FBS18 – Population Statistics

Table of Contents

1. Scope

1.1. This procedure outlines the guidelines used to calculate statistics employing population database information.

2. Background

2.1. To establish the practices for documenting the examination of evidence to conform to the requirements of the Department of Forensic Sciences (DFS) Forensic Science Laboratory (FSL) *Quality Assurance Manual*, the accreditation standards under ISO/IEC 17025:2005, and any supplemental standards.

2.2. "When a comparison of DNA profiles derived from evidence and reference samples fails to exclude an individual(s) as a contributor(s) of the evidence sample, statistical assessment and/or probabilistic reasoning are used to evaluate the significance of the association."

2.2.1. -DAB Recommendations on Statistics, July 2000 issue of Forensic Science Communications.

2.3. Statistical calculations are formulated using the allele frequencies from validated databases. The frequency of each allele present in a profile is assessed and genetic principles are applied to give an approximate frequency of the profile in a population. This number is expressed as a random match probability and is interpreted as the likelihood of selecting an unrelated individual at random having an STR profile matching the DNA acquired from the questioned specimen.

3. Safety

3.1. Not applicable

4. Materials Required

4.1. Not applicable

5. Standards and Controls

5.1. Not applicable

6. Calibration

6.1. Not applicable

7. Procedures

7.1. Random Match Probabilties are calculated for inclusions when:

7.1.1. The match is probative (e.g. the victim's DNA typing results are present on clothing recovered from a suspect, etc.) AND

7.1.2. The sample is unmixed OR

7.1.3. The sample is a mixture of two components and the source of one component is known (e.g. vaginal epithelial cell carryover into a sperm fraction) OR

7.1.4. There is a large difference in peak heights between a major and minor component and the genotypes of the major component are easily inferred.

7.1.4.1. Statistics are not necessary for expected inclusions (e.g., DNA profiles obtained from intimate samples).

7.1.4.2. Statistics for indistinguishable mixtures may be calculated following the National Research Council (NRC) recommendations and SWGDAM Interpretation Guidelines (see Mixtures section).

7.2. Population Databases:

7.2.1. Revelant populations, for which frequencies are estimated, are selected as suggested in the NRC Recommendation 4.1 and are identified in the issued report.

7.2.1.1. Frequencies will be reported using at least three major United States population groups (e.g., African-American, Caucasian, Southeast Hispanic, Southwest Hispanic).

7.2.1.2. Other racial/ethnic population groups may be relevant (e.g., Vietnamese-American) and may be reported if allele frequency data is available (NRC 4.2 and 4.3).

7.3. Genotype Frequencies:

7.3.1. Genotype frequencies will be calculated for STRs following NRC Recommendation 4.1 and are summarized below:

Heterozygotes: $2pq$

Homozygotes: $p^2 + p(1 - p)\theta$ **(See note 1)** $(\theta = 0.01)$ **(See note 2)**

Minimum allele frequencies (allele count < 5): 5/2N ***(See note 3)***

7.3.1.1. **NOTE 1:** In simplest form, the genotype frequency for a homozygote is estimated as the square of the estimated allele frequency (p^2). However, NRC recommendation 4.1 suggests using the following expression: $p^2 + p(1-p)\theta$, where **p** represents the estimated allele frequency and θ (theta) is a measure of potential population subdivision.

7.3.1.1.1. The 2p rule may be warranted, in certain situations, in order to produce a more conservative genotype frequency estimate. Factors to be reviewed before making such a decision include, but are not limited to, potential stochastic effects, degradation, zygosity, nature of the particular sample (e.g., quantity available, peak heights), and analyst's discretion with DNA Unit Supervisor/Technical Leader approval.

7.3.1.2. **NOTE 2:** A θ value of 0.01 is typical for most situations. A θ value of 0.03 is reserved for smaller, isolated populations (such as the Native American population).

7.3.1.3. **NOTE 3:** A lower limit for the frequency of rare STR alleles (i.e., alleles with fewer than five observations in the population database) produces a conservative frequency. This "minimum allele frequency" alleviates potential concerns of underestimation. The minimum allele frequency is calculated using the following

expression: $p_{min} = 5 /2N$, where **N** represents the number of individuals in the population database.

