



**DNA Interpretation Workshop 1**

**Statistical Approaches**

Michael D. Coble, PhD  
U.S. National Institute of Standards and Technology (NIST)

Bruce J. Heidebrecht, MS  
Maryland State Police Forensic Science Division  
<http://www.cstl.nist.gov/strbase/training.htm>

ISFG Pre-Conference Workshop  
Melbourne, Australia  
September 2-3, 2013

**NIST and NIJ Disclaimer**

**Funding:** Interagency Agreement between the **National Institute of Justice** and NIST Office of Law Enforcement Standards

**Points of view are mine** and do not necessarily represent the official position or policies of the US Department of Justice or the National Institute of Standards and Technology.

Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

**Our publications and presentations are made available at:**  
<http://www.cstl.nist.gov/strbase/NISTpub.htm>


**Two Parts to Mixture Interpretation**

- Determination of alleles present in the evidence and **deconvolution of mixture components** where possible
  - Many times through comparison to victim and suspect profiles
- Providing some kind of statistical answer** regarding the weight of the evidence
  - There are multiple approaches and philosophies


Software tools can help with one or both of these...

**The Three Possible Outcomes**  
of Evidence Examination (**Q-K Comparison**)

**"Suspect"**  
Known (K) Sample





**"Evidence"**  
Question (Q) Sample




Exclusion (no match)

Must provide significance of this match






Non-Exclusion (Match or Inclusion)



No result  
(or a complex mixture)



Inconclusive Result

(no decision as there is insufficient data to support a conclusion)

**Statistical Analysis of DNA Typing Results**

SWGDM Guidelines 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.**"

Buckleton, J. and Curran, J. (2008) A discussion of the merits of random man not excluded and likelihood ratios. *Forensic Sci. Int. Genet.* 2: 343-348.

**Basic Math Terms**

- When '+' is used, this means 'OR'
- When 'x' is used, this means 'AND'
- Pr. is shorthand for probability
- Therefore...
  - the probability of a 'AND' b happening together is  $Pr(a \text{ and } b) = a \times b$
  - the probability of a 'OR' b happening together is  $Pr(a \text{ or } b) = a + b$

Slide information from Peter Gill

## Conditioning

- **Probabilities are conditional**, which means that the probability of something is based on a hypothesis
- In math terms, conditioning is denoted by a vertical bar
  - Hence,  $\Pr(a|b)$  means "the probability of  $a$  given that  $b$  is true"
- The probability of an event  $a$  is dependent upon various assumptions—and these assumptions or hypotheses can change...

Slide information from Peter Gill

## Laws of Probability (3 easy rules)

- Probabilities can range from 0 to 1.
- Events can be mutually exclusive (add)

$$\Pr(G \text{ or } H|E) = P(G|E) + P(H|E)$$

## Probabilities

- What is the probability of rolling a "5" using a six-sided die?
- $P(\text{rolling a } 5) = 1/6$
- What is the probability of rolling a "5" **or** "6"?
- $P(\text{rolling a } 5) + P(\text{rolling a } 6) = 1/6 + 1/6 = 2/6$  or  $1/3$ .

## Laws of Probability (3 easy rules)

- Probabilities can range from 0 to 1.
- Events can be mutually exclusive (add)

$$\Pr(G \text{ or } H|E) = P(G|E) + P(H|E)$$

- Events can be independent (multiply)

$$\Pr(G \text{ and } H|E) = P(G|E) \times P(H|E)$$

## Probabilities

- What is the probability of rolling a "5" on the first throw **and** rolling a "6" on the second roll?
- $P(\text{rolling a } 5) * P(\text{rolling a } 6) = 1/6 * 1/6 = 1/36$ .

1,1	2,1	3,1	4,1	5,1	6,1
1,2	2,2	3,2	4,2	5,2	6,2
1,3	2,3	3,3	4,3	5,3	6,3
1,4	2,4	3,4	4,4	5,4	6,4
1,5	2,5	3,5	4,5	5,5	6,5
1,6	2,6	3,6	4,6	<b>5,6</b>	6,6

## Statistical Approaches with Mixtures

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

- **Calculation of Exclusion Probabilities** - CPE/CPI (RMNE) – The probability that a random person (unrelated individual) would be excluded as a contributor to the observed DNA mixture.
- **Inferring Genotypes of Contributors** - RMP - Separate major and minor components into individual profiles and compute the random match probability estimate as if a component was from a single source.
- **Calculation of Likelihood Ratio Estimates** – LR – Comparing the probability of observing the mixture data under two (or more) alternative hypotheses.

**RMNE** = Random Man Not Excluded (same as CPE)  
**CPE** = Combined Probability of Exclusion ( $CPE = 1 - CPI$ )  
**CPI** = Combined Probability of Inclusion ( $CPI = 1 - CPE$ )

### Statistical Approaches with Mixtures

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

<p><b>“Exclusionary” Approach</b></p> <p><b>Random Man Not Excluded (RMNE)</b></p> <p style="font-size: small;">Combined Prob. of Inclusion (CPI)</p> <p style="font-size: small;">Combined Prob. of Exclusion (CPE)</p>	<p><b>“Inferred Genotype” Approach</b></p> <p><b>Random Match Probability (RMP)</b></p> <p style="color: red; font-weight: bold; font-size: small;">(mRMP)</p> <p><b>Likelihood Ratio (LR)</b></p>
--	--

## Exclusionary Approach

### DAB Recommendations on Statistics

February 23, 2000  
*Forensic Sci. Comm.* 2(3); available on-line at  
<http://www.fbi.gov/hq/lab/fsc/backissu/july2000/dnastat.htm>

“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)
- • Devlin, B. (1993) Forensic inference from genetic markers. *Statistical Methods in Medical Research* 2: 241–262.
- Likelihood ratios (LR)
- Evett, I. W. and Weir, B. S. (1998) *Interpreting DNA Evidence*. Sinauer, Sunderland, Massachusetts.

Statistical Methods in Medical Research 1993; 2: 241–262

### Forensic inference from genetic markers

B Devlin Department of Epidemiology and Public Health, Yale University School of Medicine

Section 5.1 Exclusion probability

- Discussion about exclusion probabilities in *Paternity* cases.

Two types:

(1) Conditional Exclusion Probability - excluding a random man as a possible father, given the mother-child genotypes for a particular case.

(2) Average Exclusion Probability – excluding a random man as a possible father, given a randomly chosen mother-child pair.

