

---

## Procedure for Statistical Calculations

**1.0 Purpose** – This document provides the procedure for calculating statistical data on single source, predominant profiles and mixed DNA samples.

**2.0 Scope** – This procedure applies to Casework analysts and trainees in the Forensic Biology Section who are qualified to perform statistical analysis on DNA profiles and mixtures.

### 3.0 Definitions

- **Combined Probability of Exclusion (CPE):** The probability that a randomly chosen, unrelated person from a given population would be excluded as a potential contributor to the observed DNA mixture.
- **Combined Probability of Inclusion (CPI):** The probability that a randomly chosen, unrelated person from a given population would be included as a potential contributor to the observed DNA mixture.
- **Intimate Sample:** Refer to the Section Procedure for Casework DNA Interpretation.
- **Minimum allele frequency:** A frequency which shall be used for any allele which is observed 5 or less times in the population frequency database (to include variant and off-ladder alleles). The formula for this frequency is  $5/2N$  ( $2N = \#$  of alleles in the population database at that locus).
- **Heterozygote:** Two different alleles at a particular gene locus on homologous chromosomes.
- **Homozygote:** Same alleles at a particular gene locus on homologous chromosomes.
- **Popstats** – The computer program used to generate Random Match Probability (RMP) and CPE/CPI statistics.
- **Product Rule:** A calculation that can be used if a locus is inherited independently of the other loci. The calculation can be made by multiplying each individual genotype frequency together.
- **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.
- **Restricted:** Referring to a statistical approach conditioned on the number of contributors and with consideration of quantitative peak height information and inference of contributor mixture ratios; used to limit the genotypic combinations of possible contributors.
- **Unrestricted:** Referring to a statistical approach without consideration of quantitative peak height information and inference of contributor mixture ratios; for CPE/CPI this may or may not be conditioned on the number of contributors.

### 4.0 Equipment, Materials and Reagents

- CODIS software program (Analyst Workbench) located on the CODIS computer with Print to PDF software.

### 5.0 Procedure

#### 5.1 General

- 5.1.1** Statistics shall be calculated on questioned items if comparison to a known item results in an inclusion. The exception to this requirement is for intimate samples where this association is made to the owner or a consensual partner.
- 5.1.2** In order to perform a statistical calculation, a minimum of 3 loci where frequency data has been determined must be present and eligible for use in the question sample(s). Refer to Section Procedures for DNA Interpretation for Casework and Report Writing for reporting statements.
- 5.1.3** Population frequency data for D2S1338 and D19S433 have not been determined for the NC populations and are not used for calculations. Amelogenin frequencies are not used for calculations.
- 5.1.4** Failure of a locus to amplify in either the questioned or known sample does not preclude the generation of statistical results at loci where complete amplification occurred.

## **5.2 Random Match Probability (RMP)**

- 5.2.1** RMP shall be used for single source or single predominant profiles.
- 5.2.2** The formulas used in the calculation of the frequency of a DNA profile shall be in accordance with those published in the NRC II guidelines and in PopStats.
- 5.2.2.1** Heterozygote frequencies:  $2pq$  (p and q are the frequencies of the alleles observed).
- 5.2.2.2** Homozygote frequencies:  $p^2 + p(1-p)\theta$ , where  $\theta = 0.01$  (p is the frequency of the allele observed and  $\theta$  is a correction factor for expected levels of population substructure in a given population).
- 5.2.2.2.1** If a locus contains a homozygote allele which falls below the stochastic threshold, that locus shall not be used.
- 5.2.2.2.2** For single allele profiles where the zygosity is in question (i.e., below stochastic threshold) the locus is assigned a statistical value of 1.
- 5.2.2.3** Multi-locus frequencies: the product rule shall be used.

## **5.2.3 Procedure for performing RMP**

- 5.2.3.1** Log onto a CODIS computer.
- 5.2.3.2** Open the Analyst Workbench program.
- 5.2.3.3** Click PopStats located in the bottom left.

**5.2.3.4** In the Specimen ID box type in the full case number (note: the item number may be entered along with the case number or separately within the comments box).

