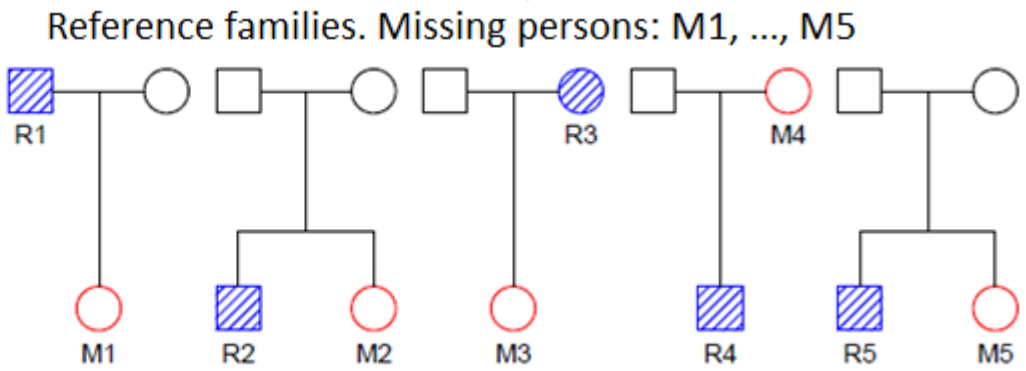
Forensic applications II



Thore Egeland

NORBIS June 16, 2022

Contents

- Power
- Bayesian approach
  - Including prior, non-DNA information
  - Controversial also in forensics
- Missing Person Identification *and* Disaster Victim Identification (DVI)
  - **library(dvir)**
  - **missingPersonIP** # IP = inclusion power
  - **missingPersonEP** # EP = exclusion power
  - **jointDVI**

