NEAFS
Probabilistic DNA Mixture Interpretation Workshop
Allentown, PA
September 25-26, 2015

Likelihood Ratios for Single Contributor Profiles

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

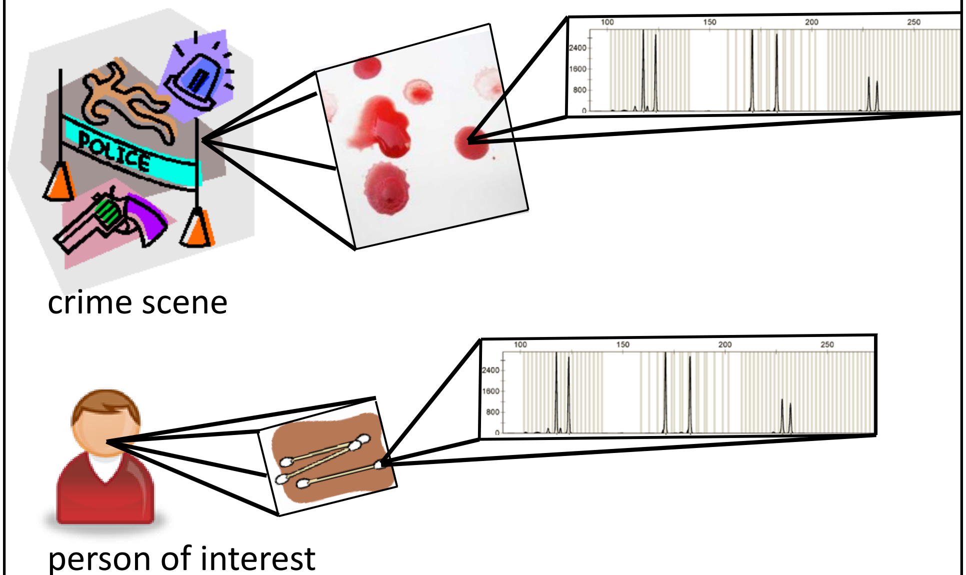
Acknowledgement

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

Disclaimer

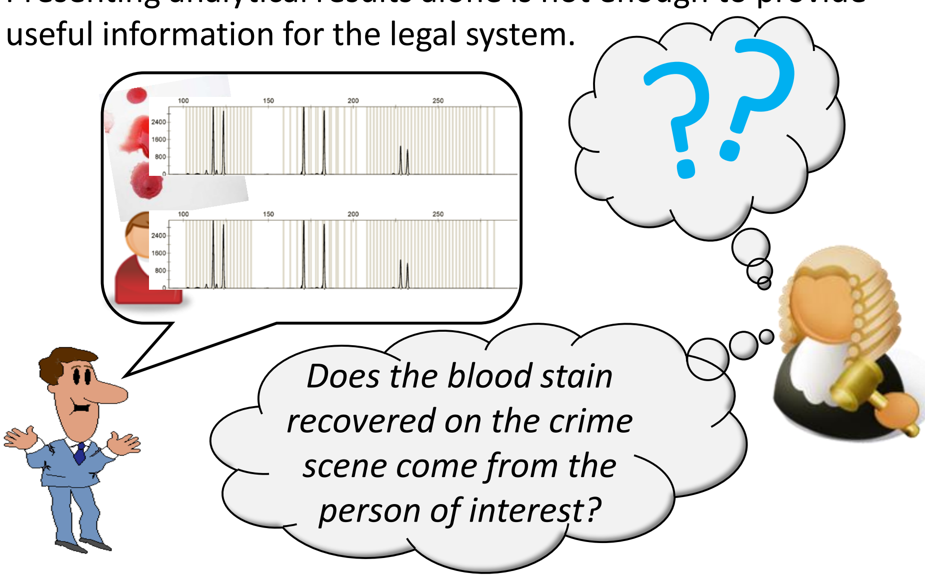
Points of view in this presentation are mine and do not necessarily represent the official position or policies of the National Institute of Standards and Technology.

Single Contributor Stain



Single Contributor Stain

Presenting analytical results alone is not enough to provide useful information for the legal system.



