- National Commission on Forensic Science
- Meeting 11, September 12-13, 2016, NIST, Gaithersburg, MD

STATISTICAL STATEMENTS THAT SATISFY *DAUBERT* (AND THE COMMISSION?)

David H Kaye
Penn State Law

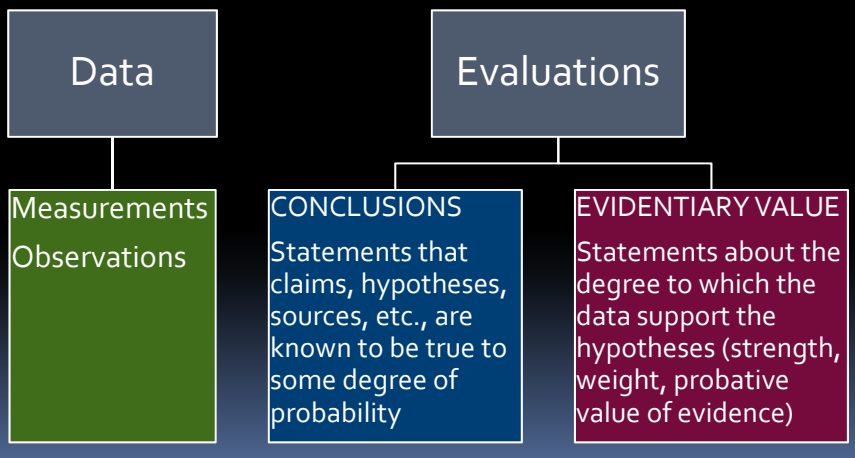
ASSIGNMENT

- Discuss how, in some specific cases, a forensic scientist in a pattern case might respond to typical *Daubert* questions in a manner that would meet acceptable statistical standards and comport with the Subcommittee's proposed report.

Daubert questions
(as framed for the panel)

- What methodology did you use to arrive at your opinion or statistical assessment?
- Does that methodology have known or generally accepted error rates? What are the false positive and false negative error rates?
- If the error rate is unknown, is that because of insufficient data, differences over methodology, too many variables to measure, too much subjectivity in the approach, or what?

Big Picture: Two Distinctions
Descriptions of Data / Evaluations or Assessments
Statement of Conclusions / Expressions of Evidentiary Value



DATA
Features only
Features similar

NONEVALUATIVE TESTIMONY

Hair

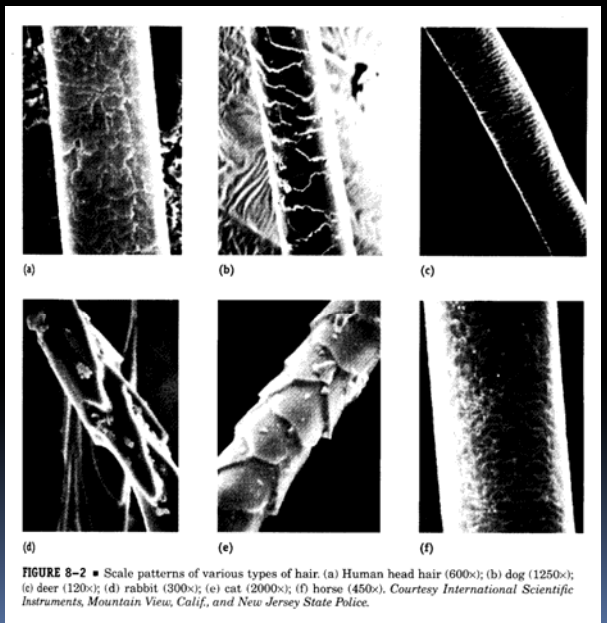
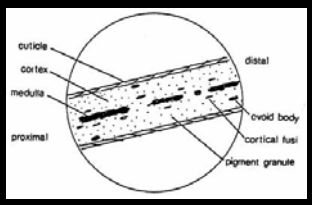


FIGURE 8–2 ■ Scale patterns of various types of hair. (a) Human head hair (600×); (b) dog (1250×); (c) deer (120×); (d) rabbit (300×); (e) cat (2000×); (f) horse (450×). *Courtesy International Scientific Instruments, Mountain View, Calif., and New Jersey State Police.*



Features-only testimony

- [A hair examiner] displayed an enlarged photograph of one of the defendant's hairs and one of the hairs recovered from the victim's clothing as they appeared side-by-side under the comparison microscope. [He] explained to the jurors how the hairs were similar and what particular features of the hairs were visible. He also drew a diagram of a hair on a courtroom blackboard for the jurors. The jurors were free to make their own determinations as to the weight they would accord the expert's testimony in the light of the photograph and their own powers of observation and comparison.

