Overcoming Bias in DNA Mixture Interpretation

American Academy of Forensic Sciences
February, 2016
Las Vegas, NV

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

American Academy of Forensic Sciences

Las Vegas, NV

February, 2016

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

DNA

Gold standard of forensic evidence

However, there may be problems with how the DNA was interpreted, such as when there are mixed samples

Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation

WILLIAM C. THOMPSON
Department of Criminology, Law and Society
University of California, Irvine, CA 92697, USA

[Received on 17 November 2006; revised on 13 April 2009; accepted on 24 April 2009]

Forensic DNA analysis tends to underestimate the frequency of matching profiles and overestimate likelihood ratios by shifting the probability of a hit from the ‘null’ or ‘innocent’ after the profile of a suspect becomes known—a process analogous to the well-known Texas sharpshooter fallacy. Using examples from casework, artificial and naturalistic experiments, and analyses of DNA databases that demonstrate how poor forensic target shifting occurs and how it can distort the frequency and likelihood ratio statistics used to characterize DNA matches, existing matches appear more probative than they actually are. It concludes by calling for broader adoption of more rigorous analytical procedures, such as sequential omnibus tests, that can reduce the sharpshooter fallacy by fixing the target before the data are seen.
Case context impact

<table>
<thead>
<tr>
<th>With context</th>
<th>Without context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Include</td>
<td>2</td>
</tr>
<tr>
<td>Exclude</td>
<td>12</td>
</tr>
<tr>
<td>Inconclusive</td>
<td>4</td>
</tr>
</tbody>
</table>
**DNA mixture**

Genotype 1  Genotype 2  Data

10, 12  +  11, 12

(oversimplified cartoon diagram)

**Interpret #1: separate**

Data  Genotype 1  Genotype 2

10, 10 @ 10%  10, 10 @ 10%
10, 11 @ 20%  10, 11 @ 10%
10, 12 @ 40%  10, 12 @ 10%
11, 11 @ 10%  11, 11 @ 10%
11, 12 @ 20%  11, 12 @ 40%
12, 12 @ 20%  12, 12 @ 10%

Unmix the mixture

**Interpret #2: compare**

Data  Genotype 2

10, 10 @ 10%
10, 11 @ 10%
10, 12 @ 10%
11, 11 @ 10%
11, 12 @ 40%
12, 12 @ 20%

Match statistic = \[ \frac{\text{Prob(match)}}{\text{Prob(coincidence)}} \] = 40% / 4% = 10

Compare with 11,12
Cognitive bias

Illogical thinking affects decisions

- Anchoring – rely on first information
- Apophenia – perceive meaningful patterns
- Attribution bias – find causal explanations
- Confirmation bias – interpretation confirms belief
- Framing – social construction of reality
- Halo effect – sentiments affect evaluation
- Oversimplification – simplicity trumps accuracy
- Self-serving bias – distort to maintain self-esteem

Contextual bias

Background information affects decisions

- Academic bias – beliefs shape research
- Educational bias – whitewash damaging evidence
- Experimenter bias – expectations affect outcomes
- Inductive bias – tilt toward training examples
- Media bias – selecting mass media stories
- Motivational bias – reaching desired outcome
- Reporting bias – under-report undesirable results
- Social desirability bias – want to be seen positively

Discard evidence
### Genotype bias

<table>
<thead>
<tr>
<th>Actual</th>
<th>Desired</th>
</tr>
</thead>
<tbody>
<tr>
<td>10, 10</td>
<td>10, 10</td>
</tr>
<tr>
<td>10, 11</td>
<td>10, 11</td>
</tr>
<tr>
<td>10, 12</td>
<td>10, 12</td>
</tr>
<tr>
<td>11, 11</td>
<td>11, 11</td>
</tr>
<tr>
<td>11, 12</td>
<td>11, 12</td>
</tr>
<tr>
<td>12, 12</td>
<td>12, 12</td>
</tr>
</tbody>
</table>

