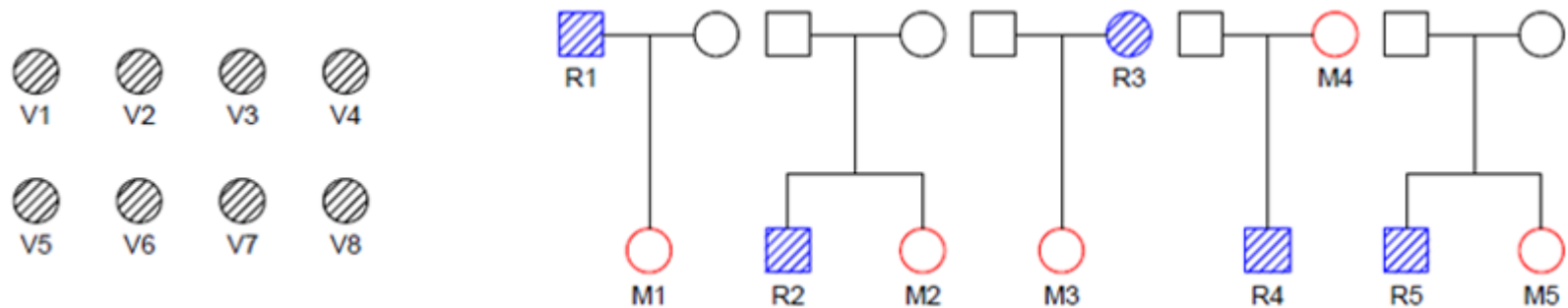




Lecture 6. Disaster victim identification and other forensic applications



Thore Egeland

Norwegian University of Life Sciences (NMBU)



The missing grandchildren of Argentina

Argentina 1976 – 1983: Dirty War

- Military dictatorship
- *War* against guerrillas
- Opponents killed or disappeared
- *About 500 children abducted*
 - kidnapped with their parents or born in captivity
 - parents killed
 - raised by police or military families

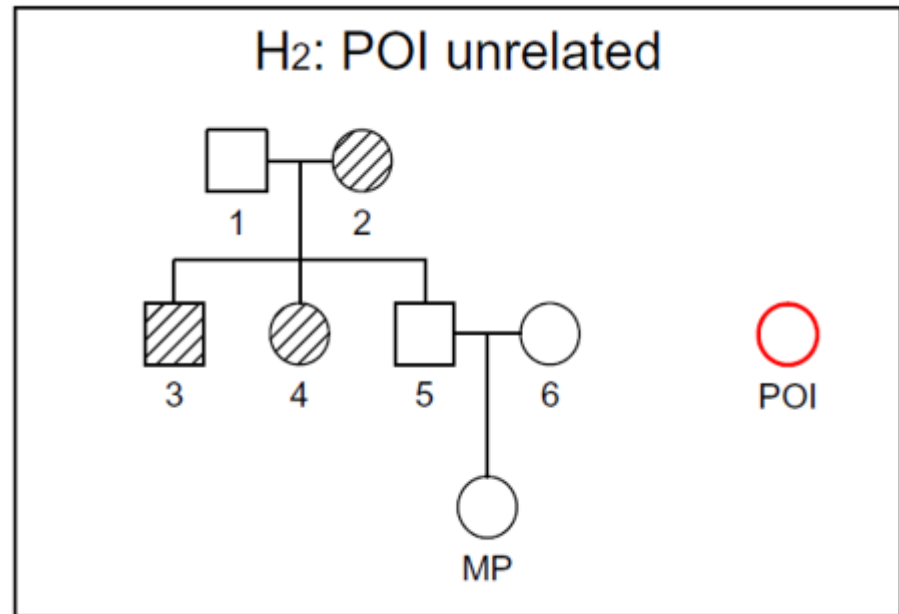
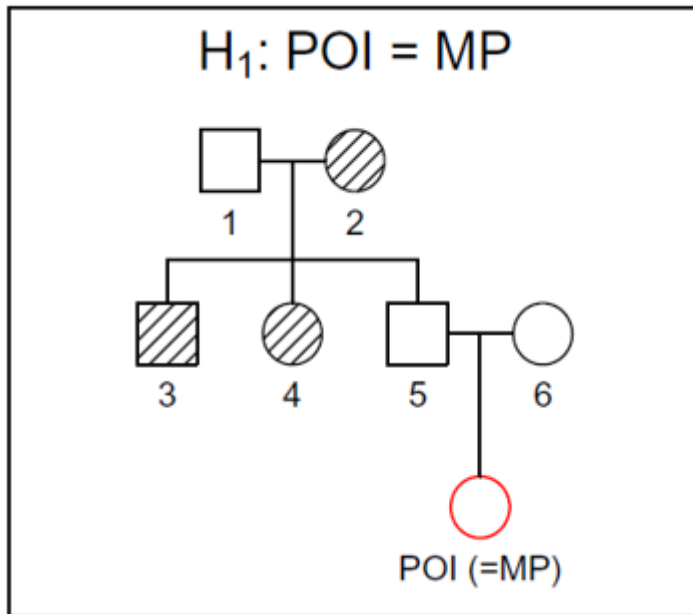


