Explaining the Likelihood Ratio in DNA Mixture Interpretation

21st International Symposium
on Human Identification
October, 2010
San Antonio, TX

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

Cybergenetics

What is the likelihood ratio?

- standard statistical measure of information
- a single number that summarizes
  the support for a simple hypothesis
- accounts for evidence in favor or against
- the match statistic in DNA identification
- forensic science's credibility in court

How the data changes our belief in a hypothesis.

LR is not yet popular in the US

- not available for most forensic disciplines
- DNA analysts find the LR hard to explain

- all DNA match statistics (eg, inclusion) are LRs
- strong LRs preserve DNA match information
- weak LRs discard considerable information
- without LR, DNA misreported as "inconclusive"

Approach: find a better way to explain the LR

History of the LR

- Thomas Bayes (1763)
  how to update our hypotheses based on data
- Alan Turing (1940's)
  updating probability for WWII code breaking
- Jack Good (1950)
  "Probability and The Weighing of Evidence"
  a scientific classic – the modern LR
- Dennis Lindley (1970's)
  "Understanding Uncertainty"
  applying the LR to forensic science
- Buckleton, Evett, Weir, … (1990's)
  interpretation of DNA evidence and mixtures

Hypothesis Form

identification hypothesis:
the suspect contributed to the evidence

$$\text{information gain in hypothesis} = \frac{\text{Odds(hypothesis | data)}}{\text{Odds(hypothesis)}}$$

after ↑ data
before

The evidence increased our belief that
the suspect contributed to the DNA
by a factor of a billion.

Likelihood Form

alternative hypothesis:
someone else contributed to the evidence

$$\text{contrast hypotheses} = \frac{\text{Prob(data | identification hypothesis)}}{\text{Prob(data | alternative hypothesis)}}$$

The probability of observing the evidence
assuming that the suspect contributed to the DNA
is a billion times greater than
the probability of observing the evidence
assuming that someone else was the contributor.

Genotype Form

At the suspect's genotype,
what is the genotype *information gain*?

$$\text{information gain in genotype} = \frac{\text{Prob(evidence genotype)}}{\text{Prob(coincidental genotype)}}$$

(evidence) after ↑ data before (coincidence)

At the suspect's genotype,
the evidence genotype
is a billion times more probable than
a coincidental genotype.

Match Form

How much more does
the suspect match the evidence
than some random person?

$$\text{information gain in DNA match} = \frac{\text{Prob(evidence match)}}{\text{Prob(coincidental match)}}$$

A match between the suspect and the evidence
is a billion times more probable than
a coincidental match.

Mixture Interpretation

$$\text{LR} = \frac{\text{Prob(evidence match)}}{\text{Prob(coincidental match)}}$$

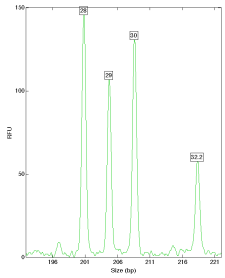
Different methods yield different DNA information

$$\text{random match} = \frac{1}{\text{Prob(coincidental match)}}$$

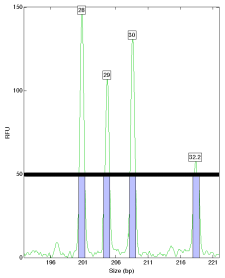
$$\text{inclusion} = \frac{\text{small matching genotype probability}}{\text{Prob(coincidental match)}}$$

$$\text{quantitative} = \frac{\text{large matching genotype probability}}{\text{Prob(coincidental match)}}$$

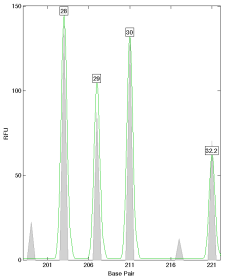
Quantitative Data



Qualitative Threshold



Quantitative Interpretation



TrueAllele® Casework

- quantitative computer interpretation
- statistical search of probability model
- preserves all identification information
- objectively infer genotype, then match

- any number of mixture contributors
- stutter, imbalance, degraded DNA
- calculates uncertainty of every peak

- created in 1999, now in version 25
- used on 100,000 evidence samples
- available as product, service or both

Commonwealth v. Foley

ScoreMethod
13 thousandinclusion
23 million obligate allele
189 billionTrueAllele

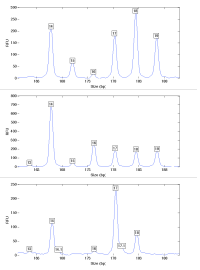
Assume that there are two contributors to the DNA mixture, including the known victim.
A match between Mr. Foley and the fingernails is 189 billion times more probable than a coincidental match to an unrelated Caucasian.

Regina v. Broughton

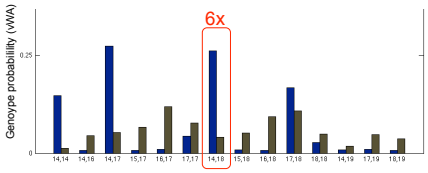
- low template mixture
- three DNA contributors
- triplicate amplification
- post-PCR enhancement

- no match score found
- TrueAllele interpretation

vWA locus data



Regina v. Broughton



Assume coancestry at a theta value of 1%, and that there are three contributors present. A match between Mr. Broughton and the fuse is 3,620,000 times more probable than a coincidental match to an unrelated Caucasian.

Explaining the Likelihood Ratio

LR Methods Vary

National Institute of Standards and Technology
Two Contributor Mixture Data, Known Victim

Some Differences in Reporting Statistics

		Case1		
LabID	Kits Used	Caucasians	African-Americans	Hispanics
90	ProPlus/Cofiler	1.18E+15	2.13E+14	3.09E+15
34	ProPlus/Cofiler	2.40E+11	7.88E+08	9.80E+10
33	ProPlus/Cofiler	2.94E+08	1.12E+08	1.74E+09
6	ProPlus/Cofiler	40,000,000	3,500,000	280,000,000
9	ProPlus/Cofiler	1.14E+07	1.97E+07	1.54E+08
79	ProPlus/Cofiler	930,000	47,366	1,350,000
16	ProPlus/Cofiler	434,600	31,710	399,100

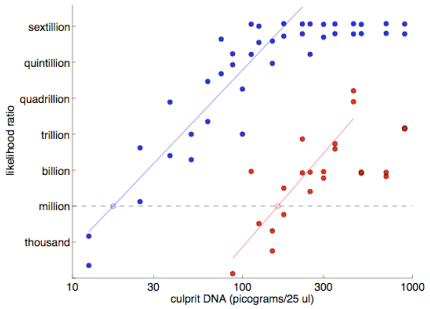
Remember that these labs are interpreting the same MIX05 electropherograms

213 trillion (14)

31 thousand (4)

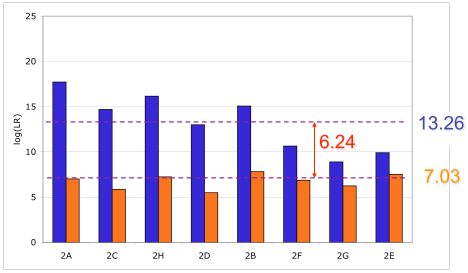
MW Perlin, A Sinelnikov. An information gap in DNA evidence interpretation. PLoS ONE, 2009.

Preserve vs. Discard

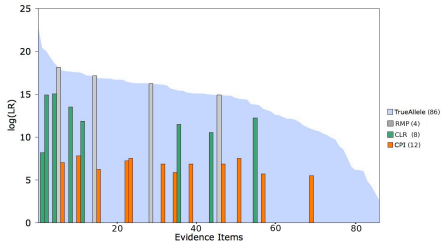


MW Perlin, MM Legler, CE Spencer, JL Smith, WP Allan, JL Belrose, BW Duceman. Validating Trueallele DNA Mixture Interpretation. Journal of Forensic Sciences, 2011.

Preserve vs. Discard



Preserve vs. Discard



Investigative DNA Databases

- "allele" approach discards information
- store & match probabilistic genotypes
- LR preserves identification information

- evidence vs. convicted offender
- disaster victim identification (WTC)
- finding missing people
- automated familial search
- customizable to states and countries

SWGDAM 2010 – Mixtures

3.2.2. If a stochastic threshold based on peak height is *not used* in the evaluation of DNA typing results, the laboratory must establish alternative criteria (e.g., quantitation values or use of a probabilistic genotype approach) for addressing potential stochastic amplification. The criteria must be supported by *empirical data and internal validation* and must be documented in the standard operating procedures.

- higher peak threshold discards information
- probability modeling preserves information

All DNA mixture methods report their match statistics as *likelihood ratios*.

Forensic Science: DNA Mission

- *preserve* DNA identification information
- provide *accurate* DNA match results
- *serve* the criminal justice system
- help *protect* the public from crime

A match between the suspect and the evidence is a billion times more probable than a coincidental match.

Conclusions

- the LR is easy to understand
- and is easy to explain in court
- need to state the appropriate form
- *numerator*: evidential match
- *denominator*: coincidental match
- *ratio*: information preserved

Cybergenetics

http://www.cybgen.com/information/presentations.shtml
perlin@cybgen.com