ASB Standard 186, First Edition
202X

Standard for Forensic Autosomal STR DNA Statistical Analyses

ASB ACADEMY STANDARDS BOARD

Standard for Forensic Autosomal STR DNA Statistical Analyses

ASB Approved Xxxxxxx 202X

ANSI Approved Xxxxxxxx 202X

410 North 21st Street
Colorado Springs, CO 80904

Foreword

Detailed and comprehensive protocols are needed so that appropriate statistical calculations for evidentiary DNA profiles are performed consistently by qualified analysts. These calculations are provided to aid in the assessment of an inclusion or association of a DNA profile with the profile of a known individual. Specific requirements for a laboratory's protocol for performing statistical analyses, protocol verification, and case record documentation are provided. These requirements include descriptions of when statistical calculations are, and are not, needed to be performed in casework; descriptions of the statistical methods available for use in the laboratory and relevant supporting information for their use; the use of assumptions in the calculations; documentation of the data used and relevant information for the calculations performed; and documented verification of and consistency of protocol use in the laboratory.

This standard includes general requirements for calculations commonly performed in forensic DNA testing laboratories. These include the likelihood ratio (LR), the random match probability (RMP), and the combined probability of inclusion/exclusion (CPI/CPE). This document applies to any manual calculations or software using fixed formulas and continuous or semi-continuous (also termed discrete) methods. This document applies to calculations resulting from the comparison of DNA profiles for identity testing (i.e., could the DNA have come from the same source?) as well as biological relationship testing (i.e., could the individuals be related?). While this standard applies directly to testing performed using the polymerase chain reaction (PCR) amplification of autosomal short tandem repeat loci (STR), many of the general requirements may also apply to other types of DNA testing and analysis. This standard applies only to statistical calculations performed at the sub-source and sub-sub-source levels in the hierarchy of propositions. Additional information regarding the application of and specific requirements for the various statistical calculation methods routinely used in forensic DNA testing laboratories may be found in Annex A and the Bibliography (Annex B). Specifics for biological relationship testing are not directly addressed in this standard; see AABB Standards for Relationship Testing Laboratories (referenced in Annex B, Bibliography) regarding statistics for biological relationship testing. This standard does not apply to any calculations performed for CODIS data entry (e.g., MME and MRE).

This standard is intended to be used in conjunction with the following ANSI/ASB Standards and Best Practice Recommendations: (1) ANSI/ASB Standard 018, *Standard for Validation of Probabilistic Genotyping Systems*, First Edition, 2020; (2) ANSI/ASB Standard 020, *Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol*, First Edition, 2018; (3) ANSI/ASB Standard 040, *Standard for Forensic DNA Interpretation and Comparison Protocols*, First Edition, 2019; (4) ANSI/ASB Best Practice Recommendation 114, *Best Practice Recommendations for Internal Validation of Software used in Forensic DNA Laboratories*, First Edition, 2022; (5) ANSI/ASB Standard 123, *Standard for Routine Internal Evaluation of a Laboratory's DNA Interpretation and Comparison Protocol*, First Edition, 2024; and (6) ANSI/ASB Standard 139, *Standard for Reporting DNA Conclusions*, First Edition, 2024; and the FBI's Quality Assurance Standards for Forensic DNA Testing Laboratories; as well as any current or future standards or recommendations that provide guidance for the appropriate use of specific statistical calculation methods and software.

The American Academy of Forensic Sciences established the Academy Standards Board (ASB) in 2015 with a vision of safeguarding Justice, Integrity, and Fairness through Consensus Based American National Standards. To that end, the ASB develops consensus based forensic standards within a framework accredited by the American National Standards Institute (ANSI), and provides training to support those standards. ASB values integrity, scientific rigor, openness, due process,

collaboration, excellence, diversity and inclusion. ASB is dedicated to developing and making freely accessible the highest quality documentary forensic science consensus Standards, Guidelines, Best Practices, and Technical Reports in a wide range of forensic science disciplines as a service to forensic practitioners and the legal system.

This document was revised, prepared, and finalized as a standard by the DNA Consensus Body of the AAFS Standards Board. The draft of this standard was developed by the Human Forensic Biology Subcommittee of the Organization of Scientific Area Committees (OSAC) for Forensic Science.

Questions, comments, and suggestions for the improvement of this document can be sent to AAFS-ASB Secretariat, asb@aafs.org or 410 N 21st Street, Colorado Springs, CO 80904.

All hyperlinks and web addresses shown in this document are current as of the publication date of this standard.

ASB procedures are publicly available, free of cost, at www.aafs.org/academy-standards-board.