The missing grandchildren

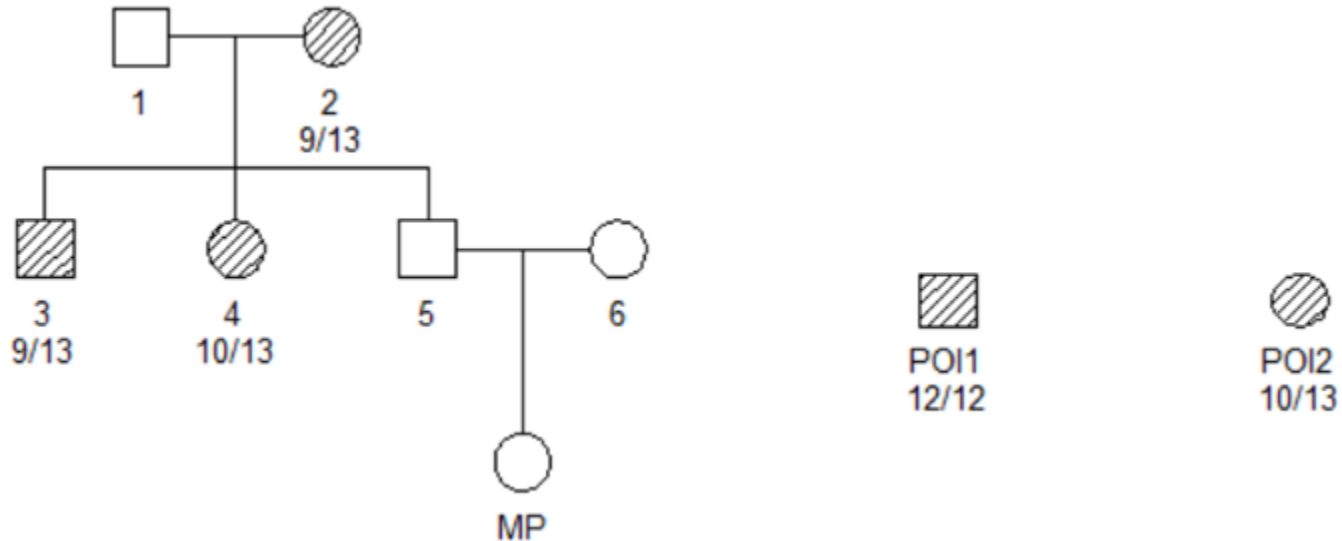
- 1977: *Grandmothers of Plaza de Mayo*
 - weekly marches ever since
 - <https://www.abuelas.org.ar>
- 1983: Mary-Claire King, US geneticist
 - Collaboration with the Grandmothers
 - HLA typing + blood groups
 - First grandchild recovered using genetics
- 1987: National Genetic Data Bank (BNDG)
- 2007: Argentine Forensic Anthropology Team
- 2021: In total 130 reunifications by BNDG



A Case Study



Exclusion and Inclusion



- POI1 = 12/12 is **excluded** if mutations are ignored since the alleles of his father are among 9/10/13
- POI2 is
 - ✓ **not excluded**
 - ✓ **included** if $LR > 10,000$ say.

