

# LR Distributions and Error Rates

Cybergenetics Webinar  
User Group Meeting  
January, 2025  
Pittsburgh, PA

Mark W Perlin, PhD, MD, PhD  
Cybergenetics, Pittsburgh, PA



Cybergenetics © 2003-2025

---

---

---

---

---

---

---

---

# Noncontributor Distribution

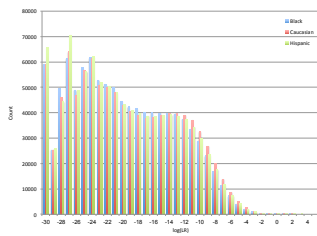
OPEN ACCESS Freely available online



TrueAllele Casework on Virginia DNA Mixture Evidence:  
Computer and Manual Interpretation in 72 Reported  
Criminal Cases

Mark W. Perlin<sup>1\*</sup>, Kiersten Dormer<sup>1</sup>, Jennifer Hornyak<sup>1</sup>, Lisa Schiermeier-Wood<sup>2</sup>, Susan Greenspoon<sup>2</sup>

<sup>1</sup>Cybergenetics, Pittsburgh, Pennsylvania, United States of America, <sup>2</sup>Department of Forensic Science, Richmond, Virginia, United States of America



What is this?  
How was it made?  
How is it used?

---

---

---

---

---

---

---

---

# Questions

1. What about individual evidence genotypes?
2. How do we construct LR distributions?
3. How do we get error rates from them?
4. How do we report a LR's error rate?

---

---

---

---

---

---

---

---

## US v Curtis Johnson

- In 2013, men robbed an armored truck outside a New Orleans bank, killing the truck guard in a shootout.
- A bandana was collected from the crime scene
- A 70 pg sample was a three-person mixture
- TrueAllele separated out bandana genotypes
- Comparing a 27% contributor with Johnson, **LR = 200**
- 2021 – *Daubert* hearing, TrueAllele admitted, first trial
- 2022 – Second trial, guilty verdict, 50-year sentence

---

---

---

---


---

---

---

---

## How to form LR distributions

**Heliyon** 

---

Received: 9 July 2018  
Revised: 18 September 2018  
Accepted: 24 September 2018

**Efficient construction of match strength distributions for uncertain multi-locus genotypes**

---

Cite as:  
Mark W. Perlin, Efficient construction of match strength distributions for uncertain multi-locus genotypes. *Heliyon* 4 (2018) e00521. doi:10.1016/j.heliyon.2018.e00521

**Mark W. Perlin\***  
Cybergenetics, Pittsburgh, PA, USA  
\*Corresponding author

---

Complete, fast, easy, accurate

[TrueAllele® VUler™ Distribution View](#)

---

---

---

---

---

---

---

---

## Reporting LR error rate

A match between the bandana and Johnson is 200 times more probable than coincidence.

For a match strength of 200, only 1 in 4.1 thousand people would match as strongly.

$$ER \leq 1/LR$$
$$1/4100 \leq 1/200$$

---

---

---

---

---

---

---

---

## Question

What about *exclusionary* error rates?

---

---

---

---

---

---

---

---

## GA v Kerry Robinson

- 17-year-old Robinson was wrongfully convicted of rape in 2002 and sentenced to 20 years in prison.
- GBI analyzed the vaginal swabs and developed STR data from the three-person mixture.
- Cybergentics' TrueAllele: exclusionary **LR = 1/103**.
- Based on this new exculpatory DNA evidence, court vacated Robinson's conviction and he was released.

---

---

---

---

---

---

---

---

## Reporting LR error rate

A match between the vaginal swabs and Robinson is 103 times *less* probable than coincidence.

For an exclusionary statistic of one over 103, only 1 in 3,730 people would be excluded as strongly.

$$\begin{aligned} ER &\leq LR \\ 1/3730 &\leq 1/103 \end{aligned}$$

---

---

---

---

---

---

---

---

## Question

What about validation studies?

Sets of genotypes from a laboratory.