Keywords: *statistics, statistical analysis, protocol, protocol verification, consistency, random match probability (RMP), combined probability of inclusion or exclusion (CPI/CPE), likelihood ratio (LR), probabilistic genotyping, DNA profile, DNA mixtures*

Table of Contents *(to be finalized prior to publication)*

Standard for Forensic Autosomal STR DNA Statistical Analyses

1 Scope

This standard provides requirements for the laboratory protocol for performing statistical analyses, including verification and consistent application of the protocol. This document also includes requirements for documentation in the case record of information regarding the statistical calculations. This standard applies to testing performed using the polymerase chain reaction (PCR) amplification of autosomal loci consisting of short tandem repeats (STR). This standard applies only to statistical calculations for sub-source and sub-sub-source levels in the hierarchy of propositions.

2 Normative References

There are no normative reference documents. Annex B, Bibliography, contains informative references.

3 Terms and Definitions

For purposes of this document, the following definitions apply.

3.1
combined probability of exclusion
CPE
The probability that a randomly selected individual would be excluded as a contributor to a DNA mixture; produced by multiplying the probabilities of inclusion from each locus chosen for inclusion and subtracting the product from 1 (i.e., 1-CPI).

3.2
combined probability of inclusion
CPI
The probability that a randomly selected individual would be included as a possible contributor to a DNA mixture; produced by multiplying the probabilities of inclusion from each locus chosen for inclusion.

3.3
conditioning
The act of assuming one or more pieces of information when assigning a conditional probability.

NOTE 1 The information might be the profile of an individual, or profiles of a set of individuals, who are assumed to have contributed DNA to the evidentiary item under a particular proposition, or it might simply be the assumption that a particular proposition is true.

NOTE 2 Any events (or information) that have been used for conditioning are placed to the right of the conditioning bar in a conditional probability expression.

3.4
likelihood ratio
LR
The ratio of two conditional probabilities of the same event under mutually exclusive hypotheses. The general formula is: $LR = Pr(E|H_1,I)/Pr(E|H_2,I)$.

NOTE For DNA testing, a statement of comparison of the probability of the evidence (E) (i.e., the DNA profile), given two competing hypotheses, [e.g., inclusionary (H_1) or exclusionary (H_2) for an individual or specific sets of individuals], and in the context of relevant information (I).

3.5
probabilistic genotyping
The use of biological modeling (i.e., statistical modeling informed by biological data), statistical theory, computer algorithms, and/or probability distributions, to infer genotypes and/or calculate likelihood ratios.

3.6
proposition
A statement that is either true or false.[a]

NOTE In the context of likelihood ratio calculations, propositions are formulated in pairs. The paired propositions are mutually exclusive (i.e., both cannot be correct at the same time) and provide one possible explanation for the evidence observed.

3.7
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.[b]

3.8
theta correction
θ
A value used to adjust statistical calculations that rely on population databases to correct for substructures within populations.[2]

4 Requirements

4.1 General

Refer to Annex A, Information on Random Match Probability (RMP), Likelihood Ratio (LR), and Combined Probability of Inclusion or Exclusion (CPI/CPE), for additional information regarding the statistical values applicable to autosomal STR DNA testing and the following requirements.

4.2 Protocol

4.2.1 The laboratory shall have and follow a protocol for performing statistical analyses derived from validation studies that includes the requirements in 4.2.2 through 4.2.14.

[a] Butler, J.M., Iyer, H., Press, R., Taylor, M.K., Vallone, P.M., and Willis, S. (2024*) DNA Mixture Interpretation: A NIST Scientific Foundation Review*. (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351.

[b] Scientific Working Group on DNA Analysis Methods (SWGDAM). *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Laboratories*.

4.2.2 The protocol shall include descriptions of scenarios where statistical analyses are, and are not, required to be performed.

NOTE 1 No statistical analysis is required for an exclusion determined manually.

NOTE 2 Statistical analyses on the evidentiary DNA profile may be performed prior to the comparison to known reference data, but are not required (e.g., to provide important or relevant information for a particular case when no reference sample is available).

NOTE 3 Statistical analyses in support of an association between two sets of evidentiary data may be calculated and provided, but are not required (e.g., evidentiary DNA profiles in common between two blood stains of unknown origin found at two different crime scenes to aid in assessing the possibility they may be from the same individual).

4.2.3 The protocol shall include a requirement that any reported association of an evidentiary DNA profile to a DNA profile from a known individual be supported by a statistical analysis that includes data from each locus used for comparison.

NOTE This does not apply to the inclusion of an individual whose DNA is reasonably expected to be present on the item of evidence based on how and from where the biological sample was collected, as defined by the laboratory protocol and/or as documented in the case record for a specific case scenario (e.g., swabbing of an area of an individual's body; clothing worn in close contact with the individual's body).

4.2.4 The protocol shall include a requirement that statistical analyses are performed only on loci deemed suitable for comparison based upon the laboratory's documented interpretation and comparison protocol (e.g., where stochastic phenomena such as allelic drop-out, allelic drop-in, or stutter are not explicitly accounted for in the statistical model being used).

NOTE This requirement is meant to eliminate the practice of omitting loci which do not exhibit the alleles of one or more individuals after a comparison has been performed to the known reference standard. Although such practice has been historically labeled as neutral or conservative, it typically is not, and can be especially problematic with interpretation methods that do not allow explicit modeling of allele drop-out or other stochastic phenomena.

