When, how, and for whom?

Anders Nordgaard, PhD, LL.D h.c., forensic specialist

Swedish Police Authority – National Forensic Centre (NFC)

NIST Workshop, Gaithersburg, Md June 25-26, 2024

Outline

Polisen

- Categorizing casework – when is evaluative reporting needed?
  - Sources of uncertainty
  - Clarifying the forensic questions
- Simplifying the expression of the value of evidence – scales of conclusions
  - Robust assignment
- Training forensic experts, investigators, prosecutors, judges and defence attorneys
  - Nordic law and normative framework
  - Who is the commissioner of forensic investigations?

Categorizing casework – when is evaluative reporting needed?

Polisen

Sources of uncertainty (in forensic investigations)

Category 1

- accuracy (measurement uncertainty)
- handling of material (contamination, mixing-up etc.)
- human factor in general

Such sources should normally not motivate evaluative reporting

- Measurement uncertainty is typically provided in technical reports
- Contamination, mixing-ups, human factor error are all unacceptable errors and should not be quantified and accompany a forensic conclusion – Should be handled by the quality assurance system.

Category 2

- only a (random) sample of the seized material is analysed

Polisen

Typical in screening analysis (e.g. identifying analysis of suspected drug material).

Conclusions should be accompanied with a statement of uncertainty reflecting the sampling error (“With 99% probability 50% of the consignment consists of Ecstasy pills”)