7.4. Genetic Profiles:

7.4.1. The significance of a genetic profile is expressed in terms of a random match probability (RMP). The random match probability is the probability that a person randomly selected from the population will have the observed genetic profile and therefore is the frequency of occurrence for that particular combination of genotypes. This frequency of occurrence, as suggested by NRC Recommendation 4.1, is simply the product of the single locus genotype frequencies (product rule), provided each locus is inherited independently. Independence, or linkage equilibrium, is expected for genetic markers physically located on different chromosomes or far apart on one chromosome, in accordance with Mendel's Law of Independent Assortment. Empirical studies have demonstrated the independence of the AmpFℓSTR STR genetic markers.

7.5. Mixtures:

7.5.1. As the NRC II Report describes, alternative methods exist for the numerical representation of the probative value of DNA evidence. Choice of statistical approach(es) applied is affected by philosophy and experience of the user, the legal system, practicality of the approach, the question(s) posed, and the available data.

7.5.1.1. When individual contributors of a DNA mixture can be distinguished, statistical analyses may be carried out for each contributor as a single contributor random match probability.

7.5.1.2. When individual contributors of a DNA mixture cannot be distinguished at each locus, a non-statistical inclusion or exclusion statement of an individual as a possible contributor may be reported.

7.5.1.3. When individual contributors of a DNA mixture cannot be distinguished at each locus and a statistical approach is desired, then either the combined probability of inclusion (CPI) and/or likelihood ratio (LR) calculations may be appropriate.

7.5.1.4. When using CPI (with no assumptions of the number of contributors), loci with alleles below the stochastic threshold may not be used for statistical purposes. These alleles may be used for comparisons and/or to establish the presence of a mixture or male DNA.

7.5.1.5. The following formulas will be used for CPI calculations:

$$\text{Probability of Inclusion} = P(I)_{locus} = \{(p_1+p_2+\ldots p_n)^2+\Theta[p_1(1-p_1)+p_2(1-p_2)+\ldots+p_n(1-p_n)]\}$$

$$\Theta = 0.01$$

Combined Probability of Inclusion (CPI)

Frequency (CPI) = $\{[P(I)_1][P(I)_2]\ldots[P(I)_n]\}$

1 in = 1/CPI

Combined Probability of Exclusion (CPE)

Percent = 100(1-CPI)

7.6. Report Wording:

7.6.1. Random match probabilities are reported in the Conclusions section of the DNA Report of Examination.

7.6.2. Single-Source Sample Example

7.6.2.1. The probability of selecting an unrelated individual at random having an STR profile matching the DNA obtained from the questioned specimen, (item #), is approximately:

Approximate Frequency	**Population Database**
1 in 1.0 million	African American
1 in 11 million	U.S. Caucasian
1 in 100,000	U.S. Hispanic

7.6.3. Mixed Sample Example (Major Contributor)

7.6.3.1. The probability of selecting an unrelated individual at random having an STR profile matching the major contributor to the mixture of DNA obtained from the questioned specimen, (item #), is approximately:

Approximate Frequency	**Population Database**
1 in 1.0 million	African American
1 in 11 million	U.S. Caucasian
1 in 120,000	U.S. Hispanic

7.6.4. Mixed Sample Example (CPI Calculation)

7.6.4.1. The probability of selecting an unrelated individual at random having an STR profile which would be included as a contributor to this mixture is approximately:

Approximate Frequency	**Population Database**
1 in 1.0 million	African American
1 in 11 million	U. S. Caucasian
1 in 120,000	U. S. Hispanic

8. Sampling

8.1. Not applicable

9. Calculations

9.1. Genotype frequencies will be calculated for STRs following NRC Recommendation 4.1 and are summarized below (see section 7.3.1 for notes):

Heterozygotes: 2pq

Homozygotes: $p^2 + p(1 - p)\theta$ **(See note 1)** $(\theta = 0.01)$ **(See note 2)**

Minimum allele frequencies (allele count < 5): 5/2N ***(See note 3)***

9.2. The following formulas will be used for CPI calculations:

$$\text{Probability of Inclusion} = P(I)_{locus} = \{(p_1+p_2+\ldots p_n)^2+\Theta[p_1(1-p_1)+p_2(1-p_2)+\ldots+p_n(1-p_n)]\}$$

$$\Theta = 0.01$$

Combined Probability of Inclusion (CPI)

$$\text{Frequency (CPI)} = \{[P(I)_1][P(I)_2]\ldots[P(I)_n]\}$$

$$1 \text{ in} = 1/\text{CPI}$$

Combined Probability of Exclusion (CPE)

$$\text{Percent} = 100(1-\text{CPI})$$

10. Uncertainty of Measurement

10.1. When quantitative results are obtained, and the significance of the value may impact the report, the uncertainty of measurement must be determined. The method used to determine the estimation of uncertainty can be found in the *FSL Quality Assurance Manual – Estimation of Uncertainty of Measurement (Section 5.4.6).*

