NOC/t: A Tool for Determining the Number of Contributors from Complex DNA Mixtures

Boston University School of Medicine
Program in Biomedical Forensic Sciences
72 E. Concord Street, Boston, MA 02118

Harish Swaminathan, Desmond Lun, Sarah Norsworthy, Genevieve Wellner, Kayleigh Rowan, Muriel Medard and Catherine Grgicak

Download from: www.bu.edu/dnamixtures

2014 CAC DNA Workshop
San Diego, CA
May 6, 2014
DNA Interpretation

Practice

- Is this peak an allele?
- Am I observing all alleles?
- How many contributors?
- Can I infer genotypes?
- Is my suspect Included/Excluded and/or what is the statistic?

Research & Validation

- Analytical Threshold
- Stochastic Threshold/Pr(D)
- When does allele counting become inaccurate
- Peak Height Considerations/Stutter Considerations
- When is CPE/CPI or LR relevant
## Inspiration for NOC/Lt

**Boston University School of Medicine**

Program in Biomedical Forensic Sciences

72 E. Concord Street, Boston, MA 02118

---

### Table: Suspect DNA Profile

<table>
<thead>
<tr>
<th>Suspect 1</th>
<th>D8S1179</th>
<th>D21S11</th>
<th>D7S820</th>
<th>CSF1PO</th>
<th>Conclusion</th>
<th>Discussion Point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14,16</td>
<td>30,30.2</td>
<td>8,9</td>
<td>12,12</td>
<td>Included: 1 in 9</td>
<td>CPI stat is not using information in the signal. (LR?)</td>
</tr>
</tbody>
</table>

Minimum number of contributors of 2 is INCORRECT < the actual number of contributors!
Inspiration for the Development of NOC/It

### Table

<table>
<thead>
<tr>
<th>Suspect 2</th>
<th>D8S1179</th>
<th>D21S11</th>
<th>D7S820</th>
<th>CSF1PO</th>
<th>Conclusion</th>
<th>Discussion Point</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>13,14</td>
<td>26,31.2</td>
<td>9,10</td>
<td>11,12</td>
<td>Excluded or Inconclusive</td>
<td>Exclusion is dependent upon the assumed NOC and interpretation procedure. If NOC = 3 → exclude NOC &gt; 3 → inconclusive</td>
</tr>
</tbody>
</table>
The Interferences that Complicate DNA Interpretation

- **Allele Overlap**
  - Increases as the actual number of contributors increases
- **Allelic Dropout**
  - Increases as template decreases
  - Increases with Analytical Threshold (AT)
- **Peak Ratio Imbalance/Peak Height**
  - Variance increases as template decreases
- **Stutter (forward and reverse)**
  - Increases as template decreases
  - Is not always detected at low-templates (i.e., stutter dropout) – AT dependent
- **False Detections of Noise/Interferences**
  - Baseline noise increases with template
  - Increases as the AT decreases
The Interferences that Complicate DNA Interpretation

- The following slides will discuss effects of the various interferences on determining the NOC (number of contributors)
### Effects of Allele Overlap and Interpretation - MLE

<table>
<thead>
<tr>
<th></th>
<th>n=1</th>
<th>n=2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Person 1</td>
<td>Person 1</td>
</tr>
<tr>
<td></td>
<td>12,13</td>
<td>12,12</td>
</tr>
<tr>
<td></td>
<td>12,13 or 13,13</td>
<td>12,13 or 12,13 or 13,13</td>
</tr>
<tr>
<td></td>
<td>13,13</td>
<td>12,12 or 12,13</td>
</tr>
</tbody>
</table>

**Equation (1)**

\[
L_A(x) = \frac{\sum_{r_1=0}^{r_n} \sum_{r_2=0}^{n-r_1} \ldots \sum_{r_{n-1}=0}^{n-r_1-r_2-\ldots-r_{n-1}} (2x)! \prod_{i=1}^{c} \prod_{j=0}^{n_i-1} [(1-\theta)p_i + j\theta] \prod_{i=1}^{c} \prod_{j=0}^{u_i-1} [(1-\theta) + j\theta]}{r! \prod_{j=0}^{2x-1} \prod_{j=0}^{2x-1} [(1-\theta) + j\theta]}
\]

**Equation (2)**

\[
\max_{j=1,2,3,\ldots} L_A(x=j)
\]

Equation (1) takes into account the variation in the subpopulation allele frequencies. When there is no need to consider population subdivision, the likelihood of the data is simply obtained by setting $\theta$ to zero.