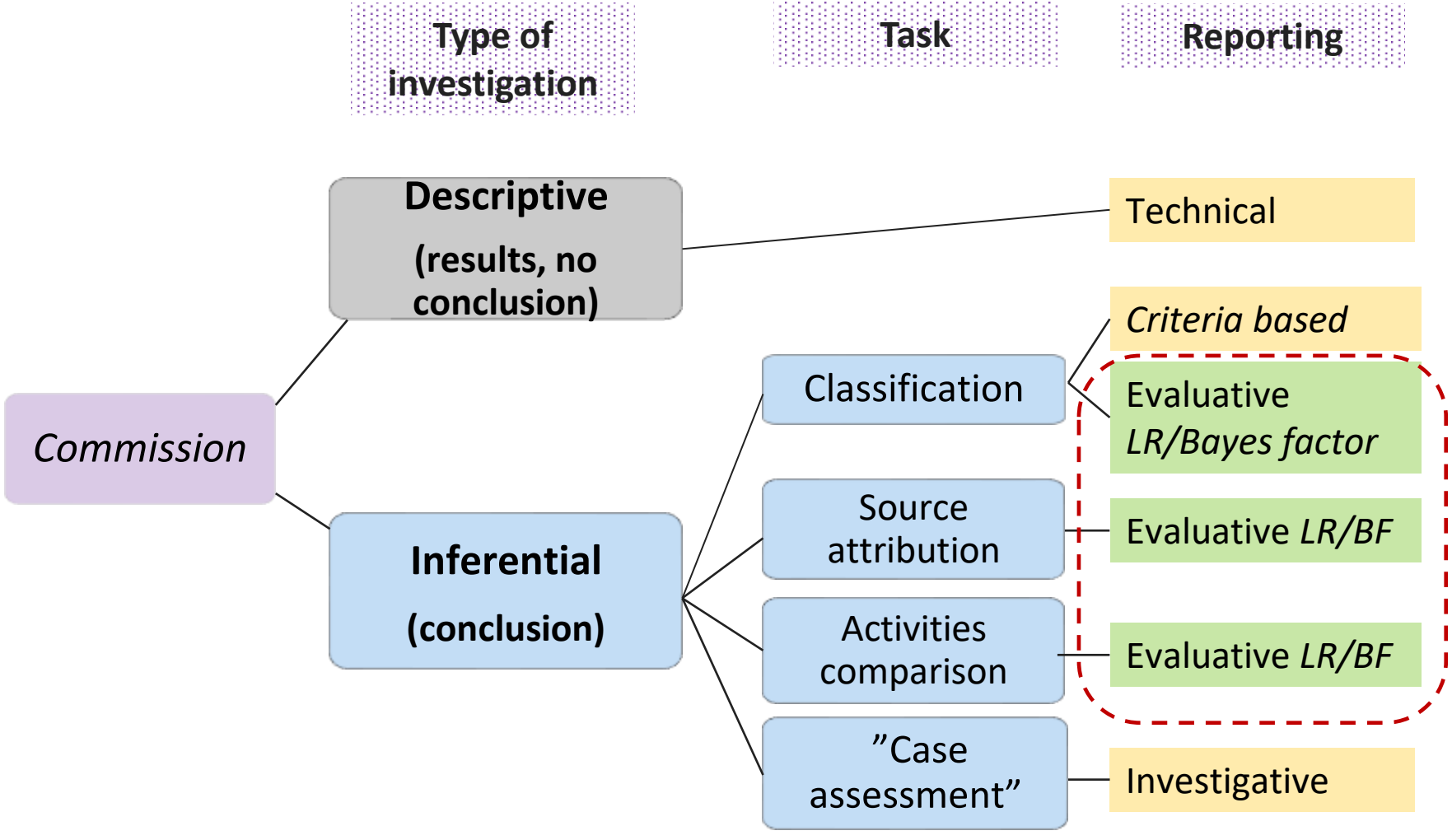
Category 3

- rarity/commonness in general of the characteristics observed/analysed
- mechanisms of transfer, persistence and background levels of material

This is the dominating source of uncertainty when the forensic question is about source attribution, competing activities or classification with no established criteria.

Requires evaluative reporting!

Clarifying the forensic questions (at NFC)



Examples

Descriptive tasks:

- What substances and with which concentrations can be found in the fire debris?
- How long is the person on the CCTV uptake?

Classification tasks with criteria-based reporting:

- Is the document a genuine Swedish passport?
- Is the electronic equipment a jammer?

Classification tasks with evaluative reporting:

- Do the fire debris contain (traces of) ignitable liquids?

Source attribution:

- Were the two scratch marks made with the same tool?
- Was part of the graffiti paint made with the spray can found with the suspect?

Activity comparison/attribution:

- Was the suspect's pullover recently in contact with the car seat?
- Did the suspect kick the victim in the face or was he just standing aside?

Simplifying the expression of the value of evidence – scales of conclusions

Polisen

The likelihood ratio (LR) (or Bayes factor (BF)) is a component of Bayes' theorem.

There are two competing hypotheses (propositions) addresses in a case:

The main hypothesis, $\boldsymbol{H_m}$ (usually forwarded by the prosecution)

The alternative hypothesis, $\boldsymbol{H_a}$

The forensic findings, $\boldsymbol{E}$ should be evaluated against these two hypotheses.

Bayes' theorem on odds form:

$$\frac{P(\boldsymbol{H_m}|\boldsymbol{E})}{P(\boldsymbol{H_a}|\boldsymbol{E})} = (\mathbf{LR/BF}) \times \frac{P(\boldsymbol{H_m})}{P(\boldsymbol{H_a})}$$

$$\frac{P(H_m|E)}{P(H_a|E)} = (\mathrm{LR/BF}) \times \frac{P(H_m)}{P(H_a)}$$

Polisen

In many accounts of forensic interpretation, the likelihood ratio is given as

$$\mathrm{LR} = \frac{P(E|H_m)}{P(E|H_a)}$$

"The probability of obtaining the forensic findings if H_m is true divided by the probability of obtaining the forensic findings if H_a is true."

However, a likelihood is not by necessity a probability. If the forensic findings are quantified on a continuous scale, *probability density functions* must be used as probative measure. Hence, a more general definition is

$$\mathrm{LR} = \frac{\mathcal{L}(H_m; E)}{\mathcal{L}(H_a; E)}$$

"The likelihood of H_m in light of the forensic findings divided by the likelihood of H_a in light of the forensic findings."

Moreover, likelihoods are defined for a particular value of a parameter or a simple hypothesis. If one or both hypotheses involved are composite, it is no longer valid to use the term likelihood ratio for the value of evidence.

The general expression for the *Bayes factor* as value of evidence is

$$BF = \frac{\sum_i \mathcal{L}(\boldsymbol{H_{m,i}}; \boldsymbol{E}) \times P(\boldsymbol{H_{m,i}} | \boldsymbol{H_m})}{\sum_k \mathcal{L}(\boldsymbol{H_{a,k}}; \boldsymbol{E}) \times P(\boldsymbol{H_{a,k}} | \boldsymbol{H_a})}$$

where $\boldsymbol{H_m} = \bigcup_i \boldsymbol{H_{m,i}}$ and $\boldsymbol{H_{m,i}} \cap \boldsymbol{H_{m,i'}} = \emptyset \ \forall\, i \neq i'$

$\boldsymbol{H_a} = \bigcup_k \boldsymbol{H_{a,k}}$ and $\boldsymbol{H_{a,k}} \cap \boldsymbol{H_{a,k'}} = \emptyset \ \forall\, i \neq i'$

Can be a challenge to assign!

But whether we address LRs or BFs, constructing a scale of conclusions is not about giving an absolute interpretation of neither of them.

Polisen

Focus on the potential posterior probabilities (or odds) they would lead to under different settings of the prior odds.

One possibility is to take as an average setting the maximum entropy.

Likelihood Ratio scale

Maximum prior entropy would be even odds, i.e. $P(\boldsymbol{H_m}) = P(\boldsymbol{H_a}) = 0.5$

Assuming exhaustive hypotheses, the posterior probability will be

$$P(\boldsymbol{H_m}|\boldsymbol{E}) = \frac{LR}{LR + 1}$$

Decide upon how many $\boldsymbol{H_m}$-supporting levels you wish to have in the scale.

Polisen

For each level (or a subset of levels), decide upon a sufficiently high posterior probability – reflecting end-users' interpretation of levels of probability with respect to their appreciation of evidentiary strength.

With even prior odds you can for each level deduce the LR as

$$P(\boldsymbol{H_m}|\boldsymbol{E}) = \frac{LR}{LR+1} \qquad LR = \frac{P(\boldsymbol{H_m}|\boldsymbol{E})}{1-P(\boldsymbol{H_m}|\boldsymbol{E})}$$

At NFC, we chose 4 levels (in 2004): +1, +2, +3, +4

Polisen

For +2 we decided that this is a level that with even prior odds should give a posterior probability of at least 99%.

99% is a probability level for which there is a wide-spread acceptance among legal professionals that something is corroborated (however, not proven).

+2 should thus – if prior odds are even or higher - be sufficient for detention.

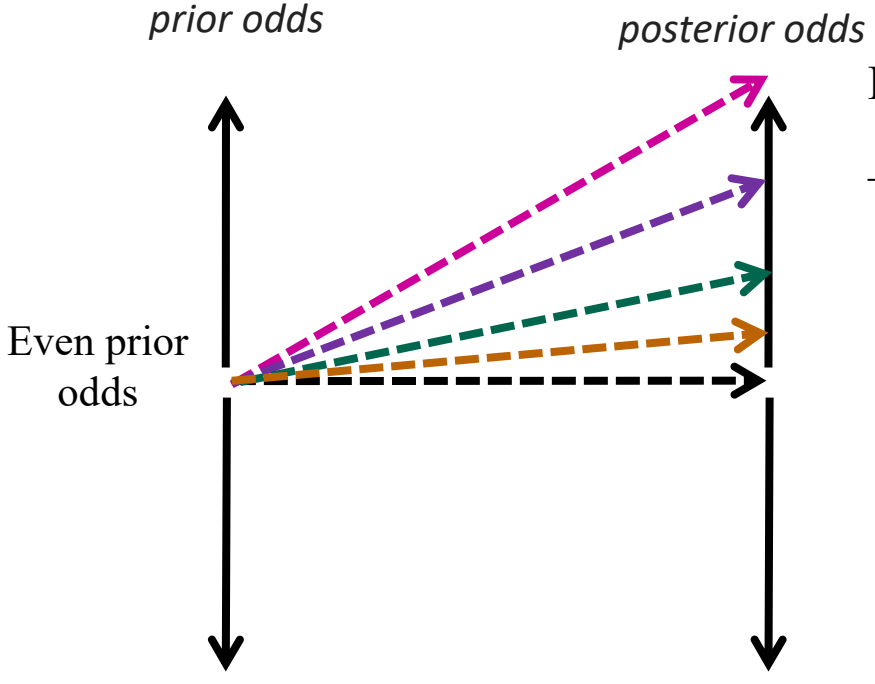
For the highest level, we set the LR to be at least one million – this magnitude was at the time a lower limit for a full match in DNA (siblings excluded).

With even prior odds a LR of one million gives a posterior probability of 0.999999.

Sufficiently high for considering $\boldsymbol{H_m}$ proven beyond reasonable doubt.

Level +1 and level +3 would correspond to LRs given a regular increase of the intervals between the levels.

The rest is math and rounding-off.



Level	Posterior probability $P(H_h \mid E)$	Lower limit for *LR*
+4 :	> 0.999999	$10^6 \leq$ LR
+3:	> 0.9998	6000 ≤ LR
+2:	> 0.99	100 ≤ LR
+1:	> 0.86	6 ≤ LR
0:	between 0.14 and 0.86	(1/6 < LR < 6)

Scale of conclusions used at NFC:

Scale level	Magnitude of the likelihood ratio (V)	"Explanation"
		The findings are deemed…
+4	at least one million	…at least one million times more probable…
+3	between 6000 and one million	…at least 6000 times more probable…
+2	between 100 and 6000	…at least 100 times more probable…
+1	between 6 and 100	…at least 6 times more probable…
0	between 1/6 and 6	… approximately equally probable…
		…if the main hypothesis is true compared to if the alternative proposition is true
–1	between 1/100 and 1/6	…at least 6 times more probable…
–2	between 1/6000 and 1/100	…at least 100 times more probable…
–3	between 1/(one million) and 1/6000	…at least 6000 times more probable…
–4	at most 1/(one million)	…at least one million times more probable…
		…if the alternative hypothesis is true compared to if the main proposition is true

Robust assignment

Polisen

It is most often the magnitude of the LR that is of interest, not its precise value.

In most forensic disciplines it is a difficult and time-consuming task to come up with a precise value of the LR – betting preferences work theoretically but are hard to imply to a community that never bets.

With a scale of conclusion, the forensic expert is instructed to report the level they are convinced is reached – but with no obligation to be more precise.

Produces a conservative report in the sense that the defence can always refer to the lower limit of an interval.

End-users *learn* successively what is a high value of evidence and what is a low value of evidence.

Training forensic experts, investigators, prosecutors, judges and defence attorneys

Polisen

Some points about Nordic Law

Nordic law is quite similar between the Nordic countries (Sweden, Denmark, Norway, Finland and Iceland)

One common thing is the free sifting of evidence – almost no evidence rules.

Sweden has a bit more adversarial system than Norway and Denmark.

NFC (and other Nordic labs) reports to the preliminary investigation with the Police (i.e. most often to the prosecutor) – not to the court.

But we must take into consideration what would be understood by the court.

Training…

Polisen

The in-lab trainee program at NFC

- (at least) 2 years of training followed by formal examination before being approved to be case responsible, sign reports and give statement of witness in court
- General part (to a large extent digital)
  - Basic module (to all personnel working to any extent with forensic evidence) – comprises one section on introductory evaluative reporting and forensic assessment
  - Add-on modules depending on function
    - Evaluative reporting for classification and source attribution task using assessed probabilities
    - Evaluative reporting for comparison/attribution of activities
    - Evaluative reporting with continuous probability distributions
    - Investigative reporting comprising evaluative steps
    - Evaluative reporting for combining evidence

- Function-specific part
  - Casework training (incl. evaluative reporting) under supervision

<u>Crime scene investigators</u>

- 1.3 years trainee program at NFC preceded by one year training at a police squad and mixed-up with such training during the program
- Investigative reporting with evaluative steps (model developed at NFC)
- 4 case studies from volume crime to severe crime
- Individual examination on an individual report on one of the case-studies (examination focussing on the investigative/evaluative part)

Police investigators

- Specific activities for target groups (comprising som evaluative reporting)
- *NEW!* Evaluative reporting part of curriculum of trainee program for investigators of severe crime

Prosecutors

- 3.5 days' basic training at NFC within the prosecutor's compulsory trainee program
  - Evaluative reporting taught via a complex case study throughout the days
- 3 days' continuing training for experienced prosecutors
  - 2-3 hours of evaluative reporting focussed on evaluation against hypotheses at activity level and combination of evidence

Judges

- 3 days' basic training at NFC within the compulsory trainee program
  - 2 hours general evaluative reporting and application in different forensic disciplines
- 3 days' continuing training for experienced judges
  - 3-4 hours of evaluative reporting focussed on the role of the judge

Defence attorneys

- Upon request, 2 days' training with 2-3 hours general evaluative reporting and application in different forensic disciplines
- Specific activities (half-day courses) at regional assemblies