ASB Standard 081, First Edition
2022

Standard for Training in the Use of Statistics in Interpretation of Forensic DNA Evidence

Standard for Training in the Use of Statistics in Interpretation of Forensic DNA Evidence

ASB Approved Xxxxx 2022

ANSI Approved Xxxxxx 2022

410 North 21st Street
Colorado Springs, CO 80904

This document may be downloaded from: www.aafs.org/academy-standards-board

Foreword

This standard defines the minimum requirements for a forensic DNA analyst training program in the application of statistics to autosomal and Y-STR DNA profiling results. The aim is to provide a framework for quality training that will result in consistency in the forensic DNA community.

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.

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.

Questions, comments, and suggestions for the improvement of this document can be sent to AAFS-ASB Secretariat, asb@aafs.org or 401 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: *random match probability, likelihood ratio, DNA interpretation, statistics, training, DNA standard*

Table of Contents *(to be finalized prior to this document's publication)*

Standard for Training in the Use of Statistics in Interpretation of Forensic DNA Evidence

1 Scope

This standard defines the minimum requirements for a training program in the use of statistical methods approved within the laboratory for interpretation of forensic DNA evidence.

2 Normative References

The following reference is indispensable for the application of the standard. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ANSI/ASB Standard 022, *Standard for Forensic DNA Analysis Training Programs*[a].

3 Terms and Definitions

For purposes of this document, the following definitions apply.

3.1
avuncular index
Likelihood ratio in which the probability of a questioned person's profile is evaluated under alternate propositions - they are an uncle/aunt/niece/nephew of a known individual versus they are unrelated to the known individual. This calculation also applies to the questions of possible half-siblings, grandparent and grandchild.

3.2
Combined Probability of Exclusion
CPE
The product of the probabilities of exclusion calculated for each DNA locus. If the single-locus exclusion probabilities are independent, and if Pj is the probability of exclusion at locus j, then the combined probability of exclusion is $1- \prod j (1 - Pj)$.

3.3
Combined Probability of Inclusion
CPI
The product of the probabilities of inclusion calculated for each DNA locus. If the single-locus exclusion probabilities are independent, and if Pj is the probability of exclusion at locus j, then the combined probability of inclusion is $\Pi j (1 - Pj)$.

3.4
counting method
A method for estimating genotype, sequence, or haplotype frequency by direct counting of the number of times a genotype, sequence or haplotype is observed in a database and dividing by the number of samples in that database. This method is commonly used for estimating frequencies in populations for mitochondrial DNA and Y STR DNA haplotype results.

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

3.5
haplotype
A set of linked DNA variations, or polymorphisms, that tend to be inherited together (e.g. commonly used for human Y-chromosome or mitochondrial analysis). A haplotype can refer to a combination of alleles or to a set of single nucleotide polymorphisms (SNPs) found on the same chromosome.

3.6
Hardy-Weinberg equilibrium
A state in which allele and single locus genotype frequencies do not change (on average) from one generation to the next in a population. When alleles in a population are independent, allele and genotype frequencies are related through the Hardy-Weinberg principle: for a locus with 2 alleles P and Q at frequencies of p and q, homozygotes for P are found at frequency p2, homozygotes for Q are found at a frequency of q2, and heterozygotes are found at a frequency of 2pq. Use of the theta correction removes the need to assume Hardy-Weinberg equilibrium in the population for which a frequency database is constructed. See **Theta Correction**.

3.7
inbreeding
Mating of two persons who are more closely related than if they were chosen at random. It increases the frequency of homozygous genotypes above the expected for a random breeding population in Hardy-Weinberg equilibrium.

3.8
kinship analysis
Comparison of genetic profiles of two or more individuals to evaluate alternative degrees of relatedness.

3.9
Likelihood Ratio
LR
The probability of the evidence under one proposition (hypothesis), divided by the probability of the evidence under an alternative, mutually exclusive proposition (hypothesis). The magnitude of its value expresses the weight of the evidence.

