, and $2P_9P_{11}$ to avoid double-counting these genotype frequencies.

Laboratories may choose to calculate the RMP as $2P_7 + 2P_9 + 2P_{11}$ for the scenario described above.

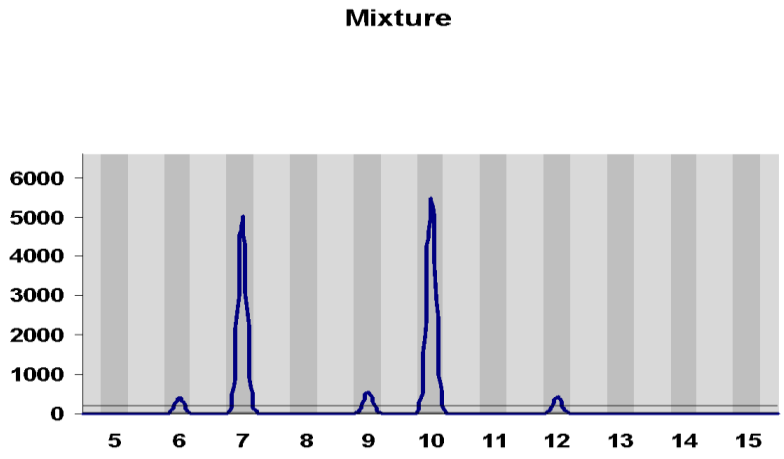
Laboratories may choose to assign the value of 1 to the scenario described above, i.e., not use the locus for statistical weight.

* For simplicity, homozygous frequencies were not adjusted using θ. With θ incorporated, the RMP could be calculated as $[P_7^2 + P_7(1-P_7)\theta + 2P_7(1-P_7)] + [P_9^2 + P_9(1-P_9)\theta + 2P_9(1-P_9)] + [P_{11}^2 + P_{11}(1-P_{11})\theta + 2P_{11}(1-P_{11})] – 2P_7P_9 – 2P_7P_{11} – 2P_9P_{11}$.

4A.8 The calculation of RMP statistics should incorporate peaks determined to be indistinguishable from stutter because the minor contributor's DNA profile may contain these alleles.

Example: *This example illustrates how peaks interpreted to be indistinguishable from stutter can be incorporated into the RMP calculation.*

Bin	RFU	Comment
6	380	Below stutter threshold
7	5000	Victim
9	550	Below stutter threshold
10	5500	Victim
12	425	Requisite to unknown contributor, above stochastic threshold and not in stutter position such that dropout of a sister allele is unreasonable.



In a sperm fraction mixture (at a locus having alleles 7, 10, and 12) assumed to be from two contributors, where the major contributor is the victim (having genotype 7,10), there remains an obligate minor contributor 12 allele above the stochastic threshold. Also present in the results are two peaks filtered as possible stutter (6 and 9). If both filtered peaks are within an RFU range that could reasonably be paired with the 12 allele as heterozygous genotypes, the sperm contributor genotypes included post-deconvolution might be 12,12 and 7,12 and 10,12 and 6,12 and 9,12. In this case, the RMP for the sperm DNA contributor could be calculated as $[P_{12}^2 + P_{12}(1-P_{12})\theta] + 2P_7P_{12} + 2P_{10}P_{12} + 2P_6P_{12} + 2P_9P_{12}$. Some laboratories may instead choose to apply a single-allele formula as discussed in section 3.4.2.4, e.g., $2P_{12}$.

Even though a laboratory may apply a $2P_{12}$ statistical calculation, the interpretation of included genotypes should still be limited to 12,12 and 7,12 and 10,12 and 6,12 and 9,12. All other genotypes should be excluded as the minor contributor.

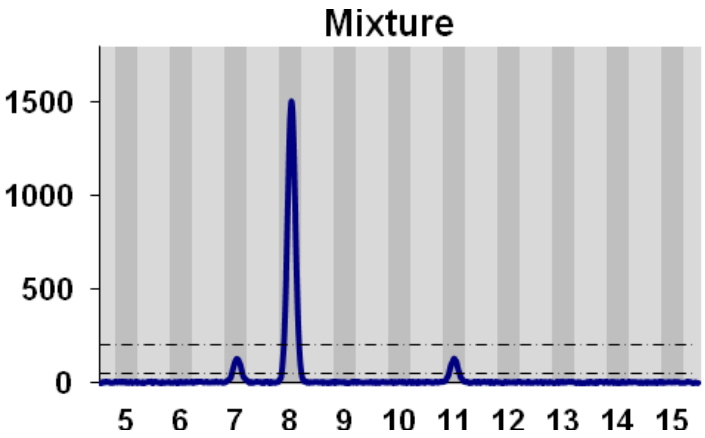
If a minimal amount of stutter occurred from allele 7 into the peak in bin 6, the peak in bin 6 could be a heterozygous allele to the allele 12 (300/425 = 71% peak height ratio balance).

If a minimal amount of stutter occurred from allele 10 into the peak in bin 9, the peak in bin 9 could be a heterozygous allele to the allele 12 (450/425 > 100% peak height ratio balance, indicating bin 9 could be both a heterozygous allele to allele 12 and some amount of stutter).

Refer to section 2.6 for additional guidance on interpreting potential stutter peaks in a mixed sample.

4A.9 Interpretations that include peaks that are indistinguishable from stutter should also include such peaks in the RMP statistics unless the interpretation incorporates the genotype with the zygosity in question (e.g., utilizing a 2p calculation is more mathematically conservative than 2pq + 2pr + 2ps).

Example: *This example illustrates how a peak interpreted to be indistinguishable from stutter need not be incorporated into the RMP calculation if the RMP calculation already accounts for the possibility of undetected data.*



Bin	RFU
7	150
8	1500
11	150

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors, at a disparate ratio of contributors. Furthermore, the evidence is intimate to the victim.

The peak in bin 7 is above the analytical threshold, but below the stutter threshold. However, since it is of similar RFU value to the unambiguous minor allele 11 and could belong in the minor contributor's genotype, the peak in bin 7 is indistinguishable from stutter.
The victim is genotype 8,8 at this locus.

Since dropout of one allele of the unknown contributor is considered reasonable, possible genotypes of the second contributor must contain the allele 11, but cannot be limited to only genotypes 8,11 and/or 11,11 and/or 7,11. As such, the RMP would be $2P_{11}$ (or $2P_{11} - P_{11}^2$ to avoid double counting the homozygous 11,11 genotype in the statistic).

Even if the person of interest (POI) is genotype 8,11 at this locus, the statistic should be driven by the interpretation and include the possible genotypes that contain the unambiguous allele 11 and any other possible allele.

Refer to section 2.6 for additional guidance on interpreting potential stutter peaks in a mixed sample.

Section 4B: Likelihood Ratio (LR)

Introduction:

The LR is a statistic for the comparison of the probability of the evidence (E), given two competing hypotheses, inclusionary (H_i) or exclusionary (H_e), for an individual, or specific sets of individuals. The general formula is:

$$LR = \frac{Pr\,(E|H_i)}{Pr\,(E|H_e)}$$

The LR may also be used to assess the DNA findings given multiple POI together in a mixture. The specific formula of the LR is dependent upon the evidence profile, the comparison reference profile(s), and the individual hypotheses. There are scenarios where choosing one hypothesis over another may be difficult. This can occur with complex mixtures (i.e., more than two contributors), non-intimate or neutral evidence (i.e., where no assumptions of known contributors can be made), and/or when there are multiple POI (e.g., suspects or victims) being considered. Depending on the hypothesis and assumptions made, different LRs will result. In these instances, multiple hypotheses and LRs may be necessary for the case file in order for the court to determine which propositions are most appropriate.

Given the myriad possible combinations of profiles and competing hypotheses, any list of examples would be incomplete. A limited set of examples is provided below which includes explanations of:

- the fundamentals of the statistical approach
- the difference between unrestricted and restricted calculations
- the application of relevant background information such as case circumstances, individuals involved, evidence type or source
- the inclusion of the possibility of undetected data
- the inclusion of peaks interpreted to be indistinguishable from stutter
- the application of the LR to a major mixture
- the application of the LR to mixtures of more than just two contributors
- the effect of the assumption of the number of contributors
- the application of the LR to mixtures of close biological relatives

4B.1 The LR for a given locus is the numerator probability divided by the denominator probability.

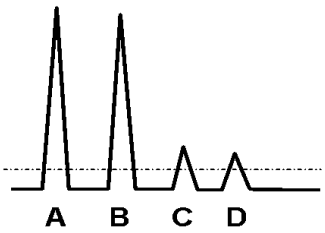
4B.2 The LR for the profile is produced by multiplying the LR values from each locus. The resultant value is the ratio of the probabilities of the evidence profile if the inclusionary hypothesis is true as opposed to if the exclusionary hypothesis is true.

4B.2.1 A final LR value greater than 1 indicates support for the inclusionary hypothesis.

4B.2.2 A final LR value less than 1 indicates support for the exclusionary hypothesis.

4B.3 When the evidence profile is determined to be single source, and the reference and evidence profiles are identical at all loci, the LR is reduced to 1 / the Random Match Probability (RMP).

4B.4 The LR for mixtures is conditioned on an assumed number of contributors, and may use peak height information (restricted approach) or not (unrestricted approach) when including genotype combinations for proposed contributors into the LR (Figure 1).



