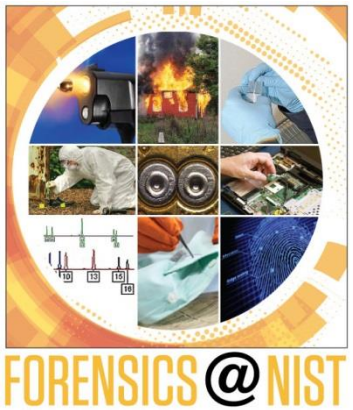


DNA Mixture Interpretation Webcast

April 12, 2013

http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm

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

Statistical Approaches

Michael D. Coble

National Institute of Standards and Technology

In every workshop presented and supported by the NIJ Training Grant (2008-DN-BX-K158)

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

This doesn't have to be a Shakespearean Tragedy!

Stats Required for Inclusions

SWGDAM Interpretation Guideline 4.1:

> "**The laboratory must perform statistical analysis in support of any inclusion** that is determined to be relevant in the context of a case, irrespective of the number of alleles detected and the quantitative value of the statistical analysis."

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

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.

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 Approaches with Mixtures

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

"Exclusionary" Approach

Random Man Not Excluded (RMNE)

Combined Prob. of Inclusion (CPI)

Combined Prob. of Exclusion (CPE)

"Inferred Genotype" Approach

Random Match Probability [modified] (mRMP)

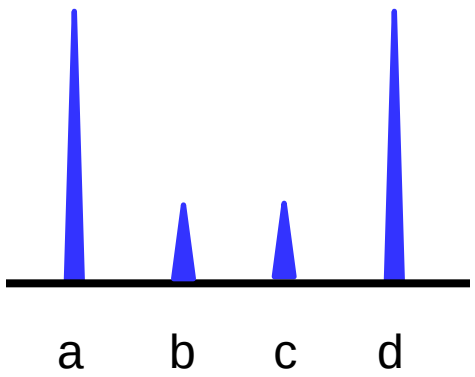
Likelihood Ratio (LR)

"Allele-centric"

"Genotype-centric"

Statistical Approaches with Mixtures

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

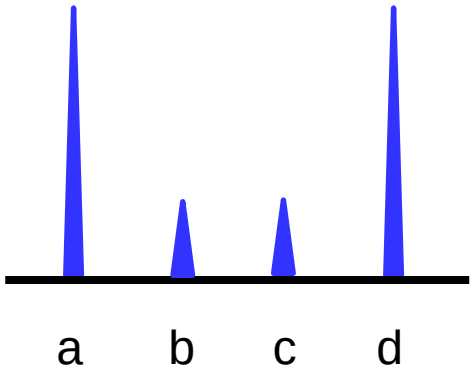


$$PI = (f(a) + f(b) + f(c) + f(d))^2$$

$$CPI = PI_{M1} \text{ X } PI_{M2} \cdots$$

$$CPE = 1 - CP1$$

Breaking down the math…



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

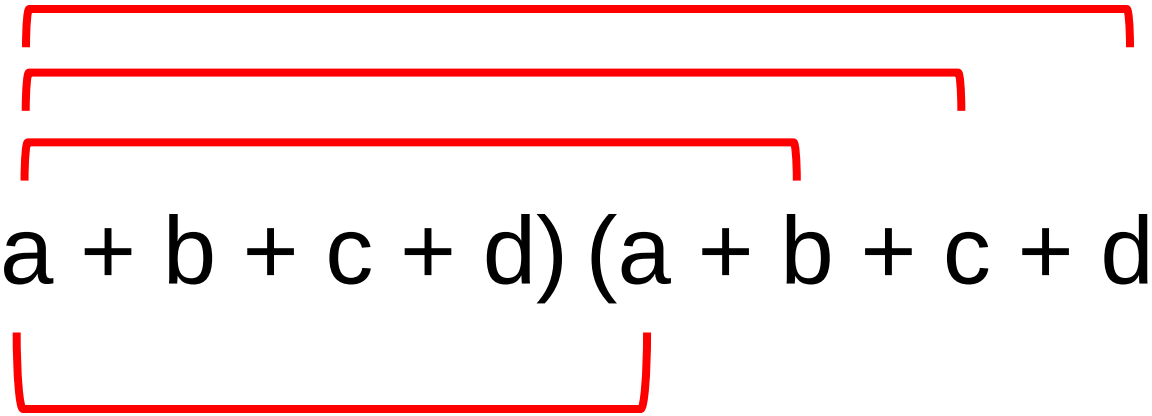
$$(a + b + c + d)^2$$

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

“FOIL”

Breaking down the math…

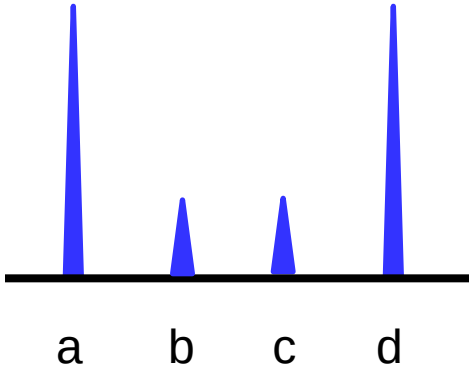
“FOIL”



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

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

RMNE Statistics

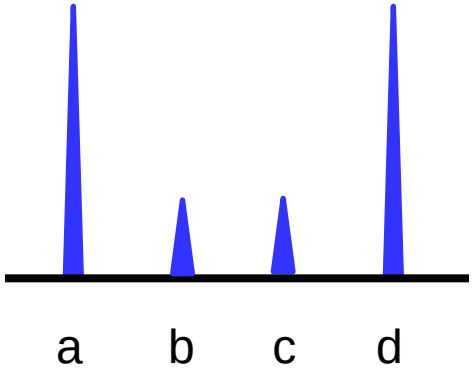


CPI – tries to find all possible “random” persons included in this mixture…

“Included Genotypes”

AA	BB	CC	DD
AB	BC	CD	
AC	BD		
AD			

RMNE Statistics



An “Illogicality” of using RMNE

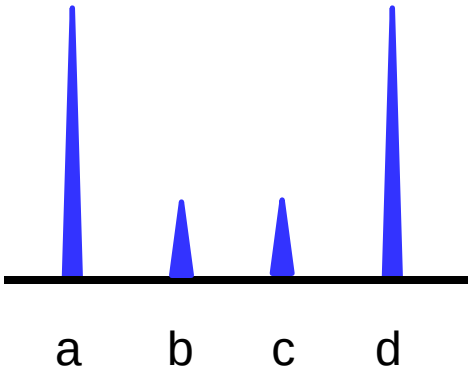
AA + BCD ???