Statistical Methods in Medical Research 1993; 2: 241–262

### Forensic inference from genetic markers

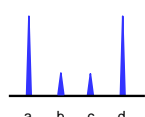
B Devlin Department of Epidemiology and Public Health, Yale University School of Medicine

**Section 5.1 Exclusion probability**

“The interpretation of conditional exclusion probability is obvious, which accounts for its value in the legal arena. Unlike [LR], however, it is not fully efficient.”

### Statistical Approaches with Mixtures

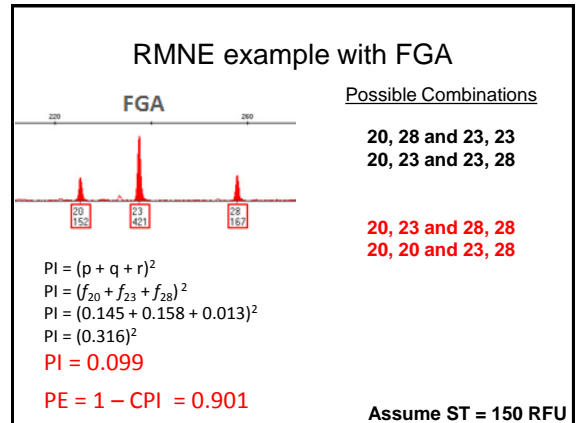
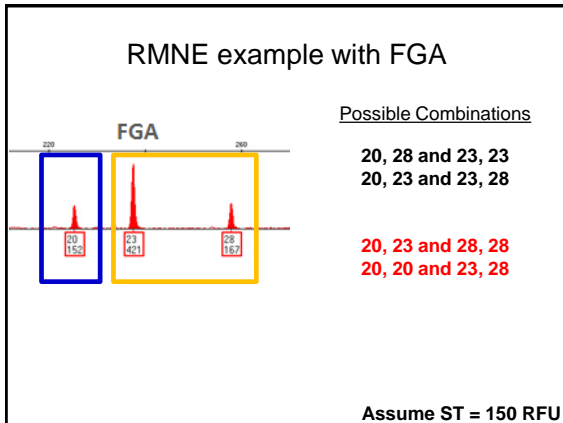
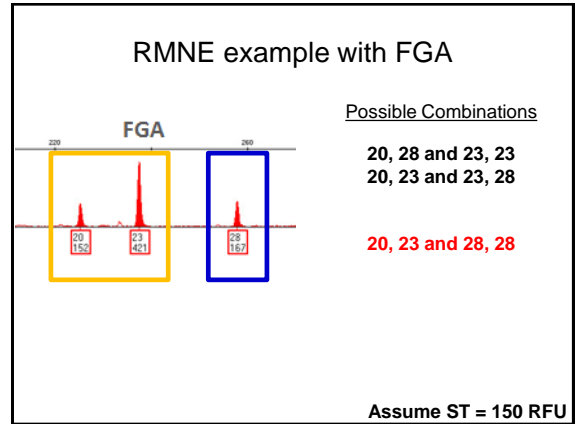
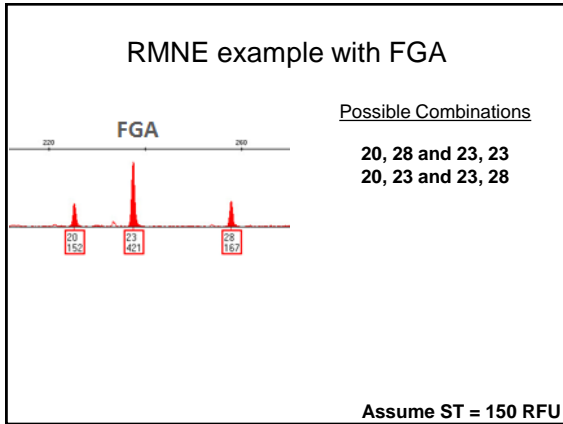
- **Random Man Not Excluded (CPE/CPI)** - The probability that a random person (unrelated individual) would be included/excluded as a contributor to the observed DNA mixture.



$$CPI = (f(a) + f(b) + f(c) + f(d))^2$$

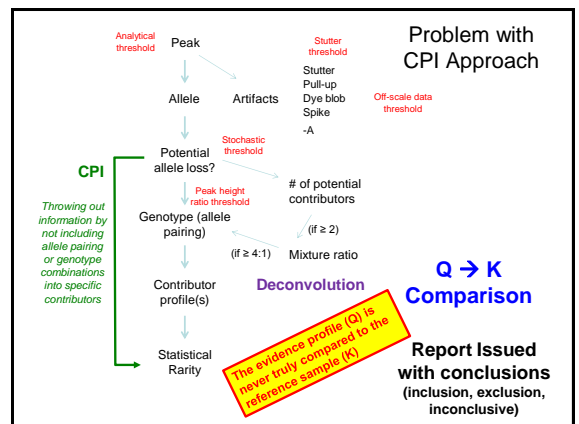
$$CPI = PI_{M1} \times PI_{M2} \dots$$

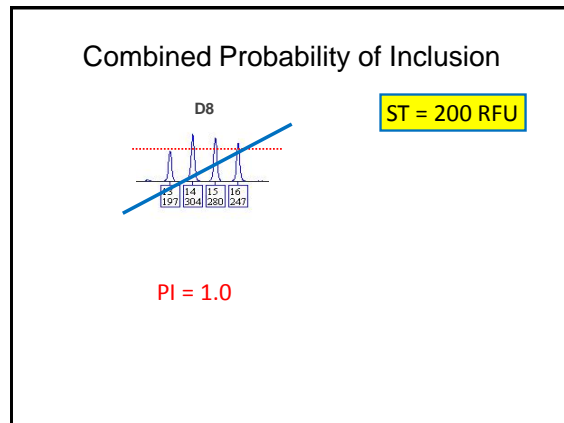
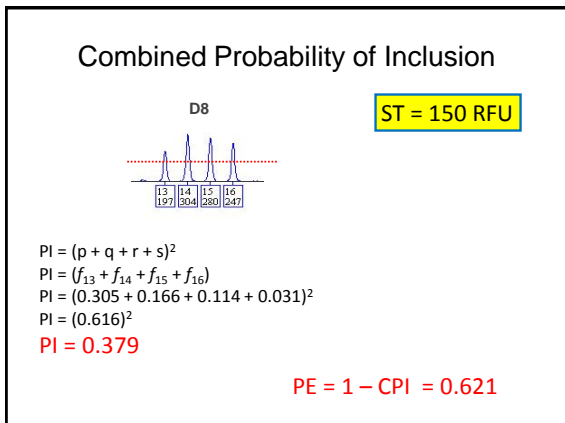
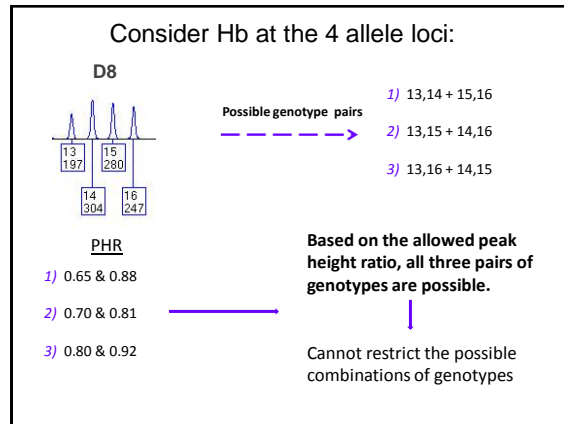
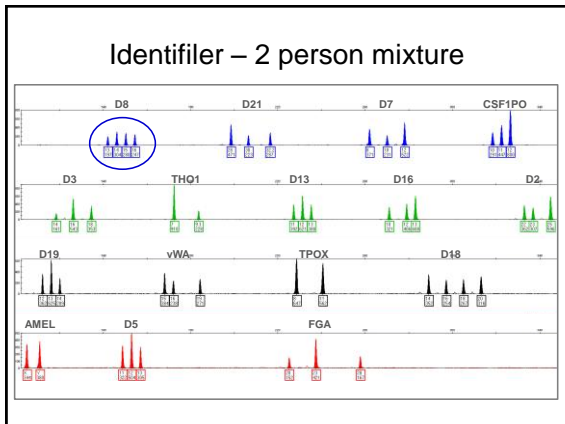
$$CPE = 1 - CPI$$



