Netherlands Forensic Institute
Ministry of Justice

Evaluating and reporting forensic evidence using the LR framework: statistical challenges

Prof. dr. Marjan Sjerps

- Netherlands Forensic Institute
- University of Amsterdam

5 May 2016

Netherlands Forensic Institute

Statistics team NFI

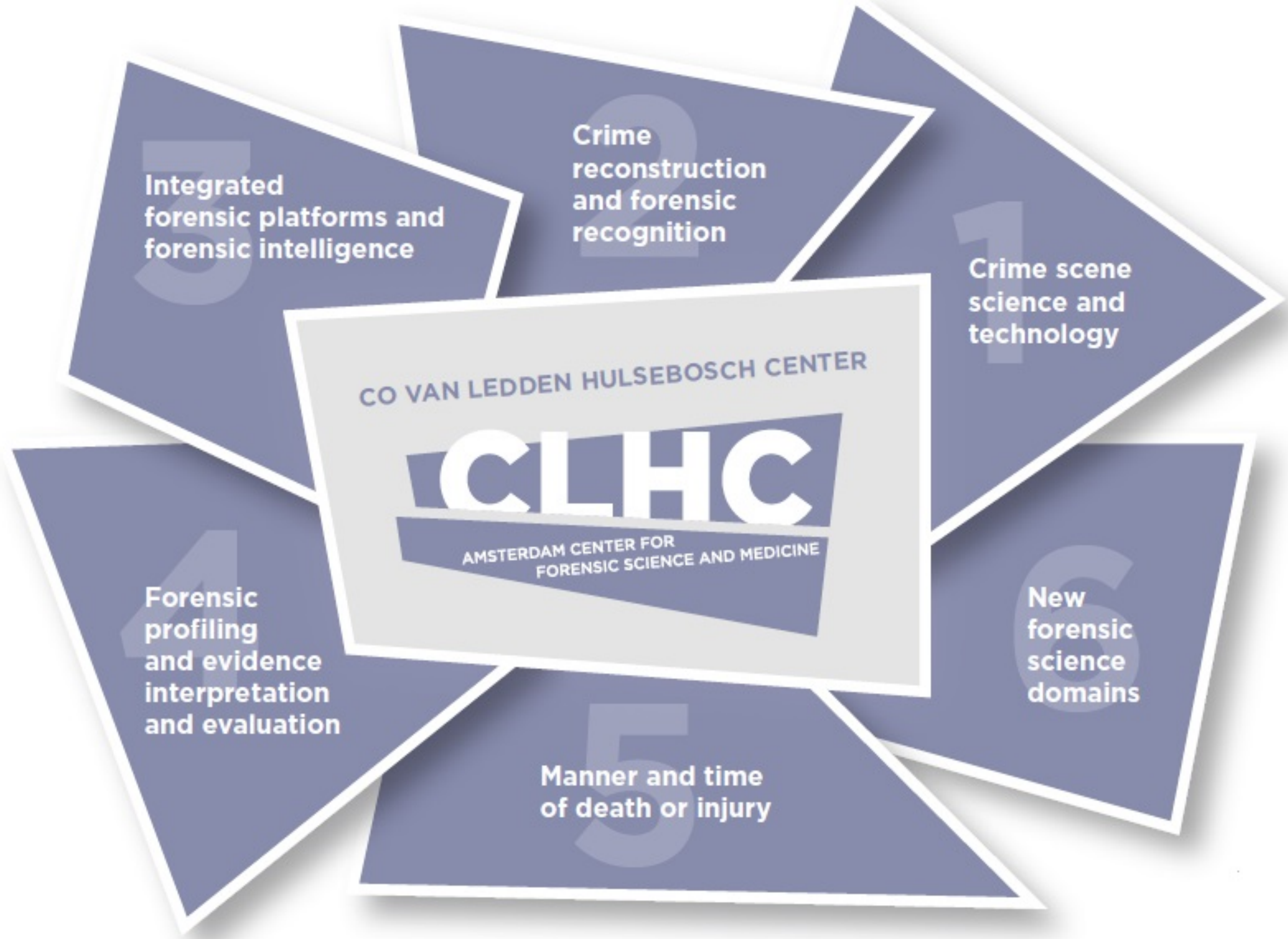





Principal scientist team NFI




Many other LR proponents in other teams


CO VAN LEDDEN HULSEBOSCH CENTER
CLHC
AMSTERDAM CENTER FOR FORENSIC SCIENCE AND MEDICINE
1
Crime scene science and technology
2
Crime reconstruction and forensic recognition
3
Integrated forensic platforms and forensic intelligence
4
Forensic profiling and evidence interpretation and evaluation
5
Manner and time of death or injury
6
New forensic science domains

Statistics in forensic science: it's everywhere!