4.2.5 For non-probabilistic genotyping (e.g., manual) methods, the protocol shall include a requirement that statistical analyses be performed only at those loci common to both profiles (e.g., when one of the profiles used for comparison has data at fewer loci than the other profile in the comparison, as in a partial, incomplete profile, or data from different multiplex kits).

4.2.6 The protocol shall include a description of statistical analysis methods available for use in the laboratory, to include the requirements in 4.2.6.1 through 4.2.6.7.

4.2.6.1 When statistical analyses are generated from manual calculations or software (e.g., RMP, CPI/CPE, and LR not from probabilistic genotyping software), the protocol shall provide all equations used in the calculations including the following:

a) for a homozygous genotype at a locus;

b) for a heterozygous genotype at a locus;

c) when a theta (θ) correction factor(s) is used and provide the value of theta used in that calculation;

d) for the possible genotype combinations when data from more than one contributor (i.e., mixture) are present at a locus;

e) for combining genotype frequencies across multiple loci in a DNA profile;

f) for minimum allele frequencies for the population database(s), if used; and

g) for biological relationships, if used.

4.2.6.2 When statistical analyses are generated using probabilistic genotyping software, the protocol shall provide the following.

a) References to the published literature, and any other relevant information (e.g., technical and/or user's manual), for the equations and the calculations used by the software for computing likelihood ratios.

b) The statistical basis for defining inclusion, exclusion, inconclusive, uninformative, and uninterpretable when those terms are used by the laboratory.

c) A requirement that when individual likelihood ratio values support an association of multiple persons to a mixed DNA profile, an analysis shall be performed using proposition pairs that test whether the DNA profile data support or fail to support the association of the multiple persons together in the mixture.

EXAMPLE

For an apparent two-person DNA mixture, if the proposition set of Person A + 1 unknown/2 unknowns generates an LR supporting an association of Person A, and Person B + 1 unknown/2 unknowns also generates an LR supporting an association of Person B, then the proposition set of Person A + Person B/2 unknowns will also be calculated to evaluate the association, or non-association, of both individuals together given the DNA profile observed. Similar proposition sets are evaluated for mixtures having more than two apparent contributors.

4.2.6.3 If replicate profile data are generated, the protocol shall define when and how the data are used.

4.2.6.4 The protocol shall include a description of when each statistical method can be employed in the laboratory.

4.2.6.5 When multiple validated methods are available in the laboratory for calculating statistical values and more than one may be appropriately used for a particular case sample scenario and/or DNA profile per requirement 4.2.6.4, the protocol shall state which statistical analysis method shall be used and/or how to determine which method will be used.

4.2.6.5.1 For single source DNA profiles, the protocol may permit the use of RMP and LR calculations; in this situation, the protocol shall clearly define which calculation is used under which scenario.

4.2.6.5.2 For a mixed DNA profile, a CPI/CPE, RMP, and/or LR calculation may be appropriate for use; the protocol shall clearly define which calculation is used.

NOTE A common scenario where this may be relevant is an assumed two-person DNA mixture obtained from a vaginal, oral, or breast swab where the DNA profile from the known female contributor is available and each of the approaches may be applicable.

4.2.6.6 If the laboratory utilizes the CPI/CPE calculation, the protocol shall include a requirement that the CPI/CPE calculation shall only be used as provided in Annex A and the Bieber et al. "Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion."

4.2.6.7 The protocol shall include a requirement that two or more conceptually different statistics used for autosomal STR data shall not be combined.

NOTE Specific examples include not multiplying a RMP with either a CPI or LR, and not multiplying a CPI with a LR.

4.2.7 The protocol shall include information regarding the appropriate scenario(s) for the use of assumptions and/or conditioning information used in the propositions that may impact the statistical analyses.

4.2.7.1 The protocol shall provide a) the types of assumptions that can be made, b) when those assumptions can be made, and c) how those assumptions are incorporated into the statistical analysis.

NOTE Such assumptions may include, but are not limited to, the number of contributors, the presence of possible artifacts (e.g., stutter) and/or stochastic effects, and the presence of assumed contributors.

4.2.7.2 The protocol shall include a requirement to document in the case record any assumptions used that may impact the statistical analyses [also see requirement 4.4.1 e)].

4.2.7.3 The protocol shall define the conditioning information that may be used in propositions to calculate likelihood ratios.

4.2.7.3.1 Before the subsequent conditioning of the evidentiary profile interpretation when using probabilistic genotyping, the laboratory shall establish a minimum likelihood ratio threshold for the conditioning profile when the DNA typing results are insufficient, as defined by the laboratory, to support conditioning based solely on a manual evaluation of the data.

4.2.7.3.2 The established minimum likelihood ratio threshold shall be based on validation studies.

NOTE A laboratory may use different minimum likelihood ratio thresholds for decisions regarding conditioning for different types of evidence.