### CPE/CPI (RMNE) Limitations

- A CPE/CPI approach assumes that all alleles are present (i.e., cannot handle allele drop-out)
- Thus, statistical analysis of low-level DNA CANNOT be correctly performed with a CPE/CPI approach because some alleles may be missing





### “Advantages and Disadvantages” RMNE

**RMNE (CPE/CPI)**

**Advantages**

- Does not require an assumption of the number of contributors to a mixture
- Easier to explain in court
- Deconvolution is not necessary

**Disadvantages**

- Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect's genotype).
- Alleles below ST cannot be used for statistical purpose
- There is a potential to include a non-contributor

Summarized from John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223  
Buckleton and Curran (2008) *FSI-G* 343-348.

### Curran and Buckleton (2010)

JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, September 2010, Vol. 55, No. 5  
doi: 10.1111/j.1556-4029.2010.01446.x  
Available online at: interscience.wiley.com

**PAPER**  
**CRIMINALISTICS; GENERAL**

James M. Curran,<sup>1</sup> M.Sc.(Hons.), Ph.D. and John Buckleton,<sup>2</sup> Ph.D.

**Inclusion Probabilities and Dropout**

### Curran and Buckleton (2010)

(1) Created 1,000  
2 person mixtures

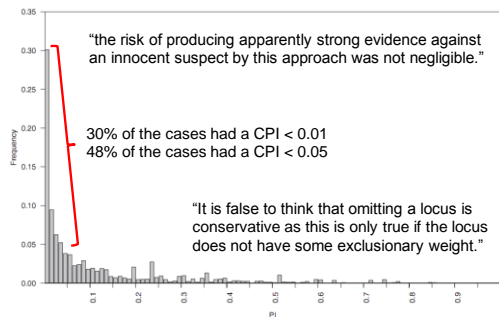


(2) Created 10,000  
Random genotypes



(3) Compared "random person" to mixture data, calculated PI for included loci, ignored discordant alleles.

### Curran and Buckleton (2010)



### CPE/CPI (RMNE) Limitations

- A CPE/CPI approach assumes that all alleles are present (i.e., cannot handle allele drop-out)
- Thus, statistical analysis of low-level DNA CANNOT be correctly performed with a CPE/CPI approach because some alleles may be missing
- Charles Brenner in his AAFS 2011 talk addressed this issue
- Research is on-going to develop allele drop-out models and software to enable appropriate calculations

### Notes from Charles Brenner's AAFS 2011 talk

The Mythical "Exclusion" Method for Analyzing DNA Mixtures – Does it Make Any Sense at All?

1. The claim that it requires **no assumption about number of contributors** is mostly wrong.
2. The supposed **ease of understanding** by judge or jury is really an illusion.
3. **Ease of use** is claimed to be an advantage particularly for complicated mixture profiles, those with many peaks of varying heights. The truth is the exact opposite. **The exclusion method is completely invalid for complicated mixtures.**
4. The exclusion method is only **conservative** for guilty suspects.

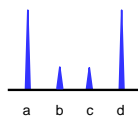
**Conclusion:** "Certainly no one has laid out an explicit and rigorous chain of reasoning from first principles to support the exclusion method. It is at best guesswork."

Brenner, C.H. (2011). The mythical "exclusion" method for analyzing DNA mixtures – does it make any sense at all? *Proceedings of the American Academy of Forensic Sciences*, Feb 2011, Volume 17, p. 79

### modified Random Match Probability

### Statistical Approaches with Mixtures


- **Random Match Probability (RMP)** – 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.



$$RMP_{\text{major}} = 2pq$$

$$= 2 \times f(a) \times f(d)$$

### 2013 JFS Article



J Forensic Sci, March 2013, Vol. 58, No. 2  
doi: 10.1111/1556-4029.12067  
Available online at: onlinelibrary.wiley.com

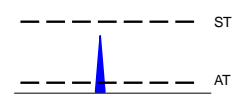
**TECHNICAL NOTE**  
**CRIMINALISTICS**

Todd Bille,<sup>1</sup> M.Sc.; Jo-Anne Bright,<sup>2</sup> M.Sc.; and John Buckleton,<sup>2</sup> Ph.D.

**Application of Random Match Probability Calculations to Mixed STR Profiles**

### 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?



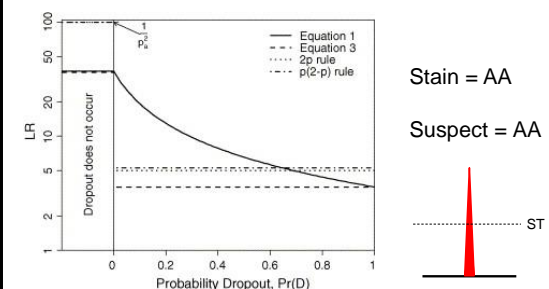
### The “2p” Rule

- “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?”