Forensic process	Statistical aspects
1. Strategy	• Sampling strategy • Design of experiments • Case pre-assessment
2. Analyse	• Statistical process control • Measurement uncertainty • Performance characteristics…
3. Conclude	• Evidence evaluation
4. Report	• Communicating probabilistic reasoning

The Likelihood Ratio (LR): general

$$\frac{\text{Probability of evidence, if hypothesis 1 is true}}{\text{Probability of evidence, if hypothesis 2 is true}}$$

$$= \quad \frac{x\%}{y\%}$$

Implementing the LR framework- European guideline (ENFSI website)/ Roadmap

Step 1
Managing the change

- Identifying **key personnel responsible** for the implementation
- Deciding on a **strategy** to approach each forensic discipline covered by the laboratory (focus groups, leaders in each discipline, etc.)
- Adopting a **project plan** with defined objectives and timeline

Step 2
Training

- Providing **training** and **workshops** on the guideline (i.e. framework of circumstance, propositions, likelihood ratio, workshops per discipline)
- Identifying what is covered by **evaluative reports** (compared to factual or investigative reports)
- Training should include **competency testing**.
- Providing **information** and **training** to the stakeholders (e.g. police officers, judiciary, mandating authority) in relation to the changes associated with the guideline in particular the exchange of information at the outset of the case and the reporting practice

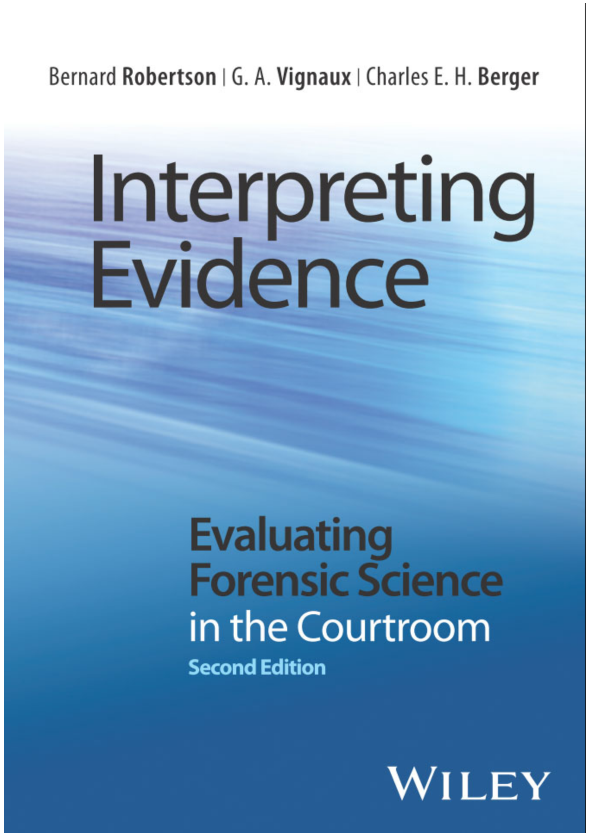
Step 3
Identifying the issues

- Implementing the mechanisms to establish the **key issues** in the submitted cases by adapting the exchange of information between the forensic laboratory and the mandating authority
- Setting an appropriate **framework of propositions** (including dealing with "no comment" interviews)
- **Identifying the levels of propositions** (source or activity level) that best help address the key issues
- If appropriate, carrying out a **pre-assessment** of cases and communicating with the mandating authority
- Identifying the ***data* requirements** (*data* as defined in the guideline) to help address the issues. If needed, undertake structured data acquisition
- **Optional**: Developing a uniform **verbal scale** to support consistent reporting within the laboratory

Step 4
Reporting according to the guideline

- Reporting on the **probability of the findings given the propositions** and relevant background information which leads to a **likelihood ratio**
- **Avoiding in reports statements that are transposing the conditional** (i.e. not reporting on the probability of the propositions given the observations)
- **Auditing** the casework using the audit template associated with the guideline

2nd edition!


Bernard Robertson | G. A. Vignaux | Charles E. H. Berger
Interpreting Evidence
Evaluating Forensic Science in the Courtroom
Second Edition
WILEY

Verbal statements expressing LR

"The findings are far more probable when the fragment comes from the window than when the fragment comes from some other glass object"

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

Communication

Nederlands Forensisch Instituut
Ministerie van Veiligheid en Justitie

Vakbijlage

De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs

Inhoudsopgave

1. De vakbijlage algemeen

Het Nederlands Forensisch Instituut (NFI) kent een groot aantal typen onderzoeken. Normaal gesproken gaat elk onderzoeksrapport van het NFI vergezeld van een vakbijlage. Deze dient als toelichting op het onderzoek en heeft een zuiver informatief karakter. Achterin de vakbijlage zijn een verklarende woordenlijst en een overzicht van bron- en literatuurverwijzingen opgenomen.