LR for Source Level Propositions

Source level propositions:

H_p: The crime stain **came from** the person of interest.

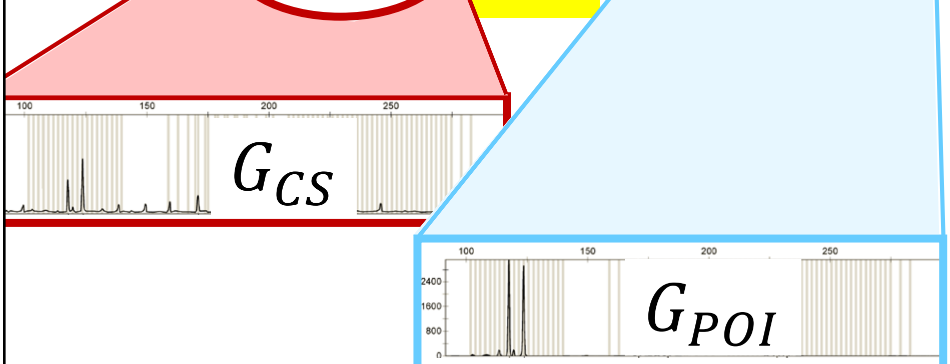
H_d: The crime stain **came from** some other person.

LR for Source Level Propositions

Source level propositions:

H_p: The crime stain **came from** the person of interest.

H crime stain **me from** s



LR for Source Level Propositions

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

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

The suspect's genotype does not depend on H_p being true or H_d being true.

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

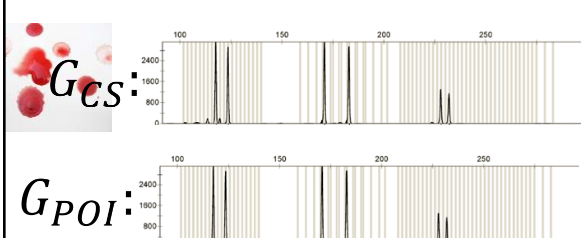
$$= \frac{\Pr(G_{CS}|G_{POI}, H_p, I)}{\Pr(G_{CS}|G_{POI}, H_d, I)} \times \underbrace{\frac{\Pr(G_{POI}|I)}{\Pr(G_{POI}|I)}}_{1}$$

LR for Source Level Propositions

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

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



$$\Pr\left(G_{CS}|G_{POI}, H_p, I\right) \approx 1$$

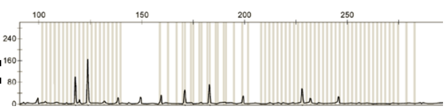
LR for Source Level Propositions

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

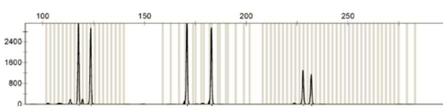
Numerator

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

G_{CS}:



G_{POI}:



$\Pr\left(G_{CS}|G_{POI}, H_p, I\right) = ?$

A) 1
B) <1
C) >1

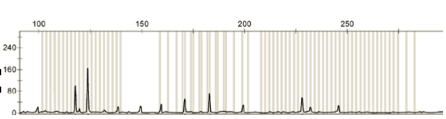
LR for Source Level Propositions

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

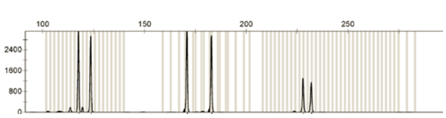
Numerator

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

G_{CS}:



G_{POI}:



$\Pr\left(G_{CS}|G_{POI}, H_p, I\right) \cancel{\approx 1}$

$\Pr\left(G_{CS}|G_{POI}, H_p, I\right) < 1$

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?

LR for Source Level Propositions

Denominator

ASSUMPTION:
The probability of observing G_{CS} is independent of the genotype obseved for G_{POI}.

