What should a forensic scientist’s likelihood ratio be?

Geoffrey Stewart Morrison

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
These slides are available from:


- This work was funded in-part by a fellowship from the Simons Foundation. The presenter would like to thank the Isaac Newton Institute for Mathematical Sciences for its hospitality during the programme Probability and Statistics in Forensic Science which was supported by EPSRC Grant Number EP/K032208/1.

- Unless otherwise stated, all opinions expressed are those of the presenter.
What should a forensic scientist do?

- National Research Council report (2009)

- *With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.*
With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.

... support conclusions about “individualization” (more commonly known as “matching” of an unknown item of evidence to a specific known source).
What should a forensic scientist do?

- Calculate a likelihood ratio

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
What should a forensic scientist do?

• National Research Council report (2009)

• With the exception of nuclear DNA analysis, however, no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source.
What should a forensic scientist’s likelihood ratio be?

• presentation of strength of evidence in court

• source level comparisons
  - same origin
  - different origin

• continuously valued data
What should a forensic scientist’s likelihood ratio be?

- How should a forensic scientist arrive at a value for their likelihood ratio?

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
What should a forensic scientist’s likelihood ratio be?

- Experience and subjective judgment

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
Experience & subjective judgement

- Association of Forensic Science Providers standard (2009)

- Wherever possible probability estimates should be made.

- *Personal data such as experience in similar cases and peer consultations may be used provided that the practitioner can justify the use of such data and demonstrate that it is soundly based.*

- *Probability assignments may be qualitative. In such cases the likelihood ratio may also be expressed in qualitative terms provided this can be justified.*
Experience & subjective judgement

• ENFSI guideline on evaluative reporting (2015)

• Regardless of the existence of sources (published or not) of numerical data, personal data such as experience in similar cases and peer consultations may be used, provided that the forensic practitioner can justify the use of such data.

• For example, if the assessment is based on experience, the forensic practitioner will be able to demonstrate the relevant and documented previous professional activity.
Experience & subjective judgement

• ENFSI guideline on evaluative reporting (2015)

• Forensic practitioners often experience difficulty in assigning and justifying probabilities when the assignments are based on expert knowledge.

• However, likelihood ratios can be informed by subjective probabilities using expert knowledge.

• Such personal probability assignment is not arbitrary or speculative, but is based on a body of knowledge that should be available for auditing and disclosure.

• Forensic practitioners should consider exploring the sensitivity of the likelihood ratio to different probabilities by examining the effect of assigning different probabilities according to their personal uncertainties.
Experience & subjective judgement

• Berger, et al. (2011) Evidence evaluation: A response to the Court of Appeal judgment in *R v T*

• *The probability that is quoted then will inevitably be a personal probability and the extent to which the data influence that probability will depend on expert judgement.*

• *We stress that the probability may be informed not only by systematic research and data but also by expert judgement, ...*

• *The recognition of subjective probabilities as a sound basis for offering expert opinion should not be misconstrued as being an open door to unstructured reasoning and ad hoc “guesses”. Any appraisal offered, based on the spectrum of knowledge an expert may call upon, should be transparent and subject to peer review and challenge.*
Experience & subjective judgement

- Lennard (2013) Fingerprint identification: how far have we come?

- The issue is that the LR is a calculated result from a mathematical model, it is not an opinion.

- The middle ground would be an expert opinion, exactly what an expert witness is expected to provide to the court, but informed, supported and backed-up by an appropriate statistical model.

- The statistics should arguably only be presented if required.

- The evidence is the expert opinion and not the statistic.

- For counterarguments see: Morrison & Stoel (2014) Forensic strength of evidence statements should preferably be likelihood ratios calculated using relevant data, quantitative measurements, and statistical models.
US Federal Rule of Evidence 702 (pre 2000)

If scientific, technical or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

Daubert (1993)

The adjective “scientific” implies a grounding in the methods and procedures of science. Similarly, the word “knowledge” connotes more than subjective belief or unsupported speculation.
Risinger (2011/2013) Reservations about likelihood ratios

it seems to me that the general modern Bayesian tradition, ..., encourages faulty expression by making it seem acceptable for people to obtain the probabilities that are incorporated into likelihood ratios by simply making their best guess from experience when more should be required. This may be a more or less educated guess, but it remains a guess nonetheless, ...

there is something about the generation of likelihood ratios with numbers from nowhere that tends to cover up the weakness of the ingredients, I believe, so my reservation remains.

my own emphasis is not on the subjective nature of probabilities vel non in this general sense, but on the empirical warrant that is, or ought to be, necessary to derive and use a probability statement in different contexts.
Experience & subjective judgement

- PCAST report (2016)

- *neither experience, nor judgment, nor good professional practices (such as certification programs and accreditation programs, standardized protocols, proficiency testing, and codes of ethics) can substitute for actual evidence of foundational validity and reliability.*

- The frequency with which a particular pattern or set of features will be observed in different samples, which is an essential element in drawing conclusions, *is not a matter of “judgment.”*

- *It is an empirical matter for which only empirical evidence is relevant.*
Similarly, an expert’s expression of confidence based on personal professional experience or expressions of consensus among practitioners about the accuracy of their field is no substitute for error rates estimated from relevant studies.