Unrestricted

All combinations of alleles are deemed possible (relative peak height differences are not utilized)

AB + AC + AD + BC + BD + CD

Restricted

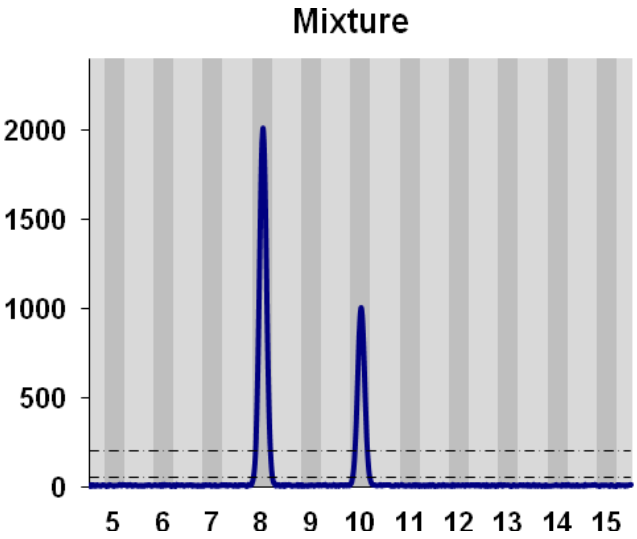
Based on relative peak heights, alleles are paired only where specific combinations of alleles are deemed possible

AB + ~~AC~~ + ~~AD~~ + ~~BC~~ + ~~BD~~ + CD

Figure 1. Illustration of “restricted” versus “unrestricted” approaches based on relative peak heights (using an assumption of two donors with all peaks above the stochastic threshold).

4B.4.1 An "unrestricted" LR is the LR calculated without consideration of the relative peak heights observed during interpretation. As such, the unrestricted LR may include genotypes in the statistic that have been excluded during the interpretation and comparison steps of the analysis. The unrestricted LR may less accurately reflect the interpretation of the evidence than the restricted LR.

Example: *This example illustrates the use of an unrestricted LR.*



Bin	RFU
8	2000
10	1000

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors. Across the entire profile, the mixture appears to have a ratio of contributors of approximately 2:1 with the POI as the minor contributor.

At this locus, a mixture with alleles 8 and 10 displays no indications of allelic dropout. No further considerations of peak heights are undertaken for statistical purposes. The POI in question is genotype 8,8, and no other reference standards are being considered for inclusion. The LR might be for the propositions that the mixture is comprised of the POI and an unknown individual, as opposed to two unknown individuals.

The unrestricted LR numerator would assume the presence of the POI, meaning that the probability of observing results consistent with his or her profile would be 1. However, the numerator must also include genotypes for the second contributor that must include the allele 10. The numerator will not utilize the assumptions of the ratio of contributors or the assumption of the POI as the lesser of the two contributors. Therefore, the genotypes included in the numerator of the unrestricted LR would be 10,10 and 8,10.

Unrestricted LR numerator = $[P_{10}^2 + P_{10}(1-P_{10})\theta] + 2P_8P_{10}$.

The denominator of the unrestricted LR would assume that the mixture is a combination of two unknown contributors with no alleles other than 8 or 10, and the combination of their genotypes

must complete the detected mixture of alleles 8 and 10. Without consideration to the relative peak height ratios and the overall mixture ratio, the unrestricted LR denominator might be limited to taking into account the following pairs of genotypes:

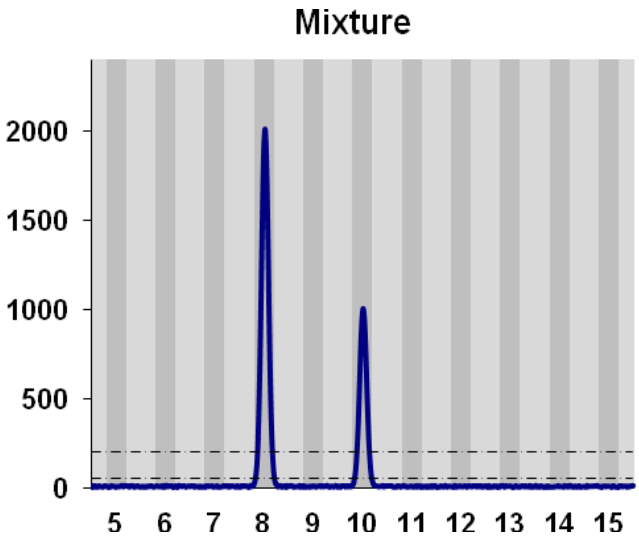
First Contributor	Second Contributor	Combined Probability
8,8	8,10	$[P_8^2 + P_8(1-P_8)\theta] * 2P_8P_{10}$
8,8	10,10	$[P_8^2 + P_8(1-P_8)\theta] * [P_{10}^2 + P_{10}(1-P_{10})\theta]$
8,10	8,8	$2P_8P_{10} * [P_8^2 + P_8(1-P_8)\theta]$
8,10	8,10	$2P_8P_{10} * 2P_8P_{10}$
8,10	10,10	$2P_8P_{10} * [P_{10}^2 + P_{10}(1-P_{10})\theta]$
10,10	8,8	$[P_{10}^2 + P_{10}(1-P_{10})\theta] * [P_8^2 + P_8(1-P_8)\theta]$
10,10	8,10	$[P_{10}^2 + P_{10}(1-P_{10})\theta] * 2P_8P_{10}$

Therefore the denominator of the unrestricted LR at this locus = the sum of the probabilities of the possible combinations of genotypes. Simplified equivalents of these calculations are available (e.g., Weir, 1997).

The LR for the locus is the numerator divided by the denominator.

4B.4.2 A "restricted" LR is the LR calculated once relative peak heights are taken into consideration. The restricted LR more precisely reflects the interpretation of the evidence than the unrestricted LR; the restricted LR will not include genotypes that would be considered unreasonable to be present in the mixture.

Example: *This example illustrates how the use of a restricted LR eliminates some of the genotype combinations included in the unrestricted LR example above.*



Bin	RFU
8	2000
10	1000

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors.

The POI in question is genotype 8,8, and no other reference standards are being considered for inclusion. Across the entire profile, the mixture appears to have a ratio of contributors of approximately 2:1 with the POI as the lesser contributor.

The LR might be for the propositions that the mixture is comprised of the POI and an unknown individual, as opposed to two unknown individuals. The numerator of the LR would assume the POI's contribution, meaning that the probability of observing results consistent with his genotype would be 1. Assuming the presence of the POI for use in the numerator of the LR, the deconvolution of the locus indicates that the unknown contributor must have genotype 8,10 to complete the mixture.

Restricted LR numerator = $2P_8P_{10}$.

The denominator of the restricted LR would assume that the mixture is a combination of two unknown contributors with a ratio of contributors of approximately 2:1. The unknown contributors must have no alleles other than 8 or 10, and the combination of their genotypes must complete the detected mixture of alleles 8 and 10. Based upon the relative peak height ratios and the overall mixture ratio, the restricted LR denominator might be limited to taking into account the following pairs of genotypes:

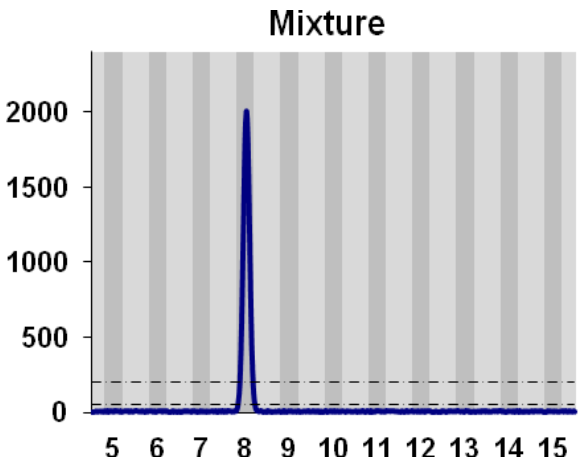
Major Contributor	Minor Contributor	Combined Probability
8,8	10,10	$[P_8^2 + P_8(1-P_8)\theta] * [P_{10}^2 + P_{10}(1-P_{10})\theta]$
8,10	8,8	$2P_8P_{10} * [P_8^2 + P_8(1-P_8)\theta]$

Therefore the denominator of the restricted LR at this locus = the sum of the probabilities of the possible combinations of genotypes.

The LR for the locus is the numerator divided by the denominator.

4B.4.3 Within an STR profile, some loci may have results that give identical restricted and unrestricted LRs.

Example: *This example illustrates how both the restricted and unrestricted LR would result in the same statistic.*



Bin	RFU
8	2000

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a close ratio of contributors with no reasonable assumption of dropout. Furthermore, the evidence is intimate to the victim.

The victim is genotype 8,8 at this locus. The POI is also genotype 8,8 at this locus.