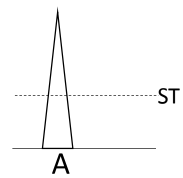
$$\Pr(G_{CS}|G_{POI}, H_d, I) = \Pr(G_{CS}|H_d, I)$$

NRC II

Formula 4.1a

Homozygote genotypes:

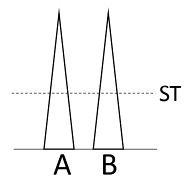
$$p_A^2$$



Formula 4.1b

Heterozygote genotypes:

$$2p_Ap_B$$



National Research Council (NRCII) Committee on DNA Forensic Science. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington DC, 1996.

LR for Source Level Propositions

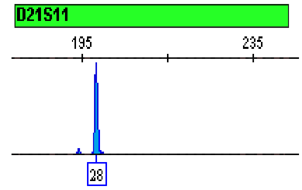
Denominator

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 interesst share an allele passed down from a common ancestor.

G_{CS} G_{POI}

$$\Pr(G_{CS}|G_{POI}, H_d, I) \neq \Pr(G_{CS}|H_d, I)$$

Subpopulations



allele 28 is

identical by state

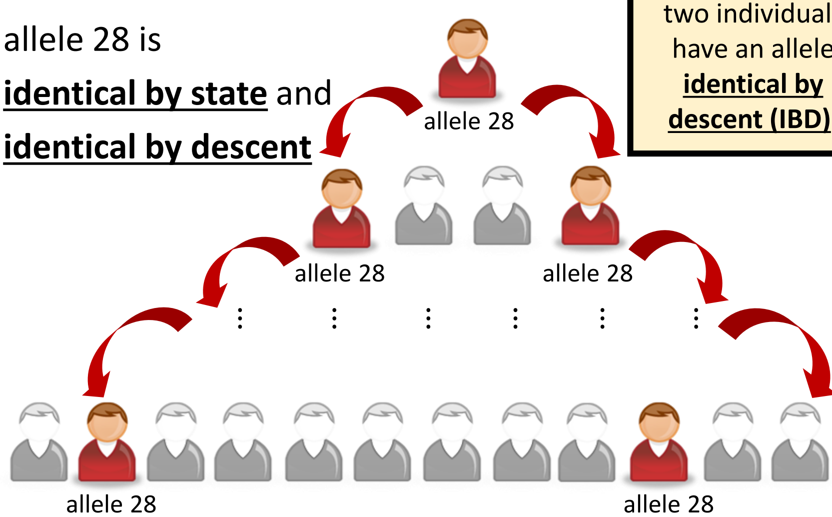
Subpopulations

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

allele 28 is

identical by state and

identical by descent

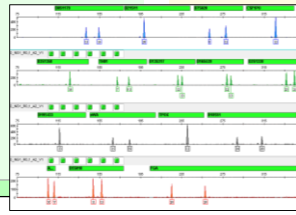


profile probability

probability of observing this profile in a population

match probability

probability of observing this profile in a population **knowing that this profile has already been observed in one individual in this population**



What is the probability of observing this profile in this population?

this individual has this profile

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 11: pages 301-302.

profile probability

probability of observing this profile in a population

match probability

probability of observing this profile in a population knowing that this profile has already been observed in one individual in this population

If $\theta = 0$:
profile probability = match probability

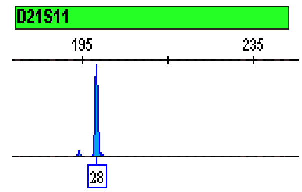
no relatives,
no coancestors

If $\theta > 0$:
profile probability < match probability

relatives,
coancestors

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 11: pages 301-302.

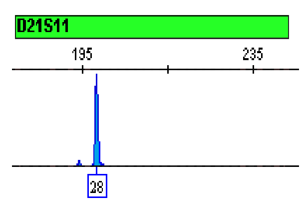
Subpopulations



General Population

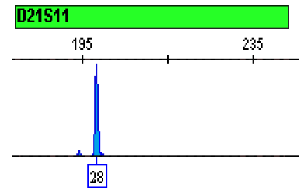
$p_{28} = 0.5$

Subpopulations



50% Subpopulation 1	50% Subpopulation 2
mates only with members of subpopulation 1	mates only with members of subpopulation 2
no random mating	
$p_{28} = 0.4$	$p_{28} = 0.6$