For forensic feature-comparison methods, establishing foundational validity based on empirical evidence is thus a sine qua non.

Nothing can substitute for it.
Daubert (1993)

whether the reasoning or methodology underlying the testimony is scientifically valid and ... whether that reasoning or methodology properly can be applied to the facts in issue.

whether [a theory or technique] can be (and has been) tested. ... In the case of a particular scientific technique, the court ordinarily should consider the known or potential rate of error.

“[T]he statements constituting a scientific explanation must be capable of empirical test”
Experience & subjective judgement


• A witness who is qualified as an expert by knowledge, skill, experience, training, or education may testify in the form of an opinion or otherwise if:

  (a) the expert’s scientific, technical, or other specialized knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;

  (b) the testimony is based on sufficient facts or data;

  (c) the testimony is the product of reliable principles and methods; and

  (d) the expert has reliably applied the principles and methods to the facts of the case.
Daubert (1993)

We note that scientists typically distinguish between “validity” (does the principle support what it purports to show?) and “reliability” (does application of the principle produce consistent results?).

In a case involving scientific evidence, evidentiary reliability will be based upon scientific validity.
the court must be satisfied that there is a sufficiently reliable scientific basis for the evidence to be admitted.

Therefore factors which the court may take into account in determining the reliability of expert opinion, and especially of expert scientific opinion, include:

(a) the extent and quality of the data on which the expert’s opinion is based, and the validity of the methods by which they were obtained; ...

(c) if the expert’s opinion relies on the results of the use of any method (for instance, a test, measurement or survey), whether the opinion takes proper account of matters, such as the degree of precision or margin of uncertainty, affecting the accuracy or reliability of those results;
Empirical validation

- Morrison (2014) Distinguishing between forensic science and forensic pseudoscience
- Meuwly, et al. (2016) A guideline for the validation of likelihood ratio methods
Empirical validation

- Morrison (2014) Distinguishing between forensic science and forensic pseudoscience
- Meuwly, et al. (2016) A guideline for the validation of likelihood ratio methods
What should a forensic scientist’s likelihood ratio be?

- Empirically calibrated subjective judgement

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
Empirically calibrated subjective judgement

- Lindh & Morrison (2011) Humans versus machine
- Input is same-origin pairs and different-origin pairs
- Output is human judgement expressed on an ordinal scale (score)
- Same-origin and different-origin output used to train a score to likelihood ratio conversion model

see also: Morrison (2013) Tutorial on logistic-regression calibration and fusion
What should a forensic scientist’s likelihood ratio be?

- Dichotomous likelihood ratio

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
Dichotomous likelihood ratios

- PCAST report (2016)
- Practitioner declares “match” or “non-match”
- Whether the features in an evidentiary sample and the features in a sample from a suspected source lie within a pre-specified measurement tolerance
Dichotomous likelihood ratios

- PCAST report (2016)

- If a “match” is declared, also report the results of an empirical assessment of $p(\text{“match”}|\text{same-source})$ and $p(\text{“match”}|\text{different-source})$

- The forensic examiner should report the overall false positive rate and sensitivity for the method established in the empirical studies of foundational validity and should demonstrate that the samples used in the foundational studies are relevant to the facts of the case.

- Mutatis mutandis if “non-match” is declared

- This can be considered empirical calibration of the “match”/“non-match” decision. It is not empirical validation of performance.
Dichotomous likelihood ratios

- continuously-valued data with within-source variability
- dichotomous procedure discards information that could be exploited by more appropriate statistical procedures
- suffers from a cliff-edge effect
What should a forensic scientist’s likelihood ratio be?