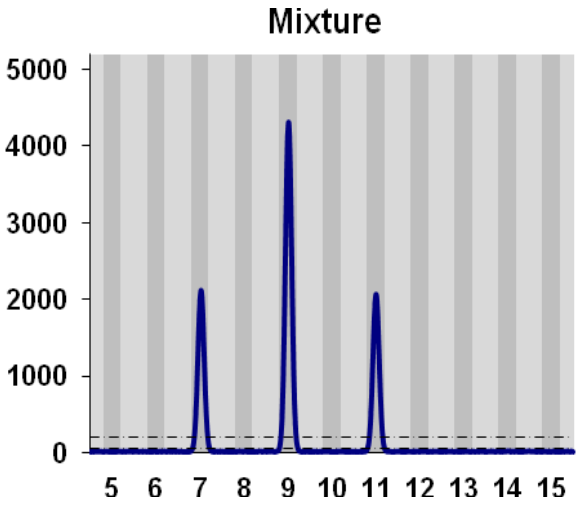
The LR might be for the propositions that the mixture is comprised of the victim and the POI, as opposed to the victim and an unknown individual. The numerator of both the unrestricted and restricted LR would assume the presence of both the victim and the POI. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

The denominator of both the unrestricted and restricted LR would assume the presence of the victim, but assume the second contributor is unknown. Since dropout is considered unreasonable, the only genotype that is reasonable for the second contributor is 8,8. The restricted LR has no further information available than the unrestricted LR with which to eliminate unreasonable genotypes. As such the denominator for both the unrestricted and restricted LR would be $[P_8^2 + P_8(1-P_8)\theta]$.

The LR for the locus is the numerator divided by the denominator.

4B.5 The LR requires that hypotheses be formulated based on relevant case information, such as case circumstances, individuals involved, evidence type or source. These hypotheses may involve different assumptions and should be evaluated and documented in the case file.

Example: *This example illustrates how the hypotheses of the LR affect the formula utilized.*



Bin	RFU
7	2100
9	4300
11	2050

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a close ratio of contributors with no reasonable assumption of dropout.

The evidence is intimate to the victim. The numerator of the unrestricted LR would assume the presence of both the victim (genotype 7,9) and the POI (genotype 9,11). The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

The denominator of the unrestricted LR would assume the presence of the victim, but assume the second contributor is unknown. Given the genotype of the victim (7,9), the second contributor must have allele 11. The unrestricted LR would include genotypes 7,11 and 9,11 and 11,11 for the second contributor. The denominator of the unrestricted LR would be $2P_7P_{11} + 2P_9P_{11} + [P_{11}^2 + P_{11}(1-P_{11})\theta]$. The restricted LR may only include genotype 9,11. The denominator of the restricted LR would therefore be $2P_9P_{11}$.

The LR for the locus is the numerator divided by the denominator.

However, if the same evidence were intimate to the POI, the numerator of the unrestricted LR would assume the presence of both the POI (genotype 9,11) and the victim (genotype 7,9). The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

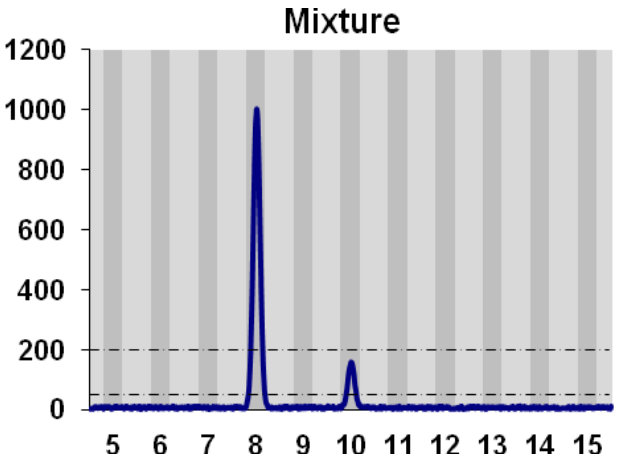
The denominator of the unrestricted LR would assume the presence of the POI, but assume the second contributor is unknown. Given the genotype of the POI (9,11), the second contributor must have allele 7. The unrestricted LR would include genotypes 7,7 and 7,9 and 7,11 for the

second contributor. The denominator of the unrestricted LR would be $[P_7^2 + P_7(1-P_7)\theta] + 2P_7P_9 + 2P_7P_{11}$. The restricted LR may only include genotype 7,9. The denominator of the restricted LR would therefore be $2P_7P_9$.

The LR for the locus is the numerator divided by the denominator.

4B.6 For both the restricted and unrestricted LR methods, when calculating the LR where only a single allele is foreign to the assumed contributor and the zygosity is in question (it is below the stochastic threshold), the formula 2p or $2p-p^2$ may be used.

Example: *This example illustrates how the LR can incorporate the possibility of undetected data.*



Bin	RFU
8	1000
10	150

The mixture is determined to have originated from only two contributors. Furthermore, the evidence is intimate to the victim.

The victim is genotype 8,8 at this locus. The POI is genotype 8,10 at this locus.
The LR might be for the propositions that the mixture is comprised of the victim and the POI, as opposed to the victim and an unknown individual.

The numerator of the LR would assume the presence of both the victim and the POI. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

The denominator of the LR would assume the presence of the victim, but assume the second contributor is unknown. Since dropout of one allele of the unknown contributor is considered reasonable, genotypes of the second contributor must contain the allele 10, but cannot be limited to only genotypes 8,10 and/or 10,10. As such, the denominator of the LR could be $2P_{10}$ or $2P_{10}$-

P_{10}^2 (to avoid double counting the homozygous 10,10 genotype in the statistic).

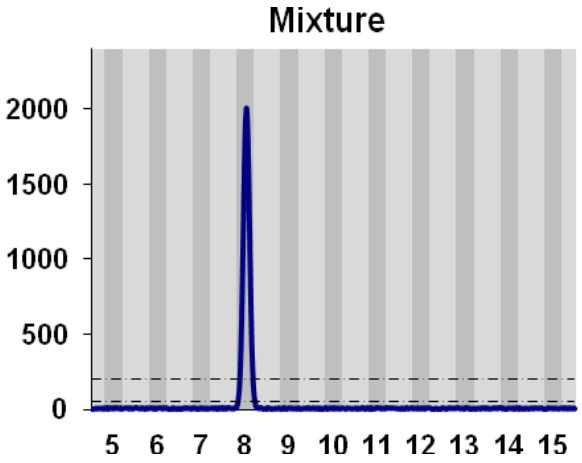
The LR for the locus is the numerator divided by the denominator.

It must be noted that since the single allele foreign to the victim is below the stochastic threshold, dropout of the second allele of the POI is considered reasonable. As such, if the POI had a genotype that included the allele 10 with an undetected second allele (e.g., genotype 10,12), then the probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator of the LR would still be 1. Since the denominator of the LR is not dependent upon the genotype of the POI, the denominator would not change, regardless of the genotype of the POI.

This example highlights a limitation of the statistics within the binary approach. In a probabilistic approach, a lower probability of the evidence may be assigned if the POI had a genotype that included the allele 10 with an undetected second allele.

4B.6.1 For both the restricted and unrestricted LR, when the interpretation of the evidence includes the possibility of a contributor whose complete genotype may be undetected, the locus may be statistically neutral (i.e., assigned a value of 1).

Example: *This example illustrates how the LR may be uninformative when accounting for a genotype that could be completely unrepresented by the detected data.*



Bin	RFU
8	2000

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a ratio of contributors that provides a clear major/minor distinction, with a reasonable assumption of dropout of the minor contributor's entire genotype at this locus. Furthermore, the evidence is intimate to the victim.

The victim is genotype 8,8 at this locus. The POI is also genotype 8,8 at this locus (though the genotype of the POI could be anything at this locus, and not be excluded, since there is a reasonable assumption of genotype dropout).

The LR might be for the propositions that the mixture is comprised of the victim and the POI, as opposed to the victim and an unknown individual.

The numerator of both the unrestricted and restricted LR would assume the presence of both the victim and the POI. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

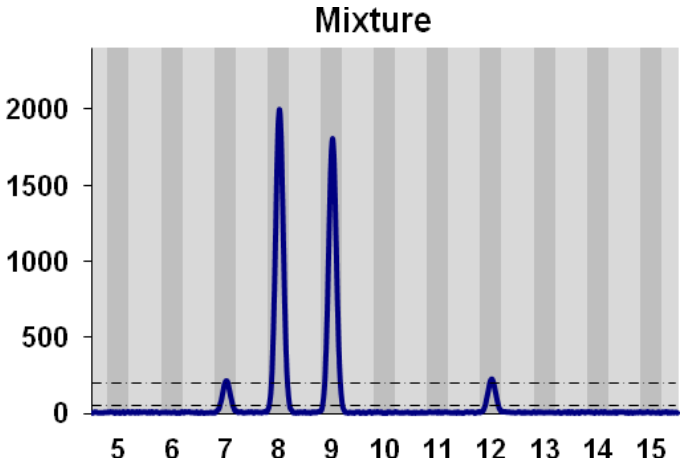
The denominator of both the unrestricted and restricted LR would assume the presence of the victim, but assume the second contributor is unknown. Since complete genotype dropout of the unknown contributor is considered reasonable, the unknown contributor could have any genotype at this locus. As such, the denominator for both the unrestricted and restricted LR would be 1 (the probability of observing this electropherogram if the contributors are the victim and an unknown contributor is equal to 1).

The LR for the locus is the numerator divided by the denominator. The LR of 1/1 provides no statistical support to either the inclusionary or exclusionary hypotheses.