10.1.1. NRC II (1996) states the following: "It is probably safe to assume that within a race, the uncertainty of a value calculated from adequate databases (at least several hundred persons) by the product rule is within a factor of about 10 above and below the true value."

11. Limitations

11.1. Not applicable.

12. Documentation

12.1. FBU Report of Results

13. References

13.1. Applied Biosystems GeneMapper ID software. Applied Biosystems, Foster City, CA, 2004.

13.2. Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keyes, K.M. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, US Caucasians, Hispanic, Bahamians, Jamaicans, and Trinidadians. Journal of Forensic Sciences (1999) 44(6): 1277-1286.

13.3. Budowle, B., et al. STR primer concordance study, Forensic Science International, (2001) 124: 47-54.

13.4. Budowle, B., Shea, B., Niezgoda, S., Chakraborty, R. CODIS STR loci data from 41 sample populations, Journal of Forensic Sciences (2001) 46(3): 453-489.

13.5. Budowle, B. Population studies on the 17 STR loci routinely used in forensic analysis. In: International Congress Series, 1028 (2002).

13.6. Budowle, B., Carmody, G., Chakraborty, R., and Monson, K. Source attribution of a forensic DNA profile. Forensic Science Communications (2000) July 2(3).

13.7. Chakraborty, R. Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. Human Biology (1992) 64(2):141-159.

13.8. Clayton, T.M., Guest, J.L., Urquhart, A.J., and Gill, P.D. A Genetic Basis for Anomalous Band Patterns Encountered During DNA STR Profiling, Journal of Forensic Sciences (2004) 49: 1207-1214.

13.9. Crouse, C., Rogers, S., Amiott, E., Gibson, S., and Masibay, A. Analysis and interpretation of short tandem repeat microvariants and three banded allele patterns using multiple allele detection systems, [published erratum appears in Journal of Forensic Sciences (1999) May 44(3)]. Journal of Forensic Sciences (1999) 44: 87-94.

13.10. DNA Advisory Board. Quality Assurance Standards for Forensic DNA Testing Laboratories. Forensic Science Communications (2000) July 2(3).

13.11. DNA Advisory Board. Statistical and Population Genetic Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated from Pertinent Population Database(s), Forensic Science Communications (2000) July 2(3).

13.12. Liebelt, C., Budowle, B., Moretti, T.R., Holt, C.L., Stauffer, C., Wallin, J.M., Lazaruk, K.D., Nguyen, T., Walsh, P.S. Practical applications of genotype surveys for forensic STR testing, Forensic Science International, (2000) 112: 91-109.

13.13. National Research Council. The Evaluation of Forensic DNA Evidence, Washington, DC: Academy Press, 1996.

13.14. Ruitberg, C.M., Reeder, D.J., Butler, J.M. STR BASE: A short tandem repeat DNA database for the human identity testing community. Nucleic Acid Research (2001) 29(1).

13.15. Stivers, D.N., and Chakraborty, R. A test of allelic independence based on distributions of allele size differences at microsatellite loci. Human Heredity (1997) 47: 66-75.

13.16. Technical Working Group on DNA Analysis Methods. Guidelines for a quality assurance program for DNA analysis. Crime Laboratory Digest (1995) 22(2):21-43.

13.17. National Research Council Committee on DNA Forensic Science (1996) *An Update: The Evaulation of Forensic DNA Evidence*, Washington, D.C., National Academy Press [aka NRC II].

13.18. Forensic Science Laboratory Quality Assurance Manual (Current Version)

13.19. FSL Departmental Operations Manuals (Current Versions)

13.20. FSL Laboratory Operations Manuals (Current Versions)

13.21. AmpF*l*STR® Identifiler Plus® PCR Amplification Kit User's Guide. Rev C. Life Technologies 2011.

13.22. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Approved 1/14/2010. www.swgdam.org.