- Ordinal scales (verbal expressions)

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
### Ordinal scales (verbal expressions)

- **ENFSI guideline on evaluative reporting (2015)**

<table>
<thead>
<tr>
<th>Values of likelihood ratio</th>
<th>Verbal equivalent (two options of phrasing are suggested)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The forensic findings do not support one proposition over the other. The forensic findings provide no assistance in addressing the issue.</td>
</tr>
<tr>
<td>2 - 10</td>
<td>The forensic findings provide weak support for the first proposition relative to the alternative. The forensic findings are slightly more probable given one proposition relative to the other.</td>
</tr>
<tr>
<td>10 - 100</td>
<td>...provide moderate support for the first proposition rather than the alternative ...are more probable given...proposition...than proposition...</td>
</tr>
<tr>
<td>100 - 1000</td>
<td>...provide moderately strong support for the first proposition rather than the alternative ...are appreciably more probable given... proposition...than proposition...</td>
</tr>
<tr>
<td>1000 - 10,000</td>
<td>...provide strong support for the first proposition rather than the alternative ...are much more probable given... proposition...than proposition...</td>
</tr>
<tr>
<td>10,000 - 1,000,000</td>
<td>...provide very strong support for the first proposition rather than the alternative ...are far more probable given... proposition...than proposition...</td>
</tr>
<tr>
<td>1,000,000 and above</td>
<td>...provide extremely strong support for the first proposition rather than the alternative ...are exceedingly more probable given... proposition...than proposition...</td>
</tr>
</tbody>
</table>
Ordinal scales (verbal expressions)

- Morrison & Enzinger (2016) What should a forensic scientist’s likelihood ratio be?
  - ranges are arbitrary
  - cliff-edge effects
  - verbal expressions are vague
  - will be interpreted differently by different individuals
  - will be interpreted differently by the same individual in different contexts
  - meaning can only be made explicit via reference to numeric likelihood ratio values

- see also: Marquis, et al. (2016) Discussion on how to implement a verbal scale in a forensic laboratory
Ordinal scales (verbal expressions)

- Morrison & Enzinger (2016) What should a forensic scientist’s likelihood ratio be?
- if a procedure was used which calculated a numeric likelihood ratio using relevant data and statistical models, report that numeric likelihood ratio
- if the level on the scale is selected on the basis of subjective judgement, empirically calibrate and report the resulting numeric likelihood ratio
What should a forensic scientist’s likelihood ratio be?

- Numeric likelihood ratio values

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
Numeric likelihood ratio values

- based on relevant data, quantitative measurements, and statistical models
- validate the system under conditions reflecting those of the case
- report the numeric likelihood ratio output of the model
  - transparent
  - replicable
  - easier to empirically calibrate
  - easier to empirically test
  - resistant to cognitive bias
What should a forensic scientist’s likelihood ratio be?

- Score-based calculation of likelihood ratio values

\[
\frac{p(E | H_p)}{p(E | H_d)}
\]
Score-based calculation of likelihood ratios

questioned source datum, $x_q$

known source data, $x_{k,r}$

relevant population data, $x_{w,r}$

score, $s_{q,k}$

feature to score model

test score, $s_{q,k}$

same origin scores, $s_{so}$

different origin scores, $s_{do}$

score to likelihood ratio model

likelihood ratio, $\lambda_{q,k}$
Score-based calculation of likelihood ratios

- difference scores or similarity scores
  - Manhattan distance
  - Euclidian distance
  - Pearson correlation
  - Kullback–Leibler divergence

- scores which take account of both similarity and typicality
  - typicality with respect to the relevant population

- attempt to calculate a likelihood ratio, but do not consider output directly interpretable
  - violated assumptions
  - attempting to estimate a large number of parameter values
Score-based calculation of likelihood ratios

- Direct
- Similarity
- Similarity & Typicality
Score-based calculation of likelihood ratios

- Morrison & Enzinger (submitted) Score based procedures for the calculation of forensic likelihood ratios
Score-based calculation of likelihood ratios

-39 -38.5 -38 -37.5 0 0.01 0.02 0.03 0.04 0.05 0.06 0.07 0.08 0.09 0.1

-0.25 -0.2 -0.15 -0.1 -0.05 0 0.05 0.1 0.15 0.2

probability density

different-origin

same-origin

similarity-only score

similarity-and-typicality score
What should a forensic scientist’s likelihood ratio be?

- Frequentist v Bayesian calculation of likelihood ratio values

\[
\frac{p(E|H_p)}{p(E|H_d)}
\]
Frequentist v Bayesian calculation of likelihood ratios

- **Frequentist**
  - sample data only
  - estimate of a true but unknown value

- **Subjective Bayesian**
  - prior distribution + sample data → posterior distribution
  - integrate out nuisance parameters
  - state of belief, subjective probability, personal probability
Frequentist v Bayesian calculation of likelihood ratios

- precision

- sources of variability
  - intrinsic at source
  - transfer process
  - sampling relevant population
  - sampling known source
  - measurement
  - statistical model training

- special issue on measuring and reporting the precision of forensic likelihood ratios
Frequentist v Bayesian calculation of likelihood ratios

trier of fact’s prior odds × trier of fact’s likelihood ratio = trier of fact’s posterior odds

forensic scientist’s likelihood ratio
Frequentist v Bayesian calculation of likelihood ratios

- **Frequentist**
  - estimate of a true but unknown value
  - should be accompanied by assessment of its precision

- **Subjective Bayesian**
  - state of belief, subjective probability, personal probability
  - already incorporates all sources of uncertainty

- reference prior

- Same practical result?

- What is best for trier of fact?
Frequentist v Bayesian calculation of likelihood ratios

- **Frequentist**
  - estimate of a true but unknown value

- **Subjective Bayesian**
  - state of belief, subjective / personal probability

- Can I use Bayesian procedures without buying into the subjectivist philosophy?
  - informative priors (empirically calculated from other data)
  - empirical Bayes
  - (non-informative) reference priors

- all approaches have some degree of subjectivity
What should a forensic scientist’s likelihood ratio be?

¿discussion?
Thank You

http://geoff-morrison.net/
http://forensic-evaluation.net/
Abstract

How should a forensic scientist arrive at a value for the strength of evidence statement that they present to the court? A number of different answers have been proposed.

One proposal is to assign probabilities based on experience and subjective judgement. This appears to be advocated in the Association of Forensic Science Providers (AFSP) 2009 standards, and the 2015 European Network of Forensic Science Institutes (ENFSI) guideline on evaluative reporting. But the warrant for such subjective judgements has been questioned. The 1993 US Supreme Court Daubert ruling and the 2016 report by the President’s Council of Advisors on Science and Technology (PCAST) argue strongly that subjective judgment is not enough, that empirical validation is needed.

If a forensic likelihood ratio is to be based on subjective judgement, it has been proposed that the judgement be empirically calibrated.

The PCAST report proposes a procedure which results in a dichotomous likelihood ratio. The practitioner applies a threshold and declares “match” or “non-match”. If a “match” is declared, the empirically derived correct acceptance rate and false acceptance rate are also provided (dividing the former by the latter would produce a likelihood ratio). Mutatis mutandis if a “non-match” is declared. This has been criticised for discarding information and thus resulting in poor performance.

The AFSP standards and ENFSI guideline propose the use of ordinal scales – each level on the scale covers a pre-specified range of likelihood ratio values, and has an associated verbal expression. These have been criticised on a number of grounds, including for having arbitrary ranges, for suffering from cliff-edge effects, and for verbal expressions being vague – they will be interpreted differently by different individuals, and differently by the same individual in different contexts.

It has also been proposed that numeric likelihood ratios be calculated on the basis of relevant data, quantitative measurements, and statistical models, and that the numeric likelihood ratio output of the statistical model be directly reported as the strength of evidence statement. Such an approach is transparent and replicable, and, relative to procedures based primarily on subjective judgement, it is easier to empirically calibrate and validate under conditions reflecting those of the case under investigation, and it is more resistant to cognitive bias.

Score based procedures first calculate a score which quantifies degree of similarity (or difference) between pairs of objects, then applies a subsequent model which converts scores to likelihood ratios (the second stage can be considered an empirical calibration stage). Scores which only take account of similarity (or difference), however, do not account for typicality with respect the relevant population for the case, and this cannot be corrected at the score to likelihood ratio conversion stage. If a score based procedure is used, the scores should take account of both similarity and typicality.

Numeric likelihood ratios can be calculated in a frequentist manner or a subjectivist Bayesian manner. Philosophically the former is an estimate of a true but unknown value, and the latter is a state of belief, a personal probability. A frequentist will assess the precision of their estimate, whereas a subjectivist Bayesian will have attempted to account for all sources of uncertainty in the assignment of the value of their likelihood ratio (a Bayes factor). The merits of the two approaches are hotly debated (including currently in a virtual special issue in Science & Justice http://www.sciencedirect.com/science/journal/13550306/vsi), but if presented with a frequentist point estimate plus degree of precision the trier of fact may decide to use a likelihood ratio closer to 1 than the point estimate (the deviation depending on the degree of precision), and (depending on the prior used) the value of the Bayes factor will be closer to 1 than a frequentist point estimate of a likelihood ratio. Can these be considered to have the same practical result? Which would be preferred by the courts? Can Bayesian procedures with empirical or reference priors be adopted without having to buy in to the subjectivist philosophy? What should a forensic scientist’s likelihood ratio be?
References


Morrison GS, Enzinger E (submitted) Score based procedures for the calculation of forensic likelihood ratios – scores should take account of both similarity and typicality.


President’s Council of Advisors on Science and Technology (2016) Forensic science in criminal courts: Ensuring scientific validity of feature-comparison methods. Washington DC: Executive Office of The President’s Council of Advisors on Science and Technology. https://www.whitehouse.gov/administration/eop/ostp/pcast/docs/reports


Legal References

Criminal Practice Directions [2014] EWCA Crim 1569