3.10
linkage equilibrium
Two loci are in linkage equilibrium if the probability an individual jointly receives particular alleles at the loci is the product of the probabilities of receiving each of the alleles separately. If both Hardy-Weinberg and linkage equilibrium hold, then match probabilities may be multiplied over loci.

3.11
mutation rate
The relative frequency at which mutations have been observed at a specific genetic locus; generally estimated as the number of mutations observed in parent-offspring pairs divided by the total number of pairs examined.

3.12
Paternity or Maternity Index
PI/MI
The likelihood ratio for observing the data in a parentage case. More specifically, the probability of observing this data if the alleged father is the biological father of the child, divided by the probability of observing the data if a random, unrelated male in the population is the biological father. The identity of the mother may or may not be known, and her genotype may or may not be included in the evaluation. One could also calculate a maternity index if the identity of the mother is in question.

Note This is not to be confused with the abbreviation "PI" that refers to the probability of inclusion in autosomal STR analysis.

3.13
Probability of Inclusion
PI
The probability a randomly selected, unrelated individual is not excluded from being a source of DNA evidence. In human forensic DNA testing, this is often referred to as the probability a random man is not excluded (RMNE). The commonly used calculation is (SPI)2, the square of the sum of the relative frequencies (PI) of the observed alleles at a locus. If the randomly selected individual is assumed to be related to the person of interest, this formula is inappropriate.

Note This is not to be confused with the abbreviation "PI" that refers to a "paternity index" commonly used in paternity testing.

3.14
Random Match Probability
RMP
The probability of an unknown individual in a given population has a particular profile. More appropriately the random match probability is computed conditioned on a known individual observed to have the profile. The unconditional probability is the profile probability.

3.15
reverse parentage
Likelihood ratio in which three individuals have been profiled - the child and two questioned biological parents. More specifically, the probability of observing the data if the child is the biological child of the alleged parents, divided by the probability of observing the data if two randomly selected people are the parents of the child.

3.16
sibship index
Likelihood ratio in which the probability of a questioned person's profile is evaluated under alternate propositions – he/she is a sibling of a known individual versus they are unrelated to the known individual.

3.17
source attribution
A decision made based on laboratory policy which identifies an individual as the source of the DNA that produced an evidentiary single-source or major contributor profile.

3.18
theta correction
A method for calculating match probabilities, first described by Balding and Nichols (1994), to allow for population structure in the population for which a frequency database is constructed. It allows match probabilities for subpopulations to be calculated from whole population allele frequencies. It avoids the need to assume Hardy-Weinberg equilibrium at the whole-population level.

4 Requirements

4.1 General

Based upon the laboratory procedures, some of the requirements in this section may be omitted from the training program.

4.2 Knowledge-Based Training

4.2.1 At a minimum, the knowledge-based portion of the training program shall require review of the following:

a) the laboratory's protocols for statistical applications;

b) the laboratory's applicable validation studies;

c) literature used to support specific calculations and their use in appropriate circumstances; and

d) applicable literature as assigned by the trainer.

4.2.2 The knowledge-based training component of the laboratory's training program shall provide the trainee with a basic understanding of statistics applied to autosomal and Y-STR data to include, at minimum, the following topics.

a) *Population Genetics*

   1) laws of Mendelian genetics (law of segregation and the law of independent assortment);

   2) Hardy-Weinberg Equilibrium;

   3) linkage equilibrium/disequilibrium;

   4) use of theta correction to adjust for inbreeding and population substructure; and

   5) frequency, probability, odds, the laws of probability (i.e. the addition rule and product rule) and Bayes' theorem.

b) *Population Allelic Frequency Databases*

   1) population database size relative to the population size;

   2) sample collection, to include:

i. number of samples,

ii. how racial origin was determined,

iii. how the database was created, maintained and reviewed;

3) population group;

4) differences in alleles frequencies observed between population databases;