2. Inleiding

In veel gevallen kan de forensisch onderzoeker de vraag die de opdrachtgever hem stelt niet met een volmondig ja of nee beantwoorden. Er is dan een bepaalde mate van onzekerheid over de conclusie. Bij voorkeur wordt deze onzekerheid getalsmatig uitgedrukt, bijvoorbeeld in de vorm van een kans of een interval. Maar in sommige onderzoeken kan de onderzoeker zijn conclusie slechts formuleren in verbale termen van waarschijnlijkheid. Hierbij gebruiken NFI-onderzoekers, waar van toepassing, een standaardreeks van termen om hun conclusie te formuleren. Deze standaardreeks is gebaseerd op inzichten die volgen uit het zogeheten 'Bayesiaanse model' voor de interpretatie van bewijs. In deze vakbijlage wordt dit model besproken aan de

Implementation numerical LR systems (ideally)

1. Ask ‘customer’ if problem is relevant and worth the effort
2. Define hypotheses and evidence, generate play data
3. Make first probability models
4. Gather small dataset (pilot experiments, small samples)
5. Refine models
6. Gather large dataset
7. Make final model
8. Gather validation dataset and validate LRs produced
9. Make user interface; arrange maintanance
10. Train experts
11. Develop case report, professional appendix etc
12. External accreditation of method
13. Publish paper in peer-reviewed international journal

Team work !

- Reporting experts
- Forensic statisticians
- Academics
- Trainees/research assistants
- Software engineers
- ‘Customer’

USE OF LR METHODS AT NFI / SOURCE LEVEL HYPOTHESES

LR methods

- Calculation of numerical LRs is promoted by management (“objectivation”)
- LR systems developed:
  - Univariate and multivariate data,
  - Discrete and continuous data
  - “Black box” and “big” data
  - Score based and feature based LRs
  - Calibrated and “raw” LRs

- The following are some examples, not a complete overview
- Most of it is work by my colleagues
- Contact me if you are interested in one of the projects

Use of LR methods in NFI evaluative reports

	conclusion	LR calculated?	Conclusion based (partly) on LR?
1	Numerical LR	yes	
2	Verbal LR	yes	yes
3	Verbal LR	yes	yes, introduction recent/soon
4	Verbal LR	yes	no, still in research phase
5	Verbal LR	no	no

LR-level 1: numerical LR reported

DNA:

- Standard DNA profile comparison: random match probability
- “Easy mixtures” : e.g. major profiles
- Relatedness analyses:
    - special LR system Bonaparte*
      developed in cooperation with Radboud University:
      http://www.bonaparte-dvi.com/
        - Standard and difficult paternity, immigration etc
        - Disaster victim identification (DVI)
        - Linking unidentified persons to missing persons
        - Population screens
        - Familial search

*: this presentation does not imply endorsement of any of the products by NIST

LR-level 2: verbal LR reported, (partly) based on numerical LR

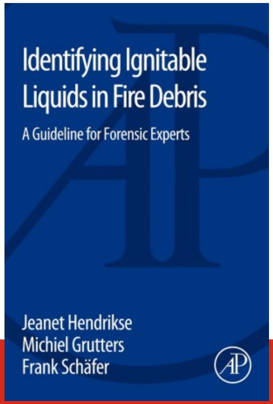
- DNA: complex profiles (mixtures, low template)
  - Special open source LR system developed: LR-mix Studio http://lrmixstudio.org/
  - Based on allele calls (not peak height)
  - Interest in continuous LR systems (using peak height)
- Camera identification based on pixel defects (PRNU pattern):
  - Special software developed to generate score based LR
- Fingermarks:
  - Special software developed (WOVI) for LR based on "black box" AFIS scores: performance not optimal, better AFIS system needed
  - Frequency of combination of minutiae configurations, core-delta distance and ground pattern class

LR-level 3: verbal LR reported, (partly) based on numerical LR, introduction recent/soon

- Speaker recognition: phone taps
  - Score based LR
- Glass
  - Multivariate feature based LR- calibrated
- Illicit drugs: XTC tablet comparison (MDMA)
  - 4 LR systems: 2 score based LR, 2 multivariate feature based LR
- Ignitable liquids: comparing fire debris with a gasoline intact source:
  - multivariate feature based LR- calibrated