4.2.8 The protocol shall define the documented source of each population database used in any statistical analyses.

4.2.9 The protocol shall define when and how an alternate population database and/or theta correction value shall be applied.

4.2.10 The protocol shall define the appropriate validated software and version number used for each type of statistical analysis.

4.2.11 The protocol shall define known limitations for the use of any formulas and any software based on external or internal validation studies, and situations where profile data cannot be used for statistical calculations.

NOTE Some possible limitations include the number of contributors that may be assumed when using certain formula(s) or software, limitations established through the laboratory validation studies, functions that have not been validated by the laboratory, and when data are insufficient for using the statistical analysis method (e.g., the inability to use CPI/CPE calculations if there is a reasonable risk that data are missing from a locus).

4.2.12 The protocol shall define when the variable input parameters may be modified and the appropriate values to be used for any parameter or input value that can be changed by the analyst in the software based on validation studies.

4.2.13 The protocol shall define a) the relevant output parameters and diagnostics that shall be evaluated, b) their acceptable values based on validation studies, and c) actions to be taken when either the parameters or diagnostics are outside the acceptable values, for each method used to generate statistical values.

4.2.14 The protocol shall include a requirement that a new statistical analysis shall be performed when subsequent review of the profile data alters how it was used in the original statistical analysis.

4.3 Protocol Verification

4.3.1 The laboratory shall verify and document that application of each protocol for performing statistical analyses generates appropriate values and that each protocol is followed consistently within the laboratory for all types of DNA profiles typically interpreted and compared by the laboratory. This verification shall include the requirements in 4.3.2 through 4.3.6.

4.3.2 Methods, equations, software, etc. shall not be used for statistical calculations without the prerequisite validation, protocol development, and protocol verification.

4.3.3 Verification of each protocol shall be performed on single source and mixed DNA samples of known origin using contributors different from those used a) in the initial validation studies for the amplification kit and statistical analysis software or b) to establish the statistical analysis protocol.

4.3.3.1 The data for all contributors used in the verification shall be known and available for review.

4.3.3.2 The data used for verification shall be generated and processed using the laboratory's validated testing procedures.

4.3.3.3 The data sets shall span the range of data anticipated to be interpreted by the laboratory.

4.3.4 The acceptable range of variability in the statistical values generated to demonstrate consistency shall be defined by the laboratory and based on the laboratory's validation studies. Verification shall demonstrate that use of the protocol:

a) returns the same value within the laboratory for the same DNA profile when using procedures without an element of randomness (e.g., PopStats or non-probabilistic genotyping software);

b) for probabilistic genotyping software having an element of randomness, results in consistent values between different runs with the same inputs, as defined by the laboratory based on validation studies for both true contributors and non-contributors; and

c) provides consistency among analysts in the laboratory for the calculated statistical values using examples representative of the range of samples handled by the laboratory.

4.3.5 Verification of each protocol shall be performed on existing, modified, and new statistical analysis protocols.

4.3.5.1 Additional validation studies and/or protocol development shall be necessary if deficiencies in the protocol or inconsistencies within the laboratory are identified through this verification process.

4.3.5.2 Any subsequent modifications to any DNA testing or data interpretation protocol shall include an evaluation for its impact on DNA statistical calculations.

4.3.6 Verification of each protocol shall be completed prior to implementation of the protocol for casework.

4.4 Case Record Documentation

4.4.1 The laboratory shall document the following in the case record for each statistical analysis performed.

a) Each population database used and the source of each database.

b) Each statistical analysis method used, and, if applicable, the software program and version number used.

c) Each theta correction factor value used.

d) The genetic loci and data used for statistical calculations.

e) The assumptions made when performing a statistical analysis, including but not limited to the number of contributors and/or assumed contributors, and in the case of paternity or kinship analysis, any alleged or assumed biological relationships.

f) All statistical analyses performed, including analyses performed using different assumptions and/or different propositions (e.g., conditioning on different DNA profiles), regardless of whether the statistical analysis is reported by the laboratory.

g) The actual value used by the analyst with each statistical analysis for any parameter or input value that can be changed in the software (e.g., random number seeds, number of Markov Chain Monte Carlo iterations, probability of drop-out and/or drop-in).

h) The reason a statistical calculation is not performed for any interpreted and compared profile.

Annex A
(informative)

Information on Random Match Probability (RMP), Likelihood Ratio (LR), and Combined Probability of Inclusion or Exclusion (CPI/CPE)

A.1 General

Additional information regarding the three major types of statistical values calculated for forensic autosomal STR DNA profiles is provided below. For Random Match Probability (RMP), Likelihood Ratio (LR), and Combined Probability of Inclusion/Exclusion (CPI/CPE):

1) these three terms refer only to statistical values and their respective calculations.