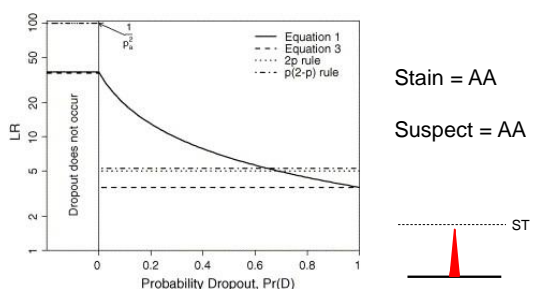
### The “2p” Rule



f(a) = 0.10   1/p<sup>2</sup> = 100   1/2p = 5

LR = 100

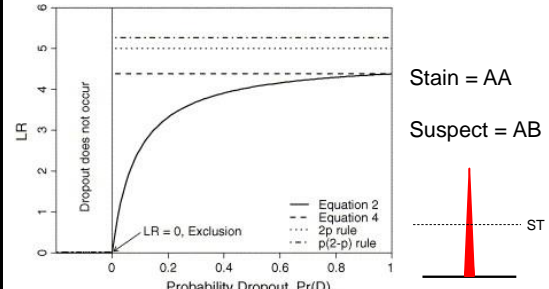
### The “2p” Rule



f(a) = 0.10   1/p<sup>2</sup> = 100   1/2p = 5

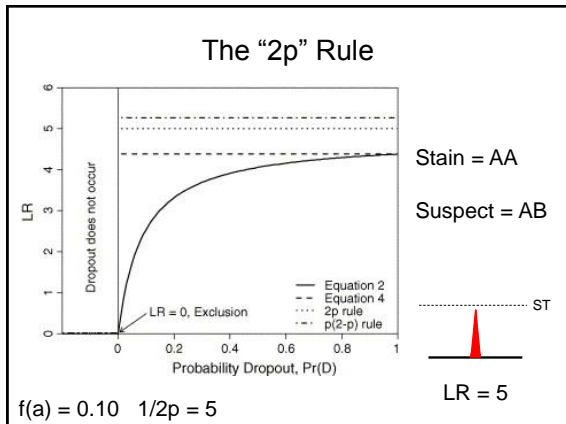
LR = 5

### The “2p” Rule



f(a) = 0.10   1/2p = 5

Exclusion



## Likelihood Ratio

- ### ISFG Recommendations on Mixture Interpretation

<http://www.isfg.org/Publication/Gill2006>

  1. The likelihood ratio (LR) is the preferred statistical method for mixtures over RMNE
  2. Scientists should be trained in and use LRs
  3. Methods to calculate LR's of mixtures are cited
  4. Follow Clayton et al. (1998) guidelines when deducing component genotypes
  5. Prosecution determines  $H_p$  and defense determines  $H_d$  and multiple propositions may be evaluated
  6. When minor alleles are the same size as stutters of major alleles, then they are indistinguishable
  7. Allele dropout to explain evidence can only be used with low signal data
  8. No statistical interpretation should be performed on alleles below threshold
  9. Stochastic effects limit usefulness of heterozygote balance and mixture proportion estimates with low level DNA

Gill et al. (2006) DNA Commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Sci. Int.* 160: 90-101

## Statistical Approaches with Mixtures

- **Likelihood Ratio** - Comparing the probability of observing the mixture data under two (or more) alternative hypotheses

## Probability Example – Will It Rain? (1)

### Defining the Event and Assumptions/Hypotheses

- Let's suppose that  $a$  is the probability of an event (e.g., **will it rain?**)
- What is the probability that it will rain in the afternoon –  $Pr(a)$ ?
- This probability is dependent upon assumptions
  - We can look at the window in the morning and observe if it is sunny (s) or cloudy (c)
  - $Pr(a)$  **if** it is sunny (s) is less than  $Pr(a)$  **if** it is cloudy (c)
- We can write this as  $Pr(a/s)$  and  $Pr(a/c)$ 
  - Since sunny or cloudy are the only possibilities,  $Pr(s) + Pr(c) = 1$
  - or  $Pr(s) = 1 - Pr(c)$

Slide information from Peter Gill

## Probability Example – Will It Rain? (2)

### Examining Available Data

- $Pr(a/s)$  and  $Pr(a/c)$  can be calculated from data
- **How often does it rain in the afternoon when its sunny in the morning?**
  - 20 out of 100 observations so  $Pr(a/s) = 0.2$
- **How often does it rain in the afternoon when it is cloudy in the morning?**
  - 80 out of 100 observations so  $Pr(a/c) = 0.8$

Slide information from Peter Gill



### Probability Example – Will It Rain? (3)

#### Formation of the Likelihood Ratio (LR)

- The LR compares two probabilities to find out which of the two probabilities is the most likely

The probability that it will rain in the afternoon when it is cloudy in the morning or  $\Pr(a|c)$  is divided by the probability that it will rain in the afternoon when it is sunny in the morning or  $\Pr(a|s)$

$$LR = \frac{\Pr(a|c)}{\Pr(a|s)} = \frac{0.8}{0.2} = 4$$

*Slide information from Peter Gill*

### Probability Example – Will It Rain? (4)

#### Explanation of the Likelihood Ratio

$$LR = \frac{\Pr(a|c)}{\Pr(a|s)} = \frac{0.8}{0.2} = 4$$

- The probability that it will rain is 4 times more likely **if** it is cloudy in the morning than **if** it is sunny in the morning.
- The word **if** is very important here. It must always be used when explaining a likelihood ratio otherwise the explanation could be misleading.

*Slide information from Peter Gill*

### Likelihood Ratios in Forensic DNA Work

- We evaluate the evidence ( $E$ ) relative to alternative pairs of hypotheses
- Usually these hypotheses are formulated as follows:
  - The probability of the evidence if the crime stain originated with the suspect or  $\Pr(E|S)$
  - The probability of the evidence if the crime stain originated from an unknown, unrelated individual or  $\Pr(E|U)$

$$LR = \frac{\Pr(E|S)}{\Pr(E|U)}$$

← The numerator  
← The denominator

*Slide information from Peter Gill*

### The Likelihood Ratio Must Be Stated Carefully

- The probability of the evidence is x times more likely **if** the stain came from the suspect Mr. Smith than **if** it came from an unknown, unrelated individual.
- It is not appropriate to say: "The probability that the stain came from Mr. Smith." because we must always include the conditioning statement – i.e., **always make the hypothesis clear in the statement.**
- Always use the word **'if'** when using a likelihood ratio to avoid this trap

