ASB Standard 081, First Edition
202X

Standard for Training in Statistical Calculations Used for Forensic Short Tandem Repeat (STR) DNA Data

Standard for Training in Statistical Calculations Used for Forensic Short Tandem Repeat (STR) DNA Data

ASB Approved Xxxxx 202X

ANSI Approved Xxxxxx 202X

410 North 21st Street
Colorado Springs, CO 80904

Foreword

This standard defines the requirements to be met in a forensic DNA analyst training program for the application of statistics to autosomal and Y-STR DNA profiling results. The aim is to provide a framework for quality training resulting in consistency within a laboratory and in the forensic DNA community.

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: *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 Statistical Calculations Used for Forensic Short Tandem Repeat (STR) DNA Data

1 Scope

This standard provides the requirements for a forensic DNA laboratory's training program for the use of statistical calculations and values reported for forensic autosomal and Y chromosome short tandem repeat (STR) DNA data.

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*[1].

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 the DNA mixture. If the single-locus exclusion probabilities are independent, and if P_j is the probability of exclusion at locus j, then the combined probability of exclusion is $1- \prod_j (1 - P_j)$.

3.2
Combined Probability of Inclusion
CPI
The probability that a randomly selected individual would not be excluded (i.e., is included) as a contributor to the DNA mixture. If the single-locus exclusion probabilities are independent, and if P_j is the probability of exclusion at locus j, then the combined probability of inclusion is $\Pi_j (1 - P_j)$.

3.3
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.

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

3.4
haplotype
A set of linked DNA variations, or polymorphisms, that tends 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 along a single chromosome.

3.5
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 p^2, homozygotes for Q are found at a frequency of q^2, and heterozygotes are found at a frequency of 2pq. Use of the theta correction allows the assumption of Hardy-Weinberg equilibrium in the population for which a frequency database is constructed.

3.6
Identical by Descent
IBD
Identical alleles that are copies of the same ancestral allele without mutation; this is a subset of identical by state (IBS).

3.7
Identical by State
IBS
Identical alleles that may or may not be copies of the same ancestral allele.

3.8
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 randomly mating population in Hardy-Weinberg equilibrium.

3.9
kinship analysis
Comparison of genetic profiles of two or more individuals to evaluate alternative degrees of biological relatedness.

3.10
Likelihood Ratio
LR
A likelihood ratio is defined by a ratio of two conditional probabilities: the probabilities of the evidence given each of two mutually exclusive and competing propositions.
ASTM E1732-12[mod][2]

[2] Available from ASTM, at www.astm.org

3.11
linkage equilibrium/disequilibrium
When two or more genetic loci appear to segregate randomly in a given population, they are at equilibrium. When the loci do not segregate randomly, they are at disequilibrium.

3.12
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.

3.13
Paternity or Maternity Index
PI/MI
Likelihood ratio in which the numerator is the probability of the DNA data given the proposition that the tested individual is the biological parent of the profile donor, and the denominator is the probability of the DNA data given the proposition that the tested individual is unrelated to the profile donor.

3.14
proposition
A statement that is true or false, associated with the standpoint, known or assumed, of one of the parties on a disputed issue of interest.

3.15
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.

3.16
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 the tested individuals are unrelated to the child.

3.17
sibship index
Likelihood ratio in which the numerator is the probability of the DNA data given the proposition that the two tested individuals are biological siblings, and the denominator is the probability of the DNA data given the proposition that the two tested individuals are unrelated.

3.18
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.

3.19
theta correction
A method used to adjust statistical calculations that rely on population databases to correct for substructures within populations.

4 Requirements

4.1 General

4.1.1 ANSI/ASB Standard 022, *Standard for Forensic DNA Analysis Training Programs* shall be used in conjunction with this document because ANSI/ASB Standard 022 provides the foundational training program requirements upon which additional specific requirements, such as this document, will be based.

4.1.2 The laboratory's training program shall include all requirements applicable to the work conducted by the laboratory and by the individual in training.

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 validation;

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

e) applicable literature as assigned by the trainer;

f) literature on the effects of cognitive bias in decision-making processes associated with statistical calculations used for forensic STR DNA data.

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

a) *Population Genetics*

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

   2) Hardy-Weinberg equilibrium, assumptions, and related evolutionary forces;

   3) linkage equilibrium/disequilibrium;

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

b) *Statistical Foundations*

1) frequency;

2) probability;

3) odds;

4) the laws of probability (e.g., the addition rule and product rule);

5) Bayes' theorem;

6) Sources of uncertainty (e.g., modelling uncertainty and sampling variability).

c) *Population Allelic Frequency Databases*

1) population database size relative to the population size;

2) sample collection, to include:

i) number of samples,

ii) how population group was determined,

iii) how the database was created, maintained and reviewed,

iv) sampling uncertainty;

3) differences in allele frequencies observed between population databases;

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

d) *Suitability of data for statistical application*

1) when to perform statistical analyses and which statistical calculation is validated for the type of data obtained and comparison performed;

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

i) no allelic data,

ii) partial allelic data,

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

iv) STR artifacts.