State v. Reid, 757 A.2d 482, 487 (Conn. 2000)

Similar features

Hair

testified that some of those hairs were consistent, meaning had the same characteristics, with known hair samples provided by [the defendant] and some of those hairs were consistent with hair samples from the victim . . . ."

- Brown v. State, 999 So.2d 853 (Miss. Ct. App. 2008)

Cord

testified that . . . a piece of cord taken from the scene of the crime [and] a piece of cord taken from the hood of a jacket 'matched each other in component structure, . . . were similar and could have . . . originated from the same jacket.'

- State v. Gomes, 881 A.2d 97 (R.I. 2005)

DNA

not error to admit "testimony that [defendant] could not be excluded as the source of the DNA obtained from the sneakers [even without] testimony explaining the statistical relevance of the nonexclusion result, such as the percentage of the population that could be excluded."

- Rodriguez v. State, 273 P.3d 845, 850–51 (Nev. 2012)

Views Document

Are there any "statistical statements" in exhibiting features?

- No, it is not "[t]o explain the value of the data in addressing hypotheses" (View 5), but it is recommended as a starting point under View 3 .

Are statements of "could have," "consistent with," "similar" or "match" statistical statements?

- Apparently not under View 6 ("should not opine on the truth, falsity, or probability of the claims themselves")

Rewrite to explain that the expert should state that because current research does not warrant evaluative statements, none are provided.

Examples of testimony on posterior probabilities Pr(H|E)
Glass
Handwriting
Kinship

EVALUATIVE: CONCLUSION

Glass: "Fit" implies Pr(H|E) = 1

An examiner may state or imply that the glass fragments were once part of the same broken object. This conclusion can only be reached when two or more pieces of broken glass physically fit together.

US Department of Justice ULTR (draft)

View 6. [F]orensic experts should not opine on the truth, falsity, or probability of the claims themselves.

Handwriting: Qualitative scale for Pr(H|E)

- identification (definite conclusion of identity)
- strong probability (highly probable, very probable)
- Probable
- Indications (evidence to suggest)
- No conclusion (totally inconclusive, indeterminable)
- Indications did not
- Probably did not
- Strong probability did not
- Elimination.

ASTM E1658-08

Kinship: Quantitative Pr(H|E)

- A hunter discovered remains of a woman and her unborn child on Fort Benning Military Reservation.
- She had been shot, and gov't charged N with murder.
- It claimed N's motive for the crime was that she was pregnant with his child.
- Q = fetal bones; K = N's cells (and the woman's?)

• United States v. Natson, 469 F.Supp.2d 1253 (M.D. Ga. 2007)

Proposed testimony: Pr(H|E)

$$L = \frac{\Pr(E|N)}{\Pr(E|{\sim}N)} = \frac{1/32}{1/832} = 26$$

$$PoP_{50} = \left[1 + \frac{1}{L}\right]^{-1} = 96.30\%$$

$$\frac{1}{PoP} = \frac{1}{\Pr(N|E)} = \frac{\Pr(N)\Pr(E|N) + \Pr({\sim}N)\Pr(E|{\sim}N)}{\Pr(N)\Pr(E|N)} = 1 + \frac{\Pr({\sim}N)}{\Pr(N)}\frac{1}{L}$$

No RDSC because *PoP* must be "at least 99.99% for the DNA scientific community to consider a DNA test to show paternity."

• *Natson*

Satisfies *Daubert* (but not View 6) if the conclusion of 96.3% follows from sufficient data, solid genetic theory, and articulated assumptions?

A more extreme example

- Christina Buettner from the Wyoming State Crime Lab first testified 'the probability of paternity' is '99.9999998638' that Mr. Snyder is the father of JL's baby.

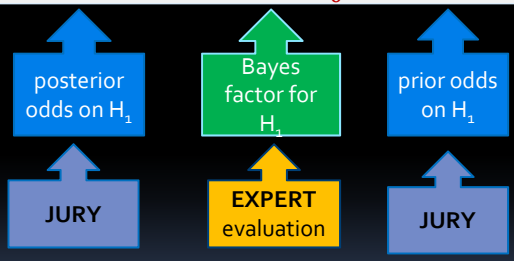
• Snyder v. State, 2015 WY 91, 353 P.3d 693, 694 (2015)

Examples of testimony on likelihoods Pr(E|H)
Kinship
Latent prints

EVALUATIVE: EVIDENCE

Bayes Rule (Simplest Case)

$$\text{Odds}(H_1|E) = \frac{\Pr(E|H_1)}{\Pr(E|H_0)} \text{Odds}(H_1)$$



- Expert can help jury assess strength of evidence by presenting this factor (= *LR*) or data that affect it.