*The Likelihood Estimator*—The maximum likelihood estimation of $x$, when a single marker $A$ is considered, satisfies:

\[
2f_{12}f_{13}
\]

\[
2(0.23)(0.14)
\]

\[
6.4 \times 10^{-2}
\]

\[
4f_{12}^3f_{13} + 6f_{12}^2f_{13}^2 + 4f_{12}f_{13}^3
\]

\[
4(0.23)^3(0.14) + 6(0.23)^2(0.14)^2 + 4(0.23)(0.14)^3
\]

\[
1.55 \times 10^{-2}
\]

---

Effects of Allele Dropout on Determining NOC: $n_c=1$

$E_{D8S1179} = 13,14$

$Pr(E|n_c = 1) = 2f_{13}f_{14}(1 - Pr(DO))^2$

If $Pr(DO) = 0$

Then

$Pr(E|n_c = 1) = 2f_{13}f_{14}$

$Pr(E|n_c = 1) = 2 \cdot 0.3393 \cdot 0.2015$

$Pr(E|n_c = 1) = 0.13675$

If $Pr(DO) = 0.4$

Then

$Pr(E|n_c = 1) = 2f_{13}f_{14}(1 - Pr(DO))^2$

$Pr(E|n_c = 1) = 2 \cdot 0.3393 \cdot 0.2015 \cdot 0.36$

$Pr(E|n_c = 1) = 0.04923$
Effects of Allele Dropout on Determining NOC: $n_c=2$

$$E_{DBS1179} = 13,14$$

$$Pr(E|n_c=2) = (4f_{13}f_{14}^3 + 6f_{13}f_{14}^2 + 4f_{13}^3f_{14}) \cdot (1 - Pr(DO))^4 + 4Pr(DO) \cdot (3f_{13}f_{14}^2 + 3f_{13}^2f_{14}) \cdot (1 - Pr(DO))^3 + 6Pr(DO)^2 \cdot (2f_{13}f_{14}) \cdot (1 - Pr(DO))^2$$

**Term #1**

<table>
<thead>
<tr>
<th>Person 1</th>
<th>Person 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,13</td>
<td>13,14 or 14,14</td>
</tr>
<tr>
<td>13,14</td>
<td>13,14 or 13,13 or 14,14</td>
</tr>
<tr>
<td>14,14</td>
<td>13,13 or 13,14</td>
</tr>
</tbody>
</table>

**Term #2**

<table>
<thead>
<tr>
<th>Person 1</th>
<th>Person 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,13</td>
<td>14,0</td>
</tr>
<tr>
<td>13,14</td>
<td>13,0 or 14,0</td>
</tr>
<tr>
<td>14,14</td>
<td>13,0</td>
</tr>
<tr>
<td>13,0</td>
<td>13,14 or 14,14</td>
</tr>
<tr>
<td>14,0</td>
<td>13,13 or 13,14</td>
</tr>
</tbody>
</table>

**Term #3**

<table>
<thead>
<tr>
<th>Person 1</th>
<th>Person 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,14</td>
<td>0,0</td>
</tr>
<tr>
<td>13,0</td>
<td>14,0</td>
</tr>
<tr>
<td>14,0</td>
<td>13,0</td>
</tr>
<tr>
<td>0,0</td>
<td>13,14</td>
</tr>
</tbody>
</table>

*If* $Pr(DO) = 0$

Then

$$Pr(E|n_c=2) = 0.008626$$

*If* $Pr(DO) = 0.4$

Then

$$Pr(E|n_c=2) = 0.2202$$
Effects of Allele Dropout on Determining NOC: $n_c=1$ v. 2 v. 3

<table>
<thead>
<tr>
<th>Locus</th>
<th>Allele 1</th>
<th>Allele 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D8S1179</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>D21S11</td>
<td>29</td>
<td>32.2</td>
</tr>
<tr>
<td>D7S820</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>CSF1PO</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>D3S1358</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>TH01</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>D13S317</td>
<td>11</td>
<td>n/a</td>
</tr>
<tr>
<td>D16S539</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>D2S1338</td>
<td>19</td>
<td>24</td>
</tr>
<tr>
<td>D19S433</td>
<td>15</td>
<td>n/a</td>
</tr>
<tr>
<td>vWA</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>TP0X</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>D18S51</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>D5S818</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>FGA</td>
<td>20</td>
<td>n/a</td>
</tr>
</tbody>
</table>

How is $Pr(\text{DO})$ determined?
Probability that the NOC is 1 decreases, even when no more than 2 alleles are observed per locus, as target amount decreases.
Effects of Stutter

Typically 5-15% of true allele in tetranucleotide repeats STR loci

Occurs less frequently (typically <2%)


**Deletion** caused by slippage on the copied (bottom) strand

**Insertion** caused by slippage of the copying (top) strand
Effects of Stutter on Interpretation

35/541 = 7%  55/314 = 18%

Is the NOC 1 or 2?