This example highlights a limitation of the LR within the binary approach. In a probabilistic approach, a lower probability of the evidence may be assigned if the POI had a genotype that was partially or fully undetected.

4B.7 The calculation of the LR should incorporate peaks determined to be indistinguishable from stutter when the numerator of the LR would assume the POI's contribution, even if the POI's DNA profile does not contain these alleles.

Example: *This example illustrates how a peak interpreted to be indistinguishable from stutter can be incorporated into the LR.*



Bin	RFU
7	210
8	2000
9	1800
12	220

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a ratio of contributors that provides a clear major/minor distinction with no reasonable assumption of dropout.

This locus of the mixture is comprised of alleles 8 and 9 above the stochastic threshold, allele 12 above the stochastic threshold, but as a distinct minor contributor, and a peak in bin 7 below the stutter threshold with an RFU value similar to allele 12. Considering the peak height ratio calculations, the following genotype combinations are possible if the mixture is assumed to be comprised of only two contributors:

Major contributor	Minor contributor	Interpretation of the peak in bin 7
8,9	12,12	7 is stutter
8,9	8,12	7 is stutter
8,9	9,12	7 is stutter
8,9	7,12	7 is an allele of the minor contributor, but the RFU value is below the stutter threshold

Upon comparison of the inferred genotype possibilities to the known reference samples of the victim and POI, it is seen that the victim who has genotype 8,9 is consistent with the major contributor. The POI who has genotype 8,12 is fully represented by the detected unambiguous alleles. The true nature of the peak in bin 7 is unknown, therefore the POI cannot be excluded as the minor contributor of this mixture since it is possible that the peak is only stutter.

The LR might be for the propositions that the mixture is comprised of the victim and the POI as opposed to the victim and an unknown individual.

Since the numerator hypothesis assumes the presence of the victim (genotype 8,9) and the POI (8,12), the peak in bin 7 is interpreted as stutter. The probability of observing this electropherogram would be 1 if this hypothesis is true. As a result, the numerator equals 1.

The denominator hypothesis would consider the assumed presence of the victim, and that the unknown contributor's profile may be any one of the inferred possible genotypes since the true nature of the peak in bin 7 is unknown. As a result, the denominator equals $[P_{12}^2 + P_{12}(1- P_{12})\theta] + 2P_8P_{12} + 2P_9P_{12} + 2P_7P_{12}$.The inclusion of genotype 7,12 into the denominator of the LR calculation is independent of the genotype of the POI since the hypothesis of the denominator assumes the POI is not a contributor to the mixture.

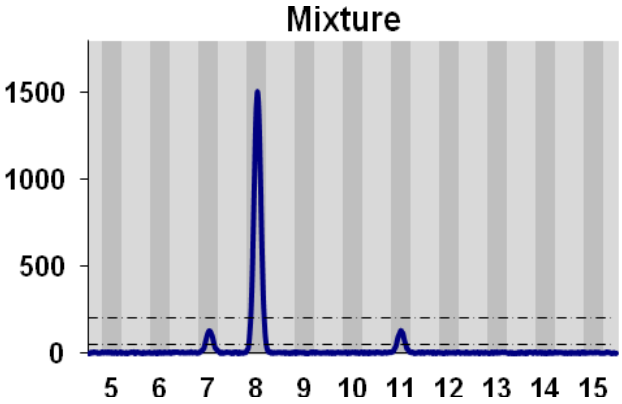
The restricted LR would likely include all the same genotypes for the second contributor as were used in the unrestricted LR, given the data. The denominator of the restricted LR would therefore be the same as the unrestricted LR.

The LR for the locus is the numerator divided by the denominator.

It must be noted that since the peak in bin 7 is considered indistinguishable from stutter, this peak is considered as reasonable to be an allele of the POI. As such, if the POI had genotype 7,12 the hypothesis of inclusion would still be supported by the electropherogram and the numerator of the LR would still be 1. Since the denominator of this example's LR is not dependent upon the genotype of the POI, the denominator would not change regardless of the genotype of the POI.

4B.7.1 If the interpretation of the evidence is that the minor contributor genotype may be experiencing allelic dropout, utilizing a LR denominator of 2p (or $2p\text{-}p^2$) would incorporate all genotypes that include allele P and no further inclusion of the peak that is indistinguishable from stutter into the statistic is necessary.

Example: *This example illustrates how a peak interpreted to be indistinguishable from stutter need not be incorporated into the LR if the LR already accounts for the possibility of undetected data.*



Bin	RFU
7	150
8	1500
11	150

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a ratio of contributors that provides a clear major/minor distinction, with the unambiguous minor allele 11 below the stochastic threshold. Furthermore, the evidence is intimate to the victim.

The peak in bin 7 is above the analytical threshold, but below the stutter threshold. However, since it is of similar RFU value to the unambiguous minor allele 11 and could belong in the minor contributor's genotype, the peak in bin 7 is indistinguishable from stutter.

The victim is genotype 8,8 at this locus. The POI is genotype 8,11 at this locus.

The LR might be for the propositions that the mixture is comprised of the victim and the POI as opposed to the victim and an unknown individual.

The numerator of the LR would assume the presence of both the victim and the POI and assume the peak in bin 7 is stutter. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

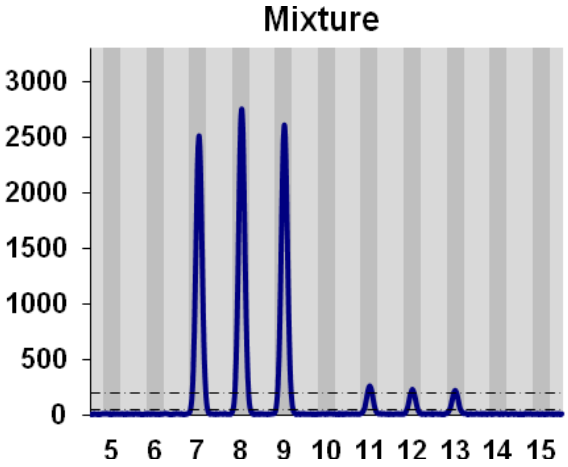
The denominator of the LR would assume the presence of the victim, but assume the second contributor is unknown. Since dropout of one allele of the unknown contributor is considered reasonable, genotypes of the second contributor must contain the allele 11, but cannot be limited to only genotypes 8,11 or 11,11 or 7,11. As such, the denominator of the LR could be $2P_{11}$ (or $2P_{11}\text{-}P_{11}^2$ to avoid double counting the homozygous 11,11 genotype in the statistic).

The LR for the locus is the numerator divided by the denominator.

It must be noted that since the single allele foreign to the victim is below the stochastic threshold, dropout of the second allele of the POI is considered reasonable. As such, if the POI had a genotype that included the allele 11 with an undetected second allele (e.g., 11,12), or second allele in bin 7 (i.e., 7,11), the hypothesis of inclusion would still be supported by the electropherogram and the numerator of the LR would still be 1. Since the denominator of this example's LR is not dependent upon the genotype of the POI, the denominator would not change, regardless of the genotype of the POI.

4B.8 The LR can be applied to "major mixtures" regardless of the interpretation of the other portion of the mixture.

Example: *This example illustrates how the LR might be applied to a major mixture, even when the total number of contributors in the mixture cannot be determined.*



Bin	RFU
7	2500
8	2750
9	2600
11	250
12	220
13	210

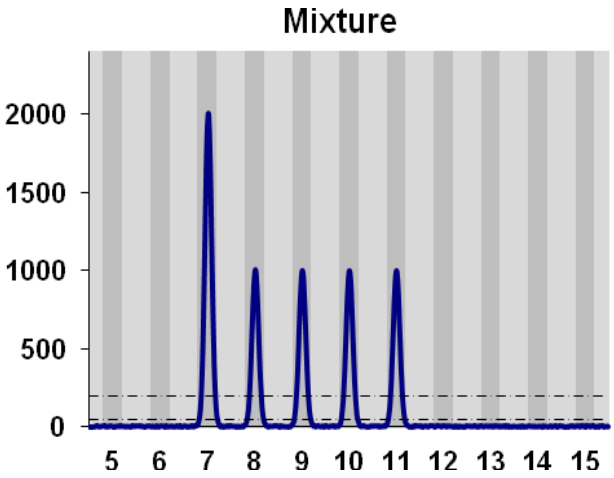
Based upon the examination of the entire DNA profile, the mixture is determined to have originated from more than three contributors, with a major mixture of only two contributors, and no reasonable assumption of dropout from either of the major contributors.

Since a finite number of total contributors cannot be assumed, the LR cannot be applied to the mixture as a whole. However, since the major mixture was defined as being only two contributors, the LR can be applied to only alleles 7, 8, and 9. The actual LR will be dependent upon the propositions utilized to formulate the LR.

The analyst must be aware during the interpretation stage of the potential impact of the minor contributor(s) on the RFU values of the major mixture if attempting to perform a restricted LR.

4B.9 The LR can be applied to mixtures of 3 or more contributors. It is requisite that a finite number of contributors be assumed in the mixture since the LR cannot be applied to an interpretation that assumes only a minimum number of contributors.

Example: *This example illustrates how the LR might be applied to a mixture of three contributors.*



Bin	RFU
7	2000
8	1000
9	1000
10	1000
11	1000

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only three contributors at a close ratio of contributors with no reasonable assumption of dropout. Furthermore, the evidence is intimate to the victim.

