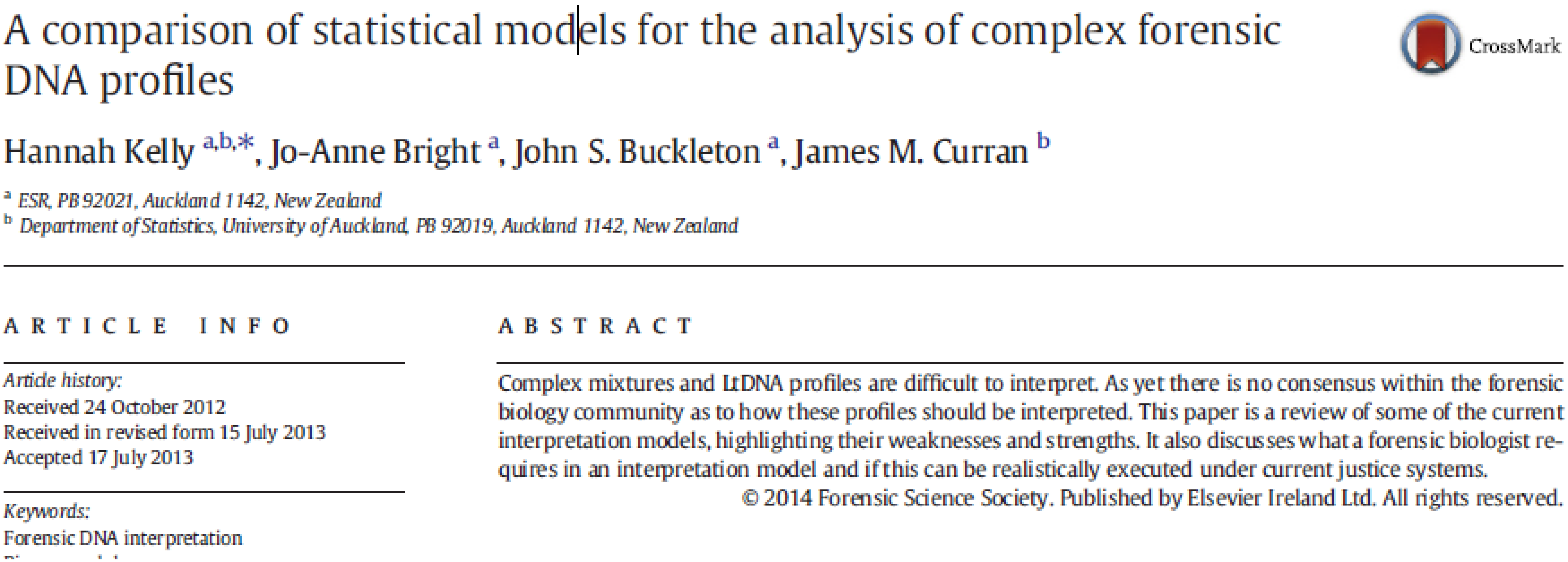
# Continuous models for mixtures

Thore Egeland

Copenhagen May 20-23 2014



# A comparison of statistical models for the analysis of complex forensic DNA profiles

Hannah Kelly [a,b,*], Jo-Anne Bright [a], John S. Buckleton [a], James M. Curran [b]

[a] ESR, PB 92021, Auckland 1142, New Zealand
[b] Department of Statistics, University of Auckland, PB 92019, Auckland 1142, New Zealand

