

Research, courtroom, and media: a forensic statistician's perspective

Mikkel Meyer Andersen

Assoc Prof (Aalborg U) and Affil Assoc Prof (U of Copenhagen)

May 5, 2026

- Since 2010: Research in statistics for forensic genetics (PhD in 2014)
- Since 2021: Subject-matter expert in evaluating DNA evidence in court cases
- Since 2022: “Expert” in media about DNA’s role in solving crime

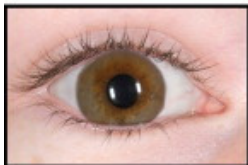
Henrik Støvring on May 4, 2026 (paraphrased?):

“If people simply provided the data required...”

Forensic statistics



1. Evidence: thief has **blue** eyes
2. Suspect has **brown** eyes



1. Evidence: thief has **blue** eyes
2. Suspect has **blue** eyes

Image source: "IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information", [10.1016/j.fsigen.2010.02.004](https://doi.org/10.1016/j.fsigen.2010.02.004)

Evidential weight

- H_p : **p**rosecutor's hypothesis: Person of interest (Pol) is donor
- H_a : **a**lternative hypothesis: A random man is the donor of the crime-scene stain
- E : Evidence

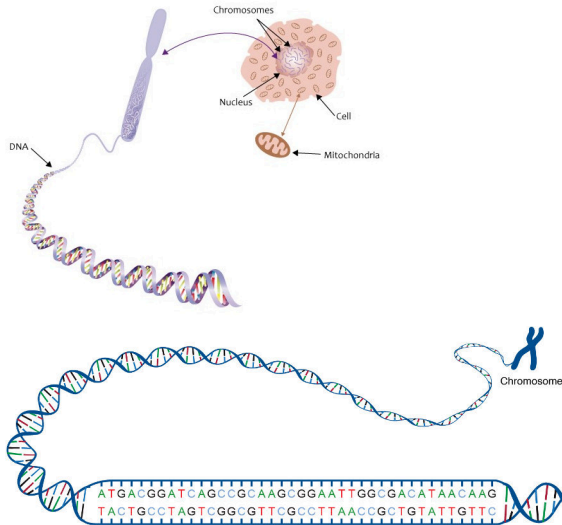
$$\underbrace{\frac{P(H_p | E)}{P(H_a | E)}}_{\text{Posterior odds}} = \underbrace{\frac{P(E | H_p)}{P(E | H_a)}}_{LR(BF)} \times \underbrace{\frac{P(H_p)}{P(H_a)}}_{\text{Prior odds}} \quad (1)$$

$LR(BF)$: Factor (weight) that perception of donorship must be increased when E is taken into account

Forensic genetics

- Trace sample left by perpetrator
- Reference sample from suspect (?)

DNA



Source: National Human Genome Research Institute, <https://www.genome.gov/>

Research, courtroom, and media: a forensic statistician's perspective



Source: Wikipedia, <https://www.wikipedia.org/>

Types of DNA profiles

- Short tandem repeat (STR)
- Single nucleotide polymorphism (SNP)

...from:

- Autosomes (chr1-22)
- Y-chromosome (males)
- mtDNA

Short tandem repeat (STR) DNA profiles

D1S1656:

Allele 10 (total length 164 bp):

...GTGTTGCTCAAGGGTCAACT... [TAGA]_10 [TG]_5 ...GAGAAATAGAATCACTAGGGAACC...
20 bp 40 bp 10 bp 25 bp

Allele 11 (total length 168 bp):

...GTGTTGCTCAAGGGTCAACT... [TAGA]_11 [TG]_5 ...GAGAAATAGAATCACTAGGGAACC...
20 bp 44 bp 10 bp 25 bp

Alleles 10-18 (?)

