Forensic Statistical Tool (FST)

The Forensic Statistical Tool (FST) computes a statistical value known as a likelihood ratio (LR). The LR value provides a statistical measurement of the strength of support for one scenario over another, i.e., one scenario being that a known person contributed to a mixture versus the scenario that an unknown, unrelated person contributed instead.

I. A comparison profile must be available in order to use FST

Whether or not the source of the comparison profile contributed to a mixture is the relevant question. Depending upon the context of the case, a comparison profile may be from a suspect or a victim or may be a single source or deconvoluted profile within a case. Profiles of known contributors to the evidence sample may be used, if available. For the majority of circumstances, a suspect should never be treated as a known contributor. Every attempt must be made to generate a full profile for a known or a comparison sample.

II. Sample Criteria for using the FST

A. The random match probability (RMP), not FST, will be used for the following samples:

1. Single source profile
2. Deconvoluted major and/or minor profiles

B. FST should be used for the following mixed samples:

1. The DNA profiles of the major and the minor contributors cannot be determined; however, the sample is informative and suitable for comparison.

2. The DNA profile(s) of the minor contributor(s) cannot be determined but the sample is informative and suitable for comparison. In this situation, the random match probability should be used to calculate the statistical value for the deconvoluted DNA profile of the major contributor and FST should be used for comparisons to the minor contributor(s). If the minor component to a mixture is not suitable for comparison, this sample should not be evaluated with FST.
Informative mixtures

a. Informative mixtures with which a comparison sample can be positively associated (qualitatively “could be a contributor” or “cannot be excluded as a contributor”) should be tested using FST. If multiple items within a case are positively associated to a suspect, FST should be applied to each mixture, as it may not be feasible to determine in advance which items will be most informative to the case. Not all mixtures generate informative results. For example, the DNA profile of a homeowner found on an item within their home is most likely not informative.

b. It may not be necessary to use FST for all informative mixtures within a case.

1. If multiple samples are taken from a single item, it may not be necessary to use FST for each one. For example, if Sample A generates a deducible mixture and Sample B from the same item generates a non-deducible mixture, statistics may not be necessary for Sample B if the comparison sample’s profile is consistent with the deconvoluted profile from Sample A, for which RMP can be calculated.

2. If related samples are taken from different items, such as sexual assault kit items or multiple stains from a crime scene, it may not be necessary to use FST for each one.

d. FST should be applied to mixtures to which a comparison sample can be positively associated. If multiple items within a case are positively associated to a suspect, FST should be applied to each mixture, as it may not be feasible to determine in advance which items will be most informative to the case.
C. Effect of relationships among the comparison sample, the known, and the unknowns.

1. There is no restriction on the relationship between the known(s) and the comparison sample.

2. FST models the unknown persons as unrelated to one another and to the known(s) and the comparison sample.
   a. In the event that it is asserted, for example, that the suspect’s brother is the source of the DNA, FST cannot account for this relationship.
   b. However, as stated in C1, FST can still be used if the comparison sample (the suspect for example) and the known contributor(s) are related because both profiles are available to be used in the calculation.
   c. If the unknown contributors are thought to be related to the comparison sample, request elimination samples from those individuals.
      i. If an elimination sample was submitted, and he/she can be positively associated (qualitatively “could be a contributor” or “cannot be excluded as a contributor”) with the evidence, that elimination sample may be used as a known. An alternative scenario should also be calculated with no known contributors.
      ii. If the elimination sample can be excluded as a contributor or the results do not support a positive association or an exclusion (qualitatively “no conclusions can be drawn”), that sample should not be used as a known.
      iii. If no elimination samples were submitted, calculate the LR with no known contributors in the model. The assumption that the unknown person(s) are unrelated must be stated.

D. Partial Profiles

1. Evidence samples may have loci with no information, which will result in blank data fields for these samples.
2. However, if a comparison or a known sample is partial, loci that are not complete or blank will be not used in the calculation. In other words, the program will only utilize loci that display allele calls for a comparison or a known sample.

   a. Samples used as a known (e.g. victim, Male Donor A, etc.) should be a full profile. Any missing loci will be omitted from the calculation, even if the evidence and the comparison sample display results.

   b. In most situations, comparison profiles should be full. Certain circumstances may dictate the use of a partial comparison profile. For example, a degraded exemplar may be used as a comparison if every attempt has been made to produce a full profile.