Loading data

mpi-example.ped

famid	id	fid	mid	sex	D3S1358	TH01
Reference	1	0	0	1	-/-	-/-
Reference	2	0	0	2	14/17	6/8
Reference	3	1	2	1	17/18	6/8
Reference	4	1	2	2	14/17	6/8
Reference	5	1	2	1	-/-	-/-
Reference	6	0	0	2	-/-	-/-
Reference	MP	5	6	2	-/-	-/-
POI1	POI1	0	0	1	15/16	8/9.3
POI2	POI2	0	0	2	14/16	7/8

mpi-example.freq

```

D3S1358
10 → 5.21594e-05
11 → 0.0046421909
12 → 5.21594e-05
13 → 0.0032860407
14 → 0.1322240269
15 → 0.2635090535
16 → 0.2366990481
17 → 0.2040480415
18 → 0.1393700283
19 → 0.0150219031
20 → 0.0009910292
21 → 0.000104319

TH01
5 → 0.0030252481

```

```

> library(forrel)
> mpi = readPed("data/mpi-example.ped")
> mpi = setFreqDatabase(mpi,
    "data/mpi-example.freq")
> summary(mpi)

```

List of 3 `ped` objects:

```

--- component 1 ---
Pedigree with 7 members.
20 attached markers.
3 typed members.

```

```

--- component 2 ---
Singleton (male) labelled `POI1`.
20 attached markers.

```

```

--- component 3 ---
Singleton (female) labelled `POI2`.
20 attached markers.

```

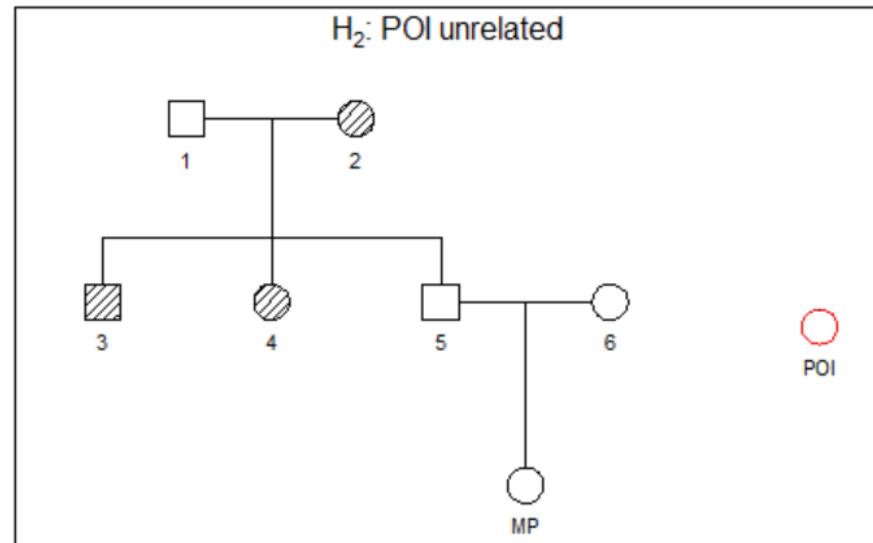
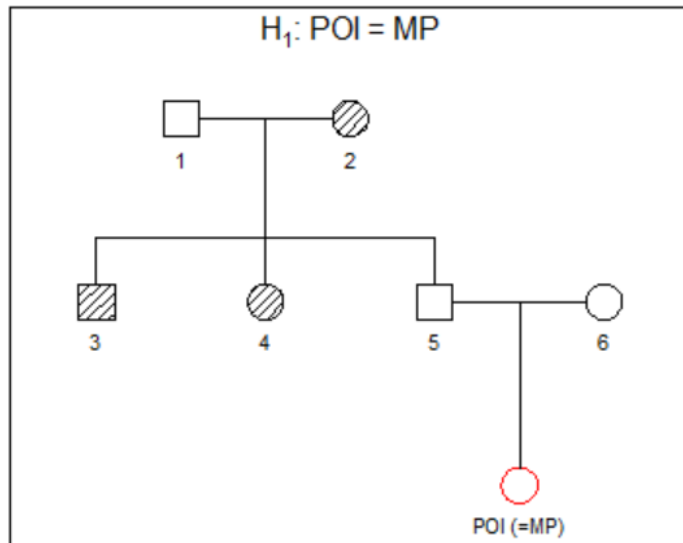
Tailor made functions