STR fragment lengths 100-400 bp

Trace sample quality

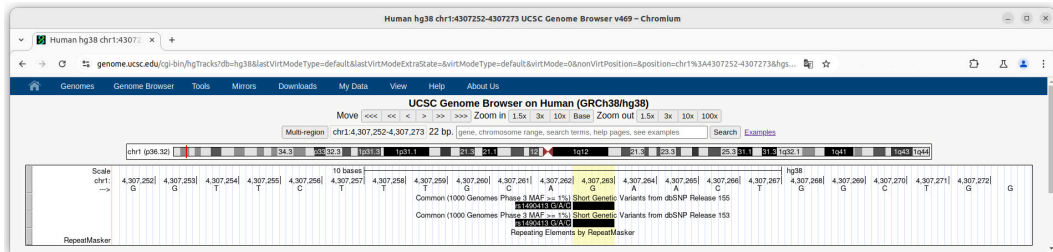
OL Meyer et. al (2025) *Nationwide study on forensic genetic analyses in criminal cases in Denmark:*

“In 2016-2022, 50 % of the samples were deemed unsuitable for standard PCR-CE analyses due to insufficient DNA quantity or quality.”

E.g. hair.

Single nucleotide polymorphism (SNP) DNA profiles

rs1490413 (chr1:4.307.263 out of 249 mio.):



Danes: C (51%) / T (49%)

Humans are diploid (one allele from mother and one from father)

E.g. shotgun sequencing (whole-genome sequencing with massively parallel sequencing)

DNA profiles from SNPs

Simple model (2024) handling low quality trace samples; collaborators: ML Kampmann, AH Jepsen, N Morling, PS Eriksen, C Børsting & JD Andersen

Binary SNPs:

- 0: Reference allele (e.g. using reference sequence GRCh38)
- 1: Alternative allele

Genotypes are denoted by the number of alternative alleles:

- 0 (0/0) with probability p_0
- 1 (0/1 or 1/0) with probability p_1
- 2 (1/1) with probability p_2

$$p_0 + p_1 + p_2 = 1$$

Errors not possible: LR for a single SNP marker

- $\mathbf{X}_T$: Observed genotype (0, 1, or 2) from **trace** donor (crime scene trace)
- $\mathbf{X}_R$: Observed genotype (0, 1, or 2) from **reference** donor (person of interest)

$$LR = \frac{P(E \mid H_p)}{P(E \mid H_a)}$$

- $E = \{\mathbf{X}_T, \mathbf{X}_R\}$: The evidence, the observed genotypes
- H_p : The Pol is the donor of the crime-scene stain
- H_a : A random man is the donor of the crime-scene stain

Errors not possible:

$\mathbf{X}_T$	$\mathbf{X}_R$	LR
0	0	$\frac{1}{p_0}$
0	1	0
0	2	0
1	0	0
1	1	$\frac{1}{p_1}$
1	2	0
2	0	0
2	1	0
2	2	$\frac{1}{p_2}$

Shotgun sequencing data from real samples

Buccal swap from same individual in duplicate

8,596,131 positions:

	0	1	2
0	8,580,543	400	61
1	671	8,317	96
2	79	61	5,903

Modelling errors in sequencing process

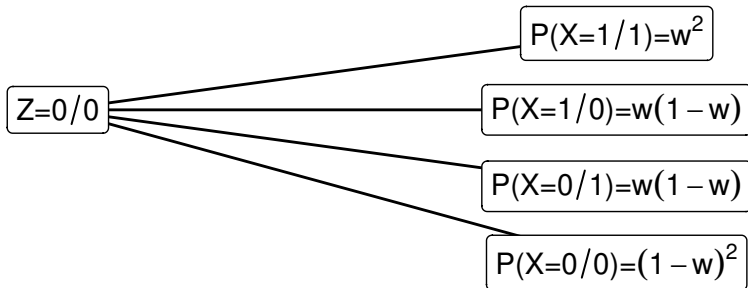
$$w = P(\text{genotyping error})$$

e.g. error in sequencing, genotype calling, ...

- $Z \in \{0, 1, 2\}$ is the true, unobserved, genotype
- $\mathbf{X} \in \{0, 1, 2\}$ is the observed genotype (potentially with error)