*Slide information from Peter Gill*

### Likelihood Ratio (LR)

- Provides ability to express and evaluate both the prosecution hypothesis,  $H_p$  (the suspect is the perpetrator) and the defense hypothesis,  $H_d$  (an unknown individual with a matching profile is the perpetrator)

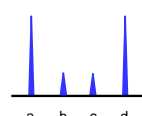
$$LR = \frac{H_p}{H_d}$$

- The numerator,  $H_p$ , is usually 1 – since in theory the prosecution would only prosecute the suspect if they are 100% certain he/she is the perpetrator
- The denominator,  $H_d$ , is typically the profile frequency in a particular population (based on individual allele frequencies and assuming HWE) – i.e., **the random match probability**

*Slide information from Peter Gill*

### Statistical Approaches with Mixtures

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

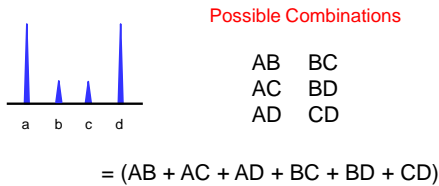


$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{1}{2pq} = 1/RMP$$

$E$  = Evidence  
 $H_1$  = Prosecutor's Hypothesis (the suspect did it) = 1  
 $H_2$  = Defense Hypothesis (the suspect is an unknown, random person)

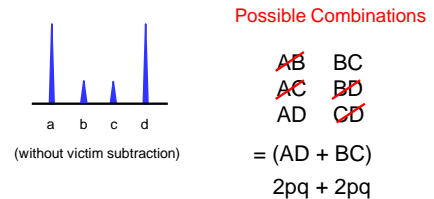
### Statistical Approaches with Mixtures

- **Unrestricted Likelihood Ratio** - All combinations of alleles are deemed possible (relative peak height differences are not utilized).



### Statistical Approaches with Mixtures

- **Restricted Likelihood Ratio** - Based on relative peak heights, alleles are paired only where specific combinations of alleles are deemed possible



### Restricted v. Unrestricted

- **Unrestricted LR:**
  - Less interpretation of the evidence (no need to determine which allele pairs are unreasonable)
  - Easier math (no need to exclude unreasonable allele pairs)
- **Restricted LR:**
  - More accurate interpretation of the evidence

### Advantages and Disadvantages RMNE and LR

#### RMNE (CPE/CPI)

##### Advantages

- Does not require an assumption of the number of contributors to a mixture
- Easier to explain in court

##### Disadvantages

- Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect's genotype)
- LR approaches are developed within a consistent logical framework

#### Likelihood Ratios (LR)

##### Advantages

- Enables full use of the data including different suspects

##### Disadvantages

- More difficult to calculate (software programs can assist)
- More difficult to present in court

Summarized from John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223  
Buckleton and Curran (2008) *FSI-G* 343-348.



Forensic Science International: Genetics 2 (2008) 343-348

A discussion of the merits of random man not excluded and likelihood ratios

John Buckleton<sup>a,\*</sup>, James Curran<sup>b</sup>

<sup>a</sup>ESR, PB 92/021, Auckland, New Zealand

<sup>b</sup>Department of Statistics, University of Auckland, PB 92019, Auckland, New Zealand  
Received 15 January 2008; received in revised form 29 April 2008; accepted 1 May 2008

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.

### Relationship between Likelihood Ratio (LR) and Random Match Probability (RMP)

- For single source samples or deduced major component profiles in a mixture...

$$LR = \frac{1}{RMP} \quad \text{or} \quad RMP = \frac{1}{LR}$$

### Basic Foundations of Likelihood Ratios

- LR calculations include assumptions:
- 1) Must be able to hypothesize the number of contributors.
    - Prosecutor and Defense are allowed to have different number of contributors in their hypotheses.
    - Prosecutor and/or Defense are even allowed to present multiple different LR calculations with varying number of hypothesized contributors.

### Example of varying number of contributors in Defense hypothesis

- The probability of the DNA profile from the bloodstain at the crime scene is approximately 'A' times more likely if it originated from Suspect and an unknown Caucasian individual than from two unknown individuals in the Caucasian population.
- The probability of the DNA profile from the bloodstain at the crime scene is approximately 'B' times more likely if it originated from Suspect and an unknown Caucasian individual than from three unknown individuals in the Caucasian population.

### Example of varying number of contributors in Defense hypothesis

- While it is possible that Prosecutor can vary the number of contributors in their theory, it is usually most beneficial for them to minimize the number of unknown contributors in their theory of the mixture to maximize their statistics.
- It is usually most beneficial for the Defense to also minimize the number of unknown contributors.
- The more unknown contributors theorized to be in the mixture, the less likely it is that all contributors would have only the limited alleles detected in the mixture.
- Increasing the number of unknown contributors increases the chances of one of the contributors having an allele other than what was detected, therefore the hypothesis is no longer supported by the evidence.

### Basic Foundations of Likelihood Ratios

- 2) May be able to include “known” contributors into the hypothesis.
  - This is common on vaginal swabs, anal swabs, oral swabs, penile swabs (etc.), and may even be extended to clothing and/or bedding depending on lab policy.
  - Prosecutor and/or Defense are even allowed to present multiple different LR calculations with varying hypothesized “known” contributors.

### Example of varying “known” contributors in Defense hypothesis

- Defense theory is that Victim was attacked, but not by Suspect:
- The probability of the DNA profile from the condom at the crime scene is approximately 'A' times more likely if it originated from Victim and Suspect than from Victim and an unknown individual in the Caucasian population.

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{V \& S}{V \& U}$$

### Example of varying “known” contributors in Defense hypothesis

- Defense theory is that Suspect left the condom, but he did not attack the Victim:
- The probability of the DNA profile from the condom at the crime scene is approximately 'B' times more likely if it originated from Victim and Suspect than from Suspect and an unknown individual in the Caucasian population.

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{V \& S}{S \& U}$$

### Example of varying “known” contributors in Defense hypothesis

- Defense theory is that neither the Suspect nor the Victim is related to the evidence:
- The probability of the DNA profile from the condom at the crime scene is approximately ‘C’ times more likely if it originated from Victim and Suspect than from two unknown individuals in the Caucasian population.

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{V \& S}{U_1 \& U_2}$$

### Basic Foundations of Likelihood Ratios

- 3) Following the recommendations of the ISFG (Gill *et al.* 2006), the Defense must be allowed to include peaks “indistinguishable from stutter” in their calculations.

### Basic Foundations of Likelihood Ratios

- 4) LR is not a probability but a ratio of probabilities
- The final calculation is NOT the probability that the Suspect left the evidence.
  - The final calculation is a comparison of the strength of two different hypotheses, given what evidence profile is present.