5) mechanisms to account for alleles not observed in the database.

c) *Suitability of data for statistical application*

1) when to perform statistical analyses; and

2) instruction on which loci to include in the statistical analyses when the following are observed, to include:

i. no allelic data,

ii. partial allelic data,

iii. tri-alleles, duplications/triplications, null alleles and mutations.

d) *Statistical Analysis for autosomal STR data*

1) general principles of autosomal STR statistical methods;

2) underlying theory of statistical method(s) in use by the laboratory, to include:

i. population substructure,

ii. mutation rates;

3) equation(s) in use by the laboratory, to include:

i. combined probability of inclusion,

ii. combined probability of exclusion,

iii. random match probability,

iv. likelihood ratio including formulating propositions;

4) the software program(s) in use by the laboratory;

5) source attribution statements, if applicable; and

6) limitations of statistical method(s) in use by the laboratory.

e) *Statistical Analysis for Y-STR data*

1) detailed instruction on the calculation of haplotype frequencies using the counting method, to include:

i. consideration of the differences between the loci that the database samples are typed with and the loci in the amplification kit used by the laboratory,

ii. instruction on confidence intervals, Y-STR profile probabilities and Y-STR match probabilities,

iii. instruction on combining statistical values from autosomal and Y-STR data;

2) the software program(s) in use by the laboratory.

f) *Kinship Analysis*

1) statistical calculations for kinship associations including derivation and use, to include:

i. the difference between alleles that are identical by state (IBS) or identical by descent (IBD),

ii. how to set-up competing propositions for kinship calculations, and

iii. how to account for mutations in the kinship calculations;

2) determination of appropriate calculation for the case (identifying the unknown in the relationship scenario), to include:

i. maternity or paternity index,

ii. reverse parentage,

iii. sibship, avuncular or single grandparent index, and

iv. complex family reconstruction.

4.3 Practical Training

4.3.1 The practical component of the laboratory's training program shall provide the trainee with sufficient practical instruction for the trainee to obtain the skills in calculating statistics used by the laboratory for the interpretation of forensic DNA evidence to include, at minimum the components in 4.3.2 through 4.3.4:

4.3.2 The protocol(s) shall be observed at least once.

NOTE This can be done by direct observation and supplemented with case file review until the protocols are clearly understood.

4.3.3 Practical exercises shall be representative of the range, type, and complexity of routine casework samples processed by the laboratory. Practical exercises shall include the following.

a) The application of statistical analysis to the laboratory's own data.

b) Hand calculations for the following, as appropriate: RMP, single source LR, CPI/CPE, and kinship analysis likelihood ratios.

c) Exercises to understand the derivation of the equations involved in the calculation for the following, as appropriate: parentage and kinship analysis likelihood ratios.

4.3.4 The practical exercises performed shall be sufficient to demonstrate the trainee's ability to follow the laboratory's protocols and produce appropriate statistical values.

4.4 Competency Component

4.4.1 General

The laboratory's training program shall include knowledge-based and practical competency in the laboratory's protocols for statistical applications. The format of the test(s) shall meet Section 4.3 of the ANSI/ASB Std 022.

4.4.2 Knowledge-Based Competency

The trainee shall successfully complete a knowledge-based test covering the critical information obtained during the training in the application of statistics. The format of the test(s) shall be at the discretion of the DNA Technical Leader. The test(s) shall cover, at a minimum, the topics outlined under 4.1.

4.4.3 Practical Competency

The trainee shall successfully complete a practical competency test covering each of the statistical applications the trainee will be independently authorized to perform.

5 Conformance

In order to demonstrate conformance with this standard, the laboratory shall meet the requirements outlined in section 5 of ANSI/ASB Std 022.

Annex A
(informative)

Bibliography

The following information provides a list of the resources that may assist the DNA Technical leader in defining the breadth and scope of the materials to be reviewed by the trainee. This list is not meant to be all inclusive.