e) *Statistical analysis for autosomal STR data*

1) proper use of statistical calculations to include derivation and applicable equations;

2) statistical calculation method(s) in use by the laboratory, to address:

i) population substructure,

ii) mutation rates,

iii) known relatedness;

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

i) combined probability of inclusion,

ii) combined probability of exclusion,

iii) random match probability (or modification),

iv) likelihood ratio including formulating propositions;

4) operation of the software program(s) in use by the laboratory, including the underlying equations and review of the output data;

5) source attribution statements

6) limitations and assumptions of statistical method(s) used by the laboratory.

f) *Statistical analysis for Y-STR data*

1) detailed instruction on the calculation of haplotype frequencies using counting methods, 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) consideration of the effect of the database source/population and size;

iv) instruction on incorporating Y-STR mutation rates;

v) instruction on combining statistical values from autosomal and Y-STR data;

2) operation of the software program(s) used by the laboratory, to include:

i) underlying equations,

ii) review of the output data,

iii) search types (e.g.,reduced, masked, transient).

g) *Kinship analysis*

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

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,

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, aunt or uncle, or single grandparent index,

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 practical instruction for the trainee to obtain the skills for calculating statistical values used by the laboratory.

4.3.2 At a minimum, the practical portion of the training program shall include the observation of a trained analyst performing the processes with exercises representative of the range, type, and complexity of DNA data from routine casework or database samples processed by the laboratory, at least once or until clearly understood.

4.3.3 Practical exercises shall be representative of the range, type, and complexity of routine DNA data from 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 (or modification), single source LR, CPI/CPE, and kinship analysis likelihood ratios;

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

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

4.4 Competency Component

4.4.1 General

The competency component of the laboratory's training program shall demonstrate knowledge-based and practical competency in the laboratory's protocols for statistical applications. The format of the test(s) and the criteria for passing the competency test(s) shall meet Section 4.3 of the ANSI/ASB Standard 022, *Standard for Forensic Training DNA Analysis Training Programs.*

4.4.2 Knowledge-Based Competency

As applicable to the trainee's job responsibilities, the trainee shall successfully complete (as defined by the laboratory's policy) a knowledge-based test covering the critical information obtained during the training in the application of statistics . The test(s) shall cover, at a minimum, the topics outlined in 4.2 and its subsections.

4.4.3 Practical Competency

The trainee shall successfully complete (as defined by the laboratory's policy) a practical competency test covering each of the statistical applications the trainee will be independently authorized to perform. DNA data from samples representative of the range, type, and complexity for which the trainee will be authorized to perform statistical calculations shall be included in the practical competency test(s).

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, *Standard for Forensic DNA Analysis Training Programs* and all the requirements set forth in this document.

Annex A
(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 publications cited informationally, and publications relevant to the standard. For undated references, the latest edition of the referenced document (including any amendments) applies.

1] ANSI/ASB Standard 022, *Standard for Forensic DNA Analysis Training Programs,* First Edition 2019.[3]

2] AABB. Standards for Relationship Testing Laboratories, current edition

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: 125(1).

7] Bille, 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, 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 Sci Int Genet*. 2012 ,vol.6(2),pp.180-4..

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.

[3] 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. NIST interlaboratory studies involving DNA mixtures (MIX13): A modern analysis, *Forensic Science International: Genetics* 37 (2018) 172-179.

18] 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.

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

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

21] 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.

22] 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.

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

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

25] Butler, J. M., Kline, M. C., & Coble, M. D. (2018). NIST interlaboratory studies involving DNA mixtures (MIX05 and MIX13): Variation observed and lessons learned. Forensic science international. Genetics, 37, 81–94.[4]

26] Butler, J.M., Vallone, P.M., Gettings, K.B., Borsuk, L.A. , Ruitberg, C.M. ,and Reeder, D.J., NIST Short Tandem Repeat DNA Internet Database, National Institute of Standards and Technology, doi:10.18434/T44G6P.

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

[4] Available from: https://doi.org/10.1016/j.fsigen.2018.07.024

28] 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.

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

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

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

32] 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)*. Forensic Science Communications, 2000, Vol. 2, no. 3.

33] 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.

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

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

36] 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.

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

38] 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.

39] "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.[5]

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

41] SWGDAM. SWGDAM Interpretation Guidelines for Autosomal and Y-STR Typing for Forensic DNA Testing Laboratories[6].

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

[6] Available from: http://media.wix.com/ugd/4344b0_da25419ba2dd4363bc4e5e8fe7025882.pdf

42] Taylor, et al. “The interpretation of single source and mixed DNA profiles.” *Forensic Science International: Genetics*, 2013, vol. 7, pp. 516–528.

43] Triggs, et al. “The sensitivity of the Bayesian HPD method to the choice of prior.” *Science and Justice*, 2006, pp. 169–178.

44] 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.

45] 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.

46] 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.

47] Weir, et al. “Interpreting DNA Mixtures.” *Journal of Forensic Science*, 1997, vol. 42(2), pp. 213–222.

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

49] 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