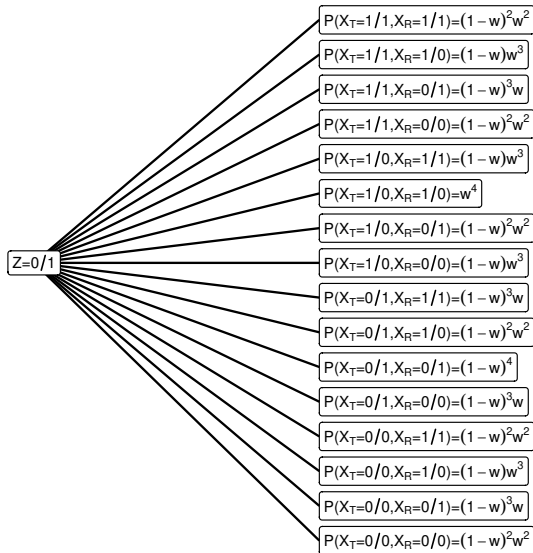
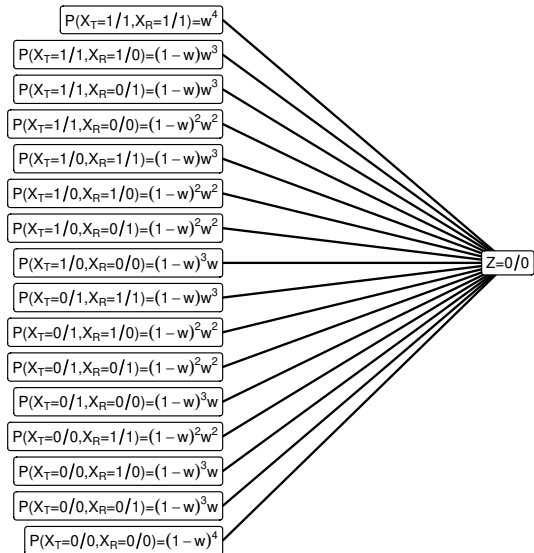
Singular sample from one individual (excerpt for $Z = 0/0$)

- $w = P(\text{genotyping error})$
- Z : true, unobserved, genotype
- X : observed genotype (potentially with error)

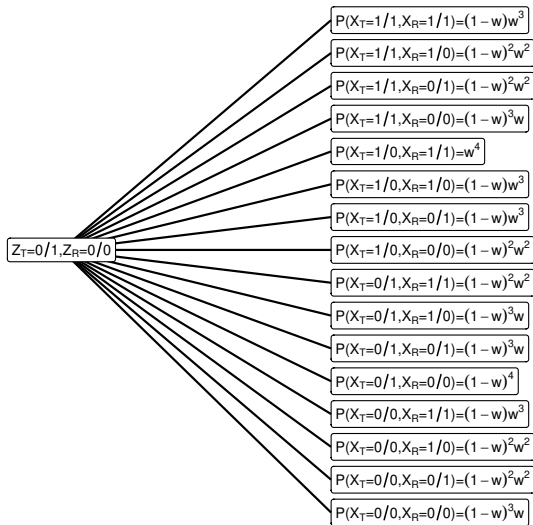
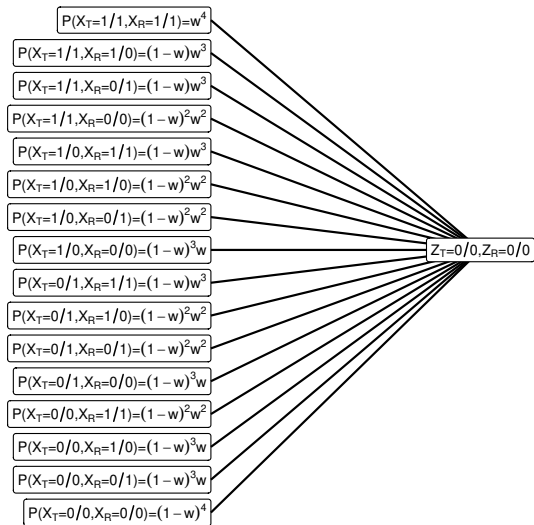


Extend to two samples from same individual (H_p) and one sample for each of two different individuals (H_a)

Duplicate sample from one individual (H_p)



Singular samples from two individuals (H_a)



LR for a single SNP marker

No errors:

$w = P(\text{genotyping error})$:

X_T	X_R	LR
0	0	$\frac{1}{p_0}$
0	1	0
0	2	0
1	0	0
1	1	$\frac{1}{p_1}$
1	2	0
2	0	0
2	1	0
2	2	$\frac{1}{p_2}$

x_T	x_R	LR
0	0	$\frac{p_0(1-w)^4 + p_1 w^2(1-w)^2 + p_2 w^4}{p_0^2(1-w)^4 + 2p_0 p_1 w(1-w)^3 + 2p_0 p_2 w^2(1-w)^2 + p_1^2 w^2(1-w)^2 + 2p_1 p_2 w^3(1-w) + p_2^2 w^4}$
0	1	$\frac{2p_0 w(1-w)^3 + p_1 w^3(1-w) + p_1 w(1-w)^3 + 2p_2 w^3(1-w)}{2p_0^2 w(1-w)^3 + 3p_0 p_1 w^2(1-w)^2 + p_0 p_1 (1-w)^4 + 2p_0 p_2 w^3(1-w) + 2p_0 p_2 w(1-w)^3 + p_1^2 w^3(1-w) + p_1^2 w(1-w)^3 + 2p_1 p_2 w^4(1-w) + p_2^2 w^4(1-w)}$
0	2	$\frac{p_0 w^2(1-w)^2 + p_1 w^2(1-w)^2 + p_2 w^2(1-w)^2}{p_0^2 w^2(1-w)^2 + p_0 p_1 w^3(1-w) + p_0 p_1 w(1-w)^3 + p_0 p_2 w^4 + p_0 p_2 (1-w)^4 + p_1^2 w^2(1-w)^2 + p_1 p_2 w^3(1-w) + p_2^2 w^2(1-w)^2}$
1	0	$\frac{2p_0 w(1-w)^3 + p_1 w^3(1-w) + p_1 w(1-w)^3 + 2p_2 w^3(1-w)}{2p_0^2 w(1-w)^3 + 3p_0 p_1 w^2(1-w)^2 + p_0 p_1 (1-w)^4 + 2p_0 p_2 w^3(1-w) + 2p_0 p_2 w(1-w)^3 + p_1^2 w^3(1-w) + p_1^2 w(1-w)^3 + 2p_1 p_2 w^4(1-w) + p_2^2 w^4(1-w)}$
1	1	$\frac{4p_0 w^2(1-w)^2 + p_1 w^4 + 2p_1 w^2(1-w)^2 + p_1 (1-w)^4 + 4p_2 w^4(1-w)}{4p_0^2 w^2(1-w)^2 + 4p_0 p_1 w^3(1-w) + 4p_0 p_1 w(1-w)^3 + 8p_0 p_2 w^2(1-w)^2 + p_1^2 w^4 + 2p_1^2 w^2(1-w)^2 + p_1^2 (1-w)^4 + 4p_1 p_2 w^3(1-w) + 4p_1 p_2 w(1-w)^3 + 4p_2^2 w^4(1-w) + 4p_2^2 w^2(1-w)^2}$
1	2	$\frac{2p_0 w^3(1-w) + p_1 w^3(1-w) + p_1 w(1-w)^3 + 2p_2 w(1-w)}{2p_0^2 w^3(1-w) + p_0 p_1 w^4 + 3p_0 p_1 w^2(1-w)^2 + 2p_0 p_2 w^3(1-w) + 2p_0 p_2 w(1-w)^3 + p_1^2 w^3(1-w) + p_1^2 w(1-w)^3 + 2p_1 p_2 w^4(1-w) + 2p_1 p_2 w^2(1-w)^2}$
2	0	$\frac{p_0 w^2(1-w)^2 + p_1 w^2(1-w)^2 + p_2 w^2(1-w)^2}{p_0^2 w^2(1-w)^2 + p_0 p_1 w^3(1-w) + p_0 p_1 w(1-w)^3 + p_0 p_2 w^4 + p_0 p_2 (1-w)^4 + p_1^2 w^2(1-w)^2 + p_1 p_2 w^3(1-w) + p_2^2 w^2(1-w)^2}$
2	1	$\frac{2p_0 w^3(1-w) + p_1 w^3(1-w) + p_1 w(1-w)^3 + 2p_2 w(1-w)}{2p_0^2 w^3(1-w) + p_0 p_1 w^4 + 3p_0 p_1 w^2(1-w)^2 + 2p_0 p_2 w^3(1-w) + 2p_0 p_2 w(1-w)^3 + p_1^2 w^3(1-w) + p_1^2 w(1-w)^3 + 2p_1 p_2 w^4(1-w) + 2p_1 p_2 w^2(1-w)^2}$
2	2	$\frac{p_0 w^4 + p_1 w^2(1-w)^2 + p_2 (1-w)^4}{p_0^2 w^4 + 2p_0 p_1 w^3(1-w) + 2p_0 p_2 w^2(1-w)^2 + p_1^2 w^2(1-w)^2 + 2p_1 p_2 w(1-w)^3 + p_2^2 (1-w)^4}$