*B.S. Weir,<sup>1</sup> C.M. Triggs,<sup>2</sup> L. Starling,<sup>3</sup> L.I. Stowell,<sup>3</sup> K.A.J. Walsh,<sup>3</sup> and J. Buckleton,<sup>3</sup>*

### Interpreting DNA Mixtures\*

Classic paper that describes the use of LR's

**REFERENCE:** Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J. Interpreting DNA mixtures. *J Forensic Sci* 1997;42(2):213–222.

### Reading the formulas for LR calculations

- “Interpreting DNA Mixtures” Weir, et al., *JFS* 1997;42(2):213-222.
- $L = \Pr(E | C) / \Pr(E | \hat{C})$
- ‘ If a jury is being asked to make a choice between explanations C and  $\hat{C}$ , it can be told “The profile E is L times more likely to have arisen under explanation C than under explanation  $\hat{C}$ .” ‘

### Reading the formulas for LR calculations

- $\Pr(E | C)$
- Probability of the evidence given the hypothesis ‘C’
- If the evidence is fully supported by the hypothesis, then the probability is 100%.
- E.G.: if Victim and Suspect complete the mixture at a given locus, Prosecutor’s hypothesis is: that happened because Victim and Suspect are the only two donors to the mixture....
- Mixing the two people the Prosecutor considers to be “known” to the mixture completes the evidence profile.

Reading the formulas for LR calculations

- $\Pr(E | \hat{C})$
- Probability of the evidence given the hypothesis  $\hat{C}$
- The Defense may not agree that the arrested Suspect should be considered "known" to the mixture.
- The Defense hypothesis may be that while the Victim can be considered "known" to the mixture, there is still one unknown contributor – the True Attacker.

Reading the formulas for LR calculations

- $P_x(U | E)$
- Probability of having X unknown contributors
- Who have U alleles
  - unknown, or unaccounted for by "known" contributor
- And nothing greater than E alleles
  - evidence in its entirety

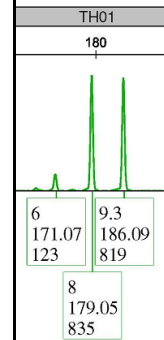
Reading the formulas for LR calculations

- Both the Prosecutor's hypothesis (numerator), and Defense's hypothesis (denominator) use the same formula

$$P_x(U | E)$$

What will provide different strength to each hypothesis is the variation in X, U, and even E

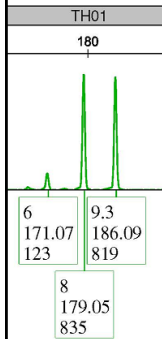
Example #1



- "Known" Victim = 8,9.3
- "Known" Suspect = 6,8  
(stochastic = 100rfu)
- $C = P_x(U | E)$
- $C = P_0(\Phi | 6,8,9.3)$
- Reads as "What is the probability of finding zero unknown contributors who have no alleles unaccounted for by the known contributors"

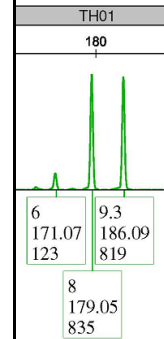
$\Phi = \text{empty set}$

Example #1



- "Known" Victim = 8,9.3
- "Known" Suspect = 6,8  
(stochastic = 100rfu)
- $C = P_x(U | E)$
- $C = P_0(\Phi | 6,8,9.3)$
- Evidence is 100% supported by the hypothesis that combining Victim and Suspect would produce this mixture.

Example #1



- "Known" Victim = 8,9.3
- 1 unknown contributor  
(stochastic = 100rfu)
- $\hat{C} = P_x(U | E)$
- $\hat{C} = P_1(6 | 6,8,9.3)$
- Reads as "What is the probability of finding one unknown contributor who has allele 6, and nothing other than 6, 8, or 9.3"

### Example #1

- $L = \Pr(E | C) / \Pr(E | \hat{C})$
  - $L = 100\% / (P_{6,6} + P_{6,8} + P_{6,9,3})$
  - $L = 1 / \{ 0.2266^2 + (0.2266)(1-0.2266)(0.01) + 2 \cdot 0.2266 \cdot 0.1256 + 2 \cdot 0.2266 \cdot 0.3054 \}$  NRC II formula 4.1
  - $L = 1 / \{ 0.0531 + 0.0569 + 0.1384 \}$
  - $L = 1 / 0.2484$
  - $L = 4$
- $f_6 = 0.2266$   
 $f_8 = 0.1256$   
 $f_{9,3} = 0.3054$

### Example #1

- $L = 4$
- It is four times more likely **IF** the evidence is a mixture of the victim (8,9.3) and the suspect (6,8) than **IF** the evidence is a mixture of the victim (8,9.3) and another random individual (because they are limited to being either a 6,6 or 6,8 or 6,9.3)

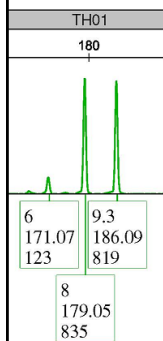
### Example #1

- $L = 4$
- It is four times more likely that pulling a random Caucasian from the general population to mix with the victim will produce a mixture that will not produce the evidence as detected.
  - Either the random person will not have the requisite allele 6, or will have alleles in addition to what was seen in the mixture.

### Example #1

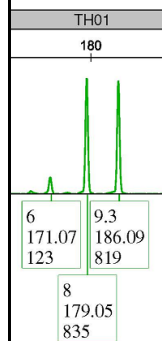
- $L = 4$
- This does NOT state that it is four times more likely that the suspect is the second contributor to the mixture. (“Prosecutor’s fallacy”)
  - Making that statement would require calculating prior probability: Chance that suspect had opportunity, etc....