Level 4: verbal LR reported; numerical LR produced in research phase

- Toolmarks: 2D or 3D microscopy
  - Score based LR system
- Fire arms:
  - score system developed by NIST used to generate LRs
- Ignitable liquids:
  - LR for GCxGC-MS in fire debris samples (thesis Martin Lopatka)
  - Fire debris analysis: current practice



Level 5-LR: verbal LR reported, no numerical LR

- Most of current evaluative NFI reports (other than previous examples and some that I forgot)

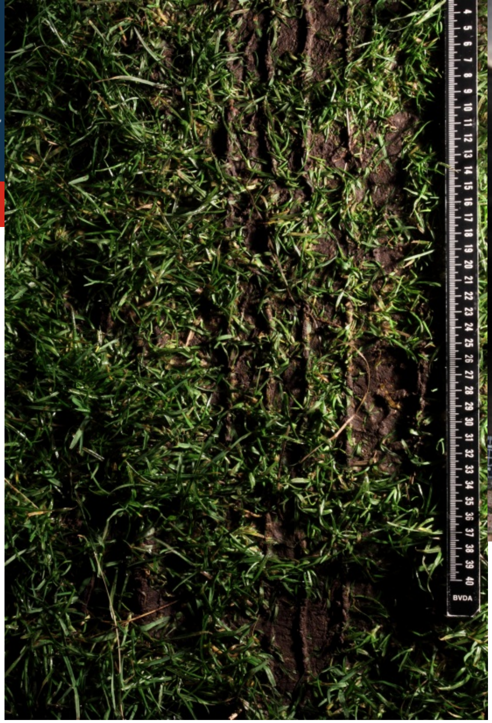
ACTIVITY LEVEL HYPOTHESES

Activity level: Bayesian networks

- DNA: how did the DNA get there?
  - Verbal LR based on expert opinion
- Interdisciplinary Cases: combining evidence of different areas, e.g. glass and DNA
  - Framework developed for identifying key forensic issues in scenarios
  - Verbal LR based on combining verbal LRs of several experts
- Combining evidence: Thesis Jacob de Zoete
- Glass activity level: (research phase)

A TIRE MARK CASE

“Cannot determine”



Defence Lawyer questions (names, dates etc adjusted)

- *2. (a) how many of the total number of (registered) (personal) cars in the Netherlands on 12 December 2013 had axes with a track width of about 144/145 centimeters (numbers and percentages).*
- *2. (b) how many of the total number of (registered) Peugeots in the Netherlands on 12 December 2013 had axes with a track width of about 144/145 centimeters (numbers and percentages);*
- *2. c. How many of the total number of (registered) Peugeots 208 in the Netherlands on 12 December 2013 had axes with a track width of about 144/145 centimeters (numbers and percentages).*
- *3. (a) how many of the total number of (mounted) tyres were in Netherlands on 12 December 2013 of the brands Uniroyal and Maxxis (numbers and percentages).*
- *3. b. How many of the total number of (mounted) Uniroyal and Maxxis tires in Netherlands on 12 December 2013 were the types of Uniroyal MS Plus 66 and Maxxis MA-P1 (numbers and percentages).*
- *3. c. How many of the total number of (mounted) Uniroyal MS Plus 55 and Maxxis MA-P1 ...*

Forensic statistics: defining relevant question

- Hypothesis 1: The Peugeot 208 (kenteken 11-XX-YY) made the marks on the lawn
- Hypothesis 2: some unknown other car made the marks on the lawn

Observations E:

1. the profile type of the front tires differs from that of the back tires
2. these two tire profiles have class characteristics *A* and *B*
3. the axes widths lie in the interval [143,146] cm.

- How probable are the observations under these hypotheses?

"What is the probability that a random car has tire profiles matching the tire tracks encountered in this case, and has corresponding axes width?"

Random match probabilities

p_{axis}=Pr(the axes widths of a random car are in the axes width interval of the tracks, i.e. [143,146] cm),

$p_{different}$=Pr(the profile type of the front tires of a random car differs from that of the back tires),

$p_{prof\_1\_match}$=Pr(some tire of a random car matches the first tire track profile found at the crime scene, i.e., has class characteristics A),

$p_{prof\_2\_match}$=Pr(some tire of a random car matches the second tire track profile found at the crime scene, i.e., has class characteristics B).