Power

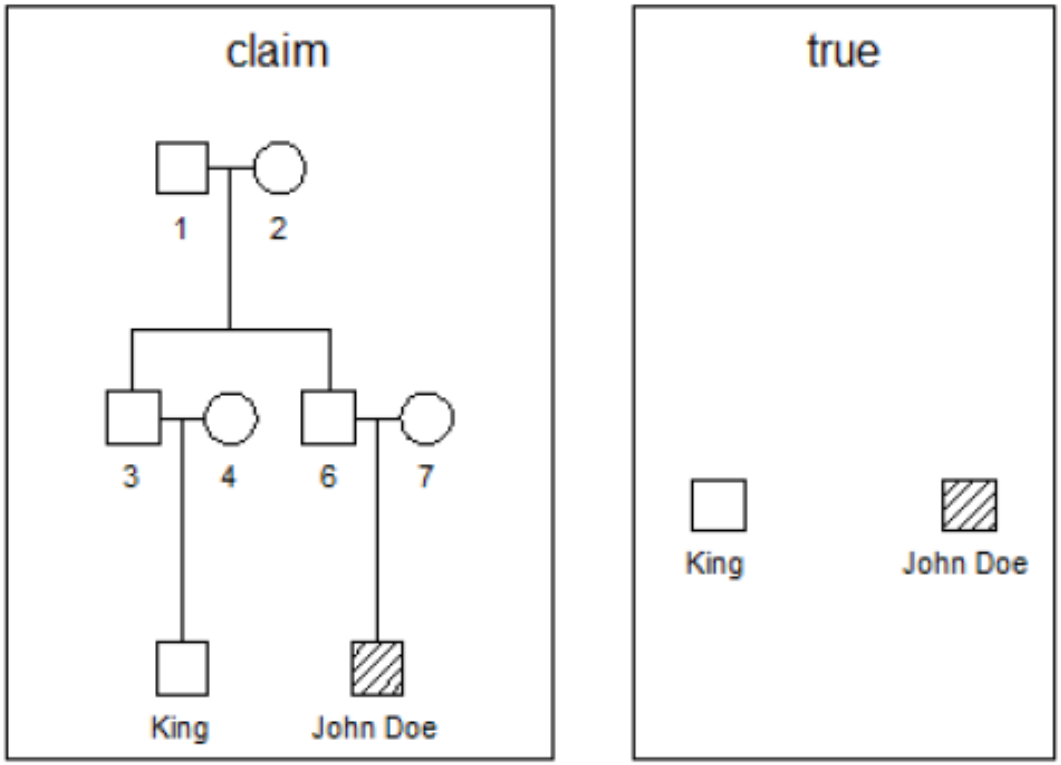
Generally

- Power calculations can be used to determine sample size

Forensic genetics

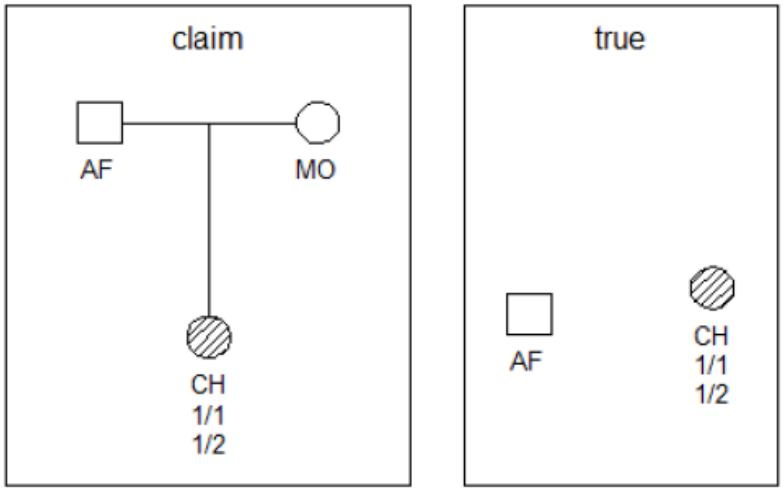
- How many and who should we genotype?
- How many, which markers should be used?

Generic example



What data do we need to exclude
John Doe as the first cousin of the King
given that he is unrelated?

ExclusionPower (EP): Two equifrequent SNPs



$$EP = P(\text{"claim" incompatible with genotypes} \mid \text{"true"})$$
$$EP_1 = P(g_{AF} = 2/2) = 0.5^2 = 0.25,\ EP_2 = 0$$
$$EP = 1 - (1 - EP_1) \cdot (1 - EP_2) = 0.25 \text{ for both markers}$$

- `forrel::exclusionPower`

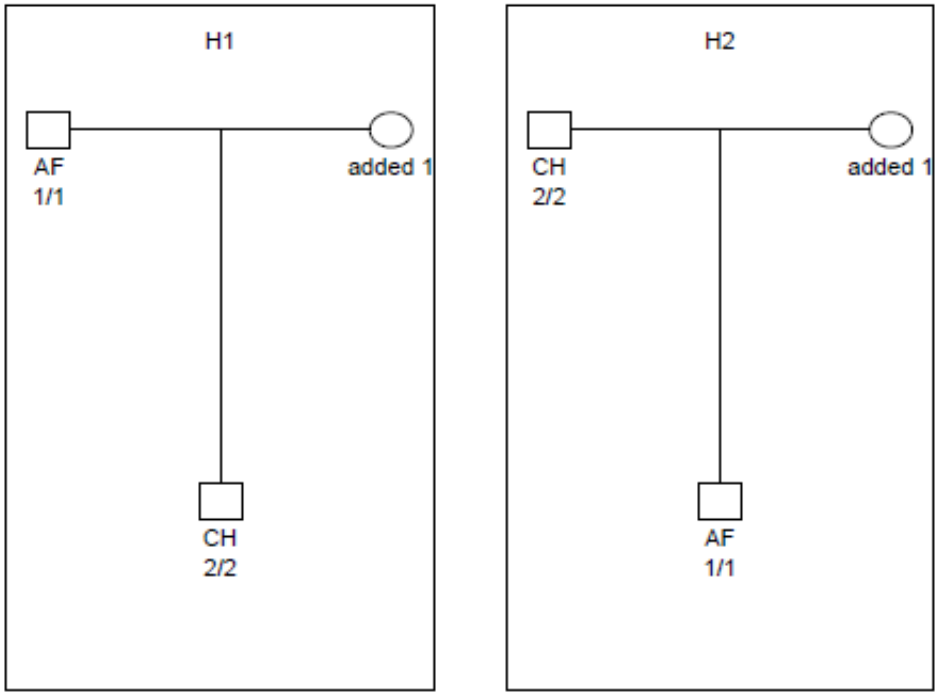
- Generally: EP = $P(\text{LR} = 0 \mid H_2)$, where LR = $P(\text{data} \mid H_1)/P(\text{data} \mid H_2)$

exclusionPower

- library(pedsuite, quietly = T)
- claim = nuclearPed(fa = "AF", mo = "MO", child = "CH", sex = 2)
- true = list(singleton("AF"), singleton("CH"))
- claim = claim |> addMarker(name = "L1", CH = "1/1", alleles = 1:2)
- claim = claim |> addMarker(name = "L2", CH = "1/2", alleles = 1:2)
- exclusionPower(claim, true, ids = "AF", verbose = F)

```
Potential mismatches: 1 (L1)
Expected mismatches: 0.25
P(at least 1 mismatch): 0.25
```