Subpopulations

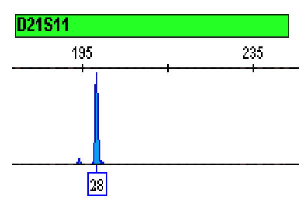


Taking into account subpopulations:

50% Subpopulation 1	50% Subpopulation 2
$\mathrm{Pr}(28,28) = 0.4^2 = 0.16$	$\mathrm{Pr}(28,28) = 0.6^2 = 0.36$

$$Pr(28,28) = \frac{1}{2} \times 0.16 + \frac{1}{2} \times 0.36 = 0.26$$

Subpopulations



Not taking into account subpopulations:

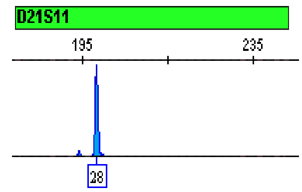
General Population

needs correction!

$$p_{28} = 0.5$$

$$Pr(28,28) = 0.5^2 = 0.25 < \mathbf{0.26}$$

Subpopulations



General Population

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

Balding & Nichols Equations

Balding D.J., Nichols R.A. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* 1994; 64: 125-40.

119

Subpopulations

The **coancestry coefficient** F_{ST}, also called θ, 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***?

allele 28

Subpopulations

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

We have seen: ***allele 28***

The probability of observing an **allele 28** is:

$$\text{either} \quad \underbrace{\theta}_{\text{allele 28 is IBD with 28}} \quad \underset{\text{or}}{+} \quad \underbrace{(1-\theta)p_{28}}_{\text{allele 28 is not IBD with any of the alleles already seen, it is observed by chance}}$$

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: ***allele 28*** and *allele 28*

The probability of observing an **allele 28** is:

$$\underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{(1-\theta)p_{28}}_{\text{allele 28 is not IBD with any of the alleles already seen, it is observed by chance}}$$

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*** and *allele 28*

The probability of observing an **allele 28** is:

$$2\theta + (1-\theta)p_{28}$$

$\mathbf{1+\theta}$

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:

$$\underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{\theta}_{\text{allele 28 is IBD with 28}} + \underbrace{(1-\theta)p_{28}}_{\text{allele 28 is not IBD with any of the alleles already seen, it is observed by chance}}$$

Subpopulations

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

The probability of observing an **allele 28** is:

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

$\mathbf{1 + 2\theta}$

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*}?

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

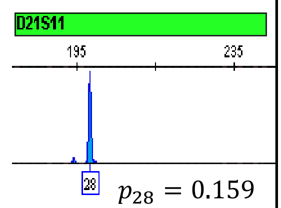
allele 28
allele 28

Balding D.J., Nichols R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International, 64*: 125-40.

Subpopulations

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

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



$p_{28} = 0.159$

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 $\boldsymbol{\theta = 0}$?

A. 0

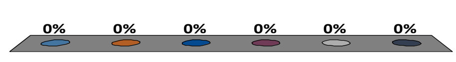
B. θ

C. p_{28}^2

D. $2p_{28}$

E. ? ? ?

Response Counter



Subpopulations

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

Subpopulations

We have seen	Divide by
1 allele	1
2 alleles	$1+\theta$
3 alleles	$1+2\theta$

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

The probability of observing an **allele 13** is:

$$\frac{\underbrace{1\times\theta}_{\text{allele 13 is IBD with 13}} + \underbrace{0\times\theta}_{\text{allele 13 is IBD with 16}} + \underbrace{(1-\theta)p_{13}}_{\text{allele 13 is not IBD with any of the alleles already seen, it is seen by chance}}}{1+\theta}$$

Subpopulations

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

The probability of observing an **allele 13** is:

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

Subpopulations

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

Subpopulations

We have seen	Divide by
1 allele	1
2 alleles	$1+\theta$
3 alleles	$1+2\theta$

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

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

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

Subpopulations

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

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

$$\frac{\theta + (1-\theta)p_{16}}{1+2\theta}$$

Subpopulations

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

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

allele 13

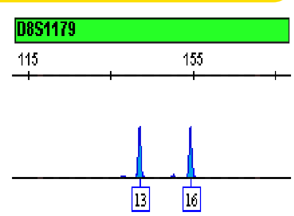
allele 16

Balding D.J., Nichols R.A. (1994). DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International, 64*: 125-40.