The victim is genotype 7,7 at this locus. The POI is genotype 8,9 at this locus and no other reference standards are being considered for inclusion.

The LR might be for the propositions that the mixture is comprised of the victim, the POI, and an unknown contributor, as opposed to the victim and two unknown individuals.

The numerator of the LR would assume the presence of the victim and the POI meaning that the probability of observing results consistent with their profiles would be 1. However, the numerator must account for the unknown contributor who must be genotype 10,11.

LR numerator = $2P_{10}P_{11}$.

The denominator of the LR would assume the presence of the victim, but assume the other two contributors are unknown. Since dropout is considered unreasonable, the unrestricted LR would include the following genotype pairs:

First unknown contributor	Second unknown contributor
8,9	10,11
8,10	9,11
8,11	9,10

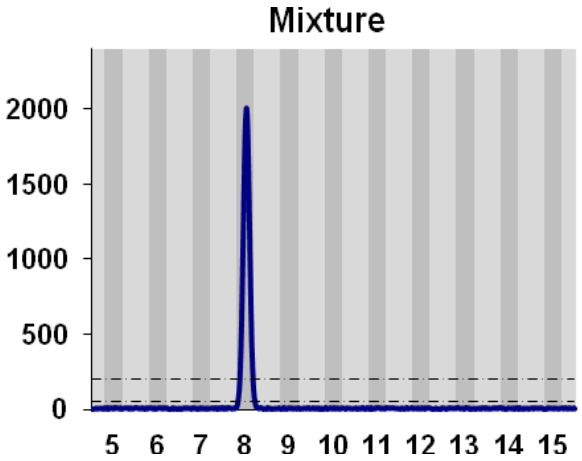
9,10	8,11
9,11	8,10
10,11	8,9

The denominator of the unrestricted LR would be $2*[(2P_8P_9*2P_{10}P_{11}) + (2P_8P_{10}*2P_9P_{11}) + (2P_8P_{11}*2P_9P_{10})]$. Given the peak heights of the unknown contributors are of such values that deconvolution of the genotypes of the two unknown contributors is not possible, the denominator of the restricted LR would be the same as the unrestricted LR in this example.

The LR for the locus is the numerator divided by the denominator.

4B.10 While most often the numerator and denominator of the LR assume the same number of contributors, this is not a requirement. However, increasing the number of unknown contributors in either the numerator or denominator generally decreases the probability of the associated hypothesis. If the number of unknown contributors is increased in only the denominator, the LR will generally increase; if the number of unknown contributors is increased in only the numerator, the LR will generally decrease.

Example: *This example illustrates how a different number of contributors can be assumed in the numerator and denominator.*



Bin	RFU
8	2000

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a close ratio of contributors with no reasonable assumption of dropout. Furthermore, the evidence is intimate to the victim.

The victim is genotype 8,8 at this locus. The POI is also genotype 8,8 at this locus.

The LR might be for the propositions that the mixture is comprised of the victim and the POI, as opposed to the victim and one unknown individual.

Using these propositions, the numerator of both the unrestricted and restricted LR would assume the presence of both the victim and the POI. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

The denominator of both the unrestricted and restricted LR would assume the presence of the victim, but assume the second contributor is unknown. Since dropout is considered unreasonable, the only genotype that is reasonable for the unknown contributor is 8,8. The restricted LR has no further information available than the unrestricted LR in which to eliminate unreasonable genotypes. As such the denominator for both the unrestricted and restricted LR would be $[P_8^2 + P_8(1-P_8)\theta]$.

The LR for the locus is the numerator divided by the denominator. Using an allele frequency of 15%, and theta value of 0.01, the LR for this locus would be $1 / [(0.15)^2 + (0.15)\ (1\text{-}0.15)\ 0.01] = 1 / 0.0238 = 42$. If requested, alternate hypotheses could be evaluated. For example, the LR could be calculated using the assumptions of the mixture being comprised of the victim and the POI, as opposed to the victim and *two* unknown individuals.

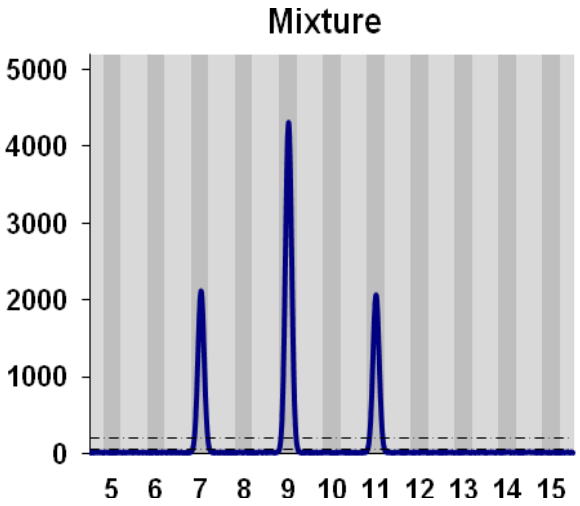
The numerator of both the unrestricted and restricted LR would assume the presence of both the victim and the POI. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

The denominator of both the unrestricted and restricted LR would assume the presence of the victim, but assume the second and third contributors are unknown. If dropout is considered unreasonable, the only genotype that is reasonable for both the second and third contributor is 8,8. The restricted LR has no further information available than the unrestricted LR in which to eliminate unreasonable genotypes. As such the denominator for both the unrestricted and restricted LR would be $[P_8^2 + P_8(1\text{-}P_8)\theta]*[P_8^2 + P_8(1\text{-}P_8)\theta]$.

The LR for the locus is the numerator divided by the denominator. Using the same allele frequency of 15%, and theta value of 0.01, the LR for this locus would be $1 / \{[(0.15)^2 + (0.15)\ (1\text{-}0.15)\ 0.01] * [(0.15)^2 + (0.15)\ (1\text{-}0.15)\ 0.01]\ \} = 1/0.000565 = 1769$.

4B.11 Likelihood ratios can be applied to mixtures of close biological relatives if the assumptions include the unknown contributor(s) are unrelated to the assumed contributors.

Example: *This example illustrates how the LR can be applied to a mixture of close biological relatives.*



Bin	RFU
7	2100
9	4300
11	2050

Based upon the examination of the entire DNA profile, the mixture is determined to have originated from only two contributors at a close ratio of contributors with no reasonable assumption of dropout.

If the evidence is intimate to the victim who has genotype 7,9, and the interpretation of the mixture has not excluded the POI who has genotype 9,11, and the POI is the biological father of the victim, the LR might be for the propositions that the mixture is comprised of the victim and the POI, as opposed to the victim and an unknown individual.

The numerator of the unrestricted LR would assume the presence of both the victim and the POI. The probability of observing this electropherogram would be 1 if this hypothesis is true. The numerator equals 1.

The denominator of the unrestricted LR would assume the presence of the victim, but assume the second contributor is unknown and unrelated to the victim and her father. Given the genotype of the victim (7,9), the second contributor must have allele 11. Since dropout is considered unreasonable, the unrestricted LR would include genotypes 7,11 or 9,11 or 11,11 for the second contributor. The denominator of the unrestricted LR would be $2P_7P_{11} + 2P_9P_{11} + [P_{11}^2 + P_{11}(1-P_{11})\theta]$. The restricted LR would likely only include genotype 9,11 as this is the most reasonable genotype of the second contributor given the data. The denominator of the restricted LR would be $2P_9P_{11}$.

The LR for the locus is the numerator divided by the denominator.

It must be noted that if the assumption in the denominator is that the unknown individual is not the biological father of the victim, but is her biological brother (or other close biological relative) whose genotype is untested, the LR as described here is not applicable. Close biological

relatives of the victim and her father may be more likely to have either genotype 7,11 or 9,11 than would a random individual given the knowledge of the genotypes of the victim and her father since the alleles 7, 9, or 11 may be common by descent.

4B.12 If the laboratory does not have the capability to perform an LR using all the available information (e.g., the possibility of undetected data or alleles that are indistinguishable from stutter) and the POI is interpreted as being included at the locus, then the LR for the locus may be assigned a neutral value of 1/1. It would be inappropriate to assign this neutral value for a POI whose genotype is interpreted as being excluded at the locus. Instead, the POI should have been excluded as part of the interpretation.

4B.13 Additional formulae for restricted and unrestricted LRs can be found in the literature (e.g., Fung and Hu (2008), Buckleton, Bright, and Taylor (2016)).

Section 4C: The Combined Probability of Inclusion (CPI) and Exclusion (CPE)

Introduction:

The CPI is an estimate of the probability that a randomly selected, unrelated individual would be included as a possible contributor to a mixture. The CPI includes all possible genotype combinations for the detected alleles at each locus. The CPE is the probability that a randomly selected, unrelated individual would be excluded as a contributor to the mixture. At a given locus, if allele dropout is reasonable based on laboratory defined criteria, all possible genotype combinations would not be represented in the CPI/CPE and therefore the locus should not be included. For additional information on using the CPI/CPE, see Bieber, et al (2016).