2) the use of all three statistical calculation methods requires prior independent interpretation of the STR DNA profile, which includes assessment of stutter and other artifacts, assessment of degradation, determination of the alleles and loci suitable for comparison, the risk of allele drop-out and drop-in at each locus and across the profile, peak heterozygosity, and the assumed number of contributors. None of these statistical calculation methods are interpretation methods and play no direct role in the interpretation of the DNA profile.

3) a single statistical calculation method is to be used across all loci that are suitable for comparison in a given profile. Per requirement 4.2.6.7, different statistical calculations for a single profile cannot be combined.

4) there may be situations where the insufficiency of the data may render the profile unsuitable for statistical calculations. There may be situations where the insufficiency of the data at one or more loci prevent those loci from being used for statistical calculations.

5) the calculated values are estimates and will vary depending on the allele frequency database used, the quality of the DNA profile, the number of loci having data, the model and formulas used, and many other variables that impact the calculations.

A.2 Random Match Probability (RMP)

Some of the key features and use of the Random Match Probability statistical calculation method for STR DNA single source and mixture profiles are provided in this section.

1) The RMP may be used for single source profiles and for some mixtures.

    a) For mixtures, the RMP may be calculated for one contributor to a mixture, a subset of contributors, or for the combined genotypes of all contributors. Within a mixture the RMP may be used for:

        i) single source profiles that may be resolved (e.g., single major or minor contributor, deduced single contributor when using the genotypes from one or more assumed contributors in the determination of possible genotypes) or

ii) multiple contributor profiles by considering the combinations of possible genotypes at a locus (e.g., two contributor profiles) and summing the probabilities for all genotypes included at the locus. This has sometimes been referred to as modified RMP or restricted RMP.

b) The assumed number of contributors to the DNA mixture and the genotypes from any assumed contributor(s) limit, or restrict, the possible genotypes at a locus that are then used for the calculation of the RMP.

c) It may be practical to limit the RMP calculation to profiles, or the portion of a profile, with a defined maximum number of contributors.

d) The RMP can be used for profiles where stochastic effects may be present.

2) The equations using Recommendation 4.1 of the NRC II (1996) for RMP calculations are listed in a) through e).

a) For homozygous loci, the equation is $p^2 + p(1-p)\theta$, where p is the frequency of allele P at a single locus and $\theta = 0.01$ (for most populations in the United States) or 0.03 (for some isolated populations).

b) For heterozygous loci, the equation is 2pq, where p is the frequency of allele P at a single locus and q is the frequency of allele Q at the same locus.

c) For single alleles at a locus for which the second allele cannot be determined (e.g., due to possible allele drop-out or allele masking at a possible shared allele), one of the three following equations may be used:

i) 2p;

ii) $2p-p^2$ or

iii) $p^2 + 2p(1-p)$, where p is the frequency of the single obligate allele P.

NOTE The equations in 2) and 3) are two notations of the same equation.

d) The product rule is used to calculate the RMP across multiple independent loci.

e) Equations using Recommendation 4.2 of the NRC II (1996) may also be used. These equations provide corrections for both homozygous and heterozygous genotypes.[c]

3) The RMP can be estimated by the frequency of occurrence for a given genotype or set of genotypes, in a particular reference population, that make up the profile of a DNA contributor

[c] In the NRC II, equations 4.1a and 4.1b are identical algebraically to equations 4.10a and 4.10b when theta is zero. However, the interpretation of these two sets of equations is fundamentally different. Equations 4.1 represent the probability of occurrence of a genotype in a population, whereas equations 4.10 represent the probability of occurrence of a genotype in a population GIVEN that this genotype has been observed in a known individual. The latter is often referred to as a conditional match probability.

among random unrelated individuals. It is commonly expressed as 1 in X number of individuals by inverting the resulting frequency after applying the product rule across all loci.

4) The RMP is calculated for the genotypes of the single source or mixed evidentiary DNA profile independently of (and even prior to) comparison to the profile from any known individual (other than assumed contributors) since the calculation is based on the evidence data alone.

a) If a subset of loci in the evidence profile is used to calculate the RMP, then the selection of those loci is determined independently of (and even prior to) comparison to any reference profile.

b) When profiles can be resolved in a DNA mixture, different RMP values are calculated for each component [e.g., one RMP for the major contributor(s), and one RMP for the minor contributor(s)].

A.3 Likelihood Ratio (LR)

Some of the key features and use of the Likelihood Ratio method for STR DNA single source and mixture profiles are provided in this section.

1) An LR may be calculated for single source profiles and mixtures.

a) A probabilistic (semi-continuous, continuous) LR can be calculated for profiles where allele drop-out and/or drop-in may have occurred.

b) A binary LR (non-probabilistic LR) cannot be used for profiles where allele drop-out and/or drop-in may have occurred.

2) An LR is a ratio of probabilities of observing the evidence (i.e., DNA profile obtained) under opposing propositions. It is NOT a measure of frequency or a probability.