LR for a single SNP marker

A numerical example:

- $p_0 = 0.0625$
- $p_1 = 0.375$
- $p_2 = 0.5625$

$w = 0$

X_T	X_R	LR
0	0	16
0	1	0
0	2	0
1	0	0
1	1	2.67
1	2	0
2	0	0
2	1	0
2	2	1.78

$w = 1 : 100,000$

X_T	X_R	LR
0	0	16
0	1	0.0002133
0	2	2.844e-09
1	0	0.0002133
1	1	2.666
1	2	7.111e-05
2	0	2.844e-09
2	1	7.111e-05
2	2	1.778

Multiple SNP markers

Independent markers chosen to enable multiplying the individual LR 's (e.g. 100-200 SNP markers)

Without error model: single mismatch gives $LR = 0$

Estimating w

Buccal swap from same individual in duplicate w/ 8,596,131 positions:

	0	1	2
0	8,580,543	400	61
1	671	8,317	96
2	79	61	5,903

- Duplicate samples from same individual (H_p) - no dependence on genotype probabilities (p_0, p_1, p_2)
 - Trinomial distribution (diagonal, off-diagonal, and corners)
- Use MLE/Bayesian handling uncertainty in genotype probabilities (p_0, p_1, p_2)

$$\hat{w} \approx 1 : 20,000$$

Asymmetric error rates

- Forensic genetics:
Different trace sample error rate (w_T) and reference sample error rate (w_R)
- w_R : known
- w_T : unknown

Extended model with parameters w_R and w_T (instead of just w)

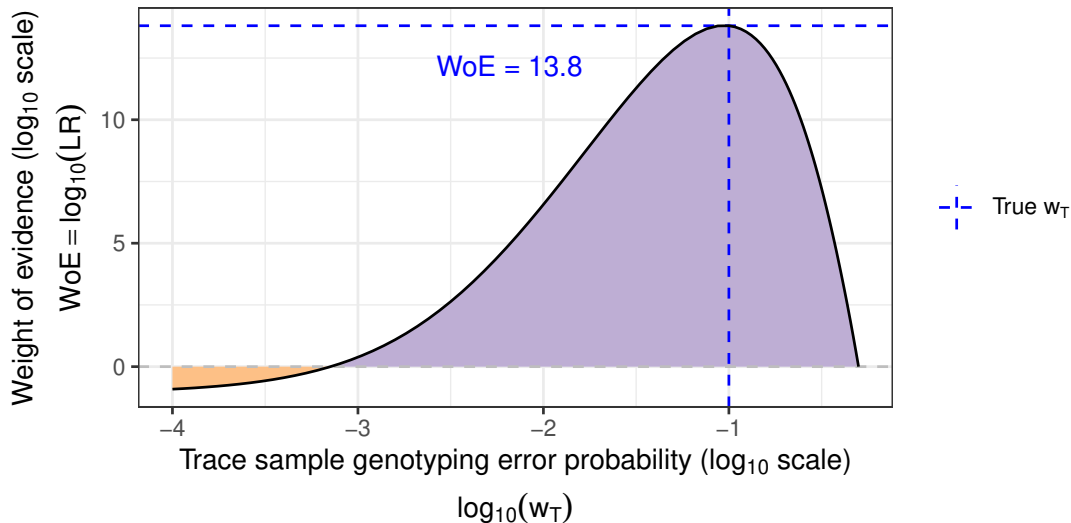
Example with 100 SNP markers

- Simulated under H_p
- $w_R = 1 : 100$ (assumed known)
- True (unknown) $w_T = 1 : 10$

X_T	X_R		
	0	1	2
0	2	2	.
1	1	37	11
2	.	5	42

(. = 0)

Example with 100 SNP markers



Handling unknown w_T

Marginalising out w_T according to a prior distribution with density $p(w_T)$ of w_T :

