



Netherlands Forensic Institute  
*Ministry of Justice*

## The use of databases in forensic casework

Prof. dr. Marjan Sjerps

- Netherlands Forensic Institute
- University of Amsterdam

5 May 2016



# Netherlands Forensic Institute





## Statisticians NFI





# WHY DO WE NEED STATISTICS?



# Searching strategy





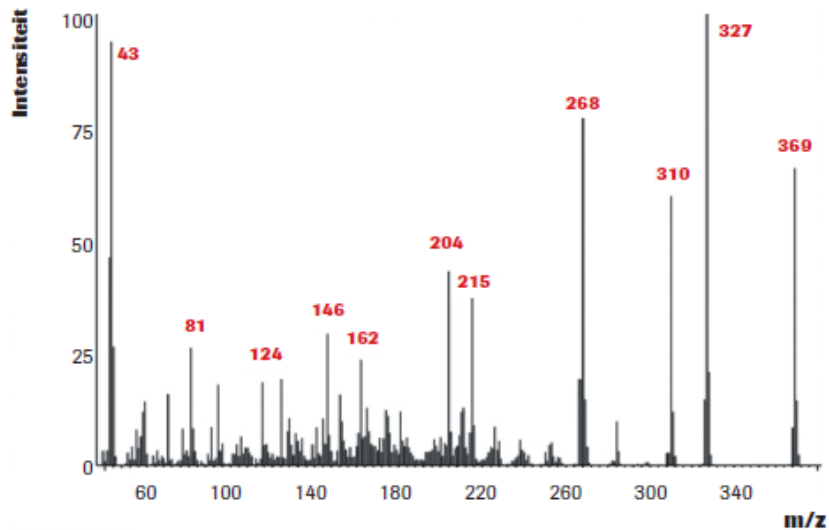


# Analysing





# Interpreting



**Figuur 4.** 70eV massaspectrum van de stof met een retentietijd van 5,50 min; op basis van tijd en spectrum geïdentificeerd als heroïne.





# Reporting







# Statistics in forensic science: it's everywhere!

## Forensic process

1. Strategy

2. Analyse

3. Conclude

4. Report

## Statistical aspects

- Sampling strategy
- Design of experiments

- Statistical process control
- Measurement uncertainty

- Evidence evaluation

- Communicating probabilistic reasoning



# THE BAYESIAN FRAMEWORK



## Example: forensic medicin

Hypothesis 1: self infliction (auto mutilation)

Hypothesis 2: attack

Evidence:

- injuries at anatomically reachable locations
- Grouped, superficial scratches in a parallel position that together form a uniform injury pattern
- One scar of comparable injury





## The Likelihood Ratio (LR)

Probability to observe this evidence, if hypothesis 1 is true  
Probability to observe this evidence, if hypothesis 2 is true

- Numerator: how well do the observations fit with self infliction?
- Denominator: how well do the observations fit with attack?

It is more probable to observe this evidence if the girl hurt herself than if she was attacked by someone else





## What is the probability that the girl was attacked?

You need to know two things:

1. The probability of observing the evidence under the two hypotheses (LR)
2. The **other information** in the case
  - How often does auto mutilation occur?
  - How often does this attack type occur?
  - What else is known?
    - Situation A: Friend testifies she witnessed girl hurting herself
    - Situation B: The subsequent three weeks three other girls, totally independent of each other, report to have been attacked in the same way







## Bayes rule

$$\text{Prior odds} \times \text{LR} = \text{posterior odds}$$





## Verbal statements expressing LR

“It is **more probable** to observe this evidence if the girl hurt herself than if she was attacked by someone else”

| conclusion                     | LR               |
|--------------------------------|------------------|
| Approximately equally probable | 1-2              |
| Slightly more probable         | 2-10             |
| More probable                  | 10-100           |
| Appreciably more probable      | 100-10,000       |
| Far more probable              | 10,000-1 million |
| Extremely probable             | > 1 million      |



Expert *assessment* based on:

- Expert knowledge (education, research, literature, experience)
- Limited data

- Verbal LR, or
- Numerical LR + basis

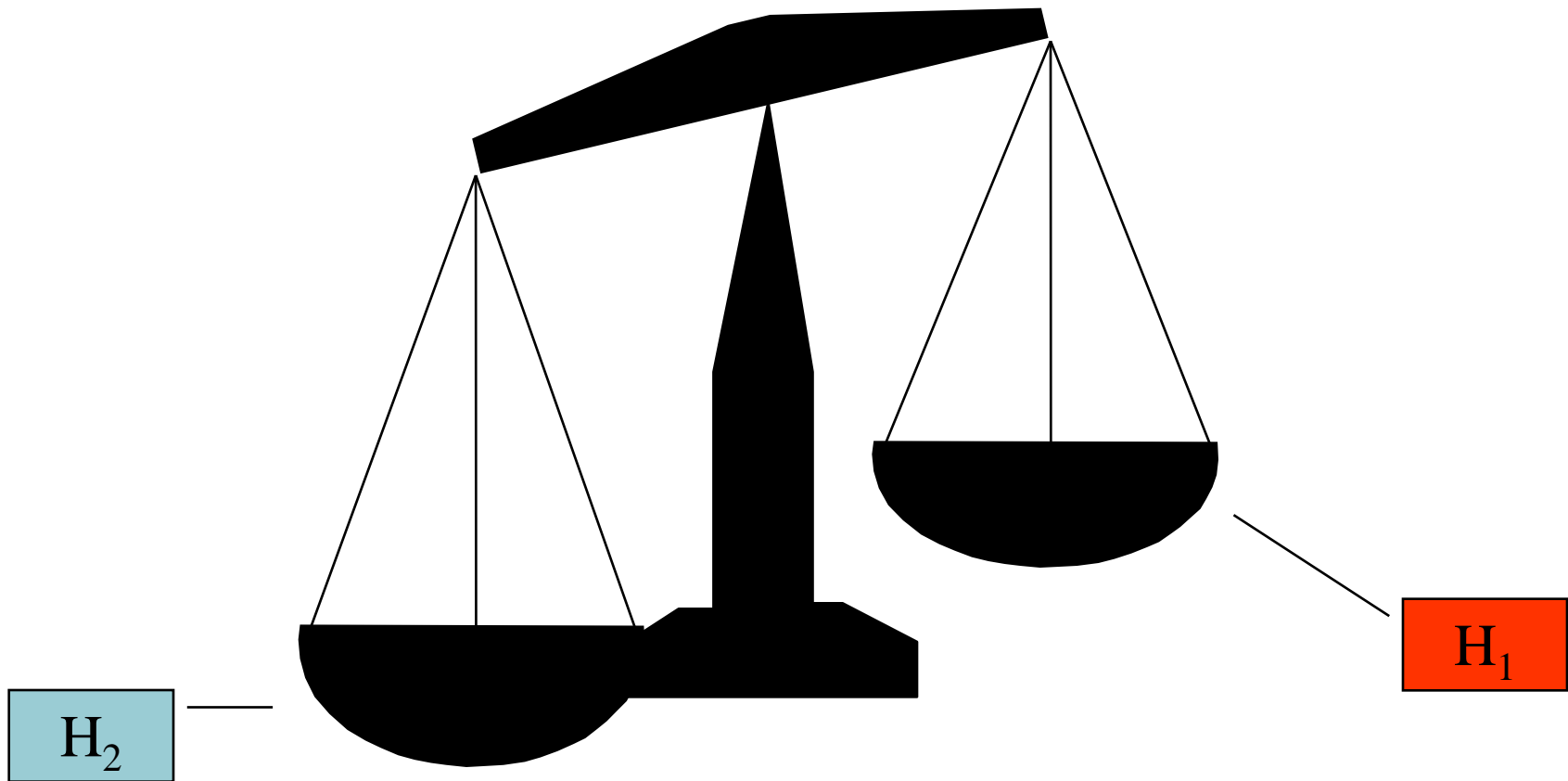
Expert *calculation* based on:

- Expert knowledge (education, research, literature, experience)
- Good data and model

- Numerical LR

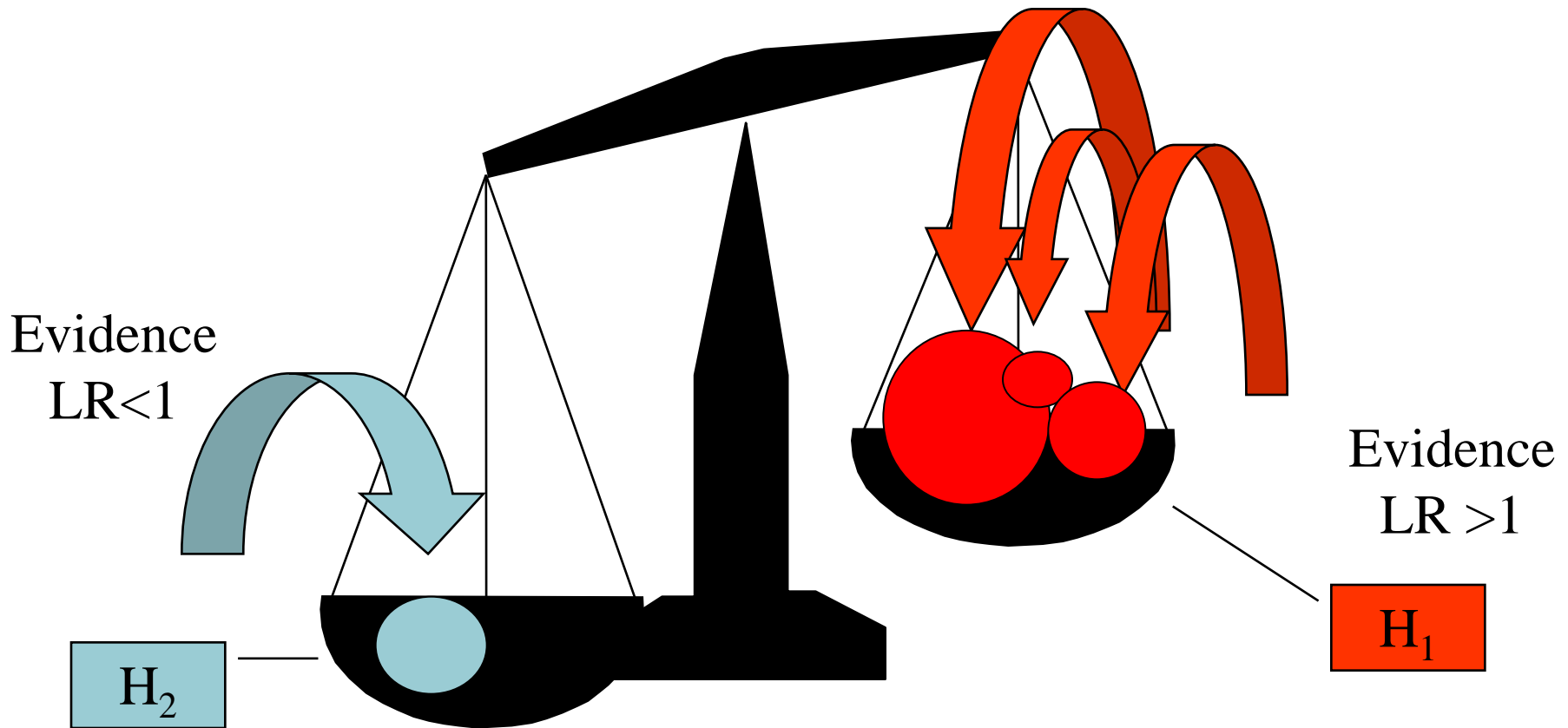


Start





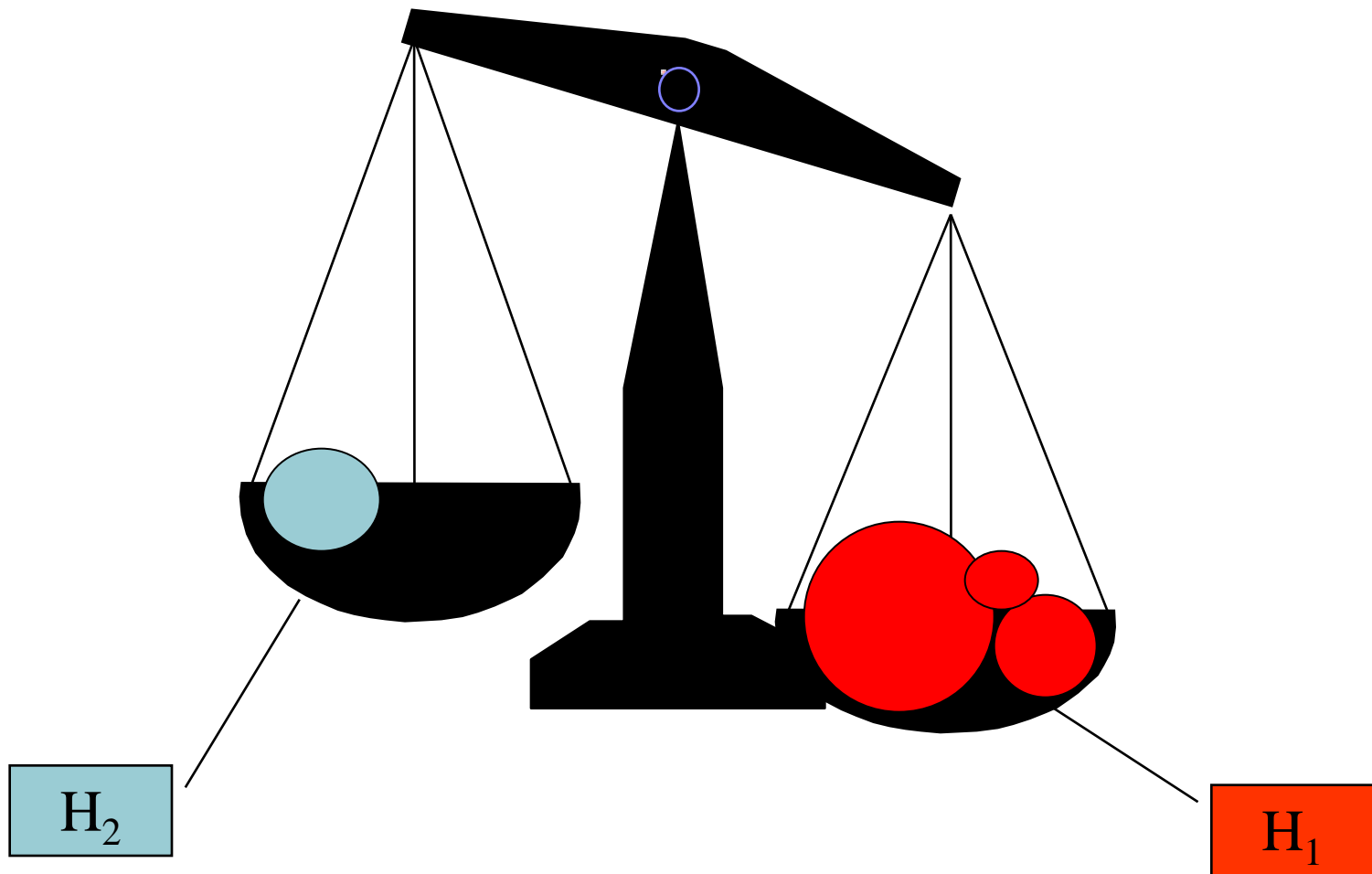
# Evidence changes probabilities of hypotheses





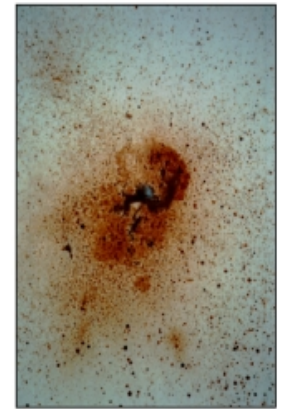


End





## Example: DNA match



What is the evidential value of the DNA match?



## Example: DNA match

Hypothesis 1: DNA in sample is from John

Hypothesis 2: DNA in sample is from someone else

Evidence:

- Match between DNA profiles of sample and of John



## The Likelihood Ratio (LR)

Probability to observe the match, if DNA is from John

Probability to observe the match, if DNA is from someone else

- Numerator: 1
- Denominator: "random match probability", e.g. 1 in 1 billion

$$LR = 1/rmp$$

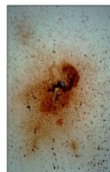
Rmp = 1 in 1 billion





## Bayes rule

$$\text{Prior odds} \times \text{LR} = \text{posterior odds}$$



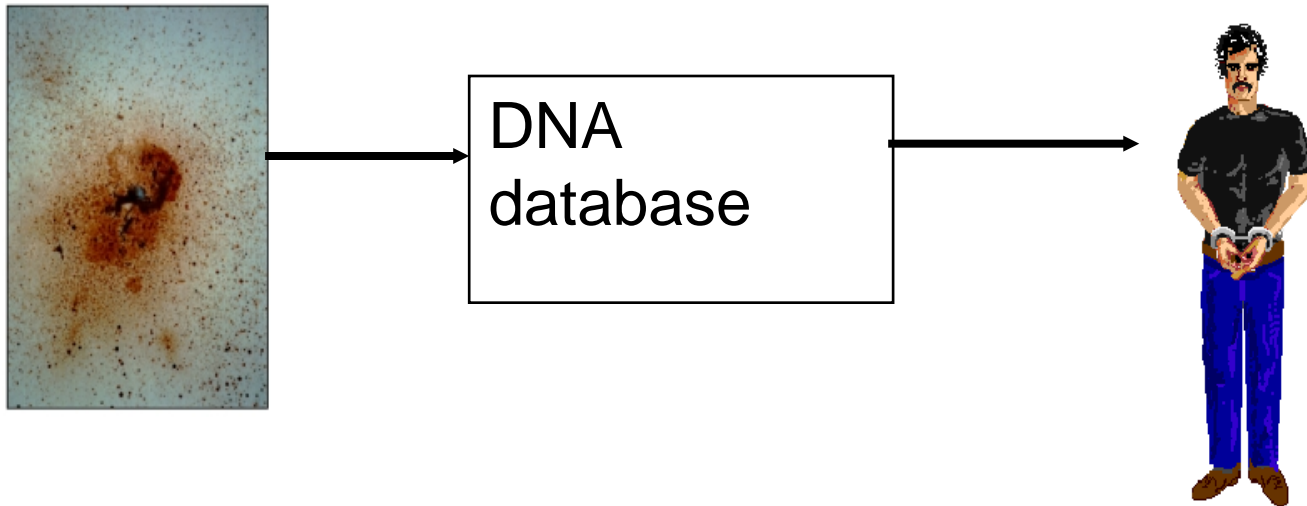




# DATABASE SEARCH



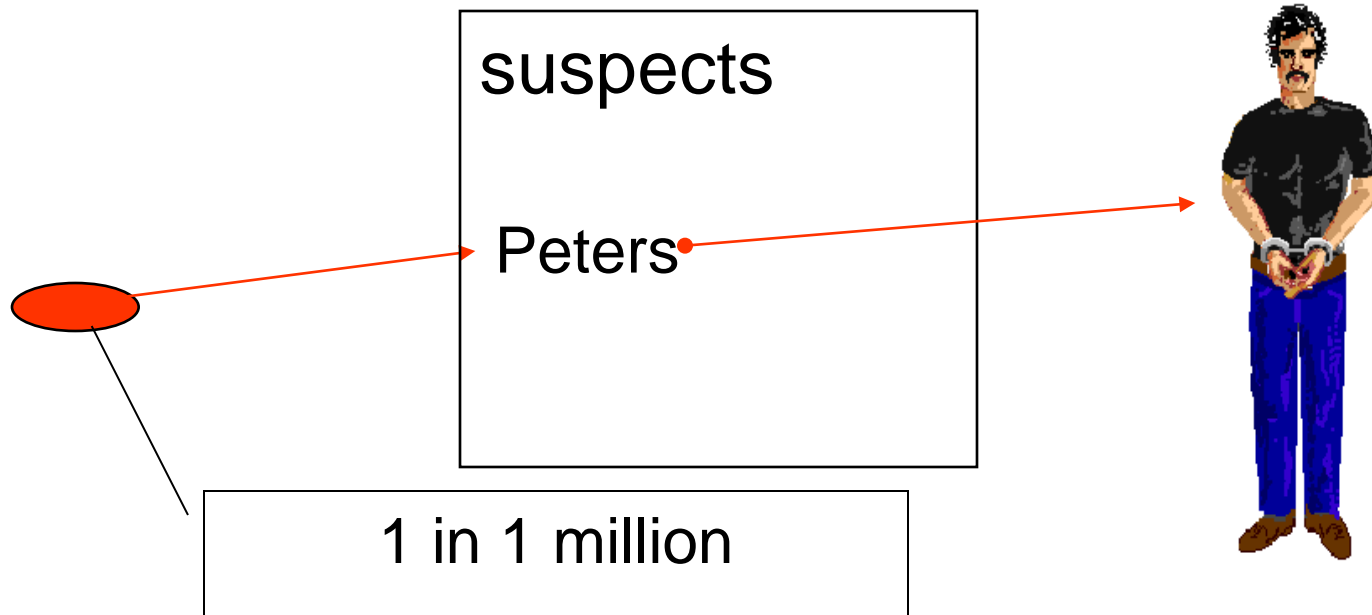
## database search



What is the effect of the database search on the evidential value?



# Match in UK database (5 million persons)





# Heuristic arguments

The more people in the database, the more people are excluded as potential donors, so the stronger the evidence.



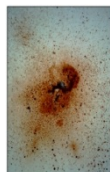
The more people in the database, the larger the probability of a fortitious match, so the weaker the evidence.





## Suspect identified by "probable cause"

$$\text{Prior odds} \times \text{LR} = \text{posterior odds}$$

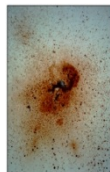






Suspect identified by "database search"

$$\text{Prior odds} \times \text{LR} = \text{posterior odds}$$





# Forensic heaven or hell?

Partial DNA profile  
/ mixture and no  
other evidence

Full DNA profile  
and other  
evidence

Fortitious hit certainly  
not excluded

Fortitious hit nearly  
excluded



## Database matches: bottom line

- Database search just excludes a lot of suspects; **evidential value of DNA does not decrease due to search**
- However, **other evidence may be completely missing**, so prior odds may be extremely small
- Lay people may not realise the effect: Expert must warn



## ENFSI DNA working group 2016: text box

### **POINT OF ATTENTION WITH REGARDS TO A DNA-DATABASE MATCH**

DNA-databases contain large numbers of DNA-profiles of known persons and of biological traces related to unsolved crimes.

When the number of DNA-profiles in a DNA-database increases, so does the chance of getting an adventitious match with a person who is not the actual donor of the trace.

This is especially true for partial DNA-profiles and mixed DNA-profiles because the chance that they would match with a randomly chosen person is greater than the chance that a full single DNA-profile would match a randomly chosen person.

If there are doubts if the matching person is the donor of the trace, for instance because there is no other tactical or technical evidence which links the person to the crime, the possibility to do additional DNA-testing can be considered.

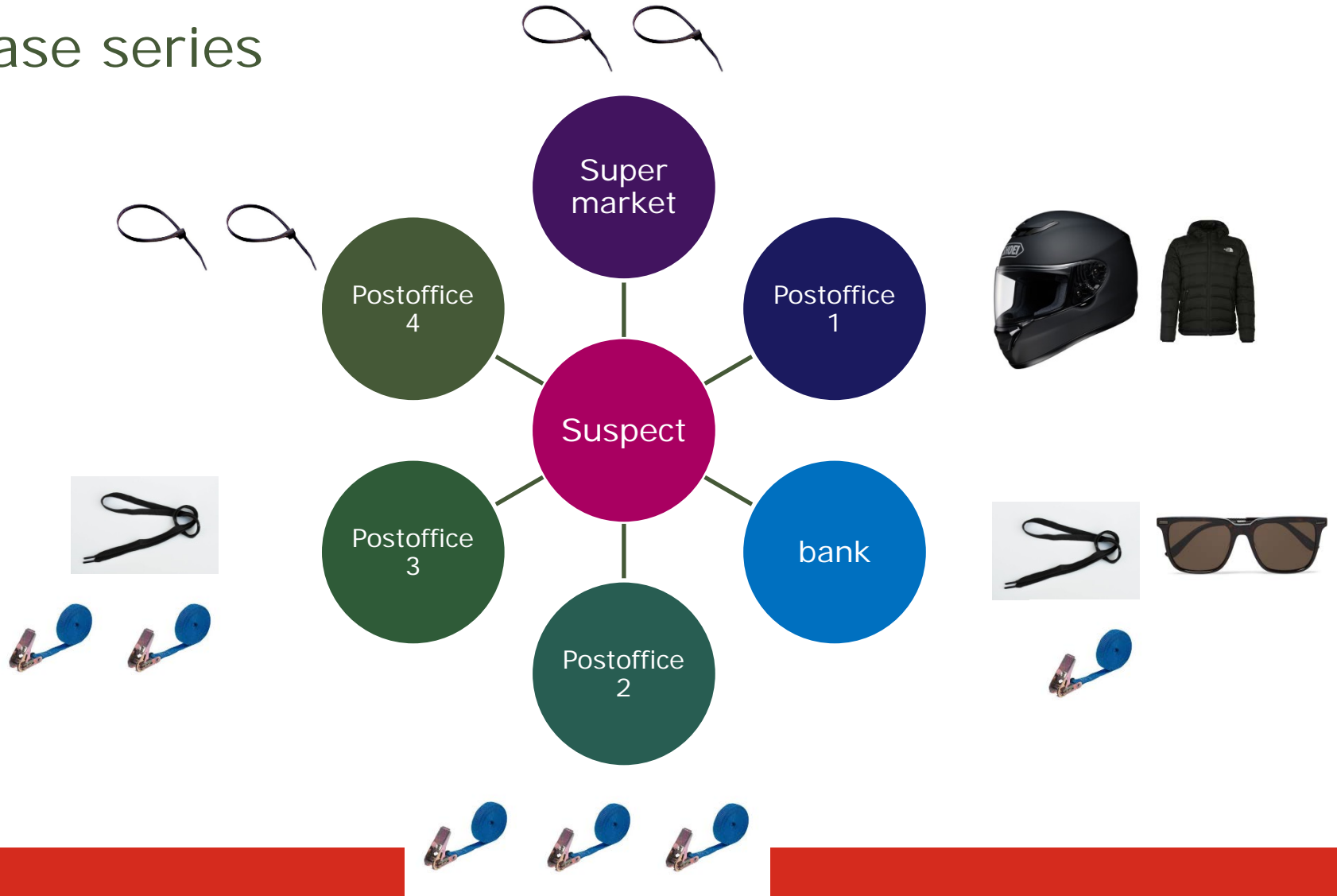
This point of attention particularly applies to matches which are found as a result of the large scale international DNA-profile comparisons based on the EU-Prüm decisions



**MORE QUESTIONS**



# Case series





## Interesting aspects: evidential value?

- Effect of error
- Database match
- Partial DNA profiles
- Moveable objects
- Combination of DNA profiles in same case
- Combination of DNA profiles in different cases
- MO: Entering through ceiling in 5 out of 6 robberies



## Need to invest in interpretation

| Equipment | Technical knowledge | Evidence interpretation | Result                |
|-----------|---------------------|-------------------------|-----------------------|
| ✓         | ✓                   | ✓                       | Good forensic science |
| ✓         | ✓                   | ✗                       | Junk science          |
| ✓         | ✗                   | ✗                       | Junk science          |
| ✗         | ✓                   | ✓                       | Junk science          |
| ✗         | ✗                   | ✓                       | Junk science          |
| ✗         | ✗                   | ✗                       | Junk science          |

- Experts and managers focus on equipment and technical skill and knowledge
- Discussion in court is mostly about interpretation!