3) The general equation for an LR is:

$$LR = \frac{\Pr(E \mid H_1, \mathrm{I})}{\Pr(E \mid H_2, \mathrm{I})} \quad \text{(A.1)}$$

where:

Pr = Probability,

E = Evidence,

H_1 = Hypothesis 1,

H_2 = Hypothesis 2, and

I = relevant

Information in formulating the propositions and assigning the probabilities. Propositions may be referred to as proposition 1/proposition 2, prosecution/alternate propositions (H_p/H_A,

respectively), prosecution/defense propositions (H_p/H_d, respectively), inclusionary propositions/exclusionary propositions, or other terms that communicate the propositions are different from one another.

a) A proposition represents the set of contributors, known and unknown, who may have contributed to the observed DNA profile. There is no requirement that a particular proposition is true.

b) The propositions depend on case information and the claims (or reasonably assumed claims) of each of the parties. The propositions may be changed at the request of either party.

c) By definition, the two propositions are mutually exclusive. At least one element of the proposition is different so that they may not both be true at the same time [e.g., Proposition 1 (H_1) states the Person of Interest (POI) is the source of the DNA and Proposition 2 (H_2) states a random, unrelated person in the population is the source of the DNA], or the value of the LR will equal 1.

d) A particular contributor genotype may be known or assumed in a proposition.

i) A conditioning profile is a profile that is assumed to be present in both propositions.

ii) A conditioning profile may be a profile assumed to be present due to the collection and/or origin of the evidence item (e.g., intimate sample) or it might simply be the assumption that a particular profile is present under a given set of propositions.

e) To prevent a major contributor from having undue influence on the weight of the evidence for a minor contributor, consideration is to be given to calculating a separate LR for each included contributor as well as an LR for the contributors together per requirement 4.2.6.2 c). Conditioning profiles may be useful in this scenario.

f) For a binary LR calculation, the weight given to a plausible genotype is 1 and the weight given to an implausible genotype is 0 (hence the name “binary”).

g) For a probabilistic LR calculation, the weight given to a genotype can vary between 0 and 1.

h) The weights of the same genotypes may differ for different propositions in the probabilistic LR calculation.

4) An LR is reported as a ratio of the probabilities of the evidence given the propositions, and not as a ratio of the probabilities of the propositions. For example, appropriate statements include: "The evidence is LR times more likely to be observed if Proposition 1 (H_1) is true rather than if Proposition 2 (H_2) is true” or “It is LR times more likely that the DNA profile would be observed if Proposition 1 (H_1) is true rather than if Proposition 2 (H_2) is true.”

a) It is expressed as an LR; it is not expressed as 1 in X number of individuals.

b) For a single source profile, when theta is equal to zero, often the LR and RMP values are numerically the reciprocal of each other; however, they answer fundamentally different questions.

5) A given LR is only for the propositions stated under the relevant information (I). A new LR calculation is needed if there are any changes to a proposition or relevant information.

   a) The value of an LR will change when the data and/or propositions change.

   b) LRs generated under the same set of propositions using probabilistic genotyping software with an element of randomness will generally vary within an expected limited range.

6) An LR calculation can return a value less than one [or negative log(LR)], which communicates that more weight of evidence is given to the proposition in the denominator.

7) A probabilistic or a binary LR calculation can return a value of 1 [or log(LR) of 0], which communicates that equal weight of evidence is given to both propositions. Neither proposition is supported over the other.

8) A probabilistic or a binary LR calculation can return a value greater than 1 [or positive log(LR)], which communicates that more weight of evidence is given to the proposition in the numerator.

A.4 Combined Probability of Inclusion (CPI) and Combined Probability of Exclusion (CPE)[d]

Some of the key features and use of the Combined Probability of Inclusion (CPI) and Combined Probability of Exclusion (CPE) statistical calculation method for mixed STR DNA profiles are provided in this section.

1) CPI/CPE is also referred to as Random Man Not Excluded (RMNE).

2) CPI/CPE is only used to provide statistical calculations for a limited subset of mixed DNA profiles. This is not used for single source DNA profiles.

   a) This calculation is most applicable for use with DNA profiles generated from the amplification of sufficiently high amounts of DNA such that stochastic effects, if present, are negligible, and have no impact on the interpretation, comparison, and ability to generate statistical frequency calculations.

   b) Generally, this is most applicable for use with DNA profiles from two-person DNA mixtures or three-person mixtures having two major contributors, where the CPI/CPE is calculated only for the two major contributors.

   c) CPI/CPE is rarely suitable for use with mixtures of three or more contributors, particularly when amplified with high sensitivity kits using recommended procedures, with the possible exception of when two distinguishable major contributor profiles are present. It can only be used with mixtures of three or more contributors when high levels of DNA are observed at

[d] While there is limited support for the continued use of CPI/CPE by forensic DNA testing laboratories, it is recognized that some laboratories are still using this calculation for their casework. Additionally, this section provides guidance for laboratories interested in assessing their prior practices and protocols for the use of CPI/CPE in older casework (e.g., cold cases, post-conviction review). CPI/CPE may be used in a case where there is no other statistical calculation available for use; however, the CPI/CPE calculation is only to be used according to the guidance provided in this Annex A and the Bieber et al. reference in Annex B, Bibliography (as stated in requirement 4.2.6.6).

the loci being interpreted and compared, and no contributor is reasonably expected to have dropped out.

d) CPI/CPE is commonly used for indistinguishable mixed DNA profiles (i.e., unable to associate alleles into genotypes for the contributors due to similarities in peak heights and the inability to assume the genotypes of one of the contributors).