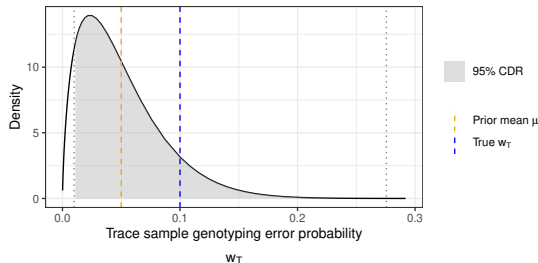
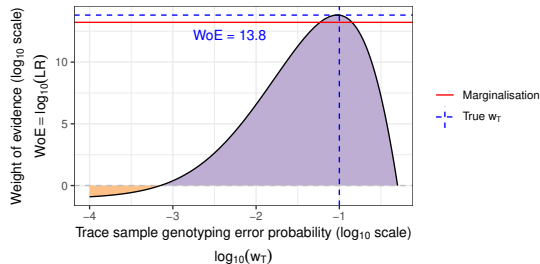
$$P(E \mid H) = \int P(E \mid H, w_T) p(w_T) dw_T \quad (2)$$

$$LR(BF) = \frac{P(E \mid H_p)}{P(E \mid H_a)} = \frac{\int P(E \mid H_p, w_T) p(w_T) dw_T}{\int P(E \mid H_a, w_T) p(w_T) dw_T} \quad (3)$$

Alternatively: Maximise likelihoods in numerator and/or denominator

Example

- Beta on $(0, \frac{1}{2})$
- Prior mean $\mu = 1 : 20$
- Prior variance $\sigma^2 = \mu^2/2 = 1 : 800$
- 95% central density region (CDR)
 - $(0.0099331; 0.2756009)$
 - $(1 : 100.67; 1 : 3.63)$



WoE by marginalisation = 13.22

References

Collaborators: ML Kampmann, AH Jepsen, N Morling, PS Eriksen, C Børsting & JD Andersen

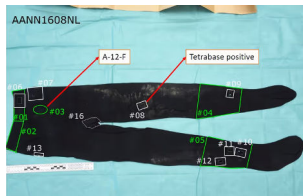
Shotgun DNA sequencing for human identification: Dynamic SNP selection and likelihood ratio calculations accounting for errors (2024),
[10.1016/j.fsigen.2024.103146](https://doi.org/10.1016/j.fsigen.2024.103146)

Shotgun DNA sequencing evidence: Sample-specific and unknown genotyping error probabilities (2026),
[10.1016/j.fsigen.2026.103474](https://doi.org/10.1016/j.fsigen.2026.103474)

wgsLR R package available @
github.com/mikldk/wgsLR

- Estimate w
 - Maximum likelihood (confidence int.)
 - Bayesian inference (posterior dist.)
- Calculate LR values
- Online documentation:
mikldk.github.io/wgsLR

Norwegian case: Birgitte Tengs (1978-1995)



Photos from 1) udeoghjemme.dk;

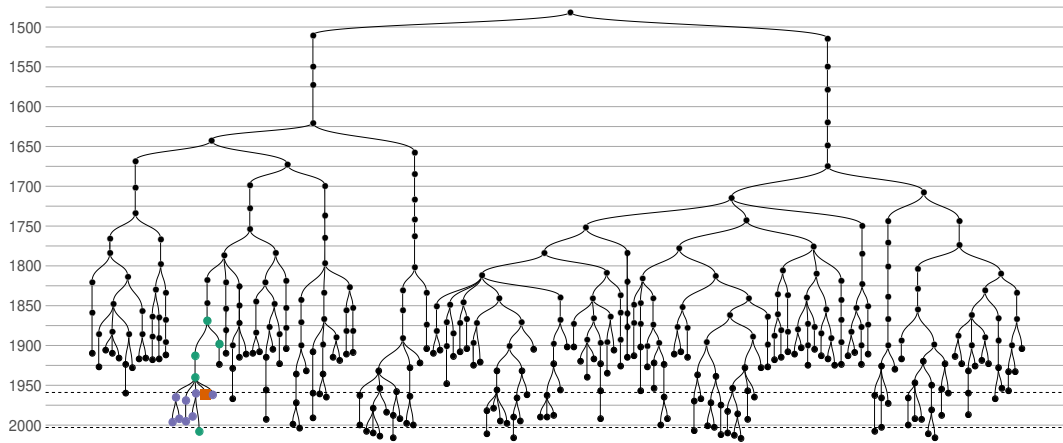
Norsk Politi via

2) nettavisen.no

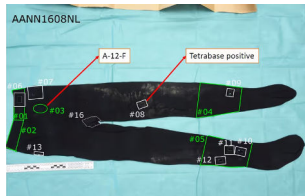
Norwegian case: Birgitte Tengs (1978-1995)

- VG (my translation), Sep 8, 2021:
Stength of DNA evidence is moderate, it has a probability rate of a bit less than 400. How is that calculate and what does it mean?
- Research with David Balding (2017, 2019)
- Norwegian Police, Sep 29, 2021: written report:
 - Y-profile in question is rare
 - $P(|\Omega| \leq 79) \geq 95\%$
 - Matching males are closely related (≤ 20 meioses)
- VG contact meanwhile - avoid writing something wrong
- Second written report, mutation discovered
 - Y-profile in question is very rare
 - $P(|\Omega| \leq 3) \geq 95\%$

Norwegian case: Birgitte Tengs (1978-1995)



Norwegian case: Birgitte Tengs (1978-1995)



Photos from 1) udeoghjemme.dk;

Norsk Politi via

2) nettavisen.no and

4) tv2.no; and

5) Stavanger Aftenblad



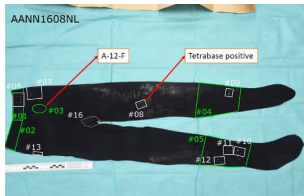
Hierarchy of propositions

- H_p : **p**rosecutor's hypothesis: Person of interest (PoI) is donor
- H_a : **a**lternative hypothesis: A random man is the donor of the crime-scene stain
- E : Evidence

$$\underbrace{\frac{P(H_p | E)}{P(H_a | E)}}_{\text{Posterior odds}} = \underbrace{\frac{P(E | H_p)}{P(E | H_a)}}_{LR} \times \underbrace{\frac{P(H_p)}{P(H_a)}}_{\text{Prior odds}} \quad (4)$$

Level	Example of H_p
Crime	The defendant raped the victim.
Activity	The defendant had sex with the victim.
Source	The semen at the crime scene originates from the defendant.
Profile	The DNA profile from the trace sample originates from the defendant.

Norwegian case: Birgitte Tengs (1978-1995)



Photos from 1) udeoghjemme.dk;

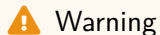
Norsk Politi via

2) nettavisen.no and

4) tv2.no; and

5) Stavanger Aftenblad





Amateur opinions incoming - nothing authoritative...

Being a researcher in 2026

Aim

- Discovering new knowledge?
- Fundamental research?
- Applied research?

“Product”

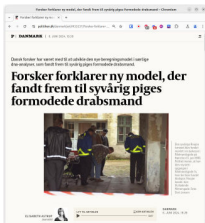
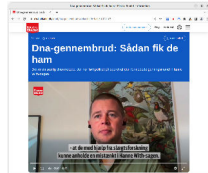
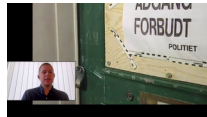
- Research (publications - what journals?)
- Teaching
- Dissemination (U's obligation at organisation level)
- Citations?
- Funding?
- Impact?

Catalysts?

- “Good” (?) journals
- Citations
- Funding
- Dissemination
 - Media
 - Talks
 - Online presence

Me as a researcher

- Do what's fun and interesting!
- Aim: Applied research, “help” society
- Product
 - Scientific publications
 - Information for the general public
 - Research/math is useful - examples
- Dissemination, for fun and for implications
 - Improve - better at explaining concepts for the general public (family gatherings...)
 - “Sell” math/stats
 - Maybe improve chances of citations/funding



Benefits

For me:

- FUN!
- Improving dissemination skills
- Help me do better (e.g. relevant) research
- Visibility / “branding” (researchers anno 2026...)
 - Journalists (Google search, ...), colleagues, bosses
- Feedback loop between the media and science (method developed and used, 3 times)

For my department, faculty, and university

- Recruitment (students / researchers)
- “Knowledge for the world”

Drawbacks

- Takes time
- Cringe? People think less of you as a researcher?
- Cannot control agenda (Norwegian appeal)
- Does it “work”? (What is the objective function...?)

My dissimulation take-aways

- Build a foundation (research + written popular science articles)
- Public talks
- Pick up the phone when journalists call...
- Do it, accept errors
- “I undervalue my contribution relative to others”
- Do it!