### Example #2



- Arrested Suspect = 6,8
- 1 unknown contributor (stochastic = 100rfu)
- $C = P_x (U | E)$
- $C = P_1 (9.3 | 6,8,9.3)$
- Reads as “What is the probability of finding one unknown contributor who has allele 9.3, and nothing other than 6, 8, and 9.3”

### Example #2



- 2 unknown contributors (stochastic = 100rfu)
- $\hat{C} = P_x (U | E)$
- $\hat{C} = P_2 (6,8,9.3 | 6,8,9.3)$
- Reads as “What is the probability of finding two unknown contributors who have alleles 6, 8, and 9.3, and nothing other than 6, 8, and 9.3”

## Example #2

- $L = P_1(9.3 | 6,8,9.3) / P_2(6,8,9.3 | 6,8,9.3)$
- $L = (P_{9.3,9.3} + P_{6,9.3} + P_{8,9.3}) /$   
 $\{ (P_{6,6} * P_{8,9.3}) + (P_{8,8} * P_{6,9.3}) + (P_{9.3,9.3} * P_{6,8})$   
 $+ (P_{6,8} * P_{9.3,9.3}) + (P_{6,8} * P_{6,9.3}) + (P_{6,8} * P_{8,9.3})$   
 $+ (P_{6,9.3} * P_{8,8}) + (P_{6,9.3} * P_{6,8}) + (P_{6,9.3} * P_{8,9.3})$   
 $+ (P_{8,9.3} * P_{6,6}) + (P_{8,9.3} * P_{6,8}) + (P_{8,9.3} * P_{6,9.3}) \}$

## Example #2

- Rearrange denominator
- $L = (P_{9.3,9.3} + P_{6,9.3} + P_{8,9.3}) /$   
 $\{ (P_{6,6} * P_{8,9.3}) + (P_{8,9.3} * P_{6,6})$   
 $+ (P_{8,8} * P_{6,9.3}) + (P_{6,9.3} * P_{8,8})$   
 $+ (P_{9.3,9.3} * P_{6,8}) + (P_{6,8} * P_{9.3,9.3})$   
 $+ (P_{6,8} * P_{6,9.3}) + (P_{6,9.3} * P_{6,8})$   
 $+ (P_{6,8} * P_{8,9.3}) + (P_{8,9.3} * P_{6,8})$   
 $+ (P_{6,9.3} * P_{8,9.3}) + (P_{8,9.3} * P_{6,9.3}) \}$

## Example #2

- Simplify denominator
- $L = (P_{9.3,9.3} + P_{6,9.3} + P_{8,9.3}) /$   
 $2 * \{ (P_{6,6} * P_{8,9.3})$   
 $+ (P_{8,8} * P_{6,9.3})$   
 $+ (P_{9.3,9.3} * P_{6,8})$   
 $+ (P_{6,8} * P_{6,9.3})$   
 $+ (P_{6,8} * P_{8,9.3})$   
 $+ (P_{6,9.3} * P_{8,9.3}) \}$ 
 $f_6 = 0.2266$   
 $f_8 = 0.1256$   
 $f_{9.3} = 0.3054$

## Example #2

- $L = (0.0953 + 0.1384 + 0.0767) /$   
 $2 * \{ (0.0531 * 0.0767)$   
 $+ (0.0169 * 0.1384)$   
 $+ (0.0953 * 0.0569)$   
 $+ (0.0569 * 0.1384)$   
 $+ (0.0569 * 0.0767)$   
 $+ (0.1384 * 0.0767) \}$
- $L = 0.3104 / 2 * \{ 0.0041 + 0.0023 + 0.0054 + 0.0079 +$   
 $0.0044 + 0.0106 \}$
- $L = 0.3104 / 0.0694$
- $L = 4.5$

## Example #2

- $L = 4.5$
- It is four and a half times more likely IF the evidence is a mixture of the suspect (6,8) and an unknown individual (who is limited to being either a 6,9.3 or 8,9.3 or 9.3,9.3) than IF the evidence is a mixture of two random unknown individuals (who must be limited to having only alleles 6,8, or 9.3)

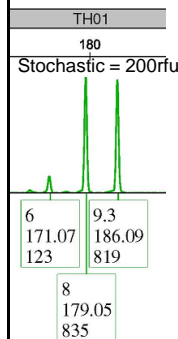
## Example #2

- $L = 4.5$
- It is four and a half times more likely that pulling two random Caucasians from the general population to mix together will produce a mixture that will not produce the evidence as detected.
  - Either the random people will not complete the mixture, or will have alleles in addition to what was seen in the mixture.

### Evidence with alleles below the stochastic threshold

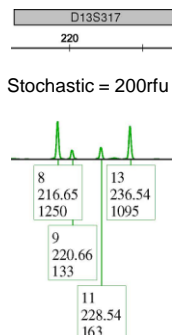
- DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. Forensic Science International 160 (2006) 90–101.
- “The advantage of the LR framework is that stutter and dropout can be assessed probabilistically..., and it is the only way to provide a meaningful calculation based on the probability of the evidence under  $H_{p(\text{prosecution})}$  and  $H_{d(\text{defense})}$ ”

### Use of a Stochastic Threshold



- Evidence is defined as being a two person mixture.
- With one allele detected below the ST, dropout may be possible.
- Use of a stochastic threshold accounts for dropout, and this locus can be included in the statistical interpretation.

### Use of a Stochastic Threshold



- Evidence is defined as being a two person mixture.
- With four detected alleles, dropout is unreasonable even though the two minor alleles are below the ST.
- Since dropout is unreasonable, locus is useable for statistics.

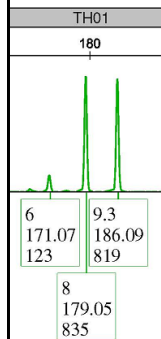
### Use of a Stochastic Threshold

- $\Pr(E | C)$
- Probability of the evidence given the hypothesis 'C'
- If the evidence is fully supported by the hypothesis, then the probability is 100%.
- e.g. if Victim and Suspect complete the mixture at a given locus, Prosecutor's hypothesis is: that happened because Victim and Suspect are the only two donors to the mixture....
- Mixing the two people the Prosecutor considers to be "known" to the mixture completes the evidence profile.

### Use of a Stochastic Threshold

- $\Pr(E | \hat{C})$
- Probability of the evidence given the hypothesis ' $\hat{C}$ '
- The Defense may not agree that the arrested Suspect should be considered "known" to the mixture.
- The Defense hypothesis may be that while the Victim can be considered "known" to the mixture, there is still one unknown contributor – the True Attacker.
- If the allele requisite to the True Attacker is below the stochastic threshold, one of their alleles may not be detected in the mixture.

### Example #3



- "Known" Victim = 8,9.3
- "Known" Suspect = 6,8 (stochastic = 200rfu)
- $C = P_x (U | E)$
- $C = P_0 (\Phi | 6,8,9.3)$
- Reads as "What is the probability of finding zero unknown contributors who have no alleles unaccounted for by the known contributors"



### Example #3

- “Known” Victim = 8,9.3
- “Known” Suspect = 6,8  
(stochastic = 200rfu)
- $C = P_x (U | E)$
- $C = P_0 (\Phi | 6,8,9.3)$
- Evidence is 100% supported by theory that combining Victim and Suspect would produce this mixture.