How do we determine the likelihood that the peak is an allele or stutter, and how does that affect our overall interpretation?
Effects of AT/Baseline Noise

Optimal analysis would be one that considered and interpreted ALL signal.
How do we determine the actual NOC using \textit{ALL} signal?

Let $E$ be the evidence, i.e. the genotyping results, and $n_c$ be the number of contributors to the stain. The APP that the number of contributors equals $i$ is defined as:

$$APP(i) := \Pr(n_c = i \mid E)$$

Using Bayes’ rule, we can calculate the APP as

$$APP(i) = \Pr(n_c = i \mid E) = \frac{\Pr(E \mid n_c = i)\Pr(n_c = i)}{\Pr(E)}$$

$$\sum_{i=1}^{\infty} APP(i) = 1$$
The NOC/t Procedure

- The next slides will discuss how to use NOC/t and the NOC/t interface
NOC It Procedure and Interface - Input
Choose a location and name for the output file and choose the DNA input value, i.e., ng, and choose the maximum number of contributors to test.
NOC\textit{It} Procedure

\begin{itemize}
\item \( n_{\text{max}} = 1 \)
\item Runtime = 1 min
\item 2Core, 2.9 GHz
\item \( n_{\text{max}} = 5 \)
\item Runtime = \( \sim 15 \) h
\item 2Core, 2.9 GHz
\end{itemize}
**NOCIt Procedure - Output**

**Example 2,**

- **Number of contributors: 0**
  - Time taken: 0.0m
  - D16S539 = 0
  - D18S51 = 0
  - D5S818 = 0
  - D2S1338 = 0
  - D7S820 = 0
  - vWA = 0
  - TPOX = 0
  - TH01 = 0
  - AMEL = 0
  - FGA = 0
  - D3S1358 = 0
  - CSF1PO = 0
  - D8S1179 = 0
  - D13S317 = 0
  - D21S11 = 0
  - D19S433 = 0
  - Likelihood: 0
  - Probability: 0

- **Number of contributors: 1**
  - Time taken: 0.32m
  - D16S539 = 3.519528450116965986746809736312419E-16 : [10.0]
  - D18S51 = 7.5795465805796717578900180707229E-18 : [20.0, 18.0]
  - D5S818 = 3.716732252940266365501621420808956E-14 : [12.0, 8.0]
  - D2S1338 = 3.01925703350527189618993844526659E-12 : [23.0, 22.0]
  - D7S820 = 4.56551017410597783682367538368022E-11 : [10.0, 8.0]
  - vWA = 8.476115635320640308481546683833E-15 : [17.0]
  - TPOX = 1.248513314172054319650746402676997E-12 : [6.0, 10.0]
  - TH01 = 3.460593759465737047605300327565956E-12 : [6.0, 8.0]
  - AMEL = 8.17827436439554541782532073396990E-7 : [Y, X]
  - FGA = 4.642919246222450769068935656819E-15 : [24.0, 22.0]
  - D3S1358 = 1.12197728615276754926697426894062E-13 : [15.0, 17.0]
  - CSF1PO = 4.7600371392264118934762281542021E-15 : [12.0, 10.0]
  - D8S1179 = 5.931571999965104372196292381570967E-18 : [12.0, 11.0]
  - D13S317 = 2.014410815183429571460921473352745E-16 : [12.0, 9.0]
  - D21S11 = 1.66266096714679812055337378154607E-23 : [28.0, 33.2]
  - D19S433 = 4.839364993287225915372863235526102E-13 : [19.0, 11.0]
  - Likelihood: 9.758249617805906487502981106556181E-222
  - Probability: 1

Result is a distribution on the Number of Contributors – the output provides 1) the most likely number and 2) whether other NOC’s are probable.
NOC/\ell t\ Results

- The following slides discuss results from mock low-template, complex DNA samples obtained when NOC/\ell t, MLE and allele counting (MAC) are used to assess the number of contributors.
Summary of NOC1t, MAC (Maximum Allele Count) and MLE (Maximum Likelihood Estimator) Results: The accuracy of (♦)MAC, (■)MLE and (▲)NOC1t for all samples/mixtures plotted against injection time.