1] ANSI/ASB Standard 022, *Standard for Forensic DNA Analysis Training Programs,* First Edition 2019.[b]

2] AABB. Standards for Relationship Testing Laboratories (14th ed.). 2019

3] Balding, D.J., P. Donnelly. "Evaluating DNA profile evidence when the suspect is identified through a database search." *Journal of Forensic Sciences,* 1996, Vol 41(4), pp. 603-607.

4] Balding, et al. "DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands." *Forensic Science International,* 1994, vol. 64, pp. 125-140.

5] Berger, C.E., P. Vergeer, J. Buckleton. "A more straightforward derivation of the LR for a database search." *Forensic Science International: Genetics,* 2015, vol. 14, pp. 156-160.

6] Bieber, et al. "Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion." *BMC Genetics,* 2016, vol. 17(1) .

7] Bill, et al. "PENDULUM – a guideline based approach to the interpretation of STR mixtures." *Forensic Science International*, 2005, vol. 148, pp. 181-189.

8] Bille, T.W., S.M. Weitz, M.D. Coble, J. Buckleton, J.-A. Bright. "Comparison of the performance of different models for the interpretation of low level mixed DNA profiles." *Electrophoresis 35*, 2014, vol. 3125--3133.

9] Bille, et al. "Application of Random Match Probability Calculations to Mixed STR Profiles." *Journal of Forensic Science*, 2013, vol. 58(2), pp. 474-485.

10] Bright, et al. "A comparison of stochastic variation in mixed and unmixed casework and synthetic samples." *Forensic Science International: Genetics,* 2011.

11] Bright, et al. "Degradation of forensic DNA profiles." *Australian Journal of Forensic Sciences*, 2013, pp. 1-5.

12] Bright, et al. "Developing allelic and stutter peak height models for a continuous method of DNA Interpretation", *Forensic Science International: Genetics,* 2013, vol. 7(2), pp. 296-304.

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

13] Bright, et al. "Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex." *Forensic Science International: Genetics,* 2010, vol. 4, pp. 111-114.

14] Bright, et al "Investigation into the performance of different models for predicting stutter." *Forensic Science International: Genetics,* 2013, vol. 7, pp. 433-427.

15] Bright, J.-A., D. Taylor, S. Gittelson, J. Buckleton. "The paradigm shift in DNA profile interpretation." *Forensic Science International: Genetics,* 2017, vol. 31, e24-e32.

16] Brookes, et al. "Characterising stutter in forensic STR multiplexes." *Forensic Science International: Genetics,* 2012, vol. 6, pp. 58-63.

17] Buckleton, et al. "Towards understanding the effect of uncertainty in the number of contributors to DNA stains." *Forensic Science International: Genetics,* 2007, vol. 1, pp. 20-28.

18] Buckleton, et al. "The interpretation of lineage markers in forensic DNA testing." *Forensic Science International: Genetics*, 2011, vol. 5, pp. 78-83.

19] Buckleton, J., et al. *Forensic DNA Evidence Interpretation*, Second ed., CRC Press, Boca Raton, 2016.

20] Budowle, et al. "Mixture Interpretation: Defining the Relevant Features for Guidelines for the Assessment of Mixed DNA Profiles in Forensic Casework." *Journal of Forensic Science*, 2009, vol. 54(4), pp. 810-821.

21] Budowle, B., Moretti, T.R., et al. "Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians." *Journal of Forensic Science*, 1999, vol. 44, No. 6, pp. 1277-1286.

22] Budowle, B., et al. "Population Data on the STR Loci D2S1338 and D19S433." *Forensic Science Communications*, 2001, Vol. 3, No. 3, pp. 1-5.

23] Butler, J.M. *Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers.* Second Edition, 2005, pages 145-174.

24] Clayton, et al. "Analysis and interpretation of mixed forensic stains using DNA STR profiling." *Forensic Science International,* 1998, vol. 91, pp. 55-70.