A limited set of examples is provided below which includes explanations of:

- Unrestricted CPI/CPE when no dropout is assumed
- Unrestricted CPI/CPE when dropout is reasonable based on the observed data
- Unrestricted CPI/CPE with assumptions
- Unrestricted CPI/CPE when dropout is reasonable based upon holistic interpretation
- Restricted CPI/CPE

4C.1 The Probability of Inclusion (PI) for a locus is calculated as (sum of allele frequencies)^2 for all detected alleles at a given locus:

> $PI = (P_A + P_B + P_C \ldots + P_N)^2$, where P_A is the frequency of allele A, P_B is the frequency of allele B, P_C is the frequency of allele C, …

4C.2 The CPI is the product of the individual locus PIs:

> $CPI = PI_1 \times PI_2 \times ... \times PI_N$

4C.3 The Probability of Exclusion (PE) for a locus has been commonly presented two ways:

> $PE = 1 - PI$
>
> $PE = q^2 + 2pq$, where p is the sum of allele frequencies and q represents all other alleles (1 – p). This is analogous to the single allele formula described in section 3.4.2.4.

4C.4 Population substructure corrections can also be applied using:

$PE = 1 - [p^2 - p(1 - p)\theta]$, where p is the sum of allele frequencies observed at that locus.

4C.5 The CPE has been commonly presented two ways:

$CPE = 1 - CPI$

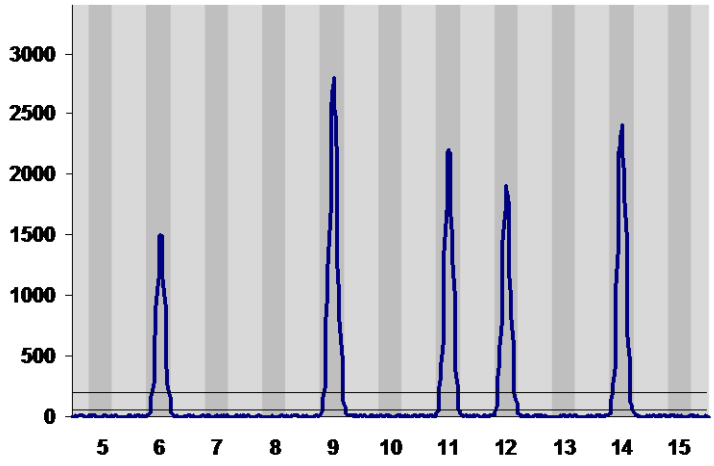
$CPE = 1 - [(1 - PE_1) \times (1 - PE_2) \times ... \times (1 - PE_N)]$

4C.6 The CPI and CPE are typically applied to all alleles detected in a mixture (i.e., an unrestricted CPI/CPE), subject to the limitations described in this section.

4C.6.1 In a mixture at a locus having alleles all above the laboratory's stochastic threshold, the interpretation might be that all potential contributors to this mixture have genotypes consisting of some combination of the detected alleles. In this case, the probability of inclusion for the locus could be calculated as (sum of allele frequencies)^2.

Example: *This example illustrates the use of an unrestricted CPI.*

The interpretation of the entire mixture is that there are three or more contributors with a close ratio of contributors. At this locus, all contributors' alleles are above the stochastic threshold, and no dropout is expected.



The PI at this locus would be calculated as: $(P_6 + P_9 + P_{11} + P_{12} + P_{14})^2$ and would include the following genotypes. All other genotypes would be excluded.

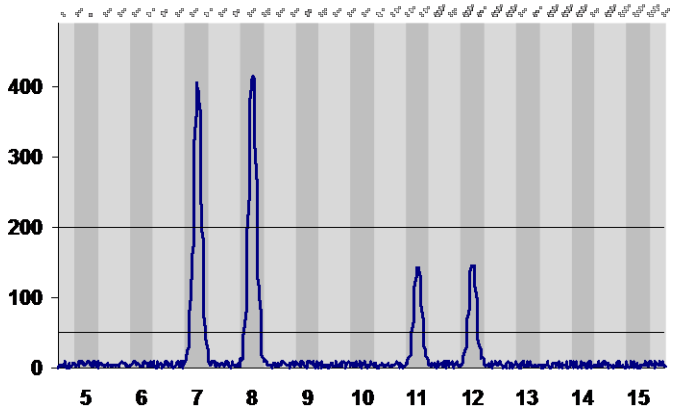
6,6	6,9	6,11	6,12	6,14
9,9	9,11	9,12	9,14	

11,11	11,12	11,14
12,12	12,14	
14,14		

4C.6.2 When the interpretation makes no assumptions of the number of contributors, loci with alleles below the stochastic threshold may not be used for statistical purposes to support an inclusion. In these instances, the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the interpreted alleles. At a locus having some alleles above the stochastic threshold and one or more alleles below that threshold, in the standard application of the CPI and CPE, no calculation would be performed at this locus. Because allelic dropout is possible at these loci, genotypes from possible contributors may not be represented in the statistical calculation.

Example: *This example illustrates the limitations of using an unrestricted CPI when some data is below the stochastic threshold.*

The interpretation of the entire mixture is that there are three or more contributors with one or more contributors below the stochastic threshold at some larger loci. At this locus, some alleles are below the stochastic threshold, and dropout is reasonable to assume.



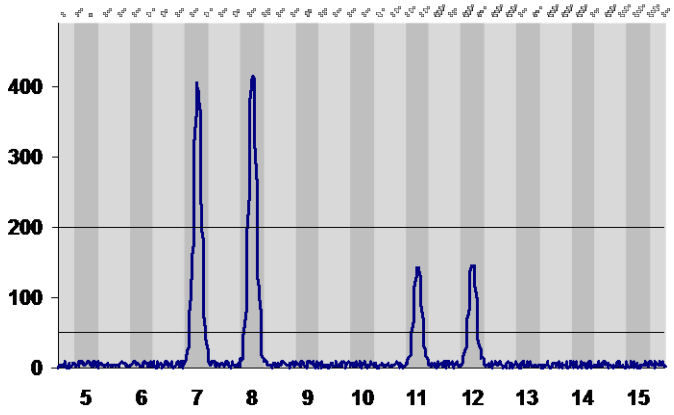
This locus is not suitable for a CPI calculation. The CPI would fail to properly support an inclusion to the locus because there is data below the stochastic threshold and dropout is reasonable to assume.

4C.6.3 If assumptions as to the number of contributors are used in the interpretation, alleles below the stochastic threshold may be used for CPI calculations if all allelic data is interpreted to be present at that locus. This determination is based on an assumed number of contributors

supported by the data. This assumption should be clearly stated in the report.

Example: *This example illustrates the use of an unrestricted CPI with assumptions when some data is below the stochastic threshold.*

Note: While the locus used in this example is the same as in example 4C.6.2, the profile in its entirety is different, and results in a different interpretation.
The interpretation of the entire mixture is that there are only two contributors, with one of the contributors below the stochastic threshold at some larger loci. At this locus, some alleles are below the stochastic threshold, but dropout is not observed based on the assumed number of contributors.



Using the assumption of two contributors, this locus fully represents both heterozygous contributors. Based on the assumption of two contributors, homozygote genotypes would be excluded as contributors.
However, the PI at this locus would be calculated as: $(P_7 + P_8 + P_{11} + P_{12})^2$ and would include the following genotypes.

7,7	7,8	7,11	7,12
8,8	8,11	8,12	
11,11	11,12		
12,12			

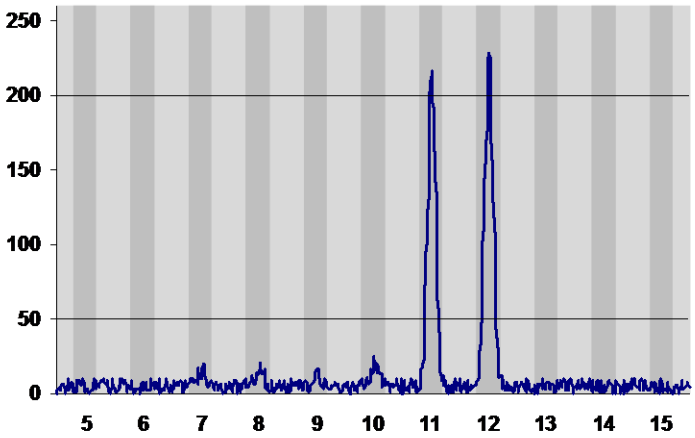
Because the heterozygous genotypes from the interpretation are included in the CPI, it is appropriate to use the CPI calculation at this locus. Further, based on the assumption of only two contributors, there are genotypes represented in the calculation that would be excluded by the interpretation. The assumption of only two contributors must be stated clearly in the report to use this locus for the CPI.
It should be noted that other statistical approaches make better use of this type of data.

4C.6.4 In a mixture at a locus with all alleles above the stochastic threshold but where dropout is possible and/or expected, no calculation would be performed at this locus. Because dropout is

possible at these loci, genotypes from one or more contributors may not be represented in the statistical calculation.

Example: *This example illustrates the use of an unrestricted CPI when all data is above the stochastic threshold, but allelic dropout is possible.*

The interpretation of the entire mixture is that there are two or more contributors with one or more contributors below the stochastic threshold at some larger loci. At this locus, all alleles are above the stochastic threshold, but dropout is possible.



Though the data is above the stochastic threshold, because dropout is possible, this locus is not suitable for the CPI. The CPI would underestimate the portion of the random population who would be included as potential contributors because it would only include individuals with genotypes 11,11 or 11,12 or 12,12. All other genotypes would be excluded. However, given the overall low RFU values, limited number of detected alleles, and assumed minimum number of contributors, one of the true contributors may be genotype 10,12 or even 8,9.