Bayesian approach: Motivation

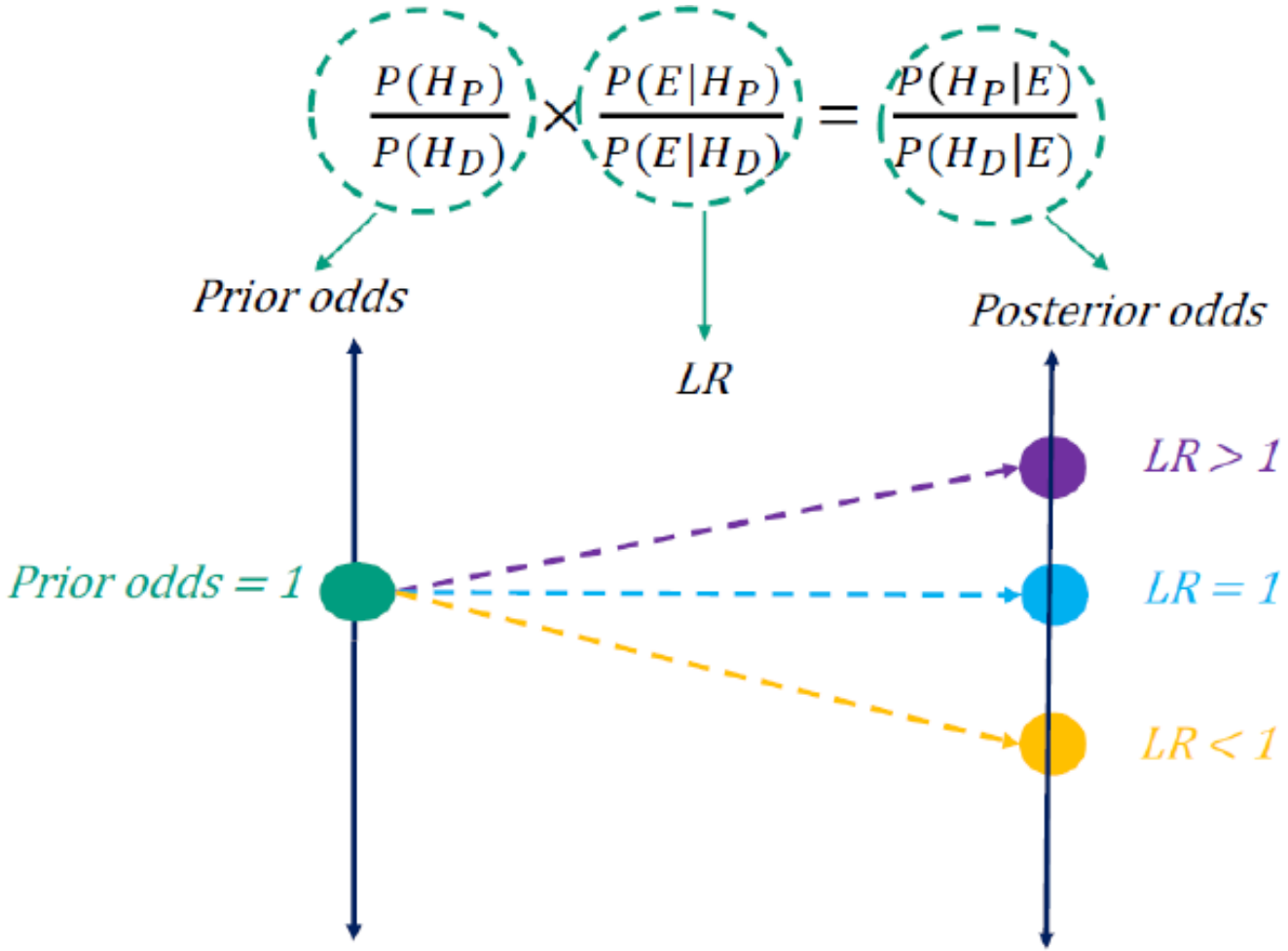


- H_1 more likely *apriori* than H_2 based on age information
- How do we include non-DNA information? Prior

Bayesian framework

- Specify $P(H_P), P(H_D)$, typically subjectively or
- Prior odds: $P(H_P)/P(H_D)$
- *Flat prior* $P(H_P) = P(H_D) = 0.5$ often used.
- I avoid using the common *uninformative prior* for flat prior.

Bayesian theorem on odds form



Prior and posterior odds: Example

Assume

- *prior odds* $\frac{P(H_1)}{P(H_2)} = 1000.$

Then

$$\text{prior odds} * \text{LR} = \text{posterior odds},$$
$$1000 * 0.66 = 666.$$

Interpretation: H_1 is 666 times more probable than H_2.

Posterior probability of paternity: Bayes Theorem

$$P(H_1 \mid E) = \frac{P(E \mid H_1)P(H_1)}{P(E \mid H_1)P(H_1) + P(E \mid H_2)P(H_2)}$$
$$= \text{"Probability of } H_1 \text{ given evidence"}$$

Important special forensic case: $P(H_1) = P(H_2) = 0.5$.
The Essen-Möller index for paternity:

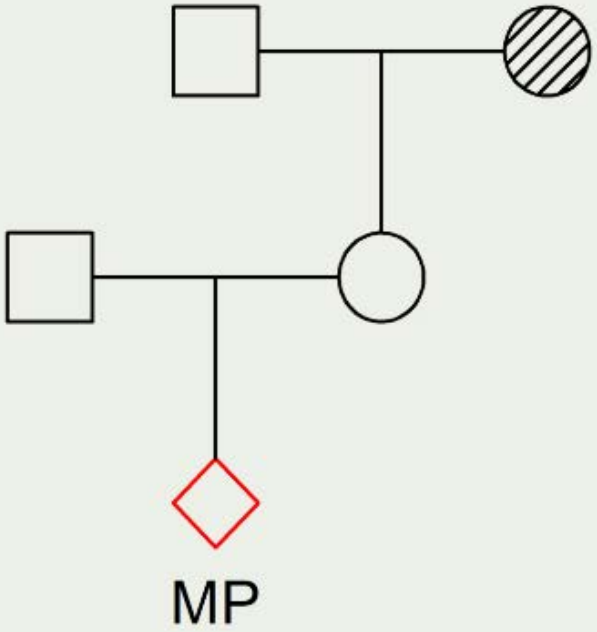
$$W = P(H_1 \mid E) = \frac{LR}{1 + LR}.$$

Allows inteligible statements like:
"The probability that he is the father is 99.73%".
Problem: the prior ...

Practical problems in forensic genetics

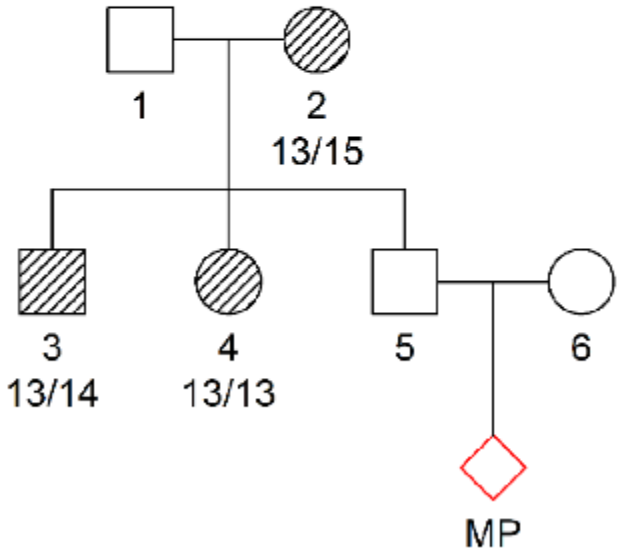
- Do we report LR, posterior probability or posterior odds?
- Or should we report on a verbal scale? Both numbers and verbal statements?
- How do we choose thresholds?

Missing Person Identification



Missing person cases: Basics

Reference family



match?

Person of interest (POI)

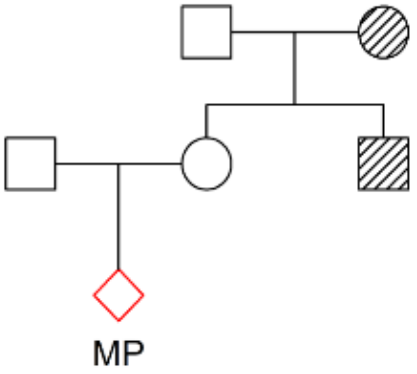
DNA-based identification

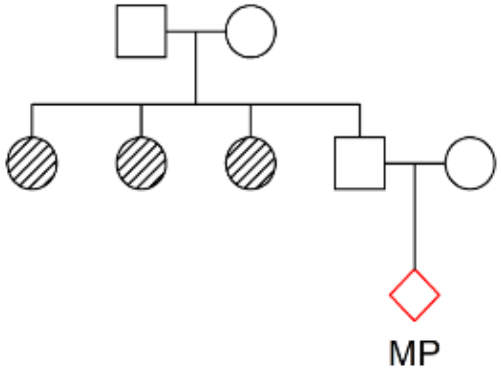
- Forensic markers
  - autosomal, X, Y, mtDNA
- Simplest with DNA from
  - the missing person
  - parents of the missing

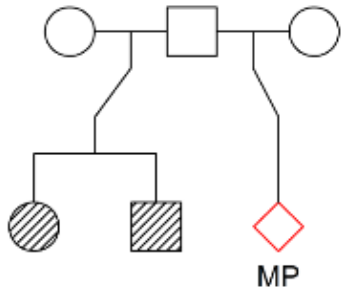
← ≈ paternity case

Standard forensic kits
- 15 - 24 autosomal STRs
- Typically 10 - 50 alleles
- Mostly unlinked

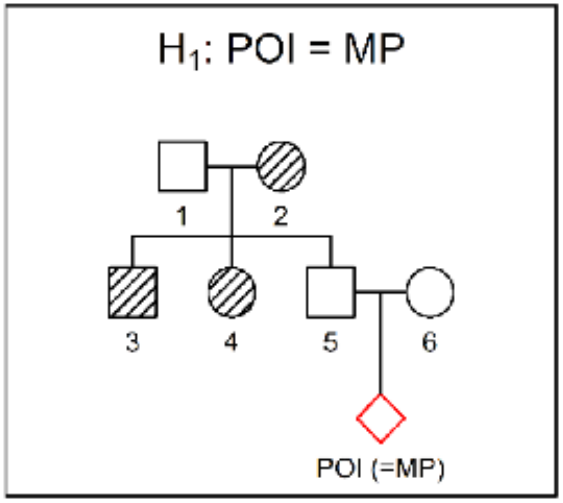
- Argentina: Parents usually unavailable

