Promega

21st International Symposium on Human Identification

Mixture Interpretation Workshop:
Principles, Protocols, and Practice

October 11, 2010 – San Antonio, TX

NIJ National Institute of Justice

Statistical Approaches (CPI, LR, RMP)

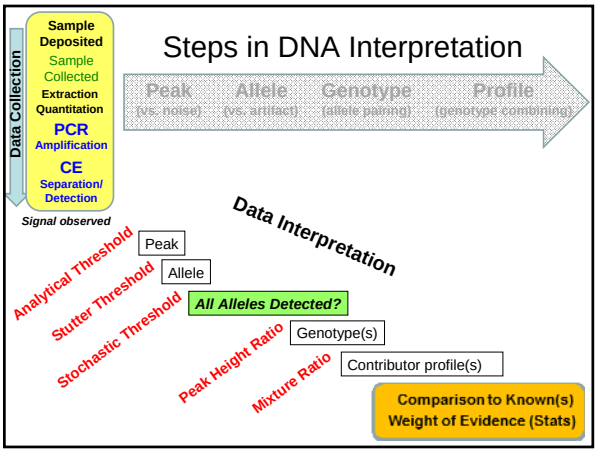
Michael D. Coble

BOSTON UNIVERSITY

NIST

Outline for **Statistical Approaches**

- **GUIDELINES**
  - **SWGDAM Guidelines Section 4 – Statistical Analysis.**
  - **SWGDAM Guidelines Section 5 – Statistical Formulae.**
  - **Table 1.**
- **PRINCIPLES**
  - **Developing the framework to perform the appropriate statistical analyses (RMP, LR, CPI) given the mixture profile.**
- **PROTOCOLS**
  - **Documentation of the procedures.**
- **PRACTICE**
  - **Worked Example.**



GUIDELINES

Statistical Analysis of DNA Typing Results

- 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.

GUIDELINES

Comparison of DNA Typing Results

- 3.6.1. The laboratory must establish guidelines to ensure that, to the extent possible, *DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors.*

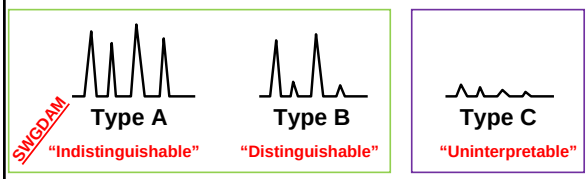
GUIDELINES

Comparison of DNA Typing Results

- 3.6.2. DNA typing results may not be obtained at all loci for a given evidentiary sample (e.g., due to DNA degradation, inhibition of amplification and/or low-template quantity); a partial profile thus results.
- 3.6.2.1. For partial profiles, the determination of which alleles/loci are suitable for comparison and statistical analysis ***should be made prior to comparison to the known profiles***.

PRINCIPLES

Schneider *et al.* (2009) and SWGDAM



A statistical analysis must be performed

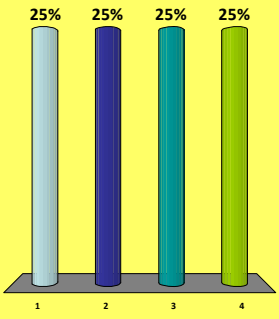
A statistical analysis should not be performed

Not all mixtures are homogeneous for Types A, B and C

e.g. Predominantly "A" with some "B" and "C" loci

Do you classify mixtures using a scheme like the German Stain Commission?

1. Yes
2. No
3. Sometimes for difficult or low level samples.
4. I use CPE or CPI stats, so a mixture is a mixture to me.



GUIDELINES

Comparison of DNA Typing Results

SWGDAM Guideline 3.6 (Intro):

- The following determinations can be made upon comparison of evidentiary and known DNA typing results (and between evidentiary samples):
- The known individual cannot be excluded (i.e., is included) as a possible contributor to the DNA obtained from an evidentiary item.
- The known individual is excluded as a possible contributor.
- The DNA typing results are inconclusive/uninterpretable.
- The DNA typing results from multiple evidentiary items are consistent or inconsistent with originating from a common source(s).

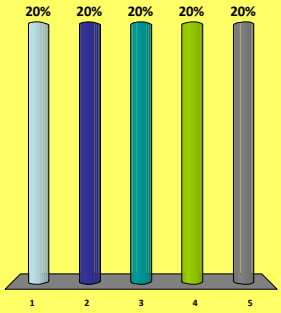
GUIDELINES

Statistical Analysis of DNA Typing Results

- 4.1. The laboratory must perform statistical analysis in support of any inclusion.
- 4.2. For calculating the CPE or RMP, any DNA typing results used for statistical analysis ***must be*** derived from **evidentiary items** and not known samples.
- 4.3. The laboratory must not use inconclusive/uninterpretable data (e.g., at individual loci or an entire multi-locus profile) in statistical analysis.

Do you interpret your evidence (lock down your inferred genotypes) independent of your alleged contributor?

1. Always
2. Most of the time
3. Sometimes
4. Rarely
5. Never



GUIDELINES

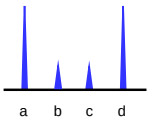
Statistical Analysis of DNA Typing Results

- 4.4. Exclusionary conclusions do not require statistical analysis.
- 4.5. The laboratory must document the source of the population database(s) used in any statistical analysis.

PRINCIPLES

Statistical Approaches with Mixtures

- **Random Match Probability (RMP)** – 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.



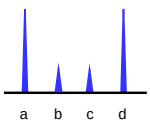
$RMP_{major} = 2pq$

$= 2 \times f(a) \times f(d)$

PRINCIPLES

Statistical Approaches with Mixtures

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



$p = f(a) + f(b) + f(c) + f(d)$
$q = 1 - p$

$PE = 2pq + q^2$

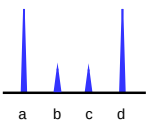
$CPE = PE_{M1} \times PE_{M2} \cdots$

$CPI = 1 - CPE$

PRINCIPLES

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



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

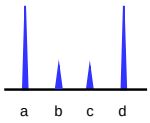
E = Evidence
H_1 = Prosecutor's Hypothesis (the suspect did it) = 1
H_2 = Defense Hypothesis (the suspect is an unknown, random person)

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

PRINCIPLES

Statistical Approaches with Mixtures

- **Unrestricted Likelihood Ratio** - All combinations of alleles are deemed possible (relative peak height differences are not utilized).



Possible Combinations

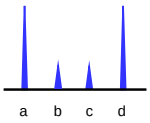
AB BC
AC BD
AD CD

= (AB + AC + AD + BC + BD + CD)

PRINCIPLES

Statistical Approaches with Mixtures

- **Restricted Likelihood Ratio** - Based on relative peak heights, alleles are paired only where specific combinations of alleles are deemed possible



(without victim subtraction)

Possible Combinations

~~AB~~ BC
~~AC~~ ~~BD~~
AD ~~CD~~

= (AD + BC)
2pq + 2pq

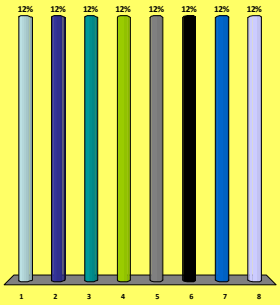
PRINCIPLES

Summary

Statistical Approach	Contributor Assumptions?	Drop-out?	Major/Minor Separation
RMNE (CPI, CPE)	NO	NO	NO
Random Match Prob.	YES	NO	YES
Restricted LR	YES	YES	NO
Unrestricted LR	YES	YES	YES

What kind of mixture statistic(s) does your lab use?

1. LR
2. CPE (RMNE, CPI)
3. RMP
4. CPE or RMP
5. Other combinations
6. Probabilistic modeling (e.g., TrueAllele)
7. We don't use stats (contradicting the new guidelines – section 4.1)
8. Don't get all mathy on me.



PRINCIPLES

Advantages and Disadvantages RMNE and LR

RMNE (CPE/CPI)

Advantages
- Does not require an assumption of the number of contributors to a mixture
- Easier to explain in court

Disadvantages
- Weaker use of the available information (robs the evidence of its true probative power because this approach does not consider the suspect's genotype)
- LR approaches are developed within a consistent logical framework

Likelihood Ratios (LR)

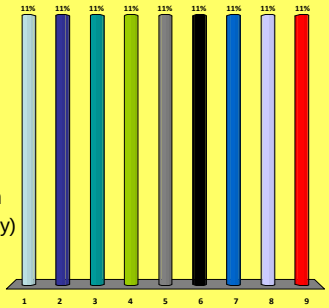
Advantages
- Enables full use of the data including different suspects

Disadvantages
- More difficult to calculate (software programs can assist)
- More difficult to present in court

Summarized from John Buckleton, *Forensic DNA Evidence Interpretation*, p. 223
Buckleton and Curran (2008) *FSI-G* 343-348.

Does your lab use any software to help calculate mixture stats?

1. PopStats
2. GMID-X
3. FSS I^3
4. GeneMarker HID
5. True Allele
6. DNA·View
7. In-house Excel program
8. On a calculator (painfully)
9. Other



PRINCIPLES

Assumptions for CPE/CPI Approach

- **There is no allele dropout** (i.e., all alleles are above stochastic threshold) – low-level mixtures can not reliably be treated with CPE
- All contributors are from the same racial group (i.e., you use the same allele frequencies for the calculations)
- All contributors are unrelated
- Peak height differences between various components are irrelevant (i.e., **component deconvolution not needed**) – this may not convey all information from the available sample data…

GUIDELINES

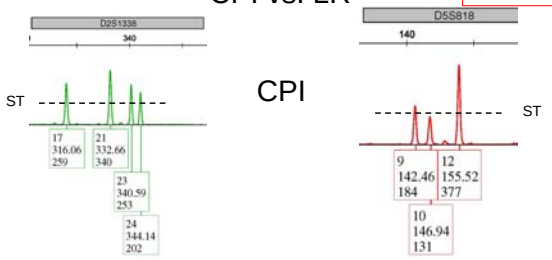
Statistical Analysis of DNA Typing Results

- 4.6.3. When using CPE/CPI (with no assumptions of number of contributors) to calculate the probability that a randomly selected person would be excluded/included as a contributor to the mixture, loci with *alleles below the stochastic threshold may not be used for statistical purposes to support an inclusion.* In these instances, the potential for allelic dropout raises the possibility of contributors having genotypes not encompassed by the interpreted alleles.

PRACTICE

CPI vs. LR

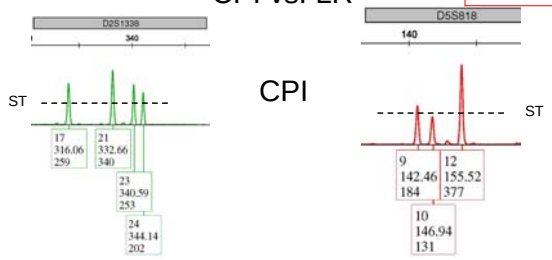
Vaginal Swab
ST = 150 RFU



PRACTICE

CPI vs. LR

Vaginal Swab
ST = 150 RFU



CPI = 17,17 + 17,21 + 17,23 + 17,24 + 21,21 + 21,23 + 21,24 + 23,23 + 23,24 + 24,24
CPI = (freq17 + freq21 + freq23 + freq24)^2
CPI = (0.1941 + 0.0197 + 0.1349 + 0.1217)^2
CPI = 0.2213 or 22.13%

CPI = 9,9 + 9,10 + 9,12 + 12,12 + 10,12 + 10,13 + 10,14 + 10,N …

CPI = (freq9 + freq10 + freq12 + Freq8 + freq11 +….)^2 = 1.0

PRINCIPLES

Curran and Buckleton (2010)

J Forensic Sci, September 2010, Vol. 55, No. 5
doi: 10.1111/j.1556-4029.2010.01446.x
Available online at: interscience.wiley.com

PAPER
CRIMINALISTICS; GENERAL

James M. Curran,[1] M.Sc.(Hons.), Ph.D. and John Buckleton,[2] Ph.D.

Inclusion Probabilities and Dropout

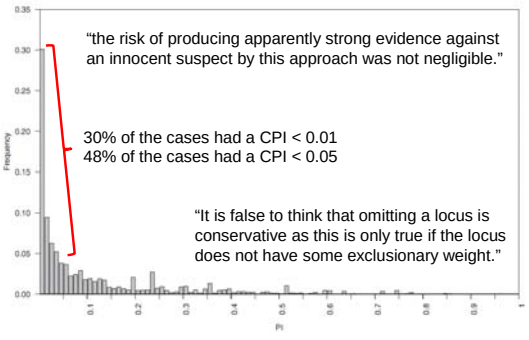
Created 1000 Two-person Mixtures (Budowle *et al.*1999 AfAm freq.).

Created 10,000 "third person" genotypes.

Compared "third person" to mixture data, calculated PI for included loci, ignored discordant alleles.

PRINCIPLES

Curran and Buckleton (2010)



"the risk of producing apparently strong evidence against an innocent suspect by this approach was not negligible."

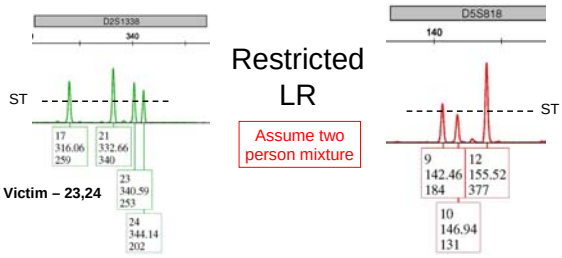
30% of the cases had a CPI < 0.01
48% of the cases had a CPI < 0.05

"It is false to think that omitting a locus is conservative as this is only true if the locus does not have some exclusionary weight."

PRACTICE UPDATED

CPI vs. LR

Vaginal Swab
ST = 150 RFU



Restricted LR

Assume two person mixture

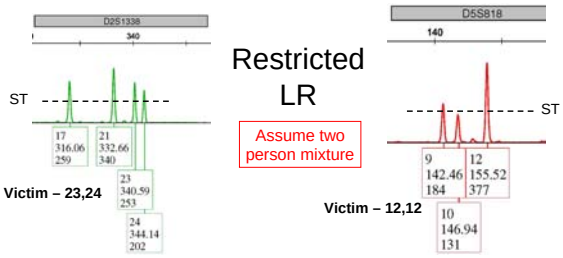
Victim – 23,24

$$\frac{P(E|H_1)}{P(E|H_2)} = \frac{V+S}{V+U} = \frac{2(f_{23})(f_{24}) + 1}{2(f_{23})(f_{24}) + 2(f_{17})(f_{21})}$$

PRACTICE UPDATED

CPI vs. LR

Vaginal Swab
ST = 150 RFU



Restricted LR

Assume two person mixture

Victim – 23,24

LR = 1/2(f_{17})(f_{21})

LR = 1/2(0.1941)(0.0197) = **130.76**

Victim – 12,12

LR = 1/2(f_9)(f_{10})

LR = 1/2(0.0308)(0.0487) = **333.34**

Combined LR = 43,587

PRINCIPLES

Suitable Statistical Analyses

Table 1 – Suitable Statistical Analyses for DNA Typing Results
The statistical methods listed in the table cannot be combined into one calculation. For example, combining RMP at one locus with a CPI calculation at a second locus is not appropriate. However, an RMP may be calculated for the major component of a mixture and a CPE/CPI for the entire mixture (as referred to in section 4.6.2).

Category of DNA Typing Result	RMP	CPE/CPI	LR (1)
Single Source	✔		✔
Single Major Contributor to a Mixture	✔		✔
Multiple Major Contributors to a Mixture	✔ (2)	✔ (2)	✔
Single Minor Contributor to a Mixture	✔	✔ (3)	✔
Multiple Minor Contributors to a Mixture	✔ (2)	✔ (3)	✔
Indistinguishable Mixture	✔ (1)	✔	✔

(1) Restricted or unrestricted
(2) Restricted
(3) All potential alleles identified during interpretation are included in the statistical calculation

PROTOCOLS

Summary

- The laboratory must perform statistical analysis in support of any inclusion (4.1).
- DNA typing results from evidentiary samples are interpreted before comparison with any known samples, other than those of assumed contributors (3.6.1).
- There are advantages and disadvantages to both RMNE and LR stats. As a general rule, RMNE does not take full advantage of all the data.
- Statistical methods cannot be combined into one calculation (e.g. combining RMP at one locus with a CPI calculation at a second locus is not appropriate).

I BRAKE FOR STATS

STUTTER HAPPENS!

Acknowledgments

- NIJ Funding to our NIST Group through NIST OLES interagency agreement 2008-DN-R-121
- Bruce Heidebrecht (Maryland State Police) for examples.
- Todd Bille (ATF) for helpful discussions.

michael.coble@nist.gov
301-975-4330