17

Kinship

Views document favors the two likelihoods and their ratio

$$L = \frac{\Pr(E|N)}{\Pr(E|{\sim}N)} = \frac{1/32}{1/832} = 26$$

- United States v. Natson, 469 F.Supp.2d 1253 (M.D. Ga. 2007)

Does that methodology have known or generally accepted error rates? What are the false positive and false negative error rates?

The thresholds for declaring alleles have good statistical properties (very few false positives and negatives).

The sampling error in the profile frequencies can be estimated with a normal error model.

The modeling uncertainty is small because the likelihoods rely on a probability model from biology (not just a statistical model).

The statistical assessment is not a classification, so there are no false positive and false negative rates. A related quantity is the probability of misleading evidence. That has been studied too.

Latent prints: Possible qualitative, subjective Pr(E|H)

- Evaluation is the formulation of an opinion on the degree to which the information gathered during analysis and comparison supports the hypothesis (S) of a common source for the exemplar and the latent print or instead supports the hypothesis (D) that the exemplar and the latent come from different individuals. An opinion of strong support for S does not necessarily eliminate the possibility another person in the world could leave a print with areas of similar agreement. It means that within the examiner's experience and knowledge, prints from the same finger would be expected to display this much similarity, whereas prints from different fingers would not be.

Qualitative *Ls*

Evett 1991

Log L	L	Verbal Tag
0 to ½	1 to 33	weak
½ to 2	33 to 100	fair
2 to 2½	100 to 330	good
2½ to 3	330 to 1000	strong
>3	>1000	very strong

		Evett et al. 2000
0 to 1	1 to 10	limited
1 to 2	10 to 100	moderate
2 to 3	100 to 1000	moderately strong
3 to 4	1000 to 10000	strong
>4	>10000	very strong

ENFSI 2015 example

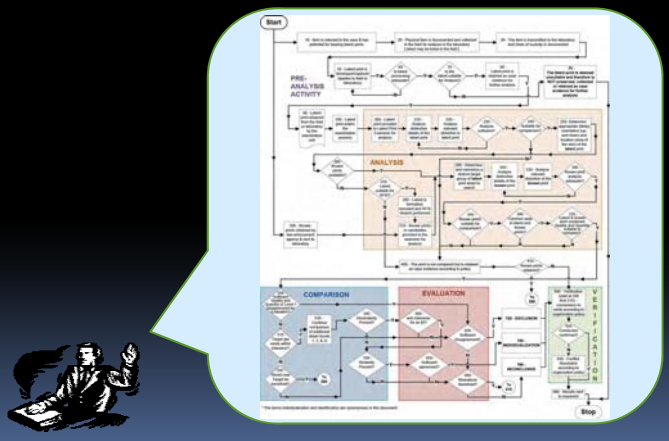
Log L	Verbal tag
0	no support
0.3 to 1	weak
2 to 3	moderate
2 to 3	strong
4 to 6	very strong
>6	extremely strong

Approved of in NRC 2009

Evaluations

Item #	Description	# of prints	Results of comparison with THOMAS SMITH
1	Demand note beginning, "I have a gun…"	2	Strong support for S
2	Bank of Los Angeles withdrawal form	1	Strong support for D
3	Pen with chain	0	N/A
4	Lift indicated as coming from customer counter	1	Strong support for D

What methodology did you use to arrive at your opinion or statistical assessment?



If the error rate is unknown, is that because of insufficient data, differences over methodology, too many variables to measure, too much subjectivity in the approach, or what?

- Subjectivity does not preclude a determination of error probabilities in classifications.
- The number of variables makes it difficult to measure error rates under all conditions of interest.
- Of course error rates are only a slice of *Daubert*. We know from controlled experiments that fingerprint features are highly variable and that examiners can detect these features.

References

- DH Kaye, DE Bernstein & JL Mnookin, The New Wigmore: A Treatise on Evidence: Expert Evidence (2d ed. 2011) (updated annually)
- DH Kaye, Likelihoodism, Bayesianism, and a Pair of Shoes, 53 Jurimetrics J. 1-9 (2012)
- DH Kaye, TM Vyvial & DL Young, Validating the Probability of Paternity, 31 Transfusion 823–828 (1991)
- G Zadora, A Martyna, D Ramos & C Aitken, Statistical Analysis in Forensic Science: Evidential Value of Multivariate Physicochemical Data 181-217 (2014)