The Forensic Statistical Tool, FST, is a statistical tool for analysis of complex forensic DNA mixtures and reference profiles, such as those from suspects or victims. The method employs a likelihood ratio, which is a standard analytical procedure, used in virtually every scientific field, including forensics. In addition, FST models allelic drop-out and drop-in. This means that FST can assign a weight to a comparison between an evidence profile and a reference profile when one or more of the reference profile's alleles are not detected in the mixture and/or when additional alleles are detected. Without this tool, it is not possible to assign a quantitative weight to such a comparison.

As a matter of background, in a conventional forensic LR framework, the probability of a crime scene DNA profile (G) is computed under two competing hypotheses: that of the prosecution (H_p) and that of the defense (H_d). The LR is the ratio of these two probabilities: $\mathrm{LR} = \frac{\Pr(G \mid H_p)}{\Pr(G \mid H_d)}$. In a typical simple scenario, the prosecution asserts that the crime scene DNA belongs to the suspect (H_p:S) and the defense asserts that the crime scene DNA belongs to an unknown, unrelated individual (H_d:U). If the crime scene profile matches the suspect profile, then $\Pr(G \mid H_p) = 1.0$ and $\Pr(G \mid H_d) = P$, where *P* is the estimated population frequency of the crime scene profile (i.e., the random match probability, RMP). Thus, in this scenario, $\mathrm{LR} = \frac{1}{RMP}$ when the evidence profile matches the suspect profile. When a crime scene profile reflects a mixture, the conventional LR can also be used by specifying the appropriate H_p and H_d. For example, the prosecution may assert that the mixture includes the victim and

suspect (H_p:V+S), while the defense may assert the mixture includes the victim and an unknown, unrelated individual (H_d:V+U).

Multiple replicate amplifications of an evidentiary sample can be considered within a LR framework. Let **R** = R_1, R_2, ..., R_n represent the alleles observed in amplification replicates 1 through *n* at a single locus in an evidentiary sample. At NYC OCME, normally, *n* = 3 for low template samples and *n* = 1 or 2 for high template samples. The replicate data are used to compute $LR = \frac{\Pr(\mathbf{R} \mid H_p)}{\Pr(\mathbf{R} \mid H_d)}$.

We have incorporated allelic drop-out and drop-in into the LR. A critical step in this process is to consider all possible genotypes for the unknown contributor(s) in the denominator (as well as the numerator for more complex scenarios). If *x* distinct alleles are observed at a locus in the evidentiary profile, there are $m = \frac{x(x+1)}{2} + x + 1$ values comprising the set of possible genotypes of an unknown contributor. This calculation treats all unobserved alleles as a single 'other' allele. That is, an unknown contributor's genotype at the locus could include any pair wise combination of the observed alleles and the unobserved 'other' allele.

If H_p: S and H_d: U, $LR = \frac{\Pr(\mathbf{R} \mid S)}{\sum_{j=1}^{m} \Pr(\mathbf{R} \mid U = G_j) \Pr(U = G_j)}$, where S represents the suspect's alleles, U represents the alleles of an unknown contributor, and G_j represents the j^{th} possibility for the genotype of the unknown contributor. $\Pr(U = G_j)$ is the expected population frequency of G_j, including a θ correction for population substructure, applied to homozygous genotypes only, as recommended in the second

National Research Council Report (National Research Council. 1996. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington DC). Note that this correction does not account for genotype correlation between individuals.

If H_p: S + V and H_d: V + U, $LR = \dfrac{\Pr(\mathbf{R} \mid S, V)}{\sum_{j=1}^{m} \Pr(\mathbf{R} \mid V; U = G_j)\Pr(U = G_j)}$.

If H_p: S + U_1 and H_d: U_2 + U_3,

$$LR = \frac{\sum_{i=1}^{m} \Pr(\mathbf{R} \mid S; U_1 = G_i)\Pr(U_1 = G_i)}{\sum_{i=1}^{m}\sum_{j=1}^{m} \Pr(\mathbf{R} \mid U_2 = G_i, U_3 = G_j)\Pr(U_2 = G_i)\Pr(U_3 = G_j)}.$$

All of the pairs of prosecutor and defense hypotheses shown in Table 1A are formulated similarly. Drop-out and drop-in rates are incorporated into $\Pr(R_1, R_2, ..., R_n \mid S)$ and $\Pr(R_1, R_2, ..., R_n \mid U = G_j)$. Separate parameters are used for partial drop-out of heterozygotes, complete drop-out of heterozygotes, and complete drop-out of homozygotes. Drop-out rates were estimated empirically as a function of locus, quantity of template DNA, number of contributors to the sample, mixture ratio (i.e., deducible or non-deducible), and level of degradation (moderate, severe, or not degraded). Drop-in rates were estimated separately for low template and high template amplification.

Option	Numerator (Prosecutor's Hypothesis)	Denominator (Defense Hypothesis)
1	Comparison	Unknown
2	Comparison + Known	Known + Unknown
3	Comparison 1 + Comparison 2	2 Unknowns
4	Comparison + Unknown	2 Unknowns
5	Comparison + 2 Unknowns	3 Unknowns
6	Comparison + Known + Unknown	2 Unknowns + Known
7	Comparison + 2 Knowns	Unknown + 2 Knowns

Table 1A. Numerator and denominator options available in LR software. "Known" refers to an elimination profile from an individual who is assumed to be a contributor to the evidence sample. "Comparison" refers to the comparison profile of interest (often the suspect). "Unknown" refers to a randomly selected individual from a population of individuals that are unrelated to the Known, Comparison, or to one another.