Sure, why not? It fits!

Risk of including individuals ***not*** in the mixture

Statistical Approaches with Mixtures

- **modified Random Match Probability (mRMP)**
  – The major and minor components can be successfully separated into individual profiles. A random match probability is calculated on the evidence as if the component was from a single source sample.

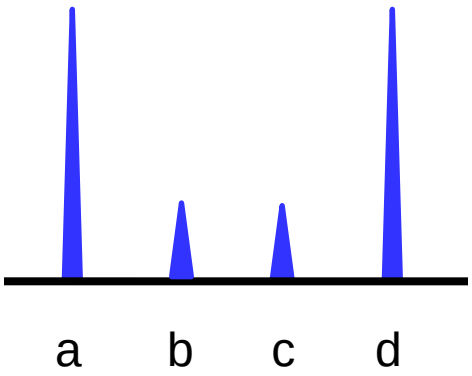


$$mRMP_{minor} = 2pq$$

$$= 2f(b)f(c)$$

Statistical Approaches with Mixtures

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



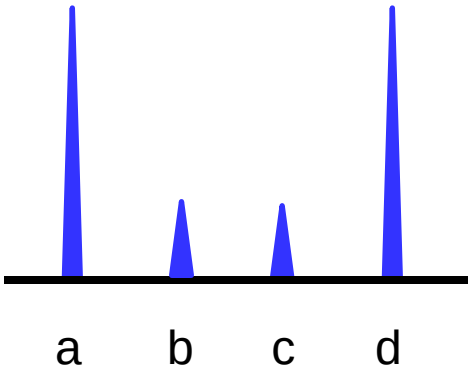
$$\frac{P(E \mid H_1)}{P(E \mid H_2)}$$

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

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

Comparison of the Methods

“Included Genotypes” RMNE



AA BB CC DD
AB BC CD AD
AC BD

“Included Genotypes” LR/mRMP

~~AA BB CC DD~~
~~AB~~ BC ~~CD~~ AD
~~AC~~ ~~BD~~

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

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

John Buckleton [a,*], James Curran [b]

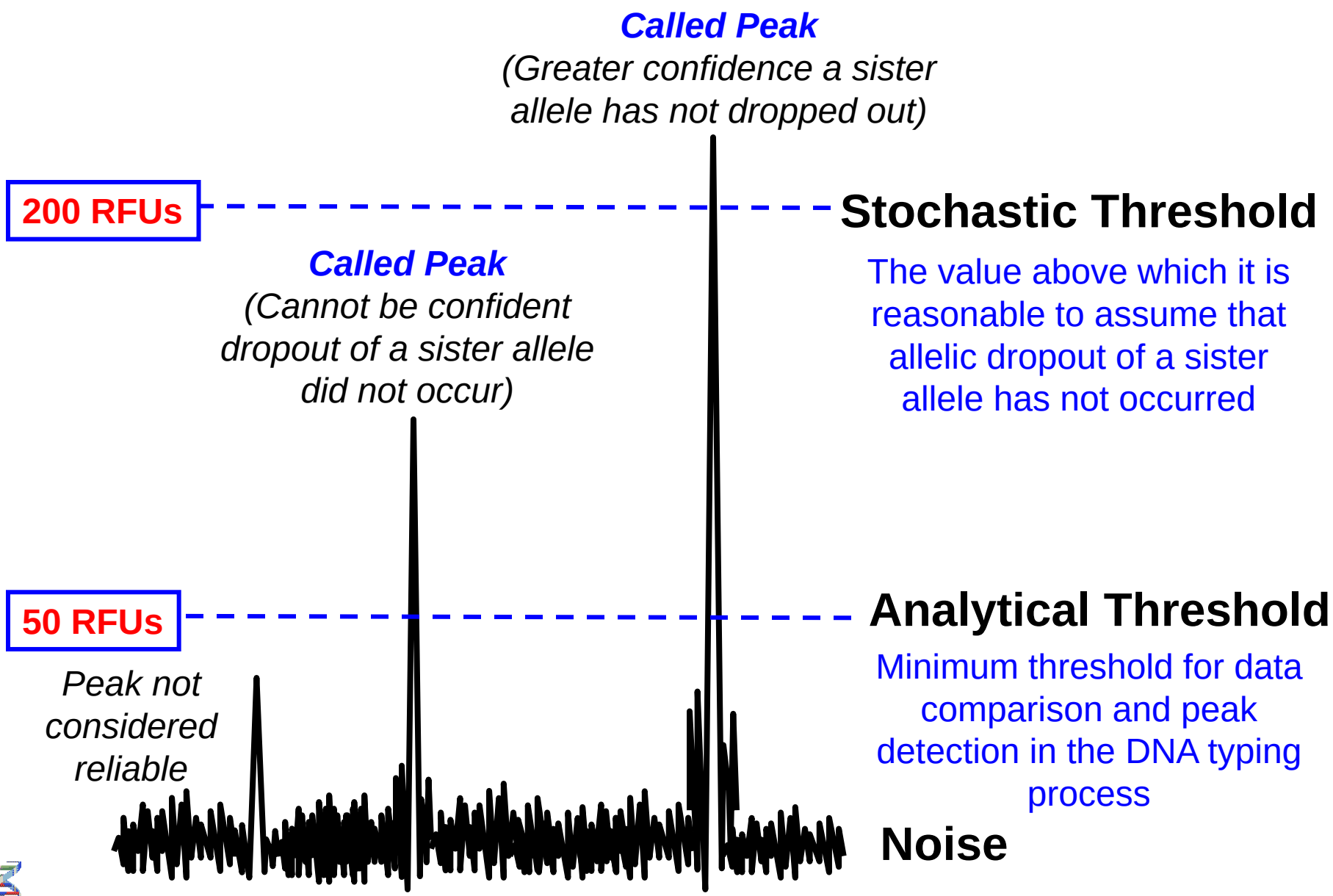
[a] *ESR, PB 92021, Auckland, New Zealand*

