DNA Mixture Interpretation Webcast
April 12, 2013


http://www.cstl.nist.gov/strbase/mixture.htm

Statistical Approaches

Michael D. Coble
National Institute of Standards and Technology
In every workshop presented and supported by the NIJ Training Grant (2008-DN-BX-K158)

• Participants said they needed more training in…
  – Mixture analysis
  – **Statistics** related to mixtures

This doesn’t have to be a Shakespearean Tragedy!
Stats Required for Inclusions

SWGDAM Interpretation Guideline 4.1: “The laboratory must perform statistical analysis in support of any inclusion that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis.”

Buckleton & Curran (2008): “There is a considerable aura to DNA evidence. Because of this aura it is vital that weak evidence is correctly represented as weak or not presented at all.”

“The DAB finds either one or both PE or LR calculations acceptable and strongly recommends that one or both calculations be carried out whenever feasible and a mixture is indicated”

- Probability of exclusion (PE)

- Likelihood ratios (LR)
Statistical Approaches with Mixtures

See Ladd et al. (2001) Croat Med J. 42:244-246

“Exclusionary” Approach

Random Man Not Excluded (RMNE)

Combined Prob. of Inclusion (CPI)

Combined Prob. of Exclusion (CPE)

“Inferred Genotype” Approach

Random Match Probability [modified] (mRMP)

Likelihood Ratio (LR)

“Allele-centric”

“Genotype-centric”
Statistical Approaches with Mixtures

- **Random Man Not Excluded (CPI)** - The probability that a random person (unrelated individual) would not be excluded as a contributor to the observed DNA mixture.

\[
PI = (f(a) + f(b) + f(c) + f(d))^2
\]

\[
CPI = PI_{M1} \times PI_{M2} \cdots
\]

\[
CPE = 1 - CP1
\]
Breaking down the math...

CPI – tries to find all possible “random” persons included in this mixture...

\[(a + b + c + d)^2\]

\[= (a + b + c + d) (a + b + c + d)\]

“FOIL”
Breaking down the math...

"FOIL"

= \((a + b + c + d) (a + b + c + d)\)

= \((a^2 + 2ab + 2ac + 2ad + b^2 + \ldots)\)
RMNE Statistics

CPI – tries to find all possible “random” persons included in this mixture…

“Included Genotypes”

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>BB</th>
<th>CC</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td></td>
<td>BC</td>
<td></td>
<td>CD</td>
</tr>
<tr>
<td>AC</td>
<td></td>
<td>BD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
RMNE Statistics

An “Illogicality” of using RMNE

AA + BCD ???

Sure, why not? It fits!

Risk of including individuals not in the mixture
• **modified Random Match Probability (mRMP)**
  – The major and minor components can be successfully separated into individual profiles. A random match probability is calculated on the evidence as if the component was from a single source sample.

\[
mRMP_{\text{minor}} = 2pq = 2f(b)f(c)
\]
Statistical Approaches with Mixtures

- **Likelihood Ratio** - Comparing the probability of observing the mixture data under two (or more) alternative hypotheses; in its simplest form LR = 1/RMP

\[
\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{1}{P(E \mid H_2)} = \frac{1}{2pq} = 1/RMP
\]

- \( E \) = Evidence
- \( H_1 \) = Prosecutor’s Hypothesis (the suspect did it) = 1
- \( H_2 \) = Defense Hypothesis (the suspect is an unknown, random person)
Comparison of the Methods

“Included Genotypes” RMNE
- AA  BB  CC  DD
- AB  BC  CD  AD
- AC  BD

“Included Genotypes” LR/mRMP
- AA  BB  CC  DD
- AB  BC  CD  AD
- AC  BD
We conclude that the two matters that appear to have real force are:
(1) LRs are more difficult to present in court and
(2) the RMNE statistic wastes information that should be utilised.
Review of Two Thresholds

**Called Peak**
(Greater confidence a sister allele has not dropped out)

200 RFUs

**Called Peak**
(Cannot be confident dropout of a sister allele did not occur)

50 RFUs

**Stochastic Threshold**
The value above which it is reasonable to assume that allelic dropout of a sister allele has not occurred

**Analytical Threshold**
Minimum threshold for data comparison and peak detection in the DNA typing process

2-Person Mixture
If CPI/CPE Stats are Used

Since exclusionary statistics cannot adjust for the possibility of dropout, and does not take the number of contributors into account, any loci with alleles below the stochastic threshold cannot be used in the CPI statistic.
If CPI/CPE Stats are Used
(ST = 200 RFU)
If CPI/CPE Stats are Used
If CPI/CPE Stats are Used

<table>
<thead>
<tr>
<th>Can use</th>
<th>Cannot use</th>
</tr>
</thead>
<tbody>
<tr>
<td>D21</td>
<td>D8 D2</td>
</tr>
<tr>
<td>CSF</td>
<td>D7 vWA</td>
</tr>
<tr>
<td>D3</td>
<td>TH01 D18</td>
</tr>
<tr>
<td>D19</td>
<td>D13 D5</td>
</tr>
<tr>
<td>TPOX</td>
<td>D16 FGA</td>
</tr>
</tbody>
</table>

Impact: discarding 2/3 of the data
If CPI/CPE Stats are Used

- CPI statistics using FBI Caucasian Frequencies
  - 1 in 71 Caucasians included
  - 98.59% Caucasians excluded
If CPI/CPE Stats are Used
(ST = 150 RFU)

The impact of changing thresholds
If mRMP/LR Stats are Used

• Since there is an assumption to the number of contributors, it is possible to use data that falls below the ST.
### mRMP - D18S51

