**Forensic validation, error and reporting:
a unified approach**

American Academy of Forensic Sciences
Jurisprudence Section
February, 2019
Baltimore, MD

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

Cybergenetics

Cybergenetics © 2003-2019

Rule 702 – Daubert reliability

- testable
- error rate
- peer reviewed
- generally accepted

Simple DNA & random match

$$\frac{1}{\text{Probability(coincidental match)}}$$
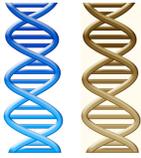
Complex DNA & likelihood ratio

4

$$\frac{\text{Probability(evidence match)}}{\text{Probability(coincidental match)}}$$

Monte Carlo for many genotypes

5

Random drawings from the human population

Gather log(LR) values
many-to-many

Validation – specificity histogram

6
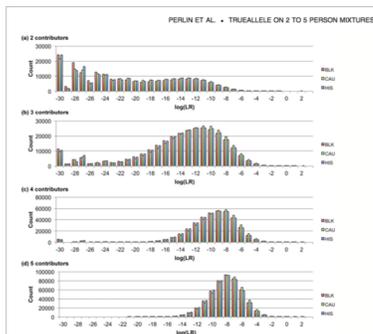
Error for validation genotypes

noncontributor

match strength
log(536,000)
= 5.729 ban

Probability

Match Strength (ban)

For a match strength of 536 thousand, only
1 in 9.65 million people would match as strongly

Monte Carlo sampling for one genotype

noncontributor

Probability

Match Strength (ban)

Gather log(LR) values
one-to-many

Direct convolution for one genotype

5

10

15

Probability
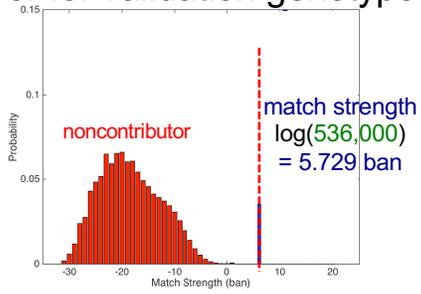
Match Strength (ban)
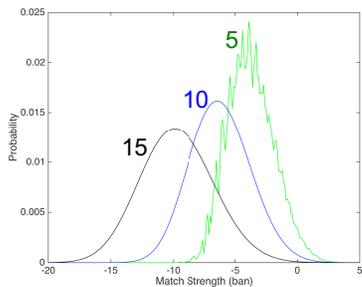
Instant log(LR) non-contributor distribution
one-to-none

Error for evidence genotype

For a match strength of 536 thousand, only
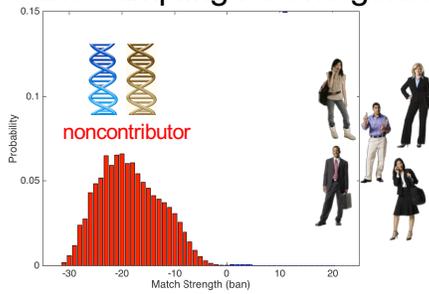1 in 7.32 million people would match as strongly
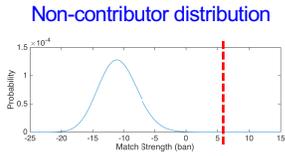
Non-contributor distribution

match strength
log(536,000)
= 5.729 ban

population
probability
1 / 7,320,000

Rule 403 – DNA match relevance

Probative value
of DNA match statistic
without error determined
from the evidence

Danger of
misleading the jury
without an error rate

How often would evidence match the wrong person as strongly as the defendant?

Evidence
information
likelihood ratio
unfamiliar concept

How often
probability
frequency
familiar concept

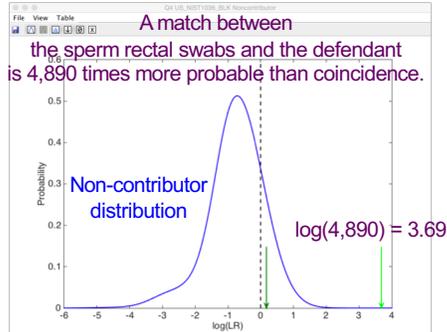## Case example – LR histogram

A match between
the sperm rectal swabs and the defendant
is 4,890 times more probable than coincidence.