**5.2.3.5** In the Specimen Details block type the alleles.

**5.2.3.6** Click Calculate.

**5.2.3.7** Print the Forensic/Broward report using the Print to PDF and add it to the FA case record object repository.

**5.2.4** See the Forensic Biology Section Procedure for Report Writing for reporting guidelines.

### **5.3 Combined Probability of Exclusion/Inclusion for DNA mixtures (CPI calculations)**

**5.3.1** Since CPI is derived from CPE, calculations will be performed for both CPE and CPI. However, only CPI shall be reported.

#### **5.3.2 Unrestricted CPI calculations**

**5.3.2.1** Shall be performed only once on a mixture profile except in the instance of multiple major profiles. (See 5.3.3.)

**5.3.2.2** If a locus has alleles that fall below the stochastic threshold, that locus shall not be used for statistical weight.

**5.3.2.3** CPI shall not be calculated at loci where allelic dropout is likely to have occurred.

#### **5.3.3 Restricted CPI may be performed on the multiple major profile of a mixture.**

**5.3.3.1** Only the alleles associated with the distinct group (multiple major) can be used for calculations.

**5.3.3.1.1** Example: A locus has alleles P, Q, R and S with alleles P, Q and R being above the stochastic threshold and S being below the stochastic threshold. Alleles P, Q and R are interpreted as being from a distinct multiple major profile. CPI calculations can be performed using only alleles P, Q and R.

**5.3.3.2** An unrestricted CPI calculation may also be performed on the mixture containing a multiple major profile if the remaining alleles are above the stochastic threshold.

**5.3.3.2.1** Example: A locus has alleles P, Q, R and S with all alleles being above the stochastic threshold; however, alleles P, Q and R are interpreted as being from a distinct multiple major profile. A restricted CPI may be calculated using alleles P, Q, and R and an unrestricted CPI calculation can be performed using all alleles (P, Q, R, S).

**5.3.4 Formulas:** If an evidence sample includes three alleles at a locus ( $A_1, A_2, A_3$ ), then:

**5.3.4.1**  $P = A_1 + A_2 + A_3$  and

**5.3.4.2**  $Q = 1 - P$

**5.3.4.3**  $PE = Q^2 + Q(1-Q)\theta + [2Q(1-Q)(1-\theta)] \quad \theta = 0.01$

**5.3.4.4** The probability of exclusion (PE) is calculated at each locus. Then the PEs from the multiple loci are combined through the following equation:

$$CPE = 1 - [(1 - PE_1) \times (1 - PE_2) \times \dots \times (1 - PE_n)]$$

$$CPI = 1 - CPE$$

### **5.3.5 Procedure for performing CPE/CPI**

**5.3.5.1** Log onto a CODIS computer.

**5.3.5.2** Open Analyst Workbench program.

**5.3.5.3** Click PopStats located in the bottom left.

**5.3.5.4** In the explorer window at the upper left, choose Forensic Mixture.

**5.3.5.5** In the Specimen ID box type in the full case number (the item number may be entered along with the case number or separately within the comments box).

**5.3.5.6** In the Mixture block type in the alleles.

**5.3.5.7** Click Calculate.

**5.3.5.8** Print the Probability of Inclusion (Short) report using the Print to PDF and add it to the FA case record object repository.

**5.3.6** See the Forensic Biology Section Procedure for Report Writing for reporting guidelines.

**6.0 Limitations** – N/A

**7.0 Safety** – N/A

**8.0 References**

Forensic Biology Section Procedure for Casework DNA Interpretation

Forensic Biology Section Procedure for CODIS

Forensic Biology Section Procedure for Report Writing

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

**9.0 Records** – N/A

**10.0 Attachments** – N/A

<b>Revision History</b>		
Effective Date	Version Number	Reason
01/03/2013	1	Original Document
03/08/2013	2	Definitions – clarified CPI and CPE definitions; 5.1.3.2 – clarified definition of theta; 5.2.5.2 – changed $Q = 1 - p$ to $Q = 1 - P$
12/18/2013	3	Header – added issuing authority
04/18/2014	4	3.0 – added Popstats definition; 5.1 – added new general stats section; 5.1, 5.2.2., 5.2.6.6 – moved redundant wording to new general section; 5.2.4 – added clarification wording; 5.2.5 – added population clock requirement
12/28/2015	5	3.0 - updated RMP definition, added definition for restricted and unrestricted; 5.1.1 – clarified wording; 5.1.4 – added clarification statement of when to perform stats; 5.2.5 – removed popclock requirement; 5.3-separated types of CPI