#### If Assume 2 Contributors....

<table>
<thead>
<tr>
<th>Major</th>
<th>Minor</th>
</tr>
</thead>
<tbody>
<tr>
<td>16,18</td>
<td>14,20</td>
</tr>
</tbody>
</table>

\[
mRMP_{\text{minor}} = 2pq = 2 \times f(14) \times f(20) = 2 \times 0.1735 \times 0.0255 = 0.00884 \text{ or } 1 \text{ in } 113
\]

\[
\text{(LR} = 113)\]
mRMP/LR

Potential for Drop-out

Diagram showing genetic markers and potential drop-out points.
If mRMP/LR Stats are Used

<table>
<thead>
<tr>
<th>Can use</th>
<th>Loci with potential D-out</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8</td>
<td>D7  D2</td>
</tr>
<tr>
<td>D21</td>
<td>TH01  vWA</td>
</tr>
<tr>
<td>D18</td>
<td>D13  D5</td>
</tr>
<tr>
<td>D3</td>
<td></td>
</tr>
<tr>
<td>D19</td>
<td></td>
</tr>
<tr>
<td>TPOX</td>
<td></td>
</tr>
<tr>
<td>FGA</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td></td>
</tr>
</tbody>
</table>
The “2p” Rule

- The “2p” rule can be used to statistically account for zygosity ambiguity – i.e. is this single peak below the stochastic threshold the result of a homozygous genotype or the result of a heterozygous genotype with allele drop-out of the sister allele?
2p – SWGDAM Guidelines

• 5.2.1.3.1. The formula $2p$, as described in recommendation 4.1 of NRCII, may be applied to this result.

• 5.2.1.3.2. Instead of using $2p$, the algebraically identical formulae $2p - p^2$ and $p^2 + 2p(1-p)$ may be used to address this situation without double-counting the proportion of homozygotes in the population.
Macbeth/Duncan Profile - TH01

Major – 7, 7

Possible Minor Contributors

7, 9.3 \hspace{1cm} (2pq)

9.3, 9.3 \hspace{1cm} p^2

9.3, ? \hspace{1cm} 2p \hspace{0.5cm} (or \hspace{0.2cm} p^2 + 2p(1 - p))
Macbeth/Duncan Profile - TH01

\[
\frac{P(E|H_1)}{P(E|H_2)} = \frac{V \& S}{V \& U} = \frac{f_7^2 + f_7(1-f_7)\theta \& 1}{f_7^2 + f_7(1-f_7)\theta \& 2p} = \frac{p^2 + 2p(1-p)}{f_{9.3}^2 + 2f_{9.3}(1-f_{9.3})} = \frac{1}{f_{9.3}^2 + 2f_{9.3}(1-f_{9.3})}
\]

\[
V = 7, 7
\]

\[
U = 7, 9.3
\]

\[
9.3, 9.3
\]

\[
9.3, ?
\]

\[
f_{9.3} = 0.3054
\]

\[
V = 7, 7
\]

\[
9.3, 9.3
\]

\[
9.3, ?
\]
Macbeth/Duncan Profile - TH01

\[
P(E|H_1) \quad \frac{V \ & \ S}{V \ & \ U} = \frac{1}{p^2 + p(1-p)\theta + 2pq}
\]

\[
V = 7, \ 7
\]

\[
U = 7, \ 9.3
\]

\[
9.3, \ 9.3
\]

Let \( ST = 125 \) RFU

\[
f_{9.3} = 0.3054
\]

\[
f_7 = 0.1724
\]

\[
\frac{1}{0.2007} = 4.98
\]
Macbeth/Duncan Profile - TH01

\[
\begin{align*}
\text{LR} & \\
\text{ST} = 200 \ (2p \text{ is used}) & 1.93 \\
\text{ST} = 125 \ (2pq \text{ is used}) & 4.98
\end{align*}
\]

2p is conservative...
The “2p” Rule

• “This rule arose during the VNTR era. At that time many smaller alleles “ran off the end of the gel” and were not visualised.”

- Buckleton and Triggs (2006)

Is the 2p rule always conservative?”
The “2p” Rule

- Stain = aa
- Suspect = aa
- \[ f(a) = 0.10 \quad 1/p^2 = 100 \quad 1/2p = 5 \]
The “2p” Rule

Stain = aa
Suspect = ab

\[ f(a) = 0.10 \quad 1/2p = 5 \]
Is there a way forward?
• “The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of \textit{probabilistic models to circumvent the requirement for a threshold} and to safeguard the legitimate interests of defendants.”
Summary of the Issues

• We need to move away from the interpretation of mixtures from an “allele-centric” point of view.
• Methods to incorporate probability will be necessary as we make this transition and confront the issues of low-level profiles with drop-out.

• “Just as logic is reasoning applied to truth and falsity, probability is reasoning with uncertainty”
  -Dennis Lindley
Summary of the Issues

• The LR is a method to evaluate evidence that can overcome many of the limitations we are facing today. ISFG Recommendations are published.

• This will require (obviously) software solutions... however, we need to better understand and be able to explain the statistics as a community.

• “But, for my own part, it was Greek to me”
  — William Shakespeare, *Julius Caesar*

• “We know what we are, but know not what we may be.” — William Shakespeare, *Hamlet*
Summary of the Issues

- Extensive training will be necessary – and a single 8 hour workshop will once a year will not suffice.
Thank you for your attention

Contact Information

Michael D. Coble
Forensic Biologist
michael.coble@nist.gov
301-975-4330

http://www.cstl.nist.gov/strbase

Additional DNA mixture information available at:
http://www.cstl.nist.gov/strbase/mixture.htm