Non-contributor
distribution

$\log(4{,}890) = 3.69$

Probability

log(LR)

Case example – match error

Noncontributor Tail Statistics

File   View

item	LR	log(LR)	left tail	L: one in	right tail	R: one in
1 N	1.52	0.1821	8.7151e-01	1.15	1.2849e-01	7.78
2	4.89 thousand	3.6896	1.0000e+00	1	3.4800e-08	28.7 million

Error = 1/28,700,000 << 1/4,890 = 1/LR

For a match strength of 4,890, only
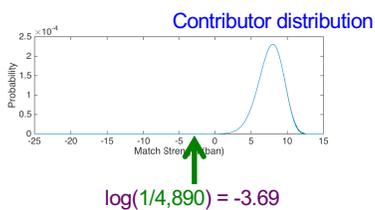1 in 28.7 million people would match as strongly

---

Exclusionary match error

Contributor distribution

Probability

Match Strength (ban)

$\log(1/4{,}890) = -3.69$

For a non-association of 1/4,890, only
1 in 28.7 million people
would be less associated with the evidence

Why are verbal equivalents unnecessary?

LR for (H$_1$) and 1/LR for (H$_2$)	Qualitative Equivalent
1	Uninformative
2 to <100	Limited Support
100 to <10,000	Moderate Support
10,000 to <1,000,000	Strong Support
≥1,000,000	Very Strong Support

• hides the real match strength information
• not what a DNA expert actually believes
• misleads the jury about "million" (Koehler)

Just report LR error, along with the LR,
when the match strength is under a million

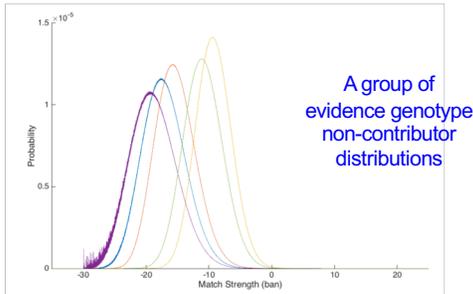## How are error rates from DNA evidence and validation studies similar?

A group of
evidence genotype
non-contributor
distributions

Validation histogram is the average

non-contributor distribution
many-to-none
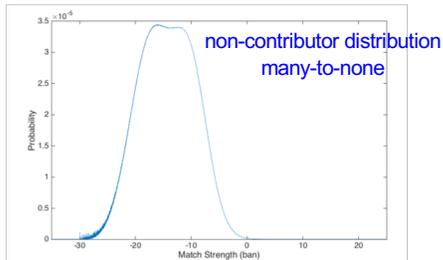
• exact: average the evidence distributions (a second)
• sample: compare evidence vs. random profiles (weeks)

Exact vs. sampled

Exact
  all – 10^{24} genotypes
  Accurate
  exact probability function
  convolution – fast

Sampled
  some – 10^4 genotypes
  Approximate
  sample using random profiles
  Monte Carlo – slow

Heliyon

ELSEVIER

Check for
updates

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

Mark W. Perlin[*]

Cybergenetics, Pittsburgh, PA, USA

[*] Corresponding author.
E-mail address: perlin@cybgen.com (M.W. Perlin).

Daubert requires match error

✓ testable
✓ error rate
✓ peer reviewed
✓ generally accepted

Conclusions

- measuring error is built into genotype probability
- always report the LR; can also report error
- verbal equivalents are not good science
- validation is easy – average the evidence curves
  - no "right" match answer is needed,
    just the evidence genotype distributions

Information theory makes forensics easy
Alternative wastes time, money & information

More information

http://www.cybgen.com/information
- Courses
- Newsletters
- Newsroom
- Presentations
- Publications
- Webinars

http://www.youtube.com/user/TrueAllele
TrueAllele YouTube channel

You Tube

Cybergenetics

perlin@cybgen.com