III. Hypothesis building

Hypotheses are built based on the data and the relevant question. For the majority of mixture comparisons no more than one, or at most two different LRs should be calculated.

A. Assuming one or more known contributors

1. If a profile is consistent with the profile of the major contributor to a mixture, the profile may be assumed as a known.

   a. If the profile reaches source attribution (refer to “Sample Comparisons” manual), only one scenario may be calculated. The full profile of the known contributor should be used for the calculation, even if only a partial profile was deconvoluted.

   b. If the profile does not reach source attribution, two scenarios should be calculated. In other words, a second scenario should be calculated that does not include the major contributor as a known. The full profile of the known contributor should be used for the calculation, even if only a partial profile was deconvoluted.

2. Other exemplar DNA profiles which are positively associated (qualitatively “could be a contributor” or “cannot be excluded as a contributor”) with the mixture may be used as a known in the calculation. An alternative scenario should also be calculated with no known contributors.
3. Under certain case scenarios, the hypothesis may assume a second suspect as a known contributor. **This circumstance is generally very rare.** If a second suspect is used as a known, a second scenario should be calculated that does not include the known.

   a. Suspects are related and both are positively associated (qualitatively “could be a contributor” or “cannot be excluded as a contributor”) to the mixture. Suspects do not need to be used as a known if they are only positively associated but are not related.

   b. One suspect is the deconvoluted major contributor to the mixture and another suspect is positively associated and therefore will be used as a comparison sample. The known profile should be the deconvoluted unknown (e.g. the deconvoluted profile of Male Donor A) and should be a full profile.

B. **Effect of the choice of number of contributors**

1. The number of contributors invoked to explain the data will have an effect on the likelihood ratio. For a given hypothesis, using the minimum possible number of contributors will usually result in the lowest possible LR.

2. Use all available information, including assumed known contributors, to determine which pair of hypotheses (with how many contributors) to use. Only in the rare instance where the data support more than one scenario, additional calculations may be performed.
IV. User defined factors that affect the drop-out and drop-in rates

A. Drop-out rates vary depending upon the amount of template DNA in a sample. The template amount is entered by the user and the program interpolates the dropout rate based on validation data. Drop-in rates depend on the number of PCR cycles used.

Note: FST may use incorrect drop-in and drop-out rates if points 2, 3 and 4 below are not explicitly followed.

1. If different template amounts were amplified in different replicates, select the replicate with the most information. Alternatively, if different information is seen among the replicates, use all replicates (up to three), but select the highest template amount amplified. In this manner, the most conservative drop-out rates are used by FST.

2. If different template amounts were amplified using different cycling parameters, select the run with the most information. Do not combine results across cycle number settings. The program uses different drop-out and drop-in rates for 28 and 31 cycle samples.

3. Drop-out rates are programmed for samples amplified with 28 cycles with template DNA amounts ranging from 101pg to 500pg per amplification. Samples amplified with more than 500pg should be entered as 500pg. Samples amplified for 28 cycles with 100pg should be entered as 101pg.

4. Drop-out rates are programmed for samples amplified with 31 cycles with template DNA amounts of 100pg per amplification and below. Samples amplified in a range between 100 and 101pg should be entered as 100pg.

B. Drop-out rates also vary depending upon the number of contributors to a mixture. Generally for a given locus and template amount, the drop-out rate is higher for a three-person mixture than a two-person mixture.

1. To determine the number of contributors to a sample, follow the OCME mixture interpretation guidelines found in the STR Results Interpretation section of the manual.
2. In some circumstances it may be useful to do more than one scenario with varying numbers of contributors. For example, if a victim/elimination sample(s) is received and upon evaluation, it is determined that more contributors are present than what was determined at initial evaluation, two scenarios should be calculated: one with two contributors and no knowns and one with three contributors using the victim/elimination sample as a known.

C. Drop-out rates vary depending upon the approximate mixture ratio of the contributors.

1. If a mixture has no major contributor, the user specifies that the mixture is “non-deducible” and the program will use drop-out rates for 1:1 (or 1:1:1) mixtures.

2. If a mixture has a major contributor whose profile can be deconvoluted according to the OCME mixture interpretation guidelines, the user specifies that the mixture is “deducible” and the program will use drop-out rates for 4:1 (or 5:1:1) mixtures.

   a. The deconvoluted profile should have no fewer than 8 alleles over 4 loci (HT-DNA) or 6 loci (LT-DNA), otherwise consider the sample non-deducible.

   b. In this situation, FST should only be used if the comparison sample is not consistent with the major contributor’s profile.