4C.6.5 If a contributor can be assumed to be below the stochastic threshold throughout the profile, the CPE/CPI statistical approach to the mixture as a whole, even if there are loci with no alleles below the stochastic threshold, is not appropriate and should not be used. In these instances, the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the detected alleles.

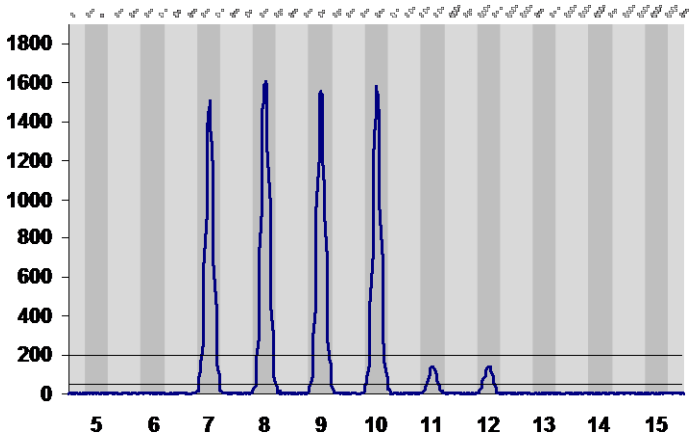
4C.7 The CPI and CPE can be applied to a subset of alleles detected in a mixture (i.e., a restricted CPI/CPE), subject to the limitations described in this Section.

4C.7.1 Given a mixture at a locus with some of the alleles significantly (as defined by laboratory validation) above the stochastic threshold, and one or more alleles below the stochastic threshold, the interpretation might be that the higher RFU alleles are a distinct group, separate

from the contributor(s) of the low-RFU allele(s). The lab may choose to calculate a restricted probability of inclusion utilizing just the group of high level alleles. A restricted CPE/CPI may be applied to multiple major contributors despite the presence of minor contributor(s) alleles below the stochastic threshold.

Example: *This example illustrates the use of a restricted CPI for a mixture of major contributors.*

The interpretation of the entire mixture is that there are three or more contributors with one or more contributors below the stochastic threshold at some larger loci. Additionally, a distinct contrast exists between the major mixture of two individuals and the one or more minor contributor(s).



Given the distinct peak height difference between alleles 7, 8, 9, 10 and alleles 11 and 12, the profile may be separated into a mixture of major contributors and one (or more) minor contributors. The interpretation is that all alleles from the major contributors have been identified, which are all above the stochastic threshold.
The restricted PI at this locus would be calculated as: $(P_7 + P_8 + P_9 + P_{10})^2$ and would include the following genotypes.

7,7	7,8	7,9	7,10
8,8	8,9	8,10	
9,9	9,10		
10,10			

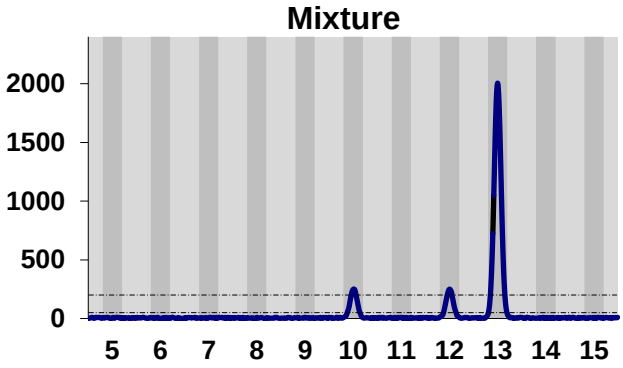
All others would be excluded from the major mixture. Additionally, because the major mixture is interpreted to be only two contributors, all homozygous genotypes should also be excluded as potential contributors even though the CPI calculation includes them. The CPI could not be used to represent all of the contributors at this locus, because the interpretation does not make the assumption of only three contributors.
In contrast, if the interpretation clearly supported only three contributors, the mixture as a whole could be represented with the unrestricted CPI with assumptions, and the assumption must be clearly stated in the report.

4C.7.2 Based on the above example, had the 11 and 12 alleles been greater than the stochastic threshold, but still identified as distinct from the higher-RFU alleles, a restricted CPI is still appropriate for the group of higher-RFU alleles. Additionally, a second general CPI or CPE could be calculated using all six alleles to represent the mixture as a whole, as long as there is no expectation of dropout (see Example 4C.6.4).

4C.8 Interpretations that include peaks that are indistinguishable from stutter should also include such peaks in the unrestricted CPI and CPE statistics. If such peaks are below the stochastic threshold, this might render the locus inappropriate for use in the CPI statistic.

Example: *This example illustrates the inclusion of peaks that are indistinguishable from stutter into a CPI calculation.*

The interpretation of the entire mixture is that there are two contributors. Additionally, a distinct contrast exists between the major contributor and the minor contributor.



Bin	RFU
10	250
12	250 (below the stutter threshold)
13	2000

The major contributor is generally represented using the random match probability (RMP; see section 4A). An individual included in the minor contributor may be represented using the unrestricted CPI. The peak in the stutter bin (i.e., bin 12) cannot be distinguished from actual alleles of the minor contributor based on the laboratory's peak height ratio expectations. As a result, this peak in the stutter bin (12) should be included in the unrestricted CPI calculation [i.e., PI = $(P_{10}+ P_{12}+ P_{13})^2$] since it could be from the minor contributor to the mixture.

4C.8.1 The application of a restricted CPI or CPE might separate the peaks that are indistinguishable from stutter as distinct from a major mixture. In such instances, the restricted CPI or CPE of the major contributors could be calculated (see Example 4C.7.1).

Glossary

Allele: a form of a gene that is located at a specific location on a specific chromosome. Alleles targeted in STR analysis vary in length.

Analytical threshold: the minimum height requirement at and above which detected peaks can be reliably distinguished from background noise; peaks above this threshold are generally not considered noise and are either artifacts or true alleles.

Artifact: a non-allelic product of the amplification process (e.g., stutter, non-templated nucleotide addition, or other non-specific product), an anomaly of the detection process (e.g., pull-up or spike), or a by-product of primer synthesis (e.g., "dye blob").

Assumed contributor: an individual whose DNA on an item of evidence is reasonably expected.

Binary model: an interpretation scheme in which there are only two values (possible or not possible) for each decision (e.g., a peak is either "an allele" or "not an allele"; a genotype is "included" or "not included").

Composite profile: a DNA profile generated by combining typing results from different loci obtained from multiple injections of the same amplified sample and/or multiple amplifications of the same DNA extract. When separate extracts from different locations on a given evidentiary item are combined prior to amplification, the resultant DNA profile is not considered a composite profile.

Conditional: an interpretation category that incorporates assumption(s) as to the number of contributors and/or the presence of specific contributor(s).

CPE: combined probability of exclusion; produced by multiplying the probabilities of inclusion from each locus and subtracting the product from 1; (i.e., 1-CPI).

CPI: combined probability of inclusion; produced by multiplying the probabilities of inclusion from each locus.

Deconvolution: separation of contributors to a mixed DNA profile based on quantitative peak height information and any underlying assumptions.

Deduced profile: inference of an unknown contributor's DNA profile after taking into consideration the contribution of a known/assumed contributor's DNA profile based on quantitative peak height information.

Differential Degradation: a DNA typing result in which contributors to a DNA mixture are subject to different levels of degradation (e.g., due to time of deposition), thereby impacting the mixture ratios across the entire profile.

Distinguishable Mixture: a DNA mixture in which relative peak height ratios allow deconvolution of the profiles of major/minor contributor(s).

Dropout: when one or more alleles present in a sample are not observed above the analytical threshold.

Evidence sample: also known as Questioned sample.

Exclusion: a conclusion that eliminates an individual as a potential contributor of DNA obtained from an evidentiary item based on the comparison of known and questioned DNA profiles (or multiple questioned DNA profiles to each other).

Exclusionary hypothesis (H_e): also referred to as H_d or H_2; the term typically used in the denominator of the likelihood ratio to represent the defense hypothesis, which does not include the person of interest as a contributor to the DNA profile.

Genotype: results of autosomal STR analysis of an individual at one or more genetic loci.

Guidelines: a set of general principles used to provide directions and parameters for decision making.

Heterozygote: an individual having different alleles at a particular locus; usually manifested as two distinct peaks for a locus in an electropherogram.

Homozygote: an individual having the same (or indistinguishable) alleles at a particular locus; manifested as a single peak for a locus in an electropherogram.

Inclusion: a conclusion for which an individual cannot be excluded as a potential contributor of DNA obtained from an evidentiary item based on the comparison of known and questioned DNA profiles (or multiple questioned DNA profiles to each other).

Inclusionary hypothesis (H_i): also referred to as H_p or H_1; the term typically used in the numerator of the likelihood ratio to represent the prosecution hypothesis, which includes the person of interest as a contributor to the DNA profile.

Inconclusive: a determination that no conclusion (i.e., inclusion/exclusion) can be drawn from the comparison of a reference sample to suitable data. This could also result from statistical analyses that fail to provide sufficient support for an inclusion or exclusion.