ARTICLE INFO

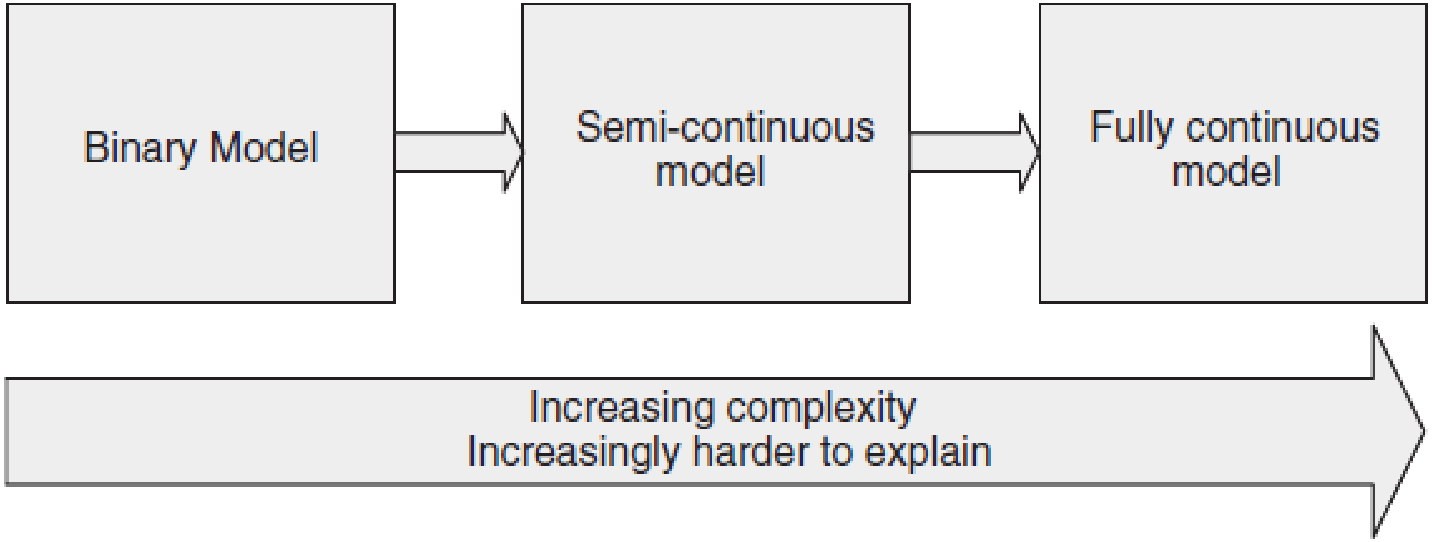
*Article history:*
Received 24 October 2012
Received in revised form 15 July 2013
Accepted 17 July 2013

*Keywords:*
Forensic DNA interpretation

ABSTRACT

Complex mixtures and LtDNA profiles are difficult to interpret. As yet there is no consensus within the forensic biology community as to how these profiles should be interpreted. This paper is a review of some of the current interpretation models, highlighting their weaknesses and strengths. It also discusses what a forensic biologist requires in an interpretation model and if this can be realistically executed under current justice systems.

http://www.cstl.nist.gov/strbase/training/ISFG2013workshops.htm



Binary Model

Semi-continuous model

Fully continuous model

Increasing complexity
Increasingly harder to explain

Fig. 3. Summary of the relationship of the different models for forensic DNA interpretation.

# Binary model

1. Qualitative binary model (aka unrestricted…)
   - Treats alleles as present or absent and does not take into account peak height information
2. Semi-quantitative binary model (aka restricted …)
   - Declares some of the combinations as possible or impossible