V. Instructions

In the sections that follow, the user is guided through instructions for setting up files, running the FST program, and interpreting the results.

A. Creating Evidence, Comparison, and Known Contributor Files for FST

Evidence, comparison, and known contributor profiles can be uploaded into FST instead of being entered manually. In order to be uploaded, files must be formatted as tab delimited text files, as shown in Tables 1 and 2 below.

For comparison and known contributor profiles, homozygous alleles must appear twice. Tri-allelic loci may not be entered, as the program assumes that there will be a maximum of two alleles per locus. Incomplete or negative loci should be left blank for comparison and known profiles as well.
To create a text file for a comparison or known contributor profile from an allele table in Excel:

1. Open “Make Suspect or Victim Profile for Upload.xlt”
2. From the allele table, copy one donor’s name and profile. Alleles can be separated by commas and/or spaces.
3. Put the cursor on cell A4 in Sheet1 of “Make Suspect or Victim Profile for Upload.xlt”.
4. Right click, choose “Paste Special”, then “values”, then “OK” to paste profile data into the row.
5. Click anywhere else in the sheet. Then press Ctrl-m to run the macro.
6. Sorted results will appear in Sheet3. Verify that the values in Sheet3 are correct.
7. Save Sheet3 as a tab-delimited text file using the donor’s name or some other identifying information as the file name. Click “OK” and “Yes” when prompted.
8. Close “Make Suspect or Victim Profile for Upload.xlt” (no need to save this time) and re-open it in order to create the next text file. If the file is not closed and re-opened, the next profile will not be sorted properly.
Table 1. Format for uploadable comparison or known contributor profiles.

To create a text file from an evidence table in Excel:

1. Open “Make Evidence File for Upload.xlt”
2. Enter up to three amplifications for an ID28 sample. Alleles can be separated by commas and/or spaces.
3. Enter all three replicates for an ID31 sample for one item. Since FST takes into account drop-in/drop-out rates, data from all loci (whether they have repeating alleles or not) should be used in the calculation. Alleles can be separated by commas and/or spaces. Do not copy the pooled sample or the composite profile.
4. Put the cursor on cell A4 in Sheet1 of “Make Evidence File for Upload.xlt”.
5. Right click, choose “Paste Special”, then “values”, then “OK” to paste evidence profile data into rows 4 and 5 for duplicate amplifications or 4, 5, and 6 for triplicate amplifications.
6. Click anywhere else in the sheet. Click on the green button to run the macro.
7. Sorted results will appear in Sheet3. Verify that the values in Sheet3 are correct.
8. Save Sheet3 as a tab-delimited text file with an appropriate file name. Click “OK” and “Yes” when prompted.
9. Close “Make Evidence File for Upload.xlt” (no need to save this time) and re-open it in order to create the next text file. If the .xlt file is not closed and re-opened, it will not sort the next profile properly.

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>REPLICATE</th>
<th>ALLELE 1</th>
<th>ALLELE 2</th>
<th>ALLELE 3</th>
<th>ALLELE 4</th>
<th>ALLELE 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>1</td>
<td>10</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S1179</td>
<td>2</td>
<td>10</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D8S1179</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D21S11</td>
<td>1</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>D21S11</td>
<td>2</td>
<td>28</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D21S11</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7S820</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7S820</td>
<td>2</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D7S820</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF1PO</td>
<td>1</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF1PO</td>
<td>2</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CSF1PO</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3S1358</td>
<td>1</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3S1358</td>
<td>2</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D3S1358</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Etc…

Table 2. Format for uploadable evidence amplifications with duplicate runs. If triplicate runs were performed, data from the third amplification would appear in rows associated with REPLICATE 3, indicated by a “3” in the second column. Off-ladder alleles are acceptable as a whole number, decimal, or “<” or “>”. The macro limits the number of alleles per locus to 8. Additional alleles must be entered manually.
B. FST Home Screen

![FST Home Screen](image)

**Figure 1.** Home Screen of Forensic Statistical Tool (FST). After deciding which model to use, as outlined in Part II, parameters are specified and files are uploaded (or profiles are manually entered) through the FST web interface.
Figure 2. Select the appropriate test scenario from the “Select Scenario” drop-down box. Options are listed in Table 3 below. The option selected here is Comparison + Unknown / 2 Unknowns, which is used for a two-person mixture with a comparison profile, but no known contributor profile.
Table 3. Numerator and denominator options available in FST. “Comparison” refers to the test profile of interest. This profile is often from a suspect, but could belong to a victim or an elimination sample. “Known” refers to an assumed known contributor. “Unknown” refers to a randomly selected individual from a population of individuals that are unrelated to the Known, Comparison or one another.