25] Clopper and Pearson. "The use of confidence or fiducial limits illustrated in the case of the binomial." *Biomatrika*, 1934, vol. 26(4), pp. 404-413.

26] Curran, et al. "Assessing uncertainty in DNA evidence caused by sampling effects." *Science and Justice*, 2002, pp. 29-37.

27] Curran, J.M., Weir B.S., et al. "Interpreting DNA Mixtures in Structured Populations." *Journal of Forensic Science*, 1999, vol. 44, no. 5, pp. 987-995.

28] Evett, et al, "Taking Account of Peak Areas when Interpreting Mixed DNA Profiles." *Journal of Forensic Science*, 1998, vol. 43(1), pp. 62-69.

29] FBI DNA Advisory Board. *Statistical and Population Genetics Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated From Pertinent Population Database(s)*. 2000.

30] Gittelson, S., T. Kalafut, S. Myers, D. Taylor, T. Hicks, F. Taroni, I.W. Evett, J.-A. Bright, 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,* 2016, vol. 61(1), pp. 186-195.

31] Hill, C.R., Duewer, D.L., Kline, M.C., Coble, M.D., Butler, J.M. "U.S. population data for 29 autosomal STR loci." *Forensic Science International: Genetics*, 2013, Vol. 7, pp. e82-e83.

32] Holt, C.L., et al. "Practical Applications of Genotypic Surveys for Forensic STR Testing." *Forensic Science International*, 2000, vol. 112, pp. 91-109.

33] Kelly, H., J.-A. Bright, J. Buckleton, J.M. Curran. "A comparison of statistical models for the analysis of complex forensic DNA profiles." *Science and Justice*, 2014, vol. 54, pp. 66-70.

34] National Research Council, *The Evaluation of Forensic DNA Evidence*. 1996.

35] Perlin, et al. "Linear Mixture Analysis: A Mathematical Approach to Resolving Mixed DNA Samples", *Journal of Forensic Science*, 2001, vol. 46(6), pp. 1372-1378.

36] "Statistical and Population Genetics Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated From Pertinent Population Database(s)." DNA Advisory Board. *Forensic Science Communications*, 2000, vol. 2, no. 3, pp 1-9.[c]

37] Steffen, S.R., Coble, M.D., Gettings, K.B., Vallone, P.M. (2017) Corrigendum to "U.S. Population Data for 29 Autosomal STR Loci." *Forensic Science International: Genetics*, 2012, vol.7, pp. e82–e83.

38] Taylor, et al. "The interpretation of single source and mixed DNA profiles." *Forensic Science International: Genetics*, 2013, vol. 7, pp. 516-528.

39] Triggs, et al. "The sensitivity of the Bayesian HPD method to the choice of prior." *Science and Justice*, 2006, pp. 169-178.

40] Tvedebrink, T., J.-A. Bright, J.S. Buckleton, J.M. Curran, N. Morling. "The effect of wild card designations and rare alleles in forensic DNA database searches." *Forensic Science International: Genetics*, 2015, vol. 16, pp. 98--104.

41] Walsh, B., Redd, A.J., and Hammer, M.F. "Joint match probabilities for Y chromosomal and autosomal markers." *Forensic Science International*, 2008, vol. 174, pp. 234-238.

42] Wang, et al. "Least Square Deconvolution: A Framework for Interpreting Short Tandem Repeat Mixtures." *Journal of Forensic Science*, 2006, vol. 51(6), pp. 1285-1294.

[c] Available from https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/july2000/dnastat.htm

43] Weir, et al. "Interpreting DNA Mixtures." *Journal of Forensic Science*, 1997, vol. 42(2), pp. 213-222.

44] Weir, B.S. *Variances of Estimated DNA Profile Probabilities*. North Carolina State University, 1999, pages 1-3.

45] Wenk, R. E. *Relationship Testing 1.0 (1st ed.)*. Bethesda, MD. AABB, 2018.

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

www.aafs.org/academy-standards-board