3) CPI/CPE is only used for profiles where there is very high confidence that all alleles, and thus all genotypes, for all contributors are present at each of the loci with data available for interpretation and comparison, and where there is no reason to expect that allele drop-out might have occurred [see A.4 2) b)].

a) Data from loci with one or more alleles below the stochastic threshold cannot be used for comparison or for calculating the CPI/CPE [with the one exception stated in A.4 3) d) below]. The assignment of the number of contributors to the DNA mixture using the entire DNA profile is critical for evaluating the prospect that all genotypes from all contributors are present at each locus.

b) The number of contributors to a DNA mixture is assigned prior to comparison of the DNA profile data to the profile from any known contributor (i.e., independently of any knowledge of data from other profiles).

c) If all alleles at a locus are above the stochastic threshold, but there are only a limited number of alleles as compared to the maximum expected allele count based on the assumed number of contributors (e.g., 1-2 alleles in 2 person mixtures; 1-4 alleles in 3 person mixtures), then the CPI/CPE calculation cannot be used if drop-out best explains the paucity of alleles. Peak heights at other loci and total peak height values at each locus are taken into account when assessing the data and the possibility of allele drop-out.

i) When the alleles from at least one contributor are below the stochastic threshold at multiple loci, it is reasonable to assume that the alleles for that individual will be below the stochastic threshold at all loci based on the mixture ratio of the contributors' DNA; thus, CPI/CPE cannot be used for this profile, even for the one or few loci with all alleles above the stochastic threshold as it is more likely that alleles are missing than the assumption that all alleles are present.

ii) If one or more alleles are missing from a locus, the CPI/CPE value resulting from the use of the existing alleles would underestimate the proportion of possible contributors as compared to the calculation using all of the alleles from all of the contributors. That is, the value calculated would give the appearance of the profile being rarer than it really is; this would overstate the value of the evidence and could be more prejudicial.

NOTE It is not generally accepted practice for rarer values to be reported or presented in testimony when providing a statistical frequency for an individual who cannot be excluded as a possible contributor.

d) Loci with one or more peaks below a defined stochastic threshold may be used in the CPI/CPE calculation only if the total number of alleles present at each locus is consistent with all alleles being present for the assumed number of contributors (e.g., six alleles are present at a locus and the assumption of three total contributors is used).

4) The equations for a CPI/CPE calculation are:

a) Probability of inclusion for a locus = (the sum of allele frequencies)^2 = (p_A + p_B + p_C + … + p_N)^2, where p_A, p_B, p_C, and p_N are the frequencies of alleles A, B, C, and N, respectively, observed at the locus, where it is assumed that all alleles from all contributors to the DNA mixture are present, based on the data observed and the assumed number of contributors to the DNA profile.

i) The value at each locus is the cumulative frequency of all possible heterozygous and homozygous genotypes.

ii) For profiles where the maximum allele count is observed based on the assumed number of contributors to the DNA mixture, the CPI/CPE calculation would still incorporate the frequencies of homozygous genotypes included at that locus; however, individuals with homozygous genotypes could be excluded definitively from that locus during interpretation and comparison based on the assumed number of contributors.

b) The CPI is the product (i.e., multiplied together) of each of the probabilities of inclusion calculated from each locus used in the interpretation.

c) CPE = (1 – CPI); other equations are available in the publications referenced in Annex B, Bibliography.

5) The CPI value is an approximation of the proportion of randomly selected individuals in a particular reference population unrelated to a true contributor in the mixture who would be expected to be included as possible contributors to the DNA mixture. It is commonly reported as 1 in X number of individuals.

a) The CPE value is an approximation of the proportion of randomly selected individuals in a particular reference population unrelated to a true contributor in the mixture who would be excluded as contributors to the DNA mixture. This value may be expressed as Y out of X individuals, but it is sometimes expressed as a percentage.

b) The CPI/CPE value is appropriate for use when related individuals are contributors to the DNA mixture.

c) The CPI/CPE calculation cannot be used to consider the possibility or compute the probability that an untested relative may be a contributor to the DNA mixture.

6) The CPI/CPE is calculated for the mixed DNA profile independently of (and even prior to) comparison of the profile from any known individual since the calculation is based on the questioned profile alone.

a) Only one CPI/CPE value can be calculated for one mixed DNA profile per reference population.

b) A CPI/CPE calculation is based on the questioned profile alone; it is never based on the profile of an individual who cannot be excluded as a contributor.