### Example #3

- “Known” Victim = 8,9.3
- 1 unknown contributor  
(stochastic = 200rfu)
- $\hat{C} = P_x (U | E)$
- $\hat{C} = P_1 (6 | 6,8,9.3,--)$
- Reads as “What is the probability of finding one unknown contributor who has allele 6, **and is allowed to have any other allele**”

### Example #3

- $L = \Pr(E | C) / \Pr(E | \hat{C})$
- $L = 1 / (P_{6,F})$
- $L = 1 / 2 * P_6$
- $L = 1 / 2 * 0.2266$
- $L = 1 / 0.4532$
- $L = 2.2$

“2p” Rule

$f_6 = 0.2266$

### Example #3

- $L = 2.2$
- It is two point two times more likely IF the evidence is a mixture of the victim (8,9.3) and the suspect (6,8) than IF the evidence is a mixture of the victim (8,9.3) and another random individual (because they are limited to having to have an allele 6)

### Peaks “indistinguishable from stutter”

- ISFG (2006) Recommendation
- “If the crime profile is a major/minor mixture, where minor alleles are the same size (height or area) as stutters of major alleles, then stutters and minor alleles are indistinguishable. Under these circumstances alleles in stutter positions that do not support  $H_p$  should be included in the assessment.”

### Peaks “indistinguishable from stutter”

- Practical application:
- The Defense is allowed to include peaks below the stutter threshold as true alleles to increase the pool of potential unknown contributors.
- The inclusion of these peaks must be based upon expected peak height ratios of heterozygous pairs, mixture ratio of contributors, number of unambiguous alleles compared to number of contributors, and the overriding possibility of dropout rendering this interpretation irrelevant.

### Peaks "indistinguishable from stutter"

- Practical application:
- The Prosecution is NOT allowed to include peaks below the stutter threshold as true alleles to increase the pool of potential unknown contributors.
- The Prosecution can only utilize such peaks to explain why alleles of "known" contributors are not reported by the software.

### Example #4

- Peak in bin 6 fell below stutter threshold.
- Cannot define peak in bin 6 as a true allele since below stutter threshold.
- However, Defense is allowed to utilize peak in bin 6 to increase their pool of potential unknown contributors.

### Example #4

- Evidence = Victim's oral swab
- Victim = 7,9.3
- Suspect = 7,8
- Prosecutor's theory is that the reported alleles of 7,8,9.3 are present because this is a mixture of Victim and Suspect
- $C = P_0 (\Phi | 7,8,9.3)$

### Example #4

- Evidence = Victim's oral swab
- Victim = 7,9.3
- One unknown contributor
- Defense's hypothesis #1 is that the alleles of 7,8,9.3 (6 is only stutter) are present because this is a mixture of Victim and True Attacker
- $\hat{C} = P_1(8 | 7,8,9.3)$

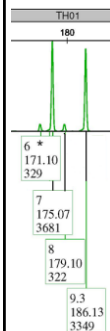
### Example #4

- Evidence = Victim's oral swab
- Victim = 7,9.3
- One unknown contributor
- Defense's theory #2 is that the alleles of 6,7,8,9.3 (6 is "unlucky" to be in stutter position, but it is a true allele) are present because this is a mixture of Victim and True Attacker
- $\hat{C} = P_1(6,8 | 6,7,8,9.3)$

### Example #4

- Evidence = Victim's oral swab
- Victim = 7,9.3
- One unknown contributor
- Defense is allowed to hold both theories as being true at the same time.
- "Peak in bin 6 is stutter, but if it's not stutter, it is present because of the True Attacker."

### Example #4



- Evidence = Victim's oral swab
- Victim = 7,9.3
- One unknown contributor
- Defense is allowed to hold both theories as being true at the same time.
- $\hat{C} = \{P_1(8 | 7,8,9.3) + P_1(6,8 | 6,7,8,9.3)\}$

### Example #4

- $L = \Pr(E | C) / \Pr(E | \hat{C})$
  - $L = 1 / \{P_1(8 | 7,8,9.3) + P_1(6,8 | 6,7,8,9.3)\}$
  - $L = 1 / \{ [(P_{8,8}) + (2*P_{7,8}) + (2*P_{8,9.3})] + (2*P_{6,8}) \}$
  - $L = 1 / \{ [0.0169 + 0.0433 + 0.0767] + 0.0569 \}$
  - $L = 1 / 0.1938$
  - $L = 5.2$
- $f_6 = 0.2266$   
 $f_7 = 0.1724$   
 $f_8 = 0.1256$   
 $f_{9.3} = 0.3054$

### Example #4

- $L = 5.2$
- It is five point two times more likely IF the evidence is a mixture of the victim (7,9.3) and the suspect (7,8) than IF the evidence is a mixture of the victim (7,9.3) and another random individual (because they are limited to having to have an allele 8, and nothing other than a 6, 7, or 9.3)

### How much difference does this really make?

- Without utilizing the peak in bin 6 as a potential allele in the Defense's hypothesis:
- Prosecutor's theory remains the same: 100%
- Defense's theory now is limited to  $P_1(8 | 7,8,9.3)$
- $L = 1 / [(P_{8,8}) + (2*P_{7,8}) + (2*P_{8,9.3})]$
- $L = 1 / [0.0169 + 0.0433 + 0.0767]$
- $L = 1 / 0.1369$
- $L = 7.3$

### How much difference does this really make?

- Without utilizing the peak in bin 6 as a potential allele in the Defense's theory:
- $L = 7.3$
- Including the peak in bin 6 as a potential allele in the Defense's theory:
- $L = 5.2$

### In Summary

- Hypothesis of number of contributors must be established before looking at any probative samples. This reduces the chance of interpretation bias.
- If the number of contributors can not be reasonably hypothesized, then LR statistics can not be applied.

### In Summary

- Using LR discounts unreasonable genotypes that are allowed in CPE.
- Using Restricted LR discounts unreasonable genotypes that are allowed in Unrestricted LR.
- Using “known” contributors discounts unreasonable genotypes that are allowed without any assumption of “known” contributors.

### In Summary

- The Defense must be allowed to use the stochastic threshold to increase their pool of potential unknown contributors.
- The Defense must be allowed to use the peaks that are “indistinguishable from stutter” to increase their pool of potential unknown contributors.
- The Prosecution is not allowed to use either to increase their pool of potential unknown contributors, only to explain why their “known” contributors are not fully represented in the reportable profile.

### Acknowledgements

- Bruce Heidebrecht (Maryland State Police)
- John Butler (NIST)
- Catherine Grgicak and Robin Cotton (Boston U.)
- Charlotte Word (Charlotte Word Consulting)