[b] *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

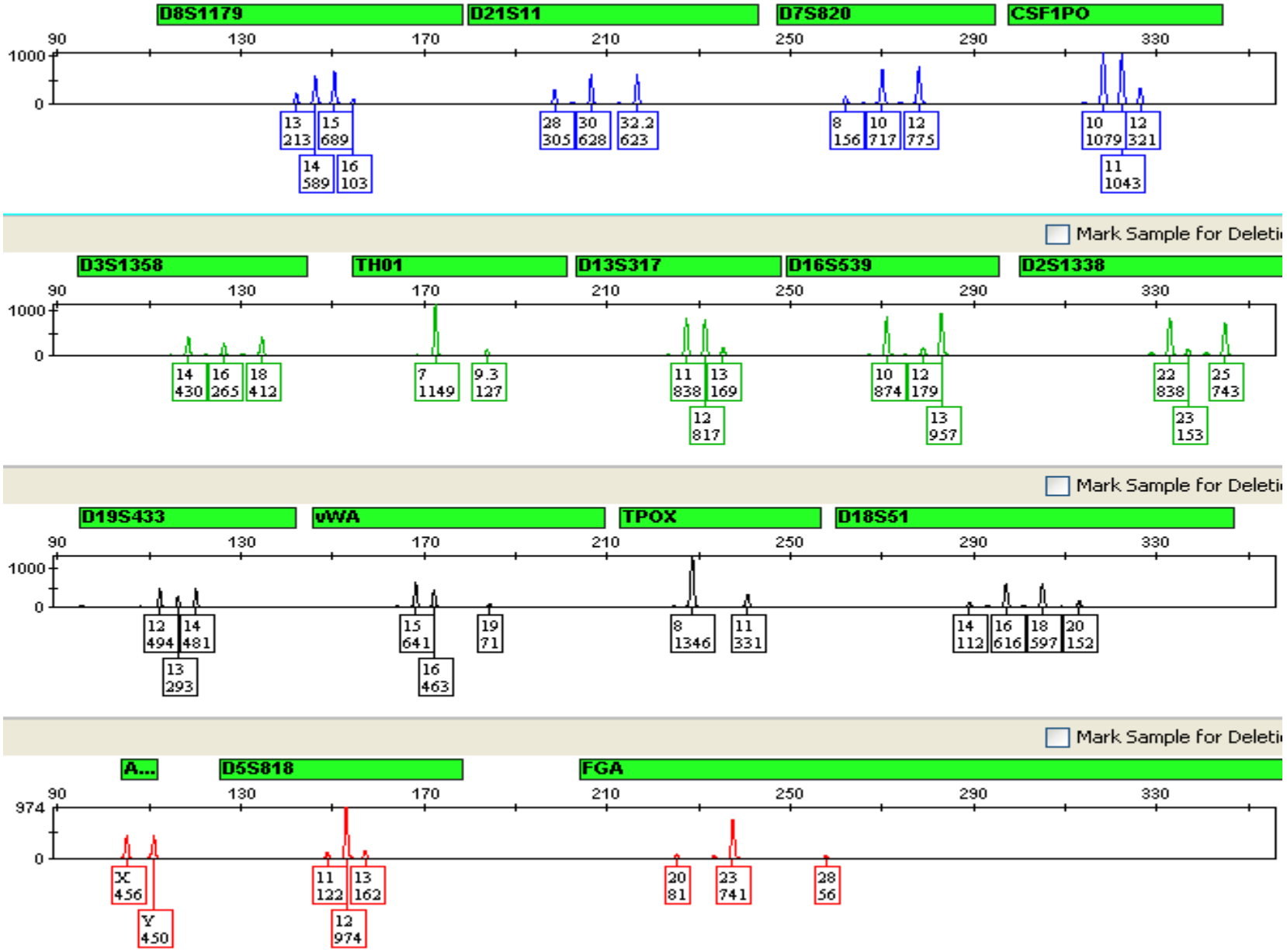
We conclude that the two matters that appear to have real force are:
(1) LRs are more difficult to present in court and
(2) the RMNE statistic wastes information that should be utilised.

Review of Two Thresholds



Butler, J.M. (2010) *Fundamentals of Forensic DNA Typing*. Elsevier Academic Press: San Diego.

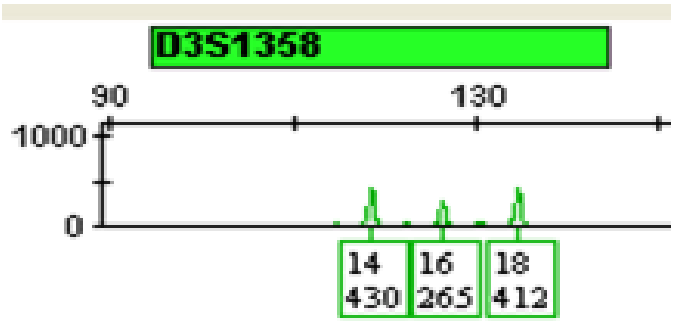
2-Person Mixture

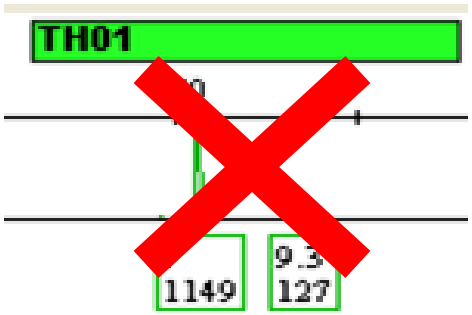


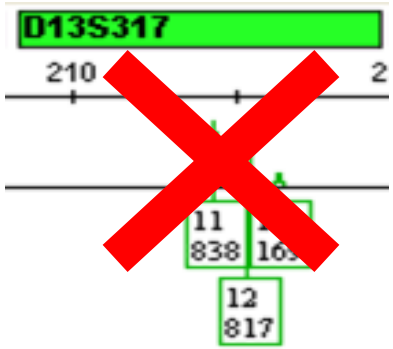
If CPI/CPE Stats are Used

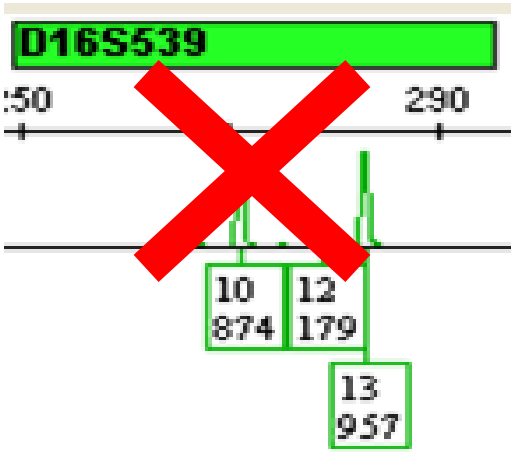
Since exclusionary statistics cannot adjust for the possibility of dropout, and does not take the number of contributors into account, any loci with alleles below the stochastic threshold cannot be used in the CPI statistic.

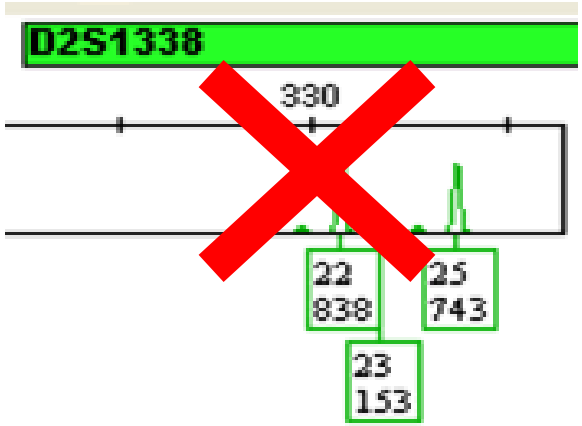
If CPI/CPE Stats are Used (ST = 200 RFU)



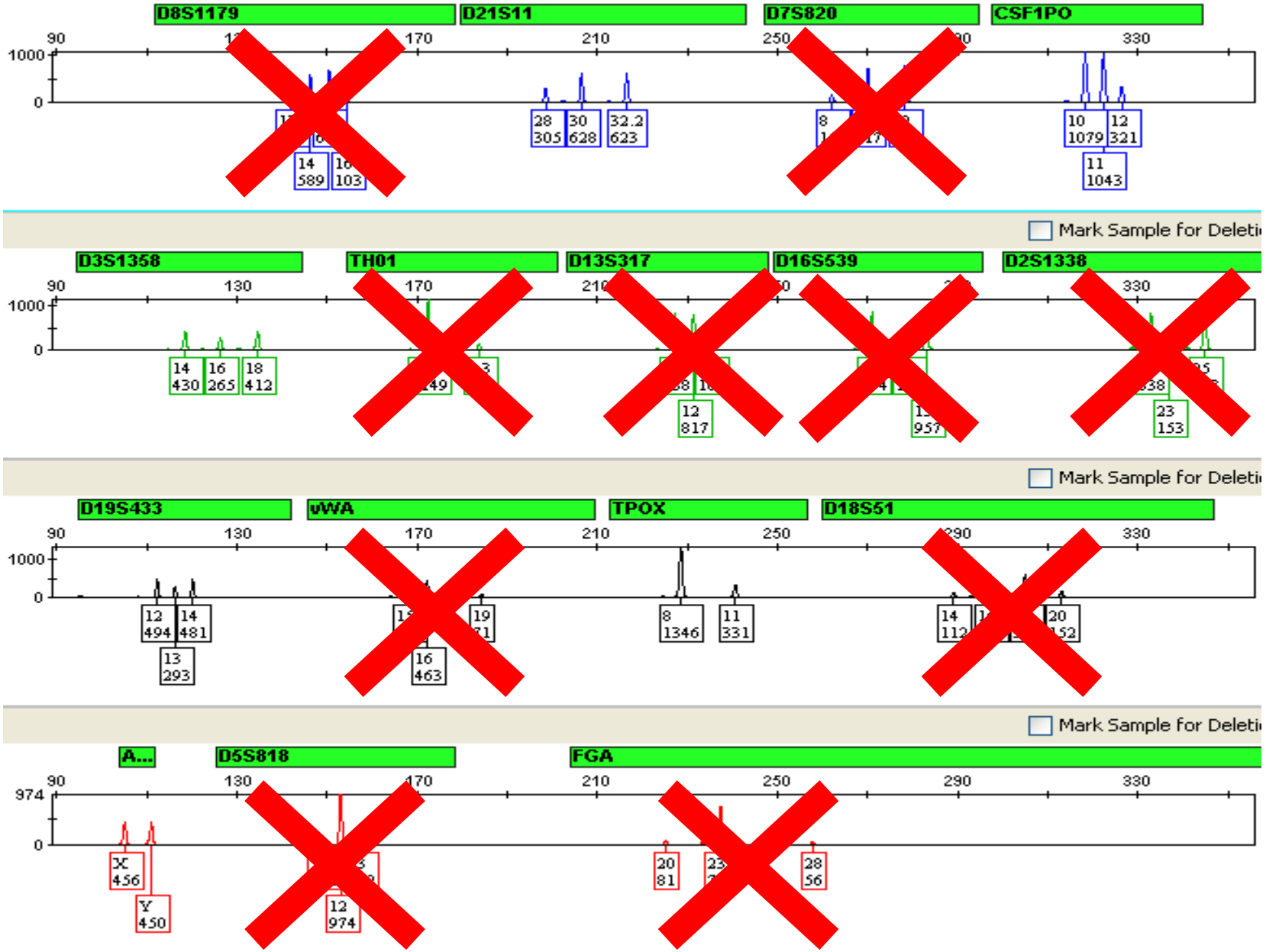








If CPI/CPE Stats are Used



If CPI/CPE Stats are Used

Can use	Cannot use	
D21	D8	D2
CSF	D7	vWA
D3	TH01	D18
D19	D13	D5
TPOX	D16	FGA

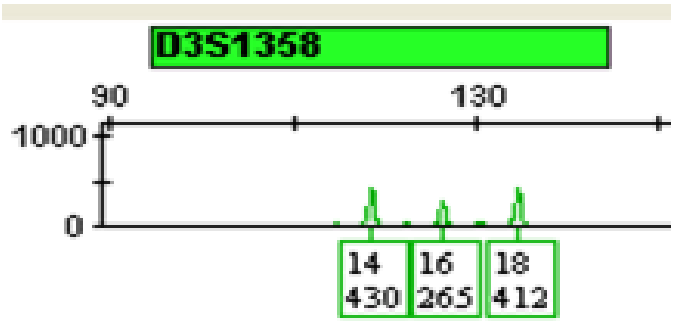
Impact: discarding 2/3 of the data

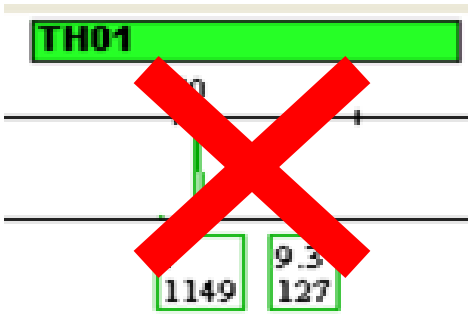
If CPI/CPE Stats are Used

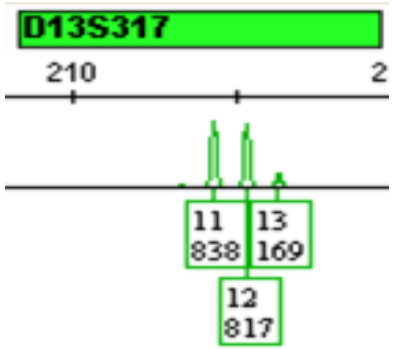
- CPI statistics using FBI Caucasian Frequencies

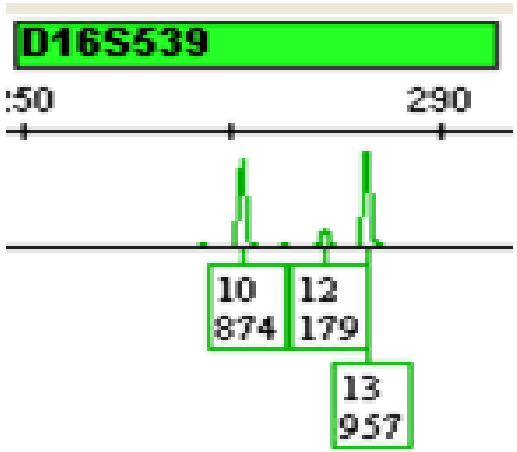
- 1 in 71 Caucasians included
- 98.59% Caucasians excluded

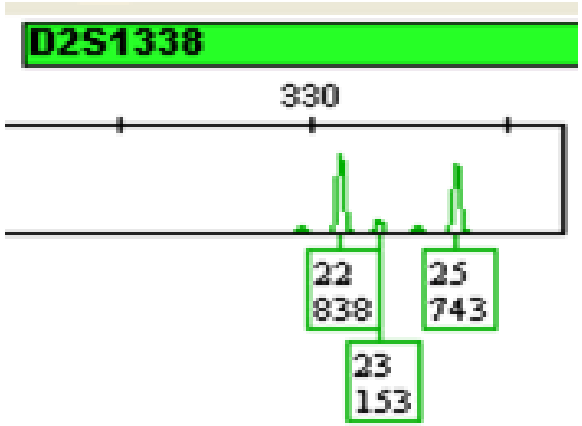
If CPI/CPE Stats are Used (ST = 150 RFU)









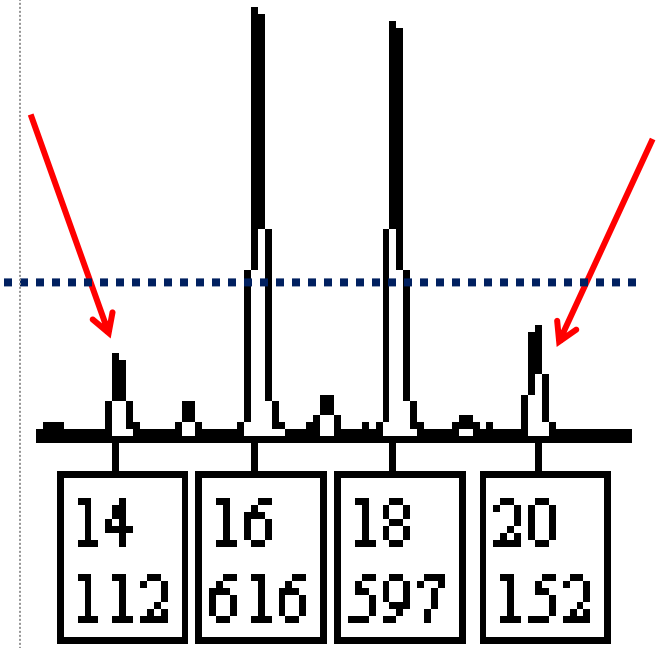


The impact of changing thresholds

If mRMP/LR Stats are Used

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

mRMP - D18S51



If Assume 2 Contributors….

Major	Minor
16,18	14,20

$mRMP_{minor} = 2pq$

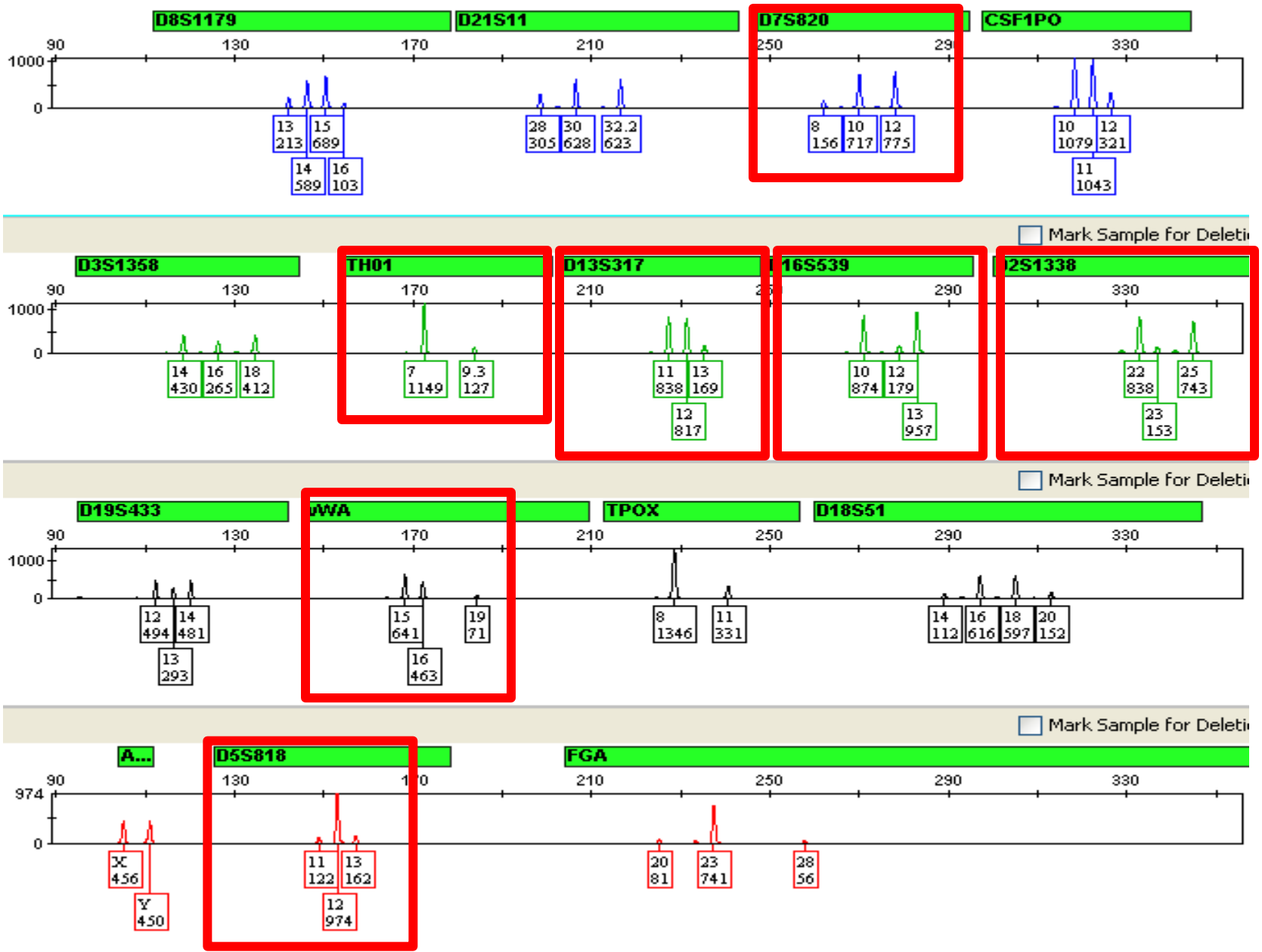
$= 2 \times f(14) \times f(20)$

$= 2 \times (0.1735) \times (0.0255)$

$= 0.00884$ or 1 in 113

(LR = 113)

Potential for Drop-out



If mRMP/LR Stats are Used

Can use

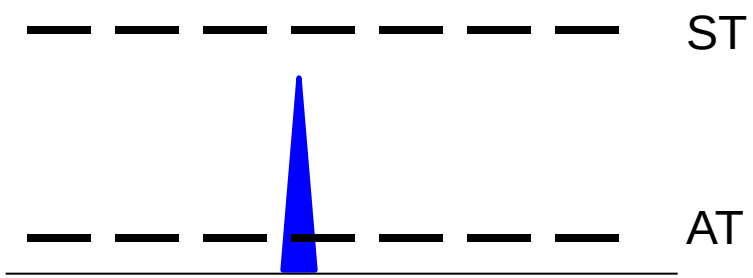
D8
D21
D18
D3
D19
TPOX
FGA
CSF

Loci with potential D-out

D7
TH01
D13
D16
D2
vWA
D5

The “2p” Rule

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

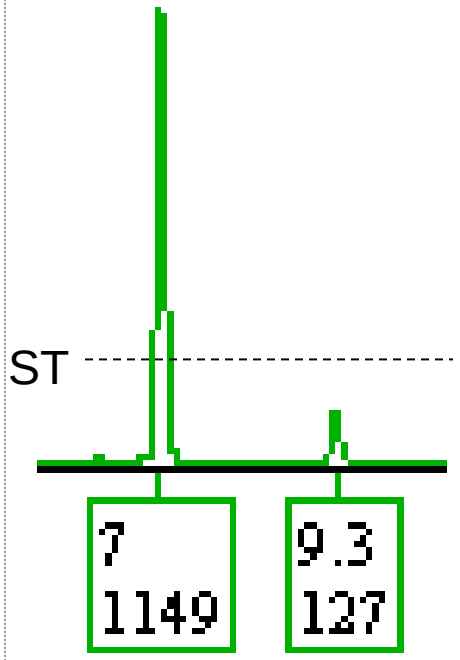


2p – SWGDAM Guidelines

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

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

Macbeth/Duncan Profile - TH01



Major – 7, 7

Possible Minor Contributors

7, 9.3	(2pq)
9.3, 9.3	p^2
9.3, ?	2p (or $p^2 + 2p(1 - p)$)

Macbeth/Duncan Profile - TH01

$$\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{V \,\&\, S}{V \,\&\, U} = \frac{f_7^2 + f_7(1-f_7)\theta \,\&\, 1}{f_7^2 + f_7(1-f_7)\theta \,\&\, 2p}$$

$$p^2 + 2p(1-p)$$

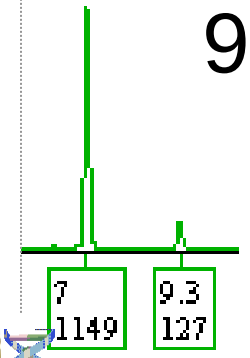
V = 7, 7

U = 7, 9.3

9.3, 9.3

9.3, ?

$$= \frac{1}{f_{9.3}^2 + 2f_{9.3}(1-f_{9.3})}$$



$f_{9.3} = 0.3054$

$= 1 / 0.5175 = \boxed{1.93}$

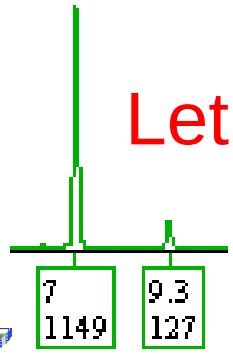
Macbeth/Duncan Profile - TH01

$$\frac{P(E \mid H_1)}{P(E \mid H_2)} = \frac{V \,\&\, S}{V \,\&\, U} = \frac{1}{p^2 + p(1-p)\theta + 2pq}$$

V = 7, 7
U = 7, 9.3
9.3, 9.3

$$= \frac{1}{f_{9.3}^2 + f_{9.3}(1-f_{9.3})\theta + 2f_{9.3}f_7}$$

Let ST = 125 RFU



$f_{9.3} = 0.3054$
$f_7 = 0.1724$

$= 1 / 0.2007 = \boxed{4.98}$

Macbeth/Duncan Profile - TH01

	LR
ST = 200 (2p is used)	1.93
ST = 125 (2pq is used)	4.98

2p is conservative…

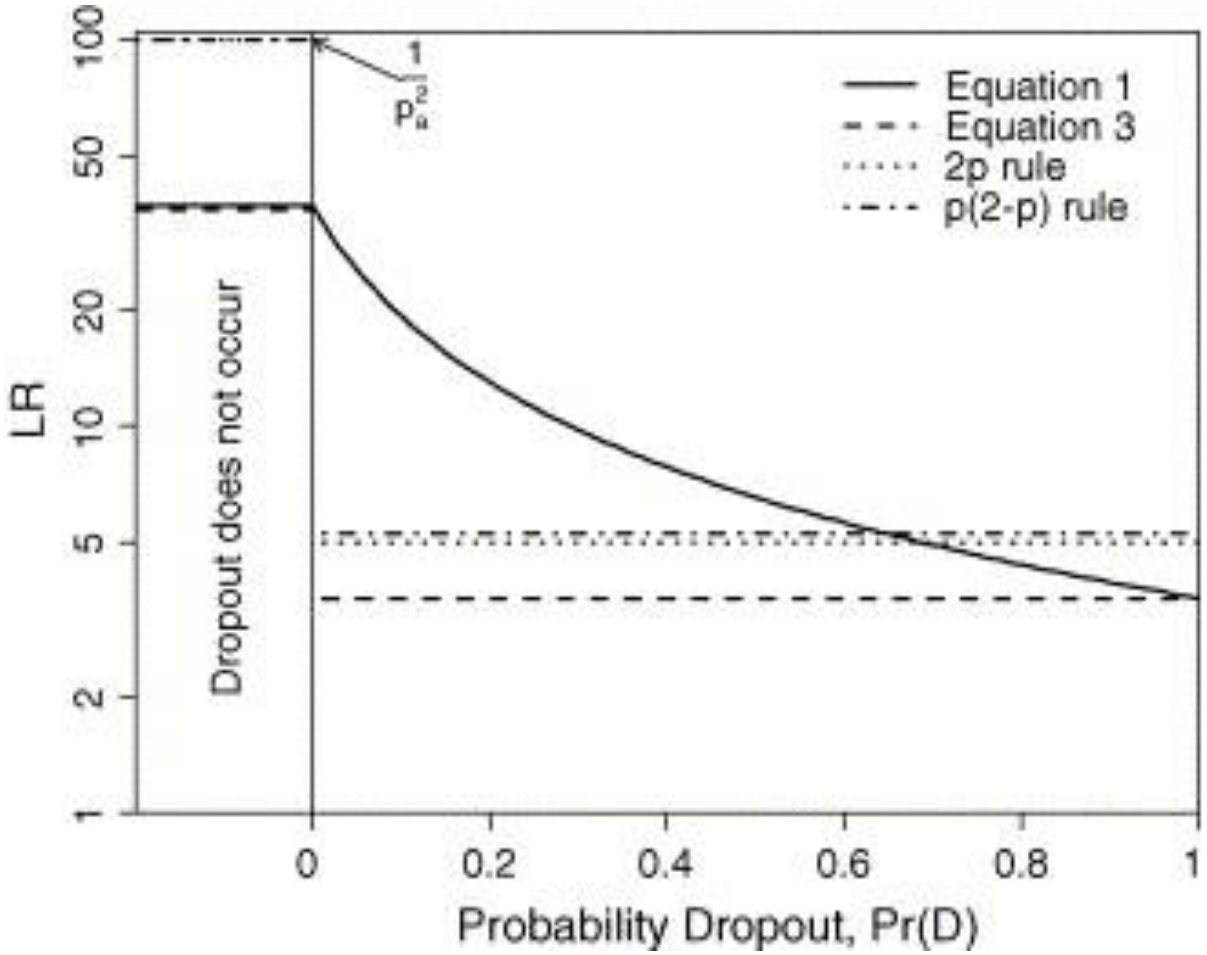
The “2p” Rule

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

- Buckleton and Triggs (2006)

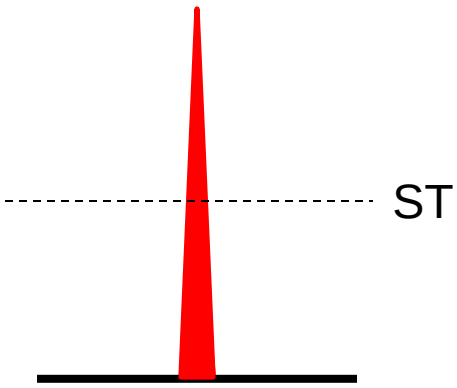
Is the 2*p* rule always conservative?”