Subpopulations

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

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



$p_{13} = 0.330$
$p_{16} = 0.033$

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 $\boldsymbol{\theta = 0}$?

A. 0

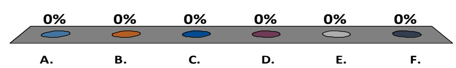
B. 2θ

C. p_{13}^2

D. $2p_{13}p_{16}$

E. ???

Response Counter



NRC II

<u>Formula 4.10a</u>

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

<u>Formula 4.10b</u>

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

National Research Council (NRCII) Committee on DNA Forensic Science. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington DC, 1996.

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

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

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

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

National Research Council Committee on DNA Forensic Science. The Evaluation of Forensic DNA Evidence. National Academy Press, Washington D.C., 1996.

Fixation indices ($\boldsymbol{F}$-statistics)

F-statistics	alternative notation	Meaning
F_{IS}	f	**Individual to Subpopulation:** the correlation of alleles within an individual within a subpopulation
F_{IT}	F	**Individual to Total population:** the correlation of alleles within an individual (``inbreeding´´)
F_{ST}	θ	**Subpopulation to Total population:** the correlation of alleles of different individuals in the same subpopulation (``coancestry´´)

J.M. Butler. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*, Chapter 10: pages 260-262.

NRC II Report Recommendations

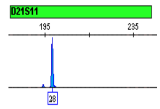
		Assumptions
	Hardy-Weinberg Law:	Assumes Hardy-Weinberg Equilibrium and Linkage Equilibrium in the population
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	Corrects for Hardy-Weinberg Disequilibrium in the population caused by population subdivision. 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 <u>sub-populations</u>.

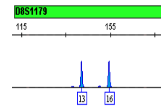
J. Buckleton, C.M. Triggs, S.J. Walsh. (2005). *Forensic DNA Evidence Interpretation.* CRC Press, London: pages 84-98.

NRC II Report Recommendations

		Homozygotes	Heterozygotes
	Hardy-Weinberg Law:	p_{28}^2	$2p_{13}p_{16}$
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	$Fp_{28} + (1-F)p_{28}^2$	$2p_{13}p_{16}$
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	$\frac{[2\theta + (1-\theta)p_{28}][3\theta + (1-\theta)p_{28}]}{(1+\theta)(1+2\theta)}$	$\frac{2[\theta + (1-\theta)p_{13}][\theta + (1-\theta)p_{16}]}{(1+\theta)(1+2\theta)}$

NRC II Report Recommendations





		Homozygotes	**Heterozygotes**
	Hardy-Weinberg Law:	0.025	0.022
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	$F = 0.01$: 0.027 $F = 0.03$: 0.029	0.022
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	$\theta = 0.01$: 0.032 $\theta = 0.03$: 0.048	$\theta = 0.01$: 0.028 $\theta = 0.03$: 0.040

NRC II Report Recommendations

		match probability for 15 loci
	Hardy-Weinberg Law:	8.9×10^{-23}
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	$F = 0.01$: 1.0×10^{-22} $F = 0.03$: 1.4×10^{-22}
Recommendation 4.2	includes possibility that an individual's alleles are IBD with each other or with other observed alleles in the population (``coancestry´´):	$\theta = 0.01$: 3.6×10^{-21} $\theta = 0.03$: 2.4×10^{-19}

NRC II Report Recommendations

		Consequences
	Hardy-Weinberg Law:	The profile seems **more rare** than it actually is.
Recommendation 4.1	includes possibility that the individual's two alleles are IBD (``inbreeding´´):	 The profile seems **a little more rare** than it actually is.
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. 

J. Buckleton, C.M. Triggs, S.J. Walsh. (2005). *Forensic DNA Evidence Interpretation.* CRC Press, London: pages 84-98.