Additional information regarding CPI/CPE calculations and uses is available in publications referenced in Annex B, Bibliography.

Annex B
(informative)

Bibliography

The following bibliography is not intended to be an all-inclusive list, review, or endorsement of literature on this topic. The goal of the bibliography is to provide examples of publications addressed in the standard.

1] ANSI/ASB Standard 018, *Standard for Validation of Probabilistic Genotyping Systems*, First Edition, 2020.[e]

2] ANSI/ASB Standard 020, *Standard for Validation Studies of DNA Mixtures, and Development and Verification of a Laboratory's Mixture Interpretation Protocol*, First Edition, 2018.[e]

3] ANSI/ASB Standard 040, *Standard for Forensic DNA Interpretation and Comparison Protocols*, First Edition, 2019.[e]

4] ANSI/ASB Best Practice Recommendations 114, *Best Practice Recommendations for Internal Validation of Software used in Forensic DNA Laboratories*, First Edition, 2022.[e]

5] ANSI/ASB Standard 123, *Standard for Routine Internal Evaluation of a Laboratory's DNA Interpretation and Comparison Protocol*, First Edition, 2024.[e]

6] ANSI/ASB Standard 139, *Standard for Reporting DNA Conclusions*, First Edition, 2024.[e]

7] Association for the Advancement of Blood & Biotherapies (AABB). *Standards for Relationship Testing Laboratories.*[f]

8] Balding, D.J. and Steele, C.D*. Weight-of-Evidence for Forensic DNA Profiles.* 2nd edition, Wiley, 2015.

9] Bieber, F.R., J.S. Buckleton, B. Budowle, J.M. Butler, and M.D. Coble. "Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion." *BMC Genetics*. Vol. 17, 125, pp. 1-15. 2016.

10] Bille, T.W., J.A. Bright, and J. Buckleton. "Application of random match probability calculations to mixed STR profiles." *Journal of Forensic Sciences*. Vol. 58(2), pp. 474-485. 2013.

11] Buckleton, J., J.A. Bright, and D.A. Taylor (Eds.). *Forensic DNA Evidence Interpretation.* 2nd edition, CRC Press, 2016.

12] Buckleton, J. and J.M. Curran. "A discussion of the merits of random man not excluded and likelihood ratios." *Forensic Science International: Genetics*. Vol. 2(4), pp. 343-348. 2008.

[e] Available from: https://www.aafs.org/academy-standards-board

[f] Available from: https://www.aabb.org/standards-accreditation/standards/relationship-testing-laboratories

13] Butler, J.M., H. Iyer, R. Press, M.K. Taylor, P.M. Vallone, and S. Willis. *DNA Mixture Interpretation: A NIST Scientific Foundation Review.* (National Institute of Standards and Technology, Gaithersburg, MD), NIST IR 8351. 2024.[g]

14] Cook, R., I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert. "A hierarchy of propositions: deciding which level to address in casework." *Science & Justice*. Vol. 38(4), pp. 231–239. 1998.

15] Curran, J.M. and J. Buckleton. "Inclusion probabilities and dropout." *Journal of Forensic Sciences.* Vol. 55(5), pp. 1171-1173. 2010.

16] Evett, I.W., G. Jackson, and J.A. Lambert. "More on the hierarchy of propositions: exploring the distinction between explanations and propositions." *Science & Justice*. Vol 40(1), pp. 3-10. 2000.

17] Federal Bureau of Investigation (FBI). *Quality Assurance Standards for Forensic DNA Testing Laboratories*.[h]

18] Forensic Science Regulator (FSR). *Guidance: DNA Mixture Interpretation*. FSR-G-222, Issue 3, 2020.

19] Gill, P., C.H. Brenner, J.S. Buckleton, A. Carracedo, M. Krawczak, W.R. Mayr, N. Morling, M. Prinz, P.M. Schneider, and B.S. Weir. "DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures." *Forensic Science International*. Vol. 160(2-3), pp. 90-101. 2006.

20] Gittelson, S., T. Kalafut, S. Myers, D. Taylor, T. Hicks, F. Taroni, I.W. Evett, J.A. Bright, and J. Buckleton. "A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in an adversarial environment." *Journal of Forensic Sciences*. Vol. 61(1), pp. 186-195. 2016.

21] National Research Council (NRC II). *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington DC, 1996.

22] Scientific Working Group on DNA Analysis Methods (SWGDAM). *Guidelines for the Use of Probabilistic Genotyping with Autosomal STR Typing Results*.[h]

23] Scientific Working Group on DNA Analysis Methods (SWGDAM). *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Laboratories*.[h]

[g] Available from: https://doi.org/10.6028/NIST.IR.8351

[h] Available from: https://www.swgdam.org/publications

Academy Standards Board
410 North 21st Street
Colorado Springs, CO 80904

www.aafs.org/academy-standards-board