Accuracy = \( \frac{\text{Most Likely NOC}}{\text{Actual NOC}} \times 100\% \)

Tested 103 samples consisting of 1-, 2-, 3-, 4- and 5- person mixtures with targets ranging from 0.008 – 0.25 ng, amplified with Identifiler® Plus (29-cycles) and run on a 3130 Genetic Analyzer, 10sec and 3 kV injection.
Allele Count, Likelihood Estimator and NOC/It results: The Breakdown

Tested 103 samples consisting of 1-, 2-, 3-, 4- and 5- person mixtures with targets ranging from 0.008 – 0.25 ng, amplified with Identifiler® Plus (29-cycles) and run on a 3130 Genetic Analyzer, 10sec and 3 kV injection.

X-axis: Number of Contributors Output from (■) MAC (AT= 50RFU), (■) MLE (AT=50 RFU), (■) MAC (AT=color, template specific), (■) MLE (AT = color, template specific, (■) NOC/It
NOC/It results: The Breakdown

Tested 103 samples consisting of 1-, 2-, 3-, 4- and 5- person mixtures with targets ranging from 0.008 – 0.25 ng, amplified with Identifiler® Plus (29-cycles) and run on a 3130 Genetic Analyzer, 10sec and 3 kV injection.

Summary of NOC/It results: The (■) accuracy of NOC/It and the (■) accuracy that NOC/It determines that there is a reasonable probability the sample originated from $n$ contributors (i.e., the APP is at least 0.01)
The following discuss the algorithm NOC/It utilizes to assess the number of contributors given the evidence.
For \( n=1, \ldots, n_{\text{max}} \). Let \( G_i, \Theta_i \) be the genotype of and fraction of total DNA mass, respectively, contributed by \( i \in \{1, \ldots, n_{\text{max}}\} \), and let \( G \) and \( \Theta \) be the \( n_{\text{max}} \)-component vectors of the \( G_i \) and \( \Theta_i \), respectively. We have

\[
\Pr(E|N = n) = \int \sum_{\Theta \in \Delta^{n-1}} \Pr(E|G = g, \Theta = \Theta, N = n) \Pr(G = g) f_{\Theta}(\Theta)
\]

where \( \Delta^{n-1} := \{(x_1, \ldots, x_n) \in \mathbb{R}^n | \sum_{i=1}^{n} x_i = 1, x_i \geq 0 \ \forall \ i \} \) is the unit \( n-1 \) simplex, \( g \) is the space of possible genotypes (for both alleles of a contributor) in the population, and \( f_{\Theta} \) is the probability density function of \( \Theta \), which we assume to be uniform over \( \Delta^{n-1} \). The distribution \( \Pr(E|G=g, \Theta=\Theta, N=n) \) is known because it is derived from calibration samples with known genotype.

We implement NOClt using a Monte-Carlo sampling algorithm. We generate random samples of \( g \) and \( \Theta \) using the background population allele frequencies and \( f_{\Theta} \) and, for each sample, we compute \( \Pr(E|G=g, \Theta=\Theta, N=n) \). After a large number of samples, we average all the computed values of \( \Pr(E|G=g, \Theta=\Theta, N=n) \) to obtain an approximation of the equation above, then we calculate the APP.
NOC/It Background Calculations, an example

NOC/It starts with determining the $L(n=1)$; 
Let's say at D7S820 the possible alleles range from 8 to 12 repeats

1. A random genotype is chosen based on the possible alleles of D7S820 as per allele frequencies in 'allele frequency table': i.e., 9, 10 $\rightarrow$ Bernoulli trial to assess dropout (allele & stutters)

2. Likelihood that $n=1$ given $G=9,10$ and $\theta=1$ 
Let us examine the peak height of the 8, 9, 10, 11 and 12 positions

3. $L(8\text{ of } H\ 144|\text{noise}) = v.\ low$

4. $L(9\text{ of } H\ 0|\text{allele}) = \text{moderate}$

5. $L(10\text{ of } H\ 12|\text{allele}) = \text{moderate}$

6. Bernoulli trial, if stutter, 
$L(11\text{ of } SR_{\text{forward}}\ 1.33|\text{stutter}_{\text{forward}}) = v.\ low$

7. $L(12\text{ of } H\ 36|\text{noise}) = v.\ low$

8. Overall: likelihood that one person, of $G=9,10$, contributed, given the data, is very low.
NOC/lt Background Calculations, 2nd iteration

NOC/lt starts with determining the L(n=1);

Let’s say at D7S820 the possible alleles range from 8 to 12 repeats.

1. A random genotype is chosen based on the possible alleles of D7S820 as per allele frequencies in ‘allele frequency table’: i.e., 8, 12 → Bernoulli trial to assess dropout (allele & stutters)
2. Likelihood that n=1 given G=8,12 and θ=1

Let us examine the peak height of the 8, 9, 10, 11 and 12 positions

3. L(8 of H 144|allele) = high

4. Bernoulli trial, if stutter, L(9 of SR_{forward} 0|stutter) = moderate

5. L(10 of H 12|noise) = v. low

6. Bernoulli trial, if stutter, L(11 of SR_{reverse} 0.44|stutter_{reverse}) = low

7. L(12 of H 36|allele) = high

8. Overall: likelihood that one person, of G=8,12, contributed, given the data, is low.
Many, many iterations are completed for the n=1 scenario and an ‘average’ likelihood is calculated. Monte Carlo random sampling stops when the ‘average’ stops changing. Loci are treated independently.

Then, NOC/It starts determining the L(n=2);

1. In the n=2 instance, 2 genotypes would be randomly chosen and a random $\theta$ (proportion) would be chosen. Based on $\theta$, we could calculate the probability of the peak of height $H$ given our 4 classifications: allele, noise, stutter$_{reverse}$, stutter$_{forward}$
2. Many, many iterations are completed for the n=2 scenario and an ‘average’ is calculated.

Then, NOC/It determines, in the same way, the L(n=3), L(n=4) and L(n=5). L(n=0) is also determined.
NOC/t Background Calculations

0.03 ng DNA amplified

<table>
<thead>
<tr>
<th>Tested NOC</th>
<th>Likelihood at D7S820</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1.0925e^{-13}</td>
</tr>
<tr>
<td>3</td>
<td>4.9051e^{-14}</td>
</tr>
<tr>
<td>4</td>
<td>1.8473e^{-14}</td>
</tr>
<tr>
<td>5</td>
<td>6.7694e^{-15}</td>
</tr>
</tbody>
</table>

\[
\Pr(n_c = 1 \mid E)_{D7S820} = \frac{0}{0 + 0 + 1.0925e^{-13} + 4.9051e^{-14} + 1.8473e^{-14} + 6.7694e^{-15}} = 0
\]

\[
\Pr(n_c = 2 \mid E)_{D7S820} = \frac{1.0925e^{-13}}{0 + 0 + 1.0925e^{-13} + 4.9051e^{-14} + 1.8473e^{-14} + 6.7694e^{-15}} = 0.59522
\]

\[
\Pr(n_c = 2 \mid E)_{15\text{Loci}} = 0.99991
\]
The following slides discuss the background models utilized by NOC/t to assess the likelihood of a peak of height $H$ (or ratio in the case of stutter) given that it is noise, stutter or peak. It also gives the models utilized to assess the probability of peak/stutter non-detection, i.e., dropout.
A linear model to describe the baseline noise for every locus is utilized in NOC/It as per,

\[ N_l = \alpha_{N,l} x + \beta_{N,l} \]

Where \( N_l \) is average or standard deviation of noise, \( x \) is the independent variable (i.e. target mass) and \( \alpha_{N,l} \) and \( \beta_{N,l} \) are the parameters (i.e., slope and y-intercept) obtained for each locus. The noise height distribution is assumed to be Gaussian.

Baseline Noise: The (◊) average noise peak heights and the (□) standard deviation of the noise heights plotted against target for samples amplified with 0.007 to 0.5 ng of DNA for the four representative loci a) CSF1P0, b) D3S1358, c) D18S51 and d) D5S818. The injection time was 10 seconds. Also presented is the trendline with an \( R^2 \) of > 0.82 and > 0.56 for the average and standard deviations, respectively.
A linear model to describe the allelic signal obtained from heterozygous loci in the Calibration file is utilized in NOC/It as per,

\[ A_I = \alpha_{A,I} x + \beta_{A,I} \]

Where \( A_I \) is average or standard deviation of allele signal (i.e., height or area) from heterozygous loci, \( x \) is the independent variable (i.e., target mass) and \( \alpha_{A,I} \) and \( \beta_{A,I} \) are the parameters obtained for each locus. The distribution is assumed to be Gaussian.

Allele Peak Height: The (◊) average and (□) standard deviations of allele peak heights (heterozygous peaks only) plotted against target mass for all loci. The linear trendline is also shown.
An exponential model to describe the probability of drop out (i.e., non-detection of allelic signal) for every locus is utilized in NOC/It. It is calculated by examining heterozygous loci as per,

$$D_l = \alpha_{D,l} e^{-\beta_{D,l}x}$$

Where $D_l$ is the probability of allele dropout at all heterozygous loci in the calibration file (i.e., #observed alleles/#expected alleles), $x$ is the independent variable (i.e., target mass) and $\alpha_{D,l}$ and $\beta_{D,l}$ are the parameters obtained for each locus.

Allele Dropout: Observed frequencies of dropout (●) for representative locus D16S519 at 7 low-template target amounts, showing the (●) logistic and (▬) exponential models. The residuals are provided on the top of the plot.
An exponential model to describe the stutter ratio for every locus is utilized in NOC/l as per,

\[ SR_l = \alpha_{SR,l} e^{-\beta_{SR,l}x} + \gamma_{SR,l} \]

Where \( SR_l \) is average or standard deviation of the stutter ratio obtained from heterozygous loci, \( x \) is the independent variable (i.e., target mass) and \( \alpha_{SR,l} \), \( \beta_{SR,l} \) and \( \gamma_{SR,l} \) are the parameters obtained for each locus. The distribution is assumed to be Gaussian.

Stutter Ratio Model: The mean and standard deviations of the reverse stutter ratios plotted against target (ng), shown with the exponential fit. \( R^2 \) values > 0.78. Forward stutter is similarly characterized.
An exponential model to describe the probability of non-detection of stutter (forward & reverse) for every locus is utilized in NOC/It as per,

\[ SD_l = \alpha_{SD,l} e^{-\beta_{SD,l}x} \]

Where \( SD_l \) is the probability of either forward or reverse stutter non-detection at all heterozygous loci in the calibration file (i.e., \#observed stutter/\#expected stutter), \( x \) is the independent variable (i.e., target mass) and \( \alpha_{SD,l} \) and \( \beta_{SD,l} \) are the parameters obtained for each locus.

Stutter Drop-out: Observed frequencies of reverse stutter dropout (○) for representative locus D8S1178 at 7 low-template target amounts and the exponential fit. Forward stutter dropout is characterized in an equivalent manner.
Electropherogram Artifacts:
The following electropherogram artifacts are not modeled in NOC/t and must be removed by the laboratory or analyst according to laboratory protocol.

1. Pull-up
2. –A
3. Spikes
4. Dye artifacts
Conclusions

Mixtures are hard to interpret.

**MAC:** 4-person contributors present as 3. When > 4 alleles are observed, we cannot assume the $NOC_{min} \approx NOC_{actual}$. Thus, MAC is usually accurate when $\leq 4$ alleles are observed.

**NOClt:** Provides the probability distribution on the number of contributors. Though it had a higher accuracy rate than other threshold-based methods, the accuracy $< 80\%$ when $NOC_{actual} \geq 4$.

Thus,
1) When calculating LRs may need to consider if 1) more than one LR should be reported OR 2) report the smallest one OR 3) use different *reasonable* assumptions of $n$ in numerator and denominator to minimize the LR OR 4) state that confident comparison to know is not possible: Discussion needed.

2) When calculating CPE/CPI need to assess whether *all* alleles from *every* contributor are observed. Stochastic thresholds don’t work well when $NOC_{actual} = NOC_{min} > 2$. 
Thank-you

- Boston University School of Medicine, Biomedical Forensic Sciences Program
- NIJ 2011-DN-BX-K558, awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings and conclusions or recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the Department of Justice
- San Diego Sheriff's Department
  - Kelly Hensley
- Ullrich Moenich and
- Viveck Cadambe

Download NOC/t, technical manual and tutorial. Sign into [www.bu.edu/dnamixtures](http://www.bu.edu/dnamixtures)
Questions? Contact Catherine M. Grgicak
E-mail: cgrgcicak@bu.edu