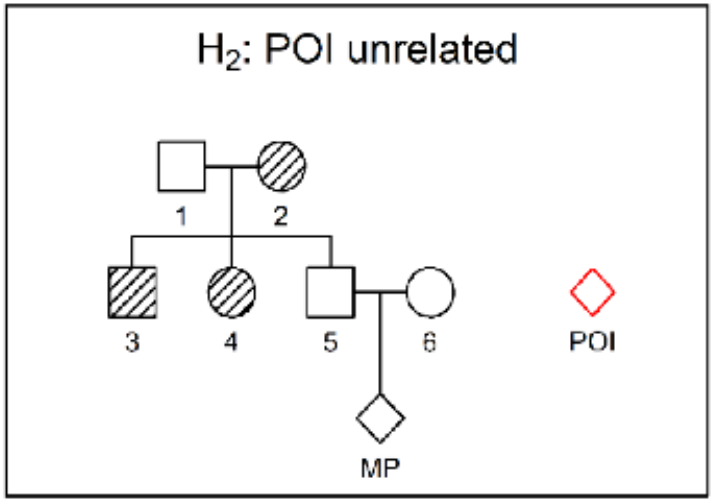






The likelihood ratio (LR)



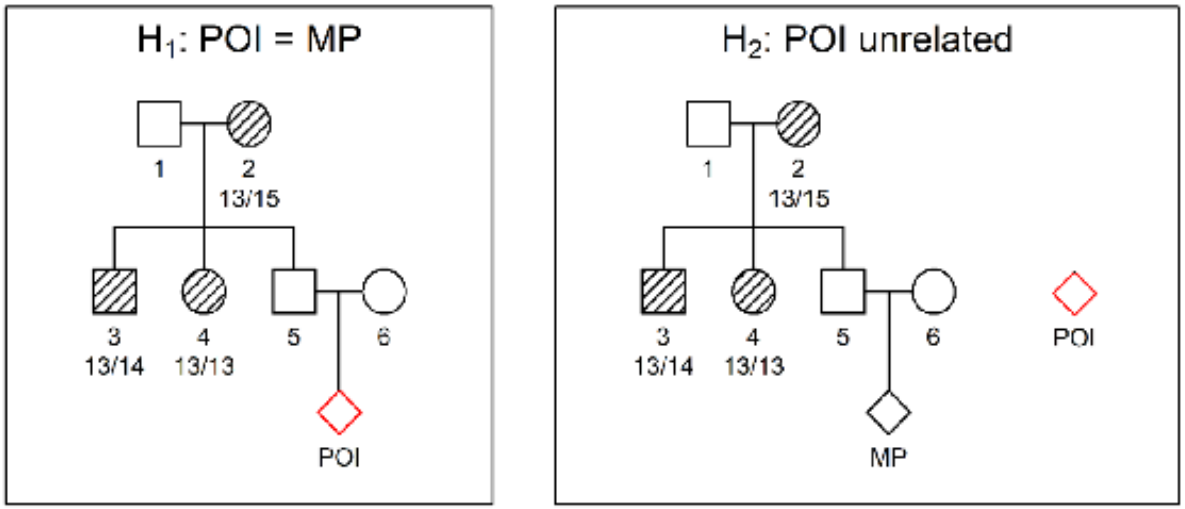


$$LR = \frac{P(\text{data} \mid H_1)}{P(\text{data} \mid H_2)}$$

Positive match if LR > 10,000*

*or other suitable threshold

Power in missing person cases

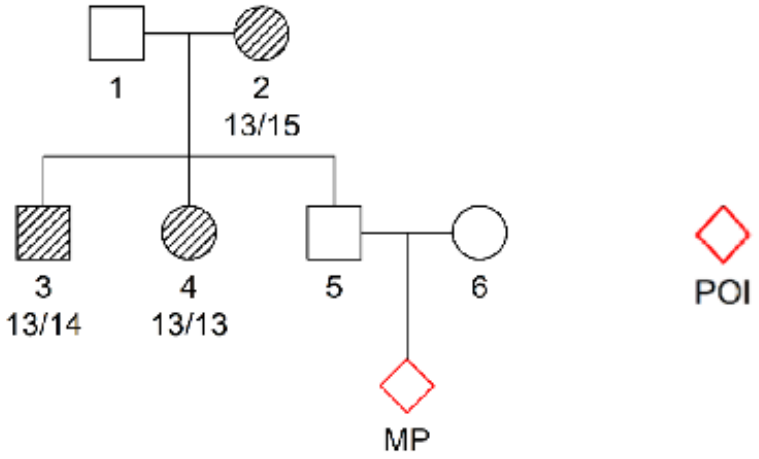


- Two complementary measures of power
  - Inclusion: The probability of recognizing the true MP
  - Exclusion: The probability of excluding an unrelated POI
- Note: Computed before POI is genotyped!

Depend on

- Reference individuals
- Reference genotypes
- Number of markers
- Allele frequencies

Inclusion power (IP)

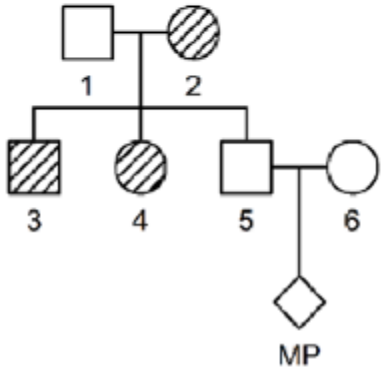


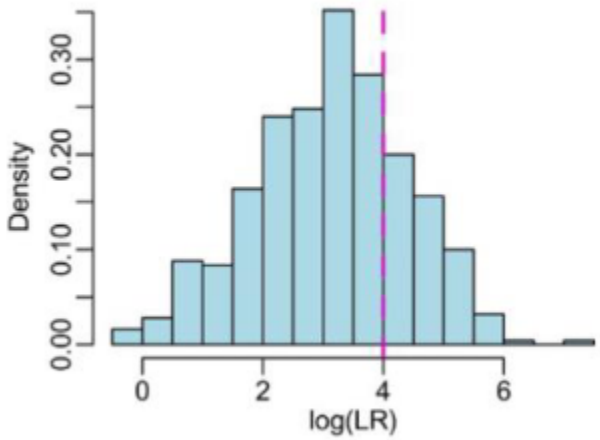
- If POI = MP: Do we have enough data to detect it?

$$IP_{10000} = P(LR > 10{,}000 \mid POI = MP)$$

- Computed by simulations of MP - condititional on the reference

Inclusion power in R





R code

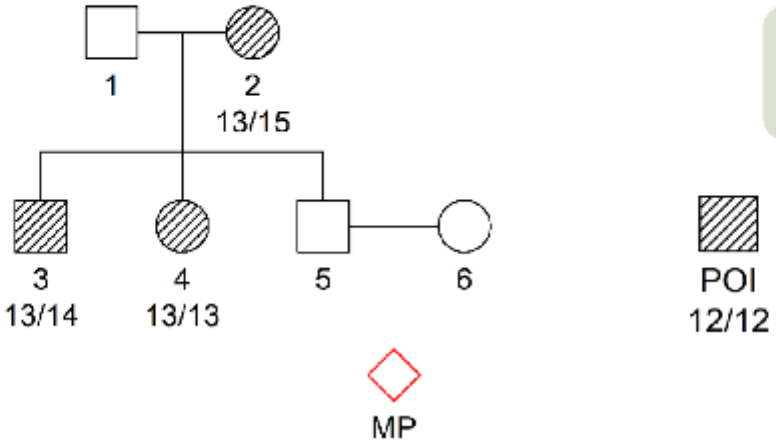
```
> library(forrel)

> ref = readFam(...)
> missingPersonIP(ref, nsim = 500,
                  threshold = 10000)

Using all 20 attached markers
Simulating 500 profiles...done
Computing likelihood ratios...done
Total time used: 9.87 secs

Mean LR: 65947.04
Mean log10(LR): 3.121
Estimated power:
  P(LR >= 10000) = 0.248
```

Exclusion power (EP)



MP cannot be 12/12!
→ POI excluded

- If POI ≠ MP: Probability of mismatch in at least 1 marker?

$$EP = P(\text{exclusion} \mid POI \text{ unrelated})$$

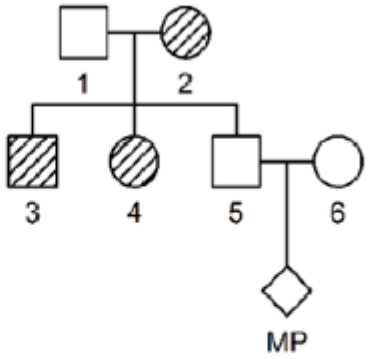
- Can be computed exactly!
  Egeland, Pinto, Vigeland (2014). *A general approach to power calculation for relationship testing*

Exclusion power in R

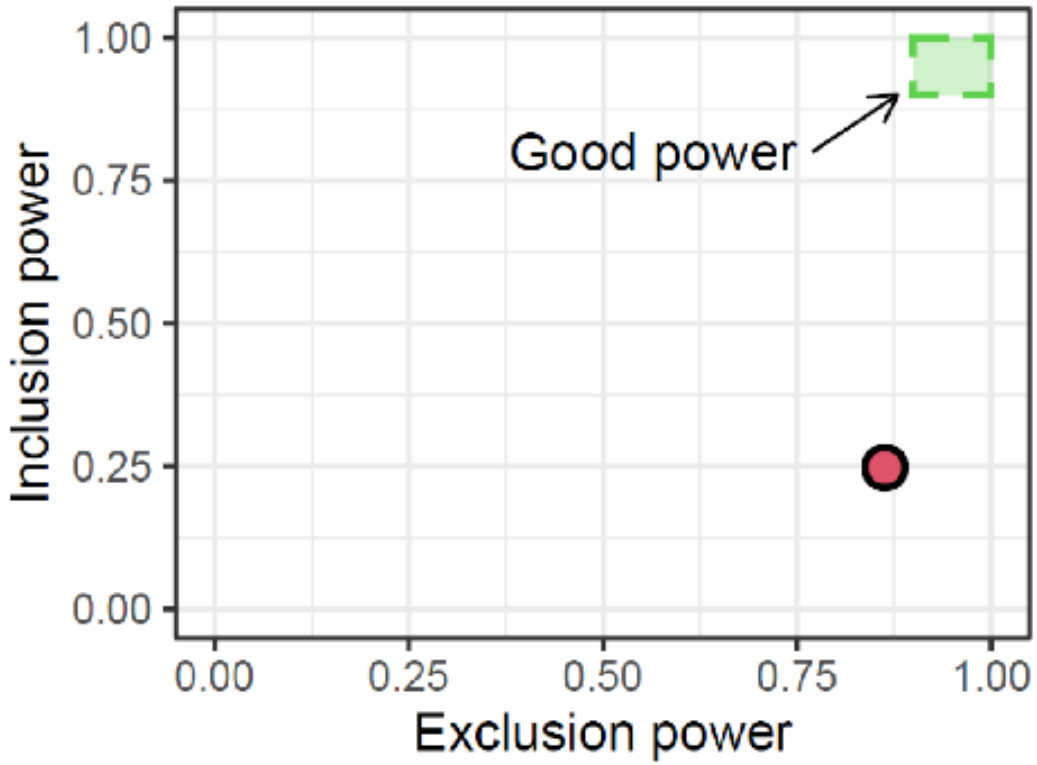
R code

```
> missingPersonEP(ref, missing = "MP")

Potential mismatches: 8 (D3S1358, D7S820, CSF1PO,
PENTA_D, VWA, TPOX, D19S433, D2S1338)
Expected mismatches: 1.679
P(at least 1 mismatch): 0.863
```

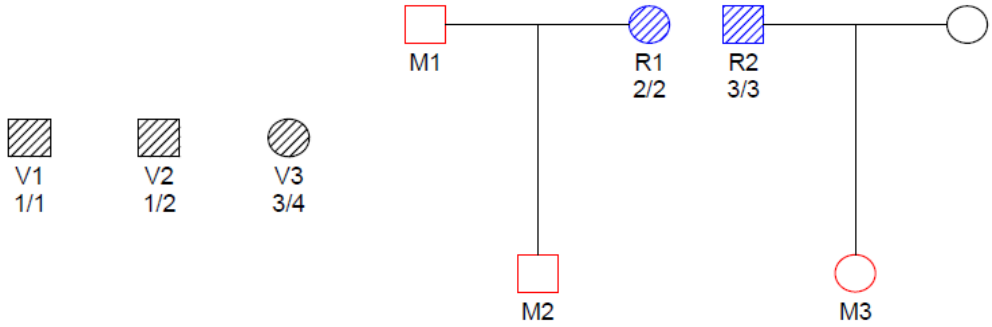


Power plot


1.00
0.75
0.50
0.25
0.00
Inclusion power
Good power
0.00
0.25
0.50
0.75
1.00
Exclusion power

Disaster Victim Identification (DVI)

- DVI
  - Match list of unidentified persons against a list of missing persons



- Cases
  - World Trade Center attack
  - Spitsbergen civil aircraft disaster
  - Balkan conflicts
  - Drowned immigrants
  - Thailand tsunami

- Traditional methods and software
  - Treat each victim or family at a time
  - Manual sequential approach
  - Vigeland, Egeland (2021): joint approach

Disaster Victim Identification in R

DVIR

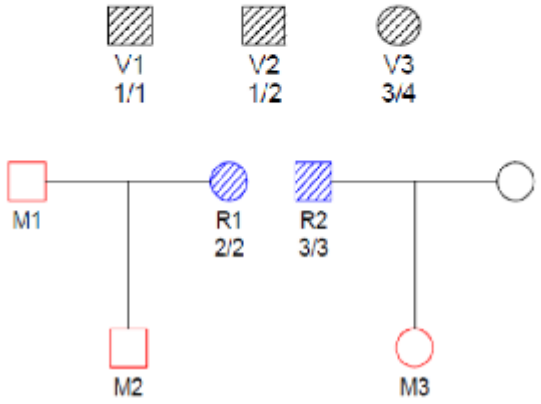
- Traditional approach/software
  - One victim or family at a time
  - Manual sequential analysis
- Further possibilities in R (**dvir**)
  - Joint analysis!
- Key functions
  - `pairwiseLR()`
  - `jointDVI()`

R code

```
> library(dvir)

> pm = example2$pm
> am = example2$am
> missing = example2$missing

> jointDVI(pm, am, missing)
```



	V_1	V_2	V_3	loglik	LR	posterior
1	M_1	M_2	M_3	-16.12	250.00	0.72
2	M_1	M_2	*	-17.73	50.00	0.14
3	*	M_2	M_3	-18.42	25.00	0.07
4	M_1	*	M_3	-20.03	5.00	0.01
5	*	M_1	M_3	-20.03	5.00	0.01
6	*	M_2	*	-20.03	5.00	0.01
7	*	*	M_3	-20.03	5.00	0.01
8	M_1	*	*	-21.64	1.00	0.00
9	*	M_1	*	-21.64	1.00	0.00
10	*	*	*	-21.64	1.00	0.00

Example: ICMP ex Jugoslavia, 9847 possible assignments

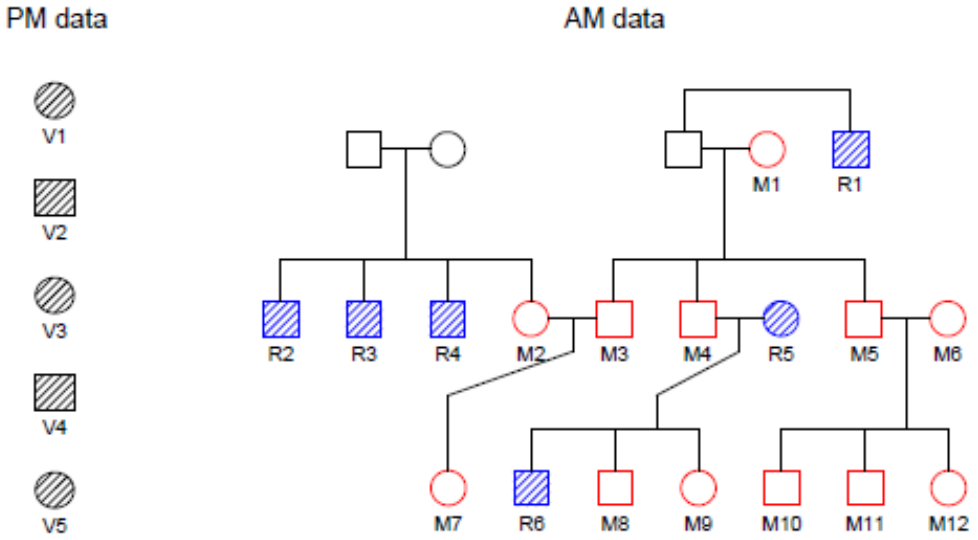


Figure 5. A large reference family with 12 missing individuals.

Sorted assignments

```
library(dvir)
pm = icmp$pm
am = icmp$am
missing = icmp$missing
jointDVI(pm, am, missing)
```

	V_1	V_2	V_3	V_4	V_5	loglik	LR	posterior
1	M_6	M_{10}	M_{12}	M_8	M_1	-312.98	1.14E+24	0.50
2	M_6	M_{11}	M_{12}	M_8	M_1	-312.98	1.14E+24	0.50
3	M_6	M_{10}	M_{12}	M_8	M_7	-327.16	7.86E+17	0.00
4	M_6	M_{11}	M_{12}	M_8	M_7	-327.16	7.86E+17	0.00
5	M_6	*	M_{12}	M_8	M_1	-327.74	4.40E+17	0.00

Table 11. The five most likely assignments for the case in Figure 5.

Posterior pairing probabilities

	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	*
V_1						1.000							
V_2										0.500	0.500		
V_3												1.000	
V_4								1.000					
V_5	1.000												

Biased selection of references

Article | Open Access | Published: 01 July 2021

Joint DNA-based disaster victim identification

Magnus D. Vigeland & Thore Egeland

Scientific Reports **11**, Article number: 13661 (2021) | Cite this article

Mariana

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Prioritising family members for genotyping in missing person cases: A general approach combining the statistical power of exclusion and inclusion

Magnus D. Vigeland [a,*], Franco L. Marsico [b], Mariana Herrera Piñero [b], Thore Egeland [c]

Daniel

Forensic Science International: Genetics

Volume 11, November 2017, Pages 57-65

Research paper

Evaluating the statistical power of DNA-based identification, exemplified by 'The missing grandchildren of Argentina'

Daniel Kling [a,b,*], Thore Egeland [b,c], Mariana Herrera Piñero [c,d], Magnus Dehli Vigeland [c,d]

Lourdes