Note: The random match probability should be routinely used for single source and deconvoluted profiles.
Figure 3. Ensure Identifiler is selected on the “Lab Kit” drop-down box.
Figure 4. Ensure 0.03 is selected on the “θ” drop-down box. For all routine testing performed at OCME, the default value of θ = 0.03 is to be used.
C. Uploading Files and Running FST

Figure 5. Enter the Comparison Profile name in the appropriate space on the Home Screen, then click “GO” on the bottom of the Home Screen to advance to the Data Entry Screen. Ignore the “BULK” option as this is reserved for quality control purposes.
Controlled versions of Department of Forensic Biology Manuals only exist in the Forensic Biology Qualtrax software. All printed versions are non-controlled copies.

© NYC OFFICE OF CHIEF MEDICAL EXAMINER
For mixtures, select “Yes” or “No” in the Deducible drop-down box. For single source samples, the Deducible option is set to “Yes” and cannot be changed.

Browse to select Comparison, Known and Evidence files.

![Forensic Statistical Tool](image)

**Figure 7.** If a model including a known contributor was selected, there will be space to upload a known profile.
Figure 8. After browsing to select Comparison, Known, and Evidence files, click “Preview” view uploaded data.
Figure 9. Profile Preview Screen. Uploaded data will be shown here. If a file was selected in error, click on “Edit” and then “Back” to re-upload the profile. Verify that the comparison and/or known(s) name(s) entered on the Home screen appear on this screen below the evidence profile entry area. Case and sample information may be entered or corrected on this screen, if necessary. If all information is correct, click “Compare” to run the analysis and generate results in a PDF file.
Figure 10. Results Screen. After clicking “Compare”, a pop-up window will provide the options to save or open the results file. Save the file as xx-xxxxx_sample name_FST in the appropriate folder and place a printout in the case file. Two person mixture results will be instantaneous. Three-person mixture results may require 10-15 minutes. Report the lowest of the four likelihood ratios shown on the bottom of the screen.
D. Interpretation of Results

It is very important that likelihood ratios are reported using the exact wording given below. Even minor deviation from this wording can lead to incorrect interpretation of results. Interpretation is always of the form “The DNA mixture found on [item] is X times more probable if the sample originated from A than if it originated from B. Therefore, there is [limited / moderate / strong / very strong] support that A contributed to this mixture, rather than B.”

Please note that the result is a “ratio” between two likelihoods and cannot be reported for just one hypothesis.

Reporting of the likelihood ratio (LR) depends on the comparison type selected and the value of the LR. Select the lowest value of the four likelihood ratios that appear at the bottom of the results page. This value will determine whether the result supports the prosecutor or the defense hypothesis. This value will also determine which descriptor (limited, moderate, strong, or very strong) to select in the second sentence. Use Table 4 to determine which descriptor to use in the second sentence. Note, only values that are equal to 1.00 should given the qualitative descriptor of “no conclusions”.

If the lowest LR is greater than one, the results are interpreted as shown below, using the example shown in Figure 10, in which the lowest value is 4.54e+08, or 4.54 x 10^8. If the lowest LR is between 10^6 and 10^14, report the result as “million”, “billion” or “trillion”. For example, report 4.54 x 10^8 as 454 million.

In the first report sentence, because the lowest LR in this example is greater than one, the DNA mixture is more probable if the prosecution hypothesis is true than if the defense hypothesis is true. In the second sentence, because 4.54 x 10^8 is greater than 1000, there is very strong support for the prosecutor’s hypothesis over the defense hypothesis.

If the comparison performed was Mr. Smith (comparison) + Unknown versus Two Unknowns (i.e., a two-person mixture with no known contributors), interpretation of the value above is:

The evidence profile is 454 million times more probable if the sample originated from Mr. Smith and one unknown, unrelated person than if it originated from two unknown, unrelated persons. Therefore, there is very strong support that Mr.
Smith and an unknown, unrelated person contributed to the mixture, rather than two unknown, unrelated persons.