Indistinguishable mixture: a DNA mixture in which relative peak height ratios are insufficient to attribute alleles to individual contributor(s).

Intimate sample: a biological sample from an evidence item that is obtained directly from an individual's body; it is not unexpected to detect that individual's allele(s) in the DNA typing results.

Known sample: biological material for which the identity of the donor is established and used for comparison purposes (referred to as a "K").

Likelihood ratio (LR): the ratio of two probabilities of the same event under different and mutually exclusive hypotheses; typically the numerator contains the prosecution's hypothesis and the denominator the defense's hypothesis.

Locus: the specific physical location of a gene on a chromosome. In forensic DNA analysis, it refers to the specific sites being tested (e.g., D3S1358, vWA or D5S818).

Major contributor(s): an individual(s) who can account for the predominance of the DNA in a mixed profile.

Masked allele: an allele of the minor contributor that may not be readily distinguishable from the alleles of the major contributor or an artifact.

Minor contributor(s): an individual(s) who can account for the lesser portion of the DNA in a mixed profile.

Mixture: a DNA typing result originating from two or more individuals.

Mixture ratio: the relative proportion of the DNA contributions of multiple individuals to a mixed DNA typing result, as determined by the use of quantitative peak height information; when expressed as a percentage it is termed a mixture proportion.

Noise: background signal detected by a data collection instrument.

Obligate allele: an allele in a mixed DNA typing result that is (a) foreign to an assumed contributor, or (b) based on quantitative peak height information, determined to be shared with the assumed contributor.

Partial profile: a DNA profile for which complete typing results are not obtained at all tested loci due, for example, to DNA degradation, inhibition of amplification and/or low- quantity template.

Peak height ratio (PHR): the relative proportion of two alleles at a given locus, as determined by dividing the peak height of an allele with a lower relative fluorescence unit (RFU) value by the peak height of an allele with a higher RFU value, and then multiplying this value by 100 to express the PHR as a percentage; used as an indication of which alleles may be heterozygous pairs and also in mixture deconvolution.

Probabilistic genotyping: the use of biological modeling, statistical theory, computer algorithms, and probability distributions to calculate likelihood ratios (LRs) and/or infer genotypes for the DNA typing results of forensic samples.

Probability of exclusion (PE): the percentage of the population that can be excluded as potential contributors to a DNA mixture at a given locus.

Probability of inclusion (PI): the percentage of the population that can be included as potential contributors to a DNA mixture at a given locus; also known as Random Man Not Excluded.

Profile probability: see RMP.

Questioned sample: biological sample recovered from a crime scene or collected from persons or objects associated with a crime (referred to as a "Q").

Random Match Probability (RMP): the probability of randomly selecting an unrelated individual from the population who could be a potential contributor to an evidentiary profile.

Reference sample: also referred to as known sample or reference standard.

Restricted: referring to a statistical approach conditioned on the number of contributors and with consideration of quantitative peak height information and inference of contributor mixture ratios; used to limit the genotypic combinations of possible contributors.

Single-source profile: DNA typing results determined to originate from one individual based on peak height ratio assessments and the number of alleles at given loci.

Source attribution: a declaration which identifies an individual as the source of the DNA that produced an evidentiary single-source or deduced contributor profile; this statement is based on a statistical estimate that meets or exceeds a laboratory defined threshold.

Stochastic effects: the observation of intra-locus peak imbalance and/or allele drop-out resulting from random, disproportionate amplification of alleles in low-quantity template samples.

Stochastic threshold: the peak height value 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.

Stutter: a minor peak typically observed one repeat unit smaller than a primary STR allele resulting from strand slippage during amplification.

Theta (θ): a value used to adjust statistical calculations that rely on population databases to correct for substructure within populations.

Uninterpretable: the determination that DNA results at one or more loci cannot be interpreted due to poor or limited data quality.

Unrestricted: referring to a statistical approach performed without consideration of quantitative peak height information and inference of contributor mixture ratios; for CPE/CPI, it may or may not be conditioned on the number of contributors.

Unsuitable (for comparison): uninterpretable results or those that fail to meet quality assurance requirements as defined by the laboratory and as a result are not usable for comparisons.

References and Suggested Readings

Ayres, K.L. (2000) *Relatedness testing in subdivided populations*. Forensic Sci. Int. 114:107-115.

Balding, D.J., and Nichols, R. (1993) *DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands*. Forensic Sci. Int. 64:125-140.

Bär, W., Brinkmann, B., Lincoln, P., Mayr, W. R., and Rossi, U. (1994) *DNA recommendations – 1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems*. Int. J. Legal Med. 107: 159-160.

Bär, W., Brinkmann, B., Budowle, B., Carracedo, A., Gill, P., Lincoln, P., Mayr, W. R., and Olaisen, B. (1997) *DNA recommendations – further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems*. Int. J. Legal Med. 110: 175-176.

Bieber, F.R., Buckleton, J.S., Budowle, B., Butler, J.M., and Coble, M.D. (December 2016) *Evaluation of forensic DNA evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion*. BMC Genetics 17:125; available at doi:10.1186/s12863-016-0429-7.

Bill, M., Curran, J., Gill, P., Clayton, T., Pinchin, R., Healy, M. and Buckleton, J. S. (2005) *Pendulum – a guideline based approach to the interpretation of STR mixtures*. Forensic Sci. Int. 148: 181-189.

Bille, T., Bright, J., and Buckleton, J. (2013) *Application of Random Match Probability Calculations to Mixed STR Profiles*. J. Forensic Sci. 58: 474-485.

Buckleton, J. S., Curran, J.M. and Gill, P. (2007) *Towards understanding the effect of uncertainty in the number of contributors to DNA stains*. FSI Genetics 1 (1) 20–28.

Buckleton, J. S. and Curran, J. M. (2008) *A discussion of the merits of random man not excluded and likelihood ratios*. FSI Genetics 2: 343-348.

Buckleton, J. S., Bright J., and Taylor, D. *Forensic DNA Evidence Interpretation. 2nd edition*. CRC Press. Boca Raton, 2016.

Butler, J.M. (2015) *Advanced Topics in Forensic DNA Typing: Methodology (2012) and Interpretation (2015).* San Diego: Elsevier Academic Press.

Clayton, T., Whitaker, J.P., Sparkes, R.L., and Gill, P. (1998) *Analysis and interpretation of mixed forensic stains using DNA STR profiling*. Forensic Sci. Int. 91:55–70.

Committee on DNA Forensic Science, National Research Council. (1996) *An Update: The Evaluation of Forensic DNA Evidence.* National Academy Press, Washington, DC.

Devlin, B. (1993) *Forensic inference from genetic markers*. Stat Methods Med Res 2(3):241-62.

DNA Advisory Board (2000) *Statistical and population genetic issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s).* Forensic Sci. Comm. 2(3).

Evett, I., Jackson, G., Lambert, J.A., and McCrossan, S. (2000) *The impact of the principles of evidence interpretation on the structure and content of statements*. Science & Justice 40(4): 233-239.

Federal Bureau of Investigation, DNA Advisory Board (2000). *Quality Assurance Standards for Forensic DNA Typing Laboratories*. Forensic Sci. Comm. 2 (3).

Federal Bureau of Investigation (2011) *Quality Assurance Standards for Forensic DNA Testing Laboratories*; available at https://www.fbi.gov/file-repository/quality-assurance-standards-for-forensic-dna-testing-laboratories.pdf/view .

Fung, W.K. and Hu, Y.-Q. (2008) *Statistical DNA Forensics: Theory, Methods and Computation*. Wiley: Hoboken, NJ.

Gill, P., Sparkes, R.L., Pinchin, R., Clayton, T., Whitaker, J.P. and Buckleton, J.S. (1998) *Interpreting simple STR mixtures using allelic peak areas*. Forensic Sci. Int. 91: 41–53.

Gill, P., Sparkes, B. and Buckleton, J.S. (1998) *Interpretation of simple mixtures when artefacts such as stutters are present – with special reference to multiplex STRs used by the Forensic Science Service*. Forensic Sci. Int. 95 (3) 213–224.

Gill, P., Brenner, C.H., Buckleton, J.S., Carracedo, A., Krawczak, M., Mayr, W.R., et al. (2006) *DNA commission of the International Society of Forensic Genetics: recommendations on the interpretation of mixtures*. Forensic Sci. Int. 160: 90–101.

Perez J., Mitchell A.A, Ducasse N., Tamariz J. and Caragine T. (2011) *Estimating the number of contributors to two-, three-, and four-person mixtures containing DNA in high template and low template amounts*. Croat Med J, Jun 52(3): 314-26.

Puch-Solis R., et. al. (2010) *Calculating likelihood ratios for a mixed DNA profile when a contribution from a genetic relative of a suspect is proposed*. Sci. Justice 50(4): 205-209.

Scientific Working Group on DNA Analysis Methods (SWGDAM) (2010) *SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories*; available at http://media.wix.com/ugd/4344b0_61b46a0e1a4c41ccb65f719a533b8e29.pdf.

Weir, B. S., Triggs, C.M., Starling, L., Stowell, L.I., Walsh, K.A.J. and Buckleton, J.S. (1997) *Interpreting DNA mixtures*. J. Forensic Sci. 42 (2) 213–222.

Other suggested readings cited on http://www.cstl.nist.gov/strbase/.