- **RMP** – random match probability
  - analyst chooses only one genotype
  - inflates DNA match statistic

### Match bias

- **CPI** – combined probability of inclusion
  - analyst begins by including the suspect
  - unrealistic, unproven model
  - random number generator
  - lacks probative value

- **LR** – likelihood ratio
  - analyst ignores much of the data
  - calculation requires suspect genotype
  - introduces “phantom” peaks (drop out)
  - considers few genotype possibilities

### Process bias

1. Choose, alter, discard, edit, and manipulate the DNA data signals
2. Compare defendant's genotype to edited data & decide if he is in the DNA evidence
3. If he is “included”, then calculate a DNA mixture statistic

**Hidden** cognitive and contextual bias largely determine the outcome

**Presented** as unbiased science

---

Software bias
Why labs choose mixture software

- Puts analyst in charge
- Results confirm belief
- Simplifies the problem
- Gets desired answer
- The FBI uses it
- Familiar process

Confirmation bias
Oversimplification
Motivational bias
Social desirability bias

Relevance (FRE 403)
Admissibility of biased DNA evidence

Rule 401
“evidence makes a fact more or less probable

Probative value inflated

Rule 403
“substantially outweighed by a danger of:

Unfair prejudice
Confusing the issues
Misleading the jury
Wasting time
Cumulative evidence inadmissible

Cross examination
Hundreds of effective questions can elicit bias

“Did you know the defendant's genotype during your analysis of the evidence?”

“Doesn't knowing your customer's desired answer bias your decisions?”

“Have any scientific studies demonstrated otherwise?”
Sequential unmasking

Human DNA review proposal (reduce bias):
1. First analyze the crime scene data, without knowing context or references
2. Then compare with reference samples

But there is potential bias in choosing data, conducting analysis, and making comparisons.

Human analysts can always introduce bias. Why is a human even involved in this process? Why not use an unbiased computer instead?

Unbiased interpretation

Use an objective computer to:
1. Examine all DNA data, without having suspect's genotype
2. Separate genotypes of each DNA mixture contributor, considering all possible solutions
3. Compare genotypes only afterwards to calculate match statistics

Eliminate all human involvement to overcome cognitive & contextual bias in DNA mixture interpretation

No data bias – use all evidence

- Learn stutter from the evidence
- No peak choice
- No thresholds
- No locus choice
- Model variation from the evidence
- Use all loci in the evidence
No genotype bias – objective

<table>
<thead>
<tr>
<th>Actual</th>
<th>Desired</th>
</tr>
</thead>
<tbody>
<tr>
<td>10, 10 @ 5%</td>
<td>Use the actual genotype probability</td>
</tr>
<tr>
<td>10, 11 @ 5%</td>
<td></td>
</tr>
<tr>
<td>10, 12 @ 75%</td>
<td></td>
</tr>
<tr>
<td>11, 11 @ 5%</td>
<td></td>
</tr>
<tr>
<td>11, 12 @ 5%</td>
<td></td>
</tr>
<tr>
<td>12, 12 @ 5%</td>
<td></td>
</tr>
</tbody>
</table>

Do not change probability

No match bias – accurate

CPI – combined probability of inclusion
- random number generator
- bad forensic science
- review all past cases

LR – likelihood ratio
- don’t ignore any data
- don’t use suspect genotype
- don’t concoct “phantom” peaks
- use all genotype possibilities

No process bias – remove analyst

1. Do not change data signals
2. Do not use defendant genotype
3. Calculate accurate DNA match statistic

Eliminate cognitive and contextual bias from the process
Present unbiased science
No software bias – true stats

Accurate, objective, thorough, validated

- Puts analyst in charge
- Results confirm belief
- Simplifies the problem
- Gets desired answer
- The FBI uses it
- Familiar process

Examine all the data without human choice
Separate genotypes consider all solutions
Compare genotypes stats decide outcome

TrueAllele® information

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

http://www.youtube.com/user/TrueAllele
TrueAllele YouTube channel
perlin@cybgen.com