If the comparison performed was Mr. Smith (comparison) + Mr. Green (known) versus Mr. Green + Unknown (i.e., a two-person mixture with one known contributor), interpretation of the value above is:

The evidence profile is 454 million times more probable if the sample originated from Mr. Smith and Mr. Green than if it originated from Mr. Green and an unknown, unrelated person. Therefore, there is very strong support that Mr. Smith and Mr. Green contributed to the mixture, rather than Mr. Green and an unknown, unrelated person.

If the lowest likelihood ratio is less than one, the DNA mixture found on the item is more probable if the defense hypothesis is true than if the prosecution hypothesis is true. In this situation, the reciprocal of the lowest LR is reported and the positions of the two hypotheses in the interpretation sentences are reversed. For example, if the four values at the bottom of the results page are:

\[
0.421 \quad 8.88e-02 \quad 1.49e-02 \quad 0.492
\]

the lowest value is 1.49e-02, or 0.0149. The reciprocal of this value is \(1 / 0.0149 = 67.114\). Report the results rounded down to three significant figures as below.

If the comparison performed was Mr. Smith (comparison) + Unknown versus Two Unknowns (i.e., a two-person mixture with no known contributors), interpretation of the value above is:

The evidence profile is 67.1 times more probable if the sample originated from two unknown, unrelated persons rather than from Mr. Smith and one unknown, unrelated person. Therefore, there is moderate support that two unknown, unrelated persons contributed to the mixture, rather than Mr. Smith and an unknown, unrelated person.

If the comparison performed was Mr. Smith (comparison) + Mr. Green (known) versus Mr. Green + Unknown (i.e., a two-person mixture with one known contributor), interpretation of the value above is:
The evidence profile is 67.1 times more probable if the sample originated from Mr. Green and one unknown, unrelated person rather than from Mr. Smith and Mr. Green. Therefore, there is moderate support that Mr. Green and an unknown, unrelated person contributed to the mixture, rather than Mr. Smith and Mr. Green.

If the LR is between $10^{-3}$ and $10^{9}$, the result will not appear in scientific notation. For example, if the results are

$$435.82 \quad 2993.882336.55 \quad 184.43$$

report a value of 184 (lowest value, rounded down to 3 significant figures), stating for example for a two-person mixture with no known contributor, “The evidence profile is 184 times more probable if the sample originated from Mr. X and one unknown, unrelated person than if it originated from two unknown, unrelated persons. Therefore, there is strong support that Mr. X and one unknown person contributed to the mixture, rather than two unknown, unrelated persons.”

<table>
<thead>
<tr>
<th>If the likelihood ratio is…</th>
<th>Then the evidence provides…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 0.001</td>
<td>Very strong support for $H_d$ over $H_p$</td>
</tr>
<tr>
<td>0.001 to 0.01</td>
<td>Strong support for $H_d$ over $H_p$</td>
</tr>
<tr>
<td>0.01 to 0.1</td>
<td>Moderate support for $H_d$ over $H_p$</td>
</tr>
<tr>
<td>0.1 to 1.0</td>
<td>Limited support for $H_d$ over $H_p$</td>
</tr>
<tr>
<td>1 to 10</td>
<td>Limited support for $H_p$ over $H_d$</td>
</tr>
<tr>
<td>10 to 100</td>
<td>Moderate support for $H_p$ over $H_d$</td>
</tr>
<tr>
<td>100 to 1000</td>
<td>Strong support for $H_p$ over $H_d$</td>
</tr>
<tr>
<td>Greater than 1000</td>
<td>Very strong support for $H_p$ over $H_d$</td>
</tr>
</tbody>
</table>

Table 4. Qualitative interpretation of likelihood ratios. Likelihood ratios provide a measure of the strength of support in favor of one hypothesis over the other. Let $H_p$ represent the prosecution hypothesis, or the hypothesis that the comparison sample did contribute to the sample. Let $H_d$ represent the defense hypothesis, or the hypothesis that the comparison sample did not contribute to the sample. Use the values suggested by Butler (2005, Forensic DNA Typing. Burlington, MA: Elsevier Academic Press, pp 513), as shown here, to describe the strength of support for either $H_p$ or $H_d$.  

© NYC OFFICE OF CHIEF MEDICAL EXAMINER