---

---

---

---

---

---

---

---

## Noncontributor Distribution

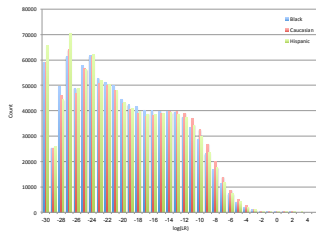
OPEN ACCESS Freely available online

PLOS ONE

**TrueAllele Casework on Virginia DNA Mixture Evidence: Computer and Manual Interpretation in 72 Reported Criminal Cases**

Mark W. Perlin<sup>1\*</sup>, Kiersten Dormer<sup>1</sup>, Jennifer Hornyak<sup>1</sup>, Lisa Schiermeier-Wood<sup>2</sup>, Susan Greenspoon<sup>2</sup>

<sup>1</sup>Cybergenetics, Pittsburgh, Pennsylvania, United States of America, <sup>2</sup>Department of Forensic Science, Richmond, Virginia, United States of America



Error rate?  
Speed?  
Exclusion?

---

---

---

---

---

---

---

---

## False Positive Rate

in over 1,000,000 comparisons per group

Tail distribution	Black	Caucasian	Hispanic
0	39	32	29
1	8	11	9
2	2	1	1
3	0	0	1
log(LR) > 0	49	44	40

For LR > 0, FPR is under 1 in 20,000 (0.005%)  
For LR > 100, FPR is 1 in 1,000,000 (0.0001)%  
For any reported LR value, look up FP error rate

---

---

---

---

---

---

---

---

## Questions

1. Why can we validate using casework data?
2. Why is "ground truth" not needed?

---

---

---

---

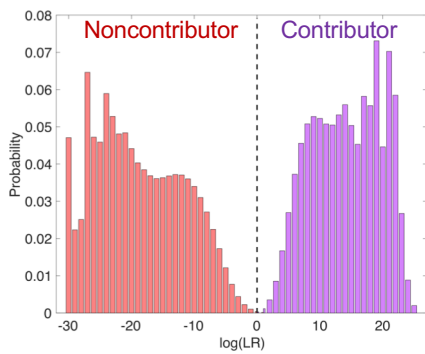
---

---

---

---

## Low ER of sister LR distributions



---

---

---

---

---

---

---

---

## Question

Why doesn't TrueAllele need or use analytical thresholds (AT)?

---

---

---

---

---

---

---

---

## Answer

1. *Bayesian modeling* accounts for baseline.
2. Bayesian computing *uses all the data*.
3. TrueAllele *explains* the STR data pattern.  
It doesn't try to *classify* the data peaks.

---

---

---

---

---

---

---

---

## Question

How do (other software's)  
analytical data thresholds affect  
LR information?

---

---

---

---

---

---

---

---

## US v Alejandro Sandoval

- Police collected a baggie containing methamphetamine
- Defense tested baggie, found a DNA mixture
- Two different PG software programs used
- STRmix™
  - AT = 90 rfu, 11 peaks, log(LR) = -0.53 ban
  - AT = 40 rfu, 24 peaks, log(LR) = -1.38 ban (lab report)
  - AT = 20 rfu, 38 peaks, log(LR) = -7.48 ban
- TrueAllele®
  - AT = none, 210 peaks, log(LR) = -6.08 ban (Cybergenetics)
- Unsuccessful *Daubert* attempt to challenge STRmix
- Plea agreement dropped the more serious charge
- JFS published a "Case Report" that speculated on TrueAllele v. STRmix

Can LR distributions explain what happened?

---

---

---

---

---

---

---

---

## Question

Why aren't LR reporting thresholds (RT) needed?

---

---

---

---

---

---

---

---

## JFS 2015



PAPER  
CRIMINALISTICS

Mark W. Perlin,<sup>1</sup> Ph.D., M.D.; Jennifer M. Hornyak,<sup>1</sup> M.S.; Garrett Sugimoto,<sup>2</sup> M.S.; and Kevin W.P. Miller,<sup>2</sup> Ph.D.

TrueAllele<sup>®</sup> Genotype Identification on DNA Mixtures Containing up to Five Unknown Contributors\*

evidence: 10 x 5-person mixtures = 50 genotypes  
reference: R = 10, 100, ..., 100,000 references

---

---

---

---

---

---

---

---

## Unattainable “zero error”

Sampling issue

R	N	positive	maximum	mean	%positive
10	500	29	2.5669	-7.5297	5.800
100	5,000	63	2.5669	-7.7570	1.260
1,000	50,000	440	3.1531	-8.1047	0.880
10,000	500,000	2,865	4.5458	-8.2003	0.573
100,000	5,000,000	28,614	4.5843	-8.1973	0.572
exact				-8.1747	0.638

Greater sampling (greater N)

More outliers (more positives)

Larger max LR (larger maximum)

Forcing higher reporting threshold

Just report all LRs and provide ER context

---

---

---

---

---

---

---

---

## Question

How to rebut “fake science” attacks?

---

---

---

---

---

---

---

---

## US v Ravel Mills

- 2020 shooting-related homicide in Washington, D.C.
- Gun and magazine recovered as evidence
- Gun: **6% component** of a **three-person** mixture
  - $\log(\text{LR}) = -7.86$ ,  $\log(\text{ER}) = -11.18$
- Magazine: **2% component** of a **four-person** mixture
  - $\log(\text{LR}) = -11.21$ ,  $\log(\text{ER}) = -14.54$
- Federal prosecutor requested *Daubert* hearing
- Typical defense expert attack: old-style binary logic
- TrueAllele won “on the papers”, no hearing needed

---

---

---

---

---

---

---

---

## Take JFS 2015 out of context

**Binary error rate (LR <> 1)**  
in the case software version

N=	Mixture Range %	count for LR<1	% for LR<1
20	1–5	7	35%
17	5–10	0	0%

- Binary error rates are simplistic and irrelevant
1. The likelihood ratio is quantitative
  2. Error rate depends on LR magnitude

---

---

---

---

---

---

---

---



## Error rate depends on LR

Source	Mixture Weight %	log(LR)	log(ER)
<b>Gun</b>	5.89	-7.86	-11.18
<b>Magazine</b>	2.40	-11.21	-14.54
	1.63	-3.49	-6.08
	1.08	-2.61	-3.84
	1.70	-2.47	-4.37
<b>Validation</b>	1.32	-1.40	-3.14
	2.26	-0.60	-2.69
	1.65	-0.54	-2.37
	1.40	-0.15	-2.53

---

---

---

---

---

---

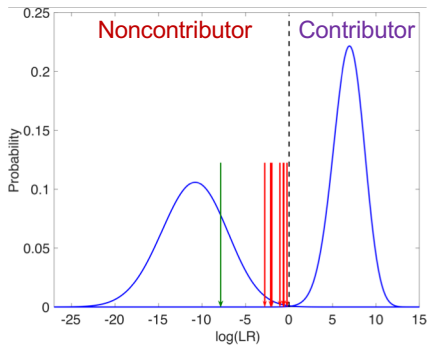
---

---

---

---

## Gun Distributions




---

---

---

---

---

---

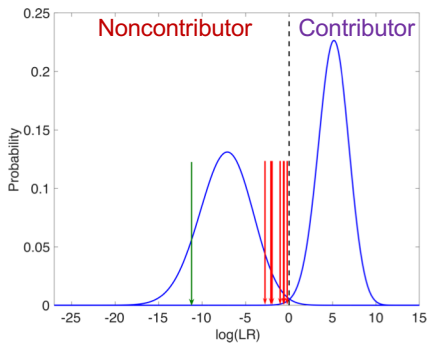
---

---

---

---

## Magazine Distributions




---

---

---

---

---

---

---

---

---

---

## Conclusions

How to always report  
the true information from DNA evidence?

1. Use a fully Bayesian model on all the STR data (no AT)
2. Have TrueAllele separate mixtures into contributor genotypes
3. Report all LRs, both inclusionary and exclusionary (no RT)
4. Use Distribution to report ER (frequency context) for every LR



---

---

---

---

---

---

---

---