# Example

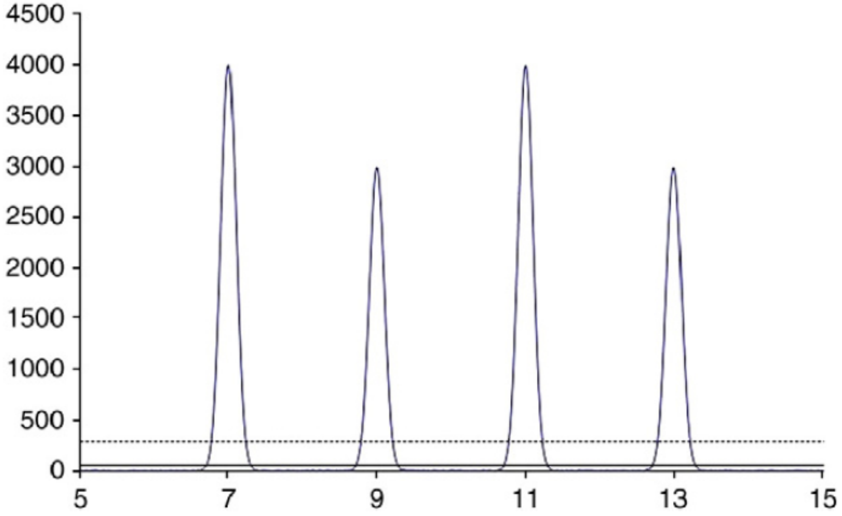


**Fig. 2.** Artificial epg of four-peak locus for a two-person mixture.

6 genotype combinations

{7/9,11/13},{7/11,9/13},{7/13,9/11}

{11/13,7/9},{9/13,7/11},{9/11,7/13}

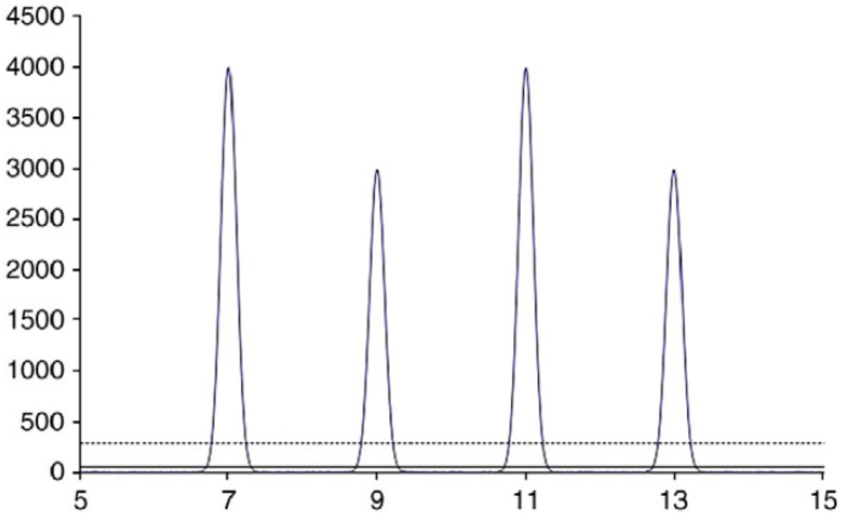


**Fig. 2.** Artificial epg of four-peak locus for a two-person mixture.

Suspect 7/11

$H_P$ : S+NN1, $H_D$ : NN1+ NN2

Assume only combinations are 7/11, 9/13

$$LR_2 = \frac{2p_9 p_{13}}{2p_7 p_9 2p_{11} p_{13} + 2p_{11} p_{13} 2p_7 p_9} = \frac{1}{4p_7 p_{11}}$$

Ignoring peak information

$$LR_1 = \frac{2p_9 p_{13}}{24p_7 p_9 p_{11} p_{13}} = \frac{1}{12p_7 p_{11}}$$

$$LR_2 = 3LR_1$$

# Semi continuous models

- Artefacts:
  - Drop-out
  - Drop-in
  - Stutters

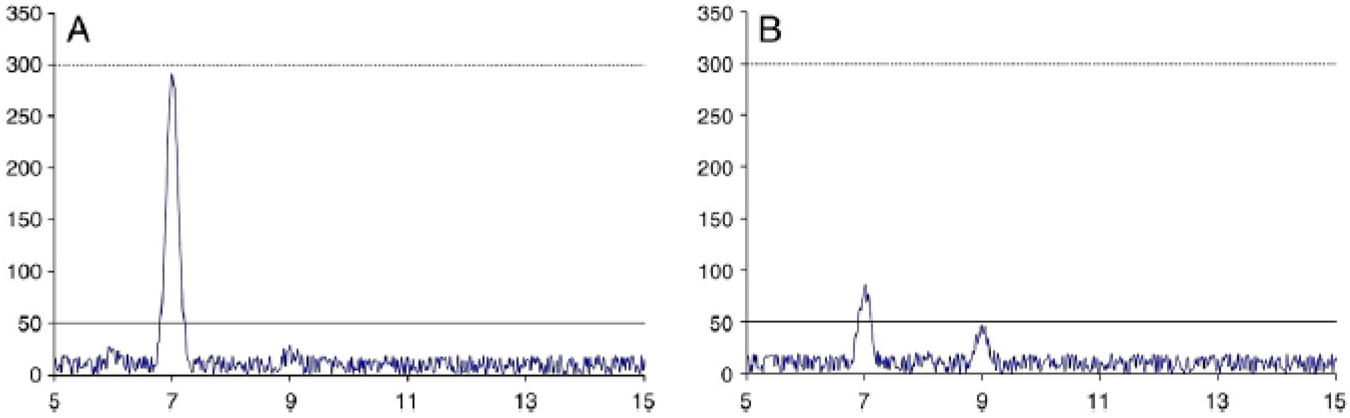


**Fig. 1.** Two examples of non-concordance where POI = 7,9. A large concordant 7 allele with no 9 peak observed (non-tolerable non-concordance) and B small concordant 7 allele with a non-concordant 9 peak visible sub-threshold (tolerable non-concordance). Stochastic threshold = 300 RFU, limit of detection 50 RFU.

Semi-continuous model includes a drop-out probability d
Software: LRmix (Haned), ....

# Continuous model

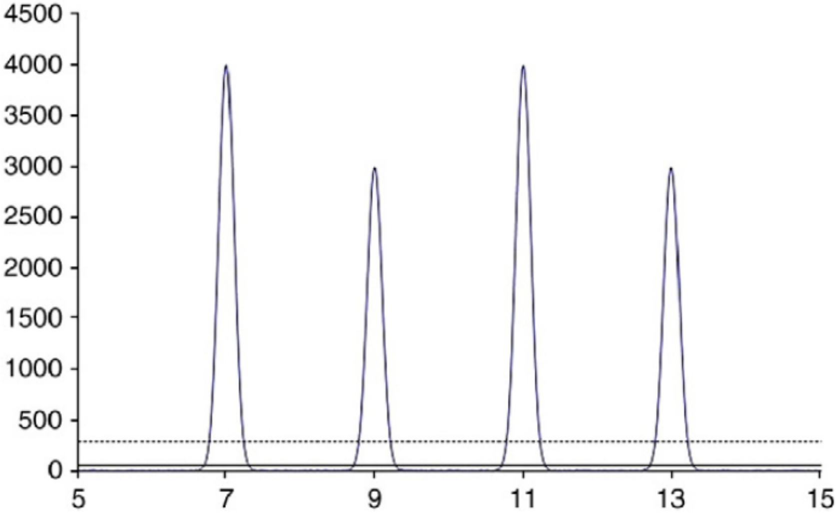


**Fig. 2.** Artificial epg of four-peak locus for a two-person mixture.

6 genotype combinations
{7/9,11/13},{7/11,9/13},{7/13,9/11}
{11/13,7/9},{9/13,7/11},{9/11,7/13}
weighed according to a
probability distribution

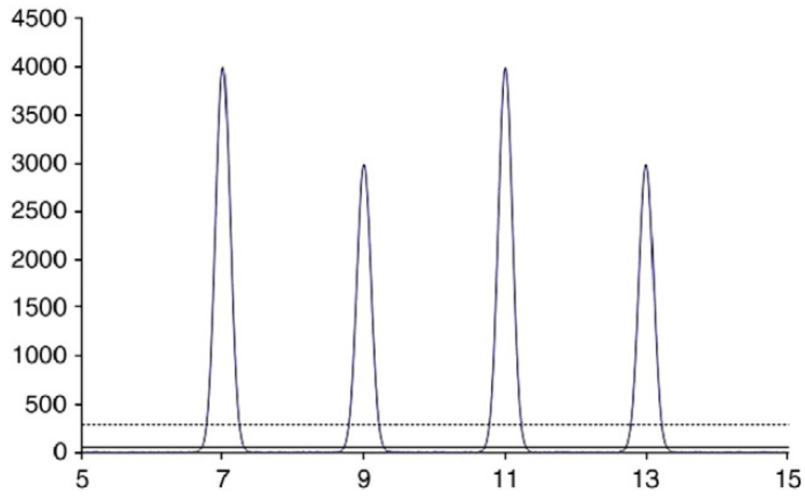


**Fig. 2.** Artificial epg of four-peak locus for a two-person mixture.

Suspect 7/11

$H_P:$ S+NN1, $H_D:$ NN1+ NN2

$$LR = \frac{w_2 2 p_9 p_{13}}{w_1 2 p_7 p_9 2 p_{11} p_{13} + w_2 2 p_7 p_{11} 2 p_9 p_{13} + \ldots + w_6 2 p_9 p_{11} 2 p_7 p_{13}}$$

Qualitative binary model: $w_1 = \ldots = w_6 = 1$

Semi – quantitative binary model: $w_2 = 1,\ w_1 = w_3 = \ldots = w_6 = 0$

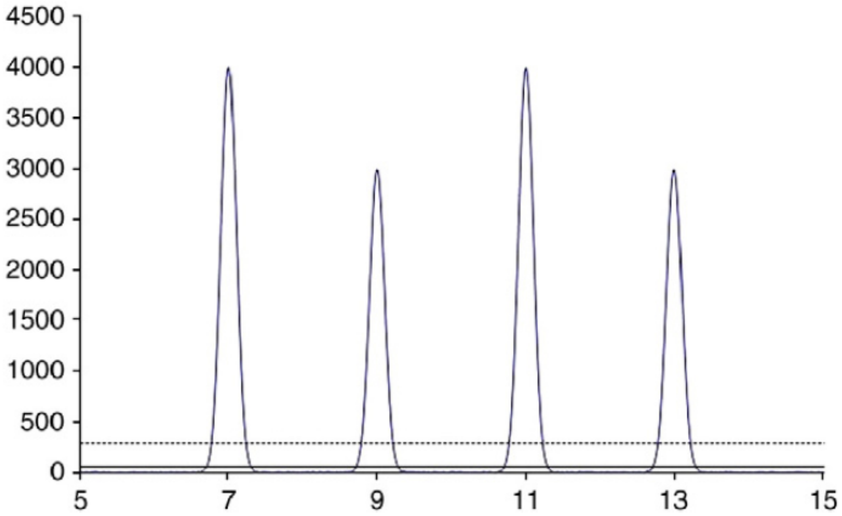


**Fig. 2.** Artificial epg of four-peak locus for a two-person mixture.

Suspect 7/11

$H_P$ : S+NN1, $H_D$ : NN1+ NN2

More terms with drop-out, drop-in ($drop$ - $terms$)

$$LR = \frac{w_2 2 p_9 p_{13}}{w_1 2 p_7 p_9 2 p_{11} p_{13} + \ldots + w_6 2 p_9 p_{11} 2 p_7 p_{13} + drop - terms}$$

# Computation

$$LR = \frac{w_2 2 p_9 p_{13}}{w_1 2 p_7 p_9 2 p_{11} p_{13} + \ldots + w_6 2 p_9 p_{11} 2 p_7 p_{13} + drop - terms}$$

- Hard part: Modelling and estimating weights

$$w_j = \Pr(\text{peak heights}|\text{known contributors})$$

  - ❖ Bayesian networks (Graversen, Cowell, Lauritzen, ...
    - o DNAmixtures
  - ❖ MCMC (Taylor, Bright,...)
    - o STRmix http://strmix.esr.cri.nz/
    - oTrueAllele http://www.cybgen.com/
  - ❖ Numerical integration

ELSEVIER

Forensic Science International: Genetics

Volume 7, Issue 5, September 2013, Pages 516–528

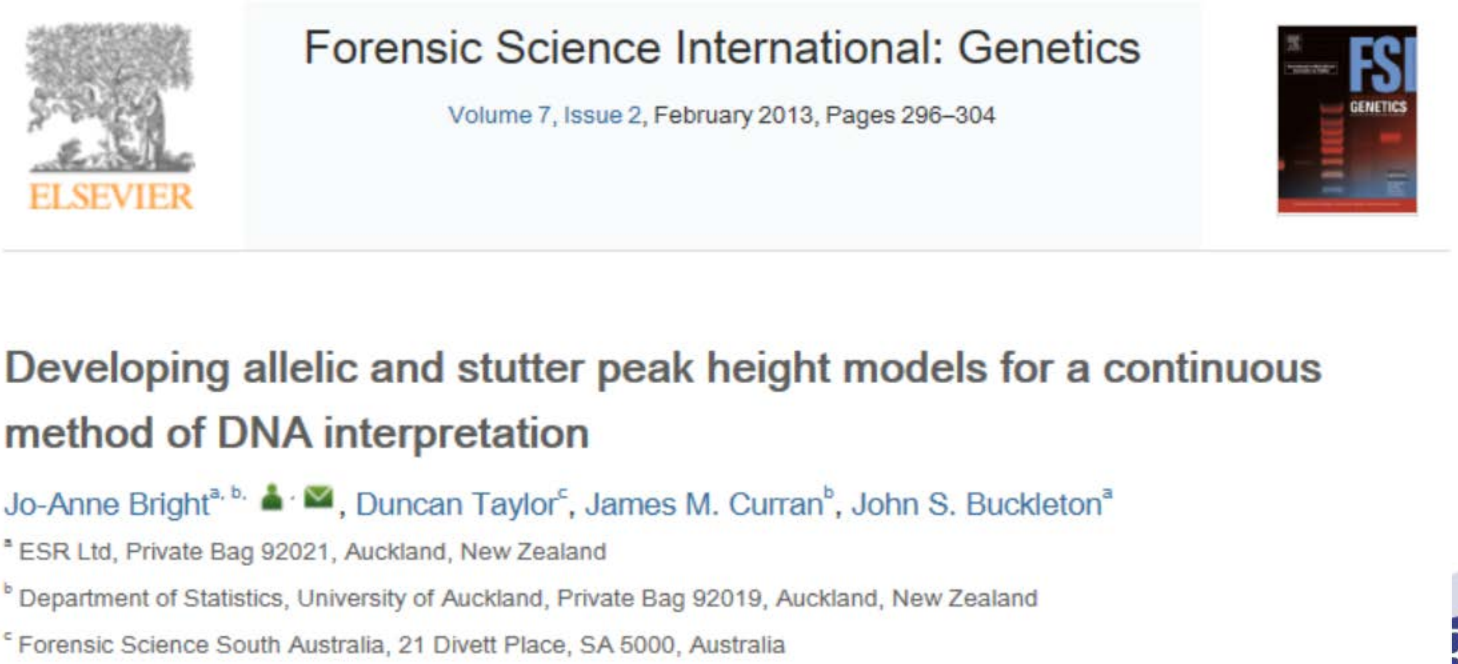
Forensic population genetics – original research

## The interpretation of single source and mixed DNA profiles

Duncan Taylor[a], Jo-Anne Bright[b], John Buckleton[b]

[a] Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia

[b] ESR Ltd, Private Bag 92021, Auckland 1142, New Zealand



ELSEVIER

Forensic Science International: Genetics

Volume 7, Issue 2, February 2013, Pages 296–304

## Developing allelic and stutter peak height models for a continuous method of DNA interpretation

Jo-Anne Bright[a, b], Duncan Taylor[c], James M. Curran[b], John S. Buckleton[a]

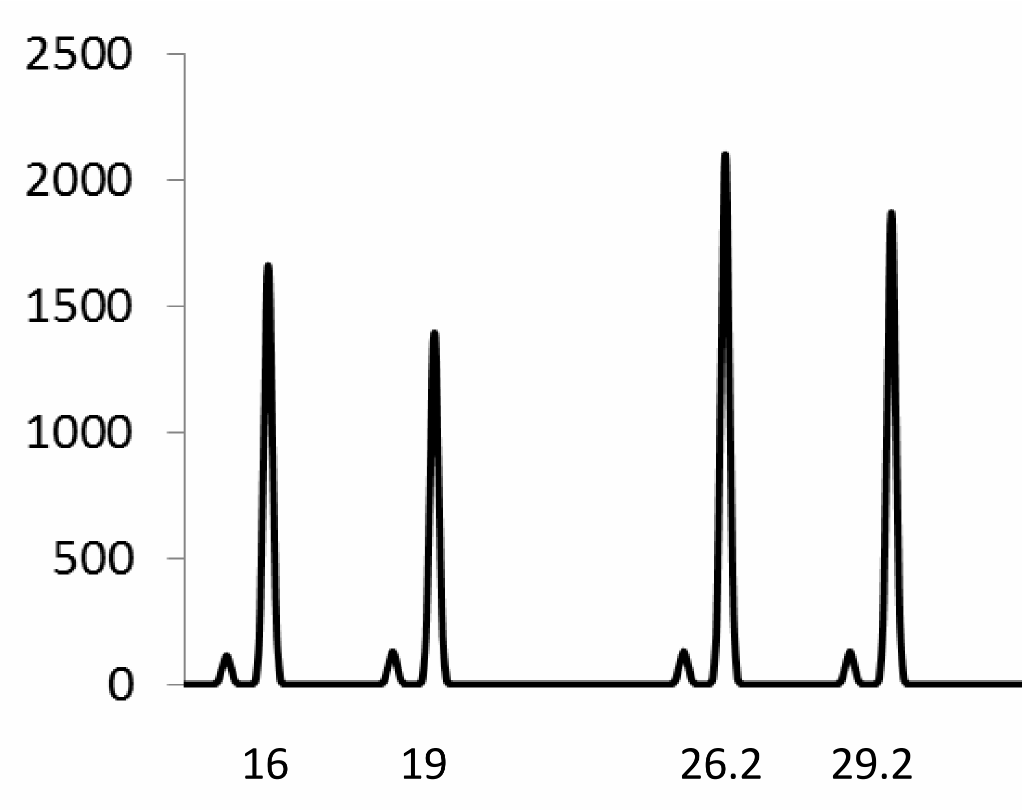
[a] ESR Ltd, Private Bag 92021, Auckland, New Zealand

[b] Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

[c] Forensic Science South Australia, 21 Divett Place, SA 5000, Australia

# SE33

- **Two person mixture**



Suspect=16/19
HP: suspect+NN
HD: NN1+NN2

# Possible genotype combinations

| Number | Contributor 1 | | Contributor 2 | |
|---|---|---|---|---|
| 1 | 19 | 29.2 | 16 | 26.2 |
| 2 | 16 | 19 | 26.2 | 29.2 |
| 3 | 16 | 29.2 | 19 | 26.2 |
| 4 | 19 | 26.2 | 16 | 29.2 |
| 5 | 16 | 26.2 | 19 | 29.2 |
| 6 | 26.2 | 29.2 | 16 | 19 |

| Number | Contributor 1 |  | Contributor 2 |  | Product $\Pr(O \mid S_j, M)$ | Expected weight |
|---|---|---|---|---|---|---|
| 1 | 19 | 29.2 | 16 | 26.2 | 26 | 0.208 |
| 2 | 16 | 19 | 26.2 | 29.2 | 76 | 0.608 |
| 3 | 16 | 29.2 | 19 | 26.2 | 7 | 0.056 |
| 4 | 19 | 26.2 | 16 | 29.2 | 12 | 0.096 |
| 5 | 16 | 26.2 | 19 | 29.2 | 3 | 0.024 |
| 6 | 26.2 | 29.2 | 16 | 19 | 1 | 0.008 |

$$LR_C = \frac{\sum_j w_j \Pr(S_j \mid H_1)}{\sum w_u \Pr(S_u \mid H_2)}$$

$u$

# Likelihood ratio

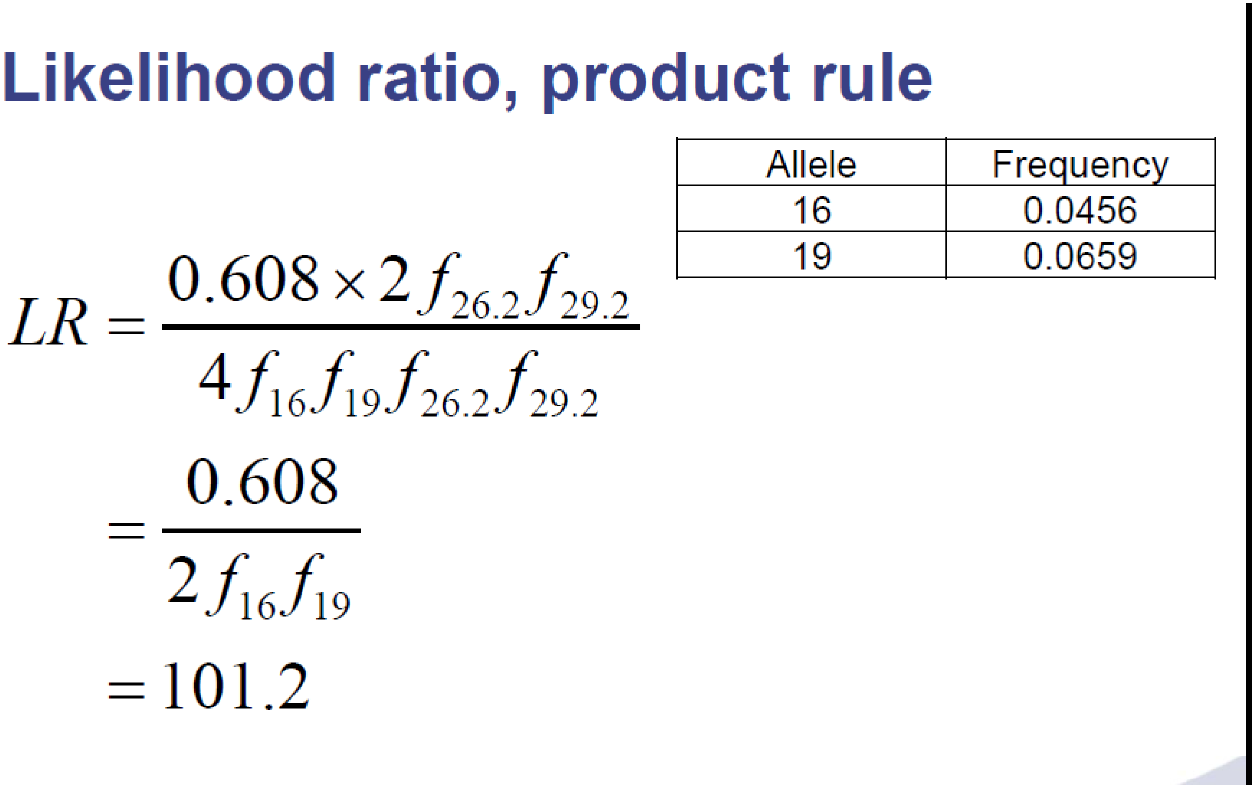
- **Assuming person of interest was 16,19**
- **Assume that according to other (unseen) loci POI must be contributor 1**

| Number | Contributor 1 | | Contributor 2 | | Product $\Pr(O \mid S_j, M)$ | Expected weight |
|---|---|---|---|---|---|---|
| 1 | 19 | 29.2 | 16 | 26.2 | 26 | 0.208 |
| 2 | 16 | 19 | 26.2 | 29.2 | 76 | 0.608 |
| 3 | 16 | 29.2 | 19 | 26.2 | 7 | 0.056 |
| 4 | 19 | 26.2 | 16 | 29.2 | 12 | 0.096 |
| 5 | 16 | 26.2 | 19 | 29.2 | 3 | 0.024 |
| 6 | 26.2 | 29.2 | 16 | 19 | 1 | 0.008 |

# Likelihood ratio

$$\Pr(E \mid H_1) = 0.608 \times 2 \times f_{26.2} \times f_{29.2}$$

$$\begin{aligned}\Pr(E \mid H_2) = {} & 0.208 \times 2 \times f_{19} \times f_{29.2} \times 2 \times f_{16} \times f_{26.2} + \\ & 0.608 \times 2 \times f_{16} \times f_{19} \times f_{26.2} \times f_{29.2} + \\ & 0.056 \times 2 \times f_{16} \times f_{29.2} \times 2 \times f_{19} \times f_{26.2} + \\ & 0.096 \times 2 \times f_{19} \times f_{26.2} \times 2 \times f_{16} \times f_{29.2} + \\ & 0.024 \times 2 \times f_{16} \times f_{26.2} \times 2 \times f_{19} \times f_{29.2} + \\ & 0.008 \times 2 \times f_{26.2} \times f_{29.2} \times 2 \times f_{16} \times f_{19} + \\ = {} & 4 f_{16} f_{19} f_{26.2} f_{29.2}\end{aligned}$$



# Likelihood ratio, product rule

| Allele | Frequency |
|---|---|
| 16 | 0.0456 |
| 19 | 0.0659 |

$$LR = \frac{0.608 \times 2 f_{26.2} f_{29.2}}{4 f_{16} f_{19} f_{26.2} f_{29.2}}$$

$$= \frac{0.608}{2 f_{16} f_{19}}$$

$$= 101.2$$

Several combinations are unlikely and so the binary model gives a lower value:

$$LR_{Binary} = \frac{1}{12 f_{16} f_{19}} \approx 50$$

# Exercise

- Describe the three methods presented and discuss pros and cons!