The “2p” Rule



Stain = aa

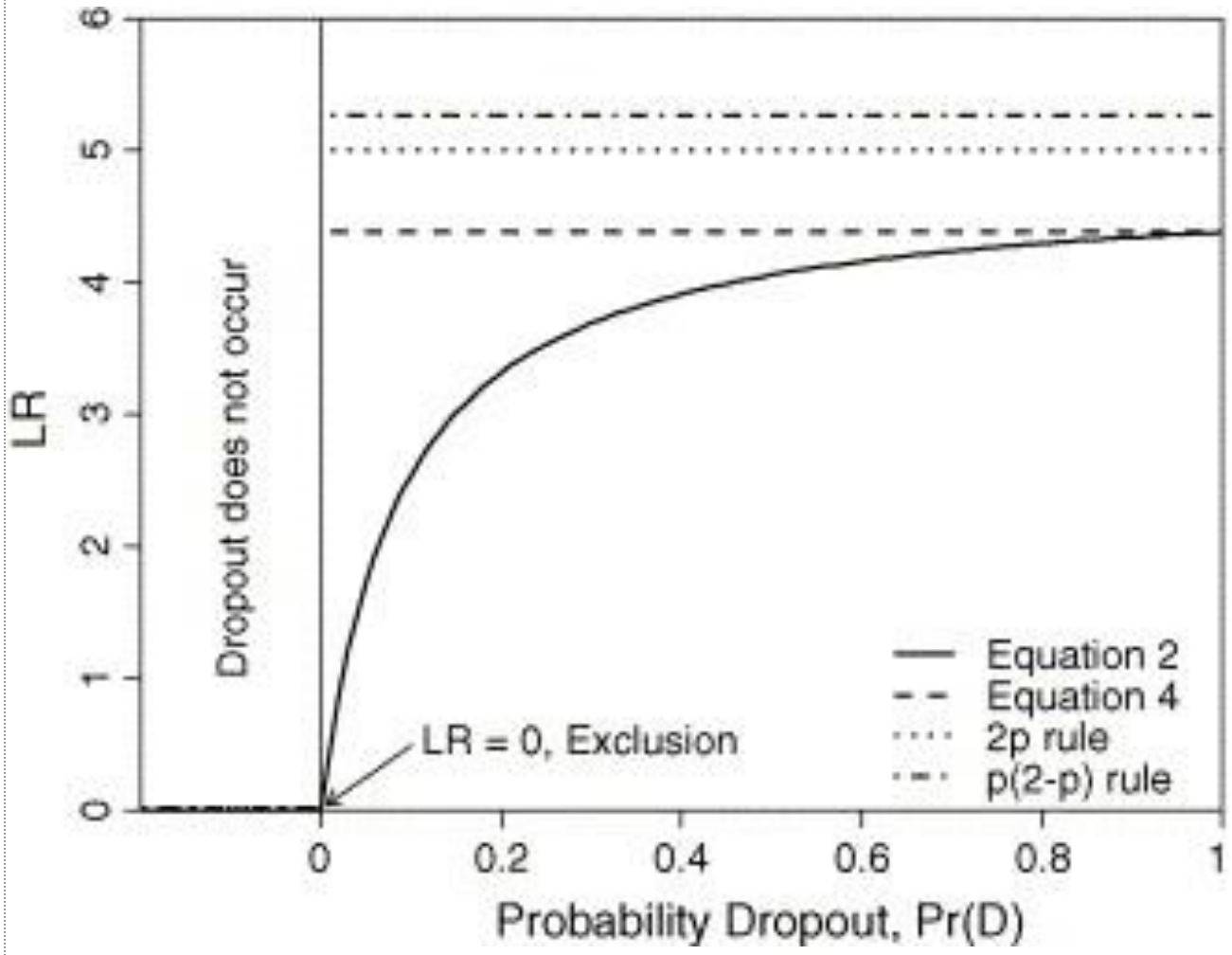
Suspect = aa



LR = 100

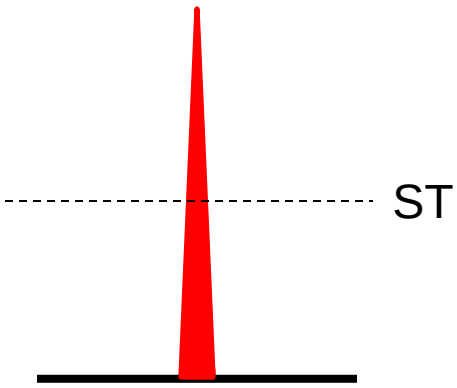
f(a) = 0.10 $1/p^2 = 100$ 1/2p = 5

The “2p” Rule



Stain = aa

Suspect = ab



Exclusion

f(a) = 0.10 1/2p = 5

Is there a way forward?

Gill and Buckleton *JFS* **55:** 265-268 (2010)

- "The purpose of the ISFG DNA commission document was to provide a way forward to demonstrate the use of ***probabilistic models to circumvent the requirement for a threshold*** and to safeguard the legitimate interests of defendants."

Summary of the Issues

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

- “Just as logic is reasoning applied to truth and falsity, probability is reasoning with uncertainty”

-Dennis Lindley

Summary of the Issues

- The LR is a method to evaluate evidence that can overcome many of the limitations we are facing today. ISFG Recommendations are published.
- This will require (obviously) software solutions… however, we need to better understand and be able to explain the statistics as a community.
- “But, for my own part, it was Greek to me”

  — William Shakespeare, *Julius Caesar*
- “We know what we are, but know not what we may be.” — William Shakespeare, *Hamlet*

Summary of the Issues

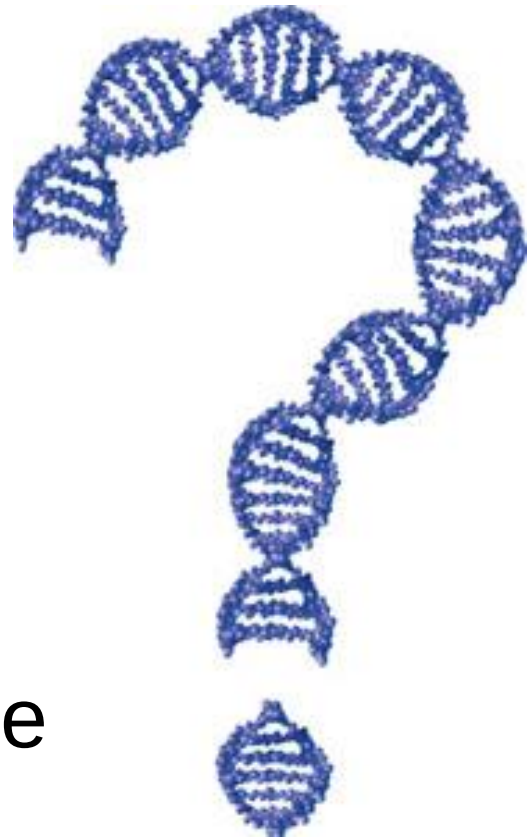
- Extensive training will be necessary – and a single 8 hour workshop will once a year will not suffice.

Thank you for your attention

Contact Information

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

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



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