- `missingPersonPlot()`
- `missingPersonLR()`
- `missingPersonEP()`
- `missingPersonIP()`

Missing Person Plot

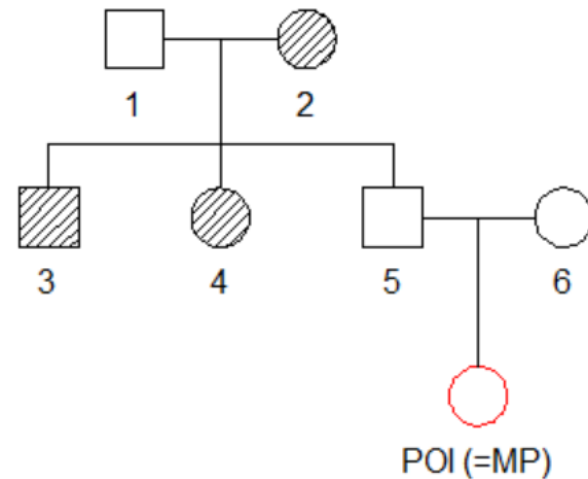
```
> ref = mpi$Reference  
> missingPersonPlot(ref, missing = "MP")
```



LR for POI1

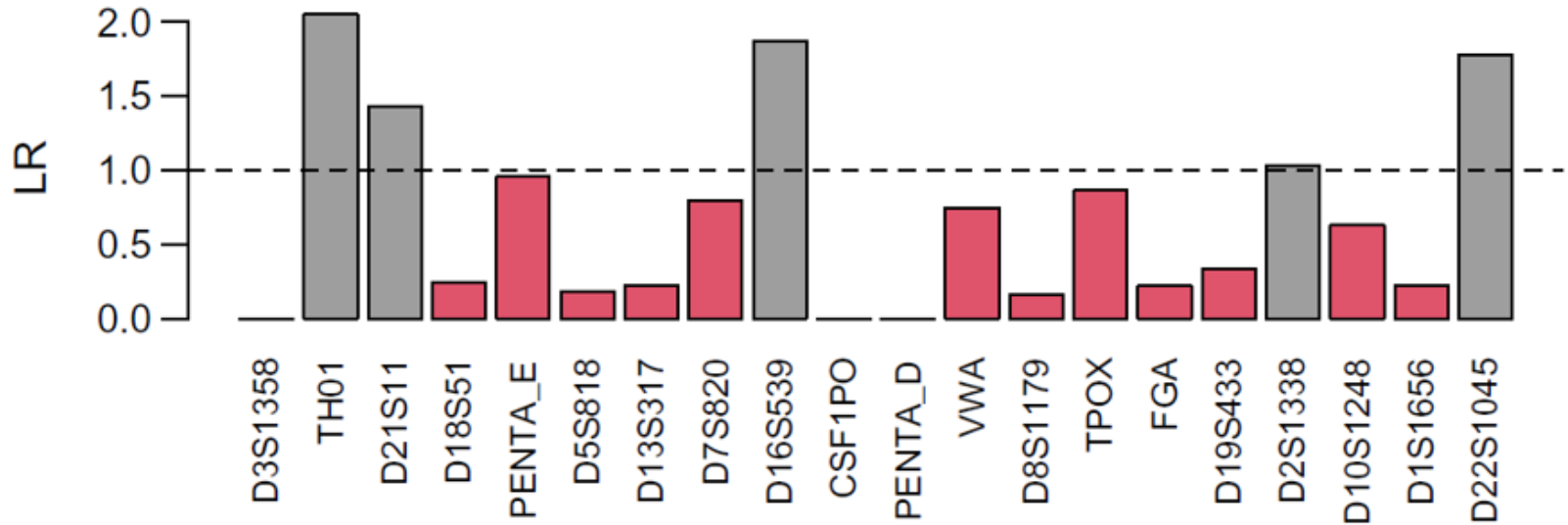
```
> ref = mpi$Reference
> poi1 = mpi$POI1
> mpiTest1 = missingPersonLR(ref, missing = "MP",
>                             poi = poi1)
> mpiTest1
```

