Likelihood Ratios for Single Contributor Profiles

Simone Gittelson, Ph.D., simone.gittelson@nist.gov
Michael Coble, Ph.D., michael.coble@nist.gov

Acknowledgement
I thank Michael Coble, Bruce Weir and John Buckleton for their helpful discussions.

Disclaimer
Points of view in this presentation are mine and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.
Presenting analytical results alone is not enough to provide useful information for the legal system.

Does the blood stain recovered on the crime scene come from the person of interest?
LR for Source Level Propositions

Source level propositions:

\( H_p \): The crime stain came from the person of interest.
\( H_a \): The crime stain came from some other person.
LR for Source Level Propositions

\[ LR = \frac{Pr(E|H_p, I)}{Pr(E|H_d, I)} \]

\[ = \frac{Pr(G_{CS}, G_{POI}|H_p, I)}{Pr(G_{CS}, G_{POI}|H_d, I)} \]

\[ = \frac{Pr(G_{CS}|G_{POI}, H_p, I)}{Pr(G_{CS}|G_{POI}, H_d, I)} \times \frac{Pr(G_{POI}|H_p, I)}{Pr(G_{POI}|H_d, I)} \]

\[ = \frac{Pr(G_{CS}|G_{POI}, H_p, I)}{Pr(G_{CS}|G_{POI}, H_d, I)} \times \frac{Pr(G_{POI}|I)}{Pr(G_{POI}|I)} \]

The suspect’s genotype does not depend on \( H_p \) being true or \( H_d \) being true.

Numerator

the probability of observing the analytical results of the crime stain if the crime stain comes from the person of interest and given the analytical results of the person of interest’s sample and other available information.

\[ \Pr(G_{CS}|G_{POI}, H_p, I) \approx 1 \]
LR for Source Level Propositions

\[ LR = \frac{Pr(G_{CS}|G_{POI}, H_p, I)}{Pr(G_{CS}|G_{POI}, H_d, I)} \]

**Numerator**

the probability of observing the analytical results of the crime stain if the crime stain comes from the person of interest and given the analytical results of the person of interest’s sample and other available information

\[ Pr(G_{CS}|G_{POI}, H_p, I) = ? \]

A) 1  
B) <1  
C) >1

---

LR for Source Level Propositions

\[ LR = \frac{Pr(G_{CS}|G_{POI}, H_p, I)}{Pr(G_{CS}|G_{POI}, H_d, I)} \]

**Numerator**

the probability of observing the analytical results of the crime stain if the crime stain comes from the person of interest and given the analytical results of the person of interest’s sample and other available information

\[ Pr(G_{CS}|G_{POI}, H_p, I) \times 1 \]

\[ Pr(G_{CS}|G_{POI}, H_p, I) < 1 \]
LR for Source Level Propositions

\[ LR = \frac{\Pr(G_{CS}|G_{POI}, H_p, I)}{\Pr(G_{CS}|G_{POI}, H_d, I)} \]

Denominator
the probability of observing the analytical results of the crime stain if the crime stain comes from some other person and given the analytical results of the person of interest’s sample and the available information

What is the probability of observing a second person with this genotype given that we have already observed one person with this genotype?

ASSUMPTION:
The probability of observing \( G_{CS} \) is independent of the genotype observed for \( G_{POI} \).

\[ \Pr(G_{CS}|G_{POI}, H_d, I) = \Pr(G_{CS}|H_d, I) \]
**NRC II**

**Formula 4.1a**
Homozygote genotypes: 
\[ p_A^2 \]

**Formula 4.1b**
Heterozygote genotypes: 
\[ 2p_Ap_B \]

---

**LR for Source Level Propositions**

**Denominator**

**ASSUMPTION:**
The probability of observing \( G_{CS} \) is not independent of the genotype observed for \( G_{POI} \). There is a probability that the crime stain’s donor and the person of interest share an allele passed down from a common ancestor.

\[
\Pr(G_{CS}|G_{POI}, H_d, I) \neq \Pr(G_{CS}|H_d, I)
\]
Subpopulations

allele 28 is
identical by state

Subpopulations

allele 28 is
identical by state and
identical by descent

The coancestry coefficient $P_{ST}$, also called $\theta$, is the probability that two individuals have an allele identical by descent (IBD).
What is the probability of observing this profile in this population?

If $\theta > 0$:
- Profile probability < match probability
- Relatives, coancestors

If $\theta = 0$:
- Profile probability = match probability
- No relatives, no coancestors

Subpopulations

General Population

\[ p_{28} = 0.5 \]

Subpopulations

50% Subpopulation 1
mates only with members of subpopulation 1

50% Subpopulation 2
mates only with members of subpopulation 2

\[ p_{28} = 0.4 \]

\[ p_{28} = 0.6 \]
Subpopulations

Taking into account subpopulations:

\[
\begin{align*}
\text{50%} & \quad \text{50%} \\
\text{Subpopulation 1} & \quad \text{Subpopulation 2} \\
Pr(28,28) & = 0.4^2 \\
& = 0.16 \\
Pr(28,28) & = 0.6^2 \\
& = 0.36 \\
Pr(28,28) & = \frac{1}{2} \times 0.16 + \frac{1}{2} \times 0.36 = 0.26
\end{align*}
\]

Subpopulations

Not taking into account subpopulations:

General Population

needs correction!

\[
p_{28} = 0.5 \\
Pr(28,28) = 0.5^2 = 0.25 < 0.26
\]
Subpopulations

General Population
We can use the coancestry coefficient $F_{ST}$, also called $\theta$, to take into account the effect of subpopulations when we use the proportion $p_{28} = 0.5$ of the general population.

Balding & Nichols Equations

Subpopulations

The *coancestry coefficient* $F_{ST}$, also called $\theta$, is the probability that two individuals have an allele *identical by descent* (IBD).

What is the probability of seeing allele 28 in this population given that we have already observed one copy of allele 28?

Subpopulations

We have seen: allele 28

The probability of observing an allele 28 is:

\[
\theta \quad + \quad (1 - \theta)p_{28}
\]

- allele 28 is IBD with 28
- allele 28 is not IBD with any of the alleles already seen, it is observed by chance
Subpopulations

Rule of Thumb

If the allele in question has not been seen previously, then it is seen by chance.

If the allele in question has already been seen, then it could be observed again by chance or because it is IBD with an allele that has already been seen.

Subpopulations

What is the probability of seeing allele 28 in this population given that we have already observed allele 28 and allele 28?
Subpopulations

We have seen: \textit{allele 28} and \textit{allele 28}

The probability of observing an \textit{allele 28} is:

\[
\frac{\theta}{1+\theta} + \frac{\theta}{1+\theta} + (1-\theta)p_{28}
\]

\textit{allele 28} is IBD with \textit{28} \hspace{1cm} \textit{allele 28} is IBD with \textit{28} \hspace{1cm} \textit{allele 28} is not IBD with any of the alleles already seen, it is observed by chance
Subpopulations

We have seen: allele 28 and allele 28

The probability of observing an allele 28 is:

\[
\frac{2\theta + (1 - \theta)p_{28}}{1 + \theta}
\]

Subpopulations

What is the probability of seeing allele 28 in this population given that we have already observed allele 28, allele 28 and allele 28?
Subpopulations

We have seen: *allele 28*, *allele 28* and *allele 28*

The probability of observing an *allele 28* is:

\[ \theta + \theta + \theta + (1 - \theta)p_{28} \]

*allele 28* is IBD with *28*

*allele 28* is IBD with *28*

*allele 28* is IBD with *28*

*allele 28* is not IBD with any of the alleles already seen, it is observed by chance

Subpopulations

We have seen: *allele 28*, *allele 28* and *allele 28*

The probability of observing an *allele 28* is:

\[ 3\theta + (1 - \theta)p_{28} \]

\[ 1 + 2\theta \]
Subpopulations

We have seen: allele 28, allele 28 and allele 28

The probability of observing an allele 28 is:

\[
\frac{3\theta + (1 - \theta)p_{28}}{1 + 2\theta}
\]

Subpopulations

What is the probability of seeing genotype \{28, 28\} in this population given that we have already observed a genotype \{28, 28\}?

\[
\frac{2\theta + (1 - \theta)p_{28}}{1 + \theta} \times \frac{3\theta + (1 - \theta)p_{28}}{1 + 2\theta}
\]

Subpopulations

What is the probability of seeing genotype \{28,28\} in this population given that we have already observed a genotype \{28,28\}? 

\[
\frac{2\theta + (1 - \theta)p_{28}}{1 + \theta} \times \frac{3\theta + (1 - \theta)p_{28}}{1 + 2\theta} 
\]

if \(\theta = 0.03\): 

\[
\frac{2(0.03) + (1 - 0.03)(0.159)}{1 + 0.03} \times \frac{3(0.03) + (1 - 0.03)(0.159)}{1 + 2(0.03)} 
\]

\[= 0.048\]

What is the genotype probability 

\[
\frac{2\theta + (1 - \theta)p_{28}}{1 + \theta} \times \frac{3\theta + (1 - \theta)p_{28}}{1 + 2\theta} 
\]

equal to if \(\theta = 0\)?

A. 0
B. \(\theta\)
C. \(p_{28}^2\)
D. 2\(p_{28}\)
E. ? ???
Subpopulations

What is the probability of seeing allele 13 in this population given that we have already observed allele 13 and allele 16?

We have seen: allele 13 and allele 16

The probability of observing an allele 13 is:

\[ 1 \times \theta + 0 \times \theta + (1 - \theta)p_{13} \]

- allele 13 is IBD with 13
- allele 13 is IBD with 16
- allele 13 is not IBD with any of the alleles already seen, it is seen by chance

\[ \frac{1}{1 + \theta} \]
Subpopulations

We have seen: \textit{allele 13} and \textit{allele 16}

The probability of observing an \textit{allele 13} is:

$$\frac{\theta + (1 - \theta)p_{13}}{1 + \theta}$$

What is the probability of seeing \textit{allele 16} in this population given that we have already observed \textit{allele 13, allele 16} and \textit{allele 13}?
Subpopulations

We have seen: **allele 13**, **allele 16** and **allele 13**

The probability of observing an **allele 16** is:

\[
\frac{0 \times \theta}{\text{allele 16 is IBD with 13}} + \frac{1 \times \theta}{\text{allele 16 is IBD with 16}} + \frac{0 \times \theta}{\text{allele 16 is IBD with 13}} + \frac{(1 - \theta)p_{16}}{\text{allele 16 is not IBD with any of the alleles already seen}} = \frac{\theta + (1 - \theta)p_{16}}{1 + 2\theta}
\]
Subpopulations

What is the probability of seeing genotype \{13, 16\} in this population given that we have already observed a genotype \{13, 16\}?

\[
2 \times \frac{\theta + (1 - \theta)p_{13}}{1 + \theta} \times \frac{\theta + (1 - \theta)p_{16}}{1 + 2\theta}
\]


Subpopulations

What is the probability of seeing genotype \{13, 16\} in this population given that we have already observed a genotype \{13, 16\}?

\[
2 \times \frac{\theta + (1 - \theta)p_{13}}{1 + \theta} \times \frac{\theta + (1 - \theta)p_{16}}{1 + 2\theta}
\]

if \(\theta = 0.03\):

\[
2 \times \frac{0.03 + (1 - 0.03)(0.33)}{1 + 0.03} \times \frac{0.03 + (1 - 0.03)(0.033)}{1 + 2(0.03)}
\]

\[
= 0.040
\]
What is the genotype probability

\[ 2 \times \frac{\theta + (1 - \theta)p_{13}}{1 + \theta} \times \frac{\theta + (1 - \theta)p_{16}}{1 + 2\theta} \]

equal to if \( \theta = 0 \)?

A. 0  
B. 2\( \theta \)  
C. \( p_{13}^2 \)  
D. 2\( p_{13}p_{16} \)  
E. ???

NRC II

**Formula 4.10a**

\[
Pr(AA|AA) = \frac{[2\theta + (1 - \theta)p_A][3\theta + (1 - \theta)p_A]}{(1 + \theta)(1 + 2\theta)}
\]

**Formula 4.10b**

\[
Pr(AB|AB) = \frac{2[\theta + (1 - \theta)p_A][\theta + (1 - \theta)p_B]}{(1 + \theta)(1 + 2\theta)}
\]

LR for Source Level Propositions

\[ LR = \frac{Pr(G_{CS} \mid G_{POI}, H_p, I)}{Pr(G_{CS} \mid G_{POI}, H_d, I)} \]

Denominator

the probability of observing the analytical results of the crime stain
if the crime stain comes from some other person and given the
analytical results of the person of interest’s sample and the
available information

homzygote: \[ Pr(G_{CS} \mid G_{POI}, H_d, I) = \frac{[2\theta + (1-\theta)p_A][3\theta + (1-\theta)p_A]}{(1+\theta)(1+2\theta)} \]

heterozygote: \[ Pr(G_{CS} \mid G_{POI}, H_d, I) = \frac{2[\theta + (1-\theta)p_A][\theta + (1-\theta)p_B]}{(1+\theta)(1+2\theta)} \]

Exercise 2:
Likelihood Ratios for Single Contributor Profiles
Exercise 2

A burglary was committed where a witness saw a Caucasian person running from the scene. The investigators believe that this was the offender. The crime scene investigators recover a blood stain from a broken window pane from a smashed window through which they presume that the offender entered the building. A forensic laboratory types this blood stain ($G_{CS}$) and a sample taken from Mr. X, a Caucasian person of interest in this case ($G_{POI}$). For locus D21S11, the laboratory obtains the following typing results:

$$G_{CS} = \{27, 32\}$$

$$G_{POI} = \{27, 32\}$$

Exercise 2

1) What is the likelihood ratio (LR) for these results with regard to the following pair of propositions?

$H_p$: The blood stain recovered on the crime scene came from Mr. X.

$H_d$: The blood stain recovered on the crime scene came from somebody else, unrelated to Mr. X.

Assume US Caucasian allele probabilities of $p_{27} = 0.026$ and $p_{32} = 0.007$ for locus D21S11, a coancestry coefficient of $\theta = 0.01$, and that the numerator of the LR is equal to 1.
Exercise 2

2) If the factfinder’s prior odds for the above propositions are \( \frac{\Pr(H_p | I)}{\Pr(H_d | I)} = \frac{1}{99} \), what should the factfinder’s posterior odds be after hearing the DNA evidence?

Exercise 2

3) What should the factfinder’s posterior probability \( \Pr(H_p | G_C, G_P, I) \) be?
NRC II Report
Recommendations

Fixation indices ($F$-statistics)

<table>
<thead>
<tr>
<th>$F$-statistics</th>
<th>alternative notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{IS}$</td>
<td>$f$</td>
<td>Individual to Subpopulation: the correlation of alleles within an individual within a subpopulation</td>
</tr>
<tr>
<td>$F_{IT}$</td>
<td>$F$</td>
<td>Individual to Total population: the correlation of alleles within an individual (&quot;inbreeding&quot;)</td>
</tr>
<tr>
<td>$F_{ST}$</td>
<td>$\theta$</td>
<td>Subpopulation to Total population: the correlation of alleles of different individuals in the same subpopulation (&quot;coancestry&quot;)</td>
</tr>
</tbody>
</table>

### NRC II Report Recommendations

#### Assumptions

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Hardy-Weinberg Law:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the population</td>
</tr>
</tbody>
</table>

#### Recommendation 4.1
- Includes possibility that the individual's two alleles are IBD ('`inbreeding``'): 
  - Corrects for Hardy-Weinberg Disequilibrium in the population caused by population subdivision.
  - Assumes Linkage Equilibrium in the population.

#### Recommendation 4.2
- Includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population ('`coancestry``'): 
  - Corrects for Hardy-Weinberg Disequilibrium and Linkage Disequilibrium in the population caused by population subdivision.
  - Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the sub-populations.

---

### NRC II Report Recommendations

#### Assumptions

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Hardy-Weinberg Law:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Homozygotes</td>
</tr>
<tr>
<td></td>
<td>Heterozygotes</td>
</tr>
<tr>
<td>4.1</td>
<td>( p_{28}^2 )</td>
</tr>
<tr>
<td></td>
<td>( 2p_{13}p_{16} )</td>
</tr>
<tr>
<td>4.2</td>
<td>( Fp_{28} + (1-F)p_{28}^2 )</td>
</tr>
<tr>
<td></td>
<td>( 2p_{13}p_{16} )</td>
</tr>
<tr>
<td></td>
<td>( \frac{[2\theta + (1-\theta)p_{28}]<em>3[3\theta + (1-\theta)p</em>{28}]}{(1+\theta)(1+2\theta)} )</td>
</tr>
<tr>
<td></td>
<td>( \frac{2[\theta + (1-\theta)p_{13}][\theta + (1-\theta)p_{16}]}{(1+\theta)(1+2\theta)} )</td>
</tr>
</tbody>
</table>

---

## NRC II Report Recommendations

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Homozygotes</th>
<th>Heterozygotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardy-Weinberg Law:</td>
<td>0.025</td>
<td>0.022</td>
</tr>
<tr>
<td>includes possibility that the individual's two alleles are IBD (&quot;inbreeding&quot;):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F = 0.01 ):</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>( F = 0.03 ):</td>
<td>0.029</td>
<td>0.022</td>
</tr>
<tr>
<td>includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (&quot;coancestry&quot;):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \theta = 0.01 ):</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td>( \theta = 0.03 ):</td>
<td>0.048</td>
<td>0.040</td>
</tr>
</tbody>
</table>

## NRC II Report Recommendations

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>match probability for 15 loci</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardy-Weinberg Law:</td>
<td>( 8.9 \times 10^{-23} )</td>
</tr>
<tr>
<td>includes possibility that the individual's two alleles are IBD (&quot;inbreeding&quot;):</td>
<td></td>
</tr>
<tr>
<td>( F = 0.01 ):</td>
<td>( 1.0 \times 10^{-22} )</td>
</tr>
<tr>
<td>( F = 0.03 ):</td>
<td>( 1.4 \times 10^{-22} )</td>
</tr>
<tr>
<td>includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (&quot;coancestry&quot;):</td>
<td></td>
</tr>
<tr>
<td>( \theta = 0.01 ):</td>
<td>( 3.6 \times 10^{-21} )</td>
</tr>
<tr>
<td>( \theta = 0.03 ):</td>
<td>( 2.4 \times 10^{-19} )</td>
</tr>
</tbody>
</table>
## NRC II Report Recommendations

<table>
<thead>
<tr>
<th>Hardy-Weinberg Law:</th>
<th>Consequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Includes possibility that the individual's two alleles are IBD (&quot;inbreeding&quot;):</td>
<td>The profile seems <strong>more rare</strong> than it actually is.</td>
</tr>
</tbody>
</table>

**Recommendation 4.1**

**Recommendation 4.2**

Includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population ("coancestry"): The profile seems **more common** than it actually is.

---