Combined Random match probability

Assumptions:

1. Given that the front and back tires of a car are from different brand/type, the two brands/types of the tires are independent of each other
2. The axes width of a car is independent of the profile of tires on its wheels

the probability p that a random car has axes widths and tires matching the tire tracks as described above, is:

$$p = 2 \times p_{\text{axis}} \times p_{\text{different}} \times p_{\text{prof\_1\_match}} \times p_{\text{prof\_2\_match}}.$$

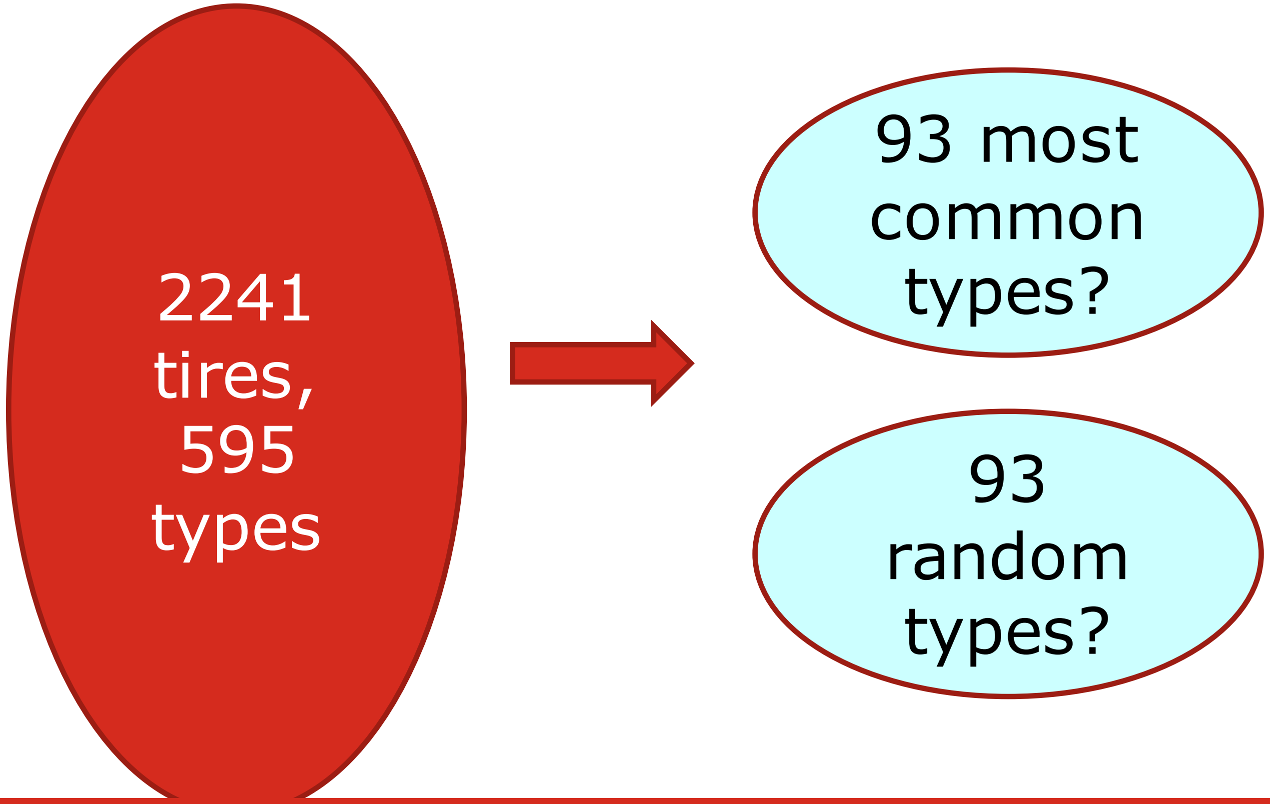
Research: Obtaining data

- RDW: axes widths of all cars in the Netherlands
- Own NFI dataset: parking lot data Amsterdam IKEA and rural area (n=30)
- Dataset representative sample Dutch tires on 40 parking lots (n=2239 cars)

Research: "blind" test expert

- Which types of tire match with the marks?
- Expert compared internet-photo's of 93 types of tires with the two marks (including tire types under suspect's car).
- Expert "blindly" stated the tire types of the suspect's car matched
- Besides these, 3 types matched for the marks on the tile, and 0 types matched the marks in the grass.

Interesting stats problem: optimal selection for expert to compare?



Conclusion: Evidential strength of combination

- The Peugeot 208 (kenteken 11-XX-YY) matches concerning axes widths and tire profile with the observations on the two marks. The probability that a random Dutch car matches is about 1 in 50 thousand. A more conservative estimate is 1 in 5 thousand.
- Therefore it is far more probable (about 5 thousand times more) to observe this combination of features when hypothesis 1 is true, than when hypothesis 2 is true.

A slide for the managers among you

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!
- To avoid junk science we need to invest in interpretation