H1:H2
0



LR per Marker

```
> lr1 = mpiTest1$LRperMarker  
> cols = ifelse(lr1 > 1, 8, 2)  
> barplot(lr1, col = cols, ylab = "LR", las = 2,  
          cex.names = 0.8)  
> abline(h = 1, lty = 2)
```

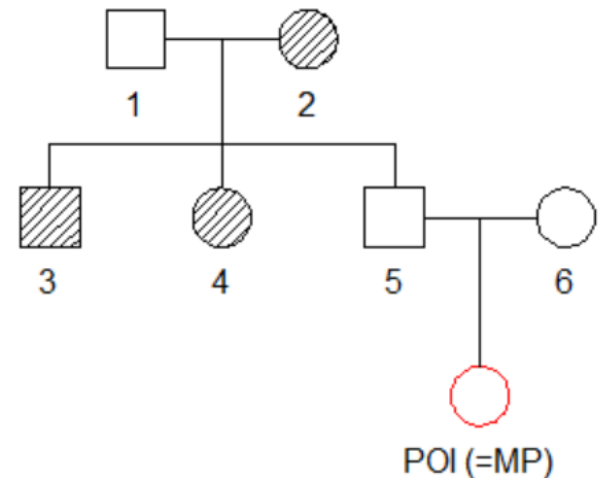


LR for POI2

```
> ref = mpi$Reference
> poi2 = mpi$POI2
> mpiTest2 = missingPersonLR(ref, missing = "MP",
                             poi = poi2)
> mpiTest2
```

H1:H2
2369.091

- LR ≈ 2400 supports POI2 = MP
- Does not quite reach the threshold of 10,000.



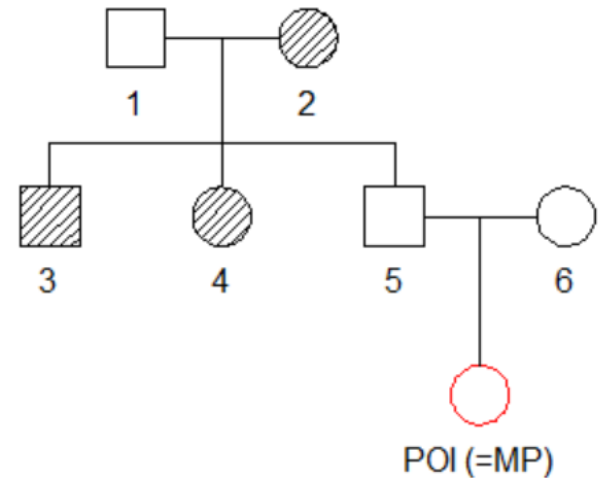
Exclusion Power

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

Potential mismatches: 8 (D3S1358, D7S820, CSF1PO,
PENTA_D, VWA, TPOX, D19S433, D2S1338)

Expected mismatches: 1.679

P(at least 1 mismatch): 0.863



Inclusion Power

```
> ip = missingPersonIP(ref, missing = "MP", nsim = 500,  
>                       threshold = 10000, seed = 42)  
> ip
```

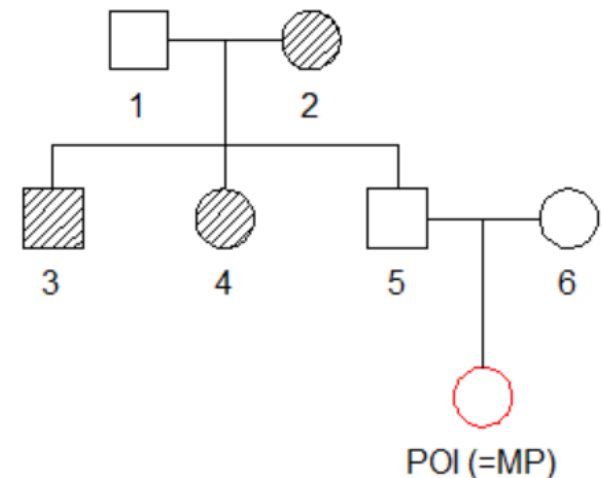
Using all 20 attached markers
Simulating 500 profiles...done
Computing likelihood ratios...done
Total time used: 9.87 secs

Mean LR: 65947.04

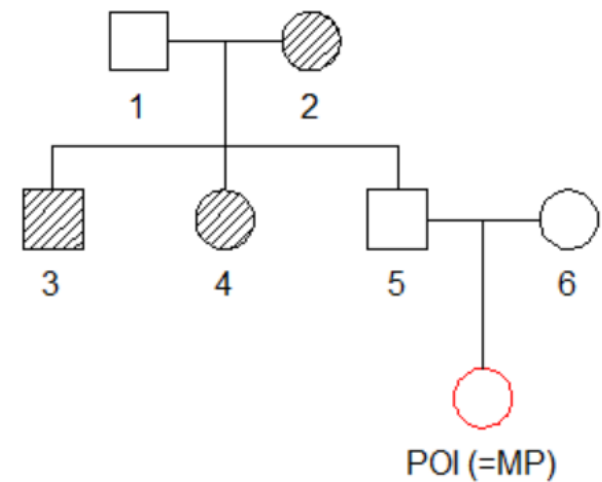
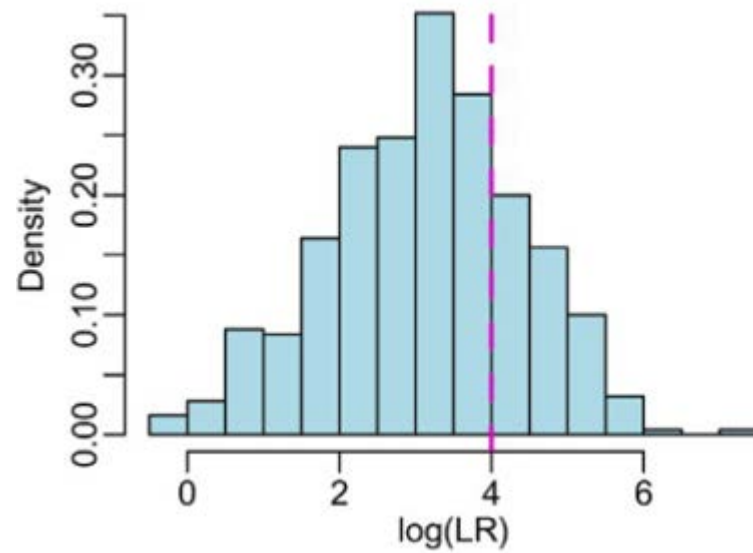
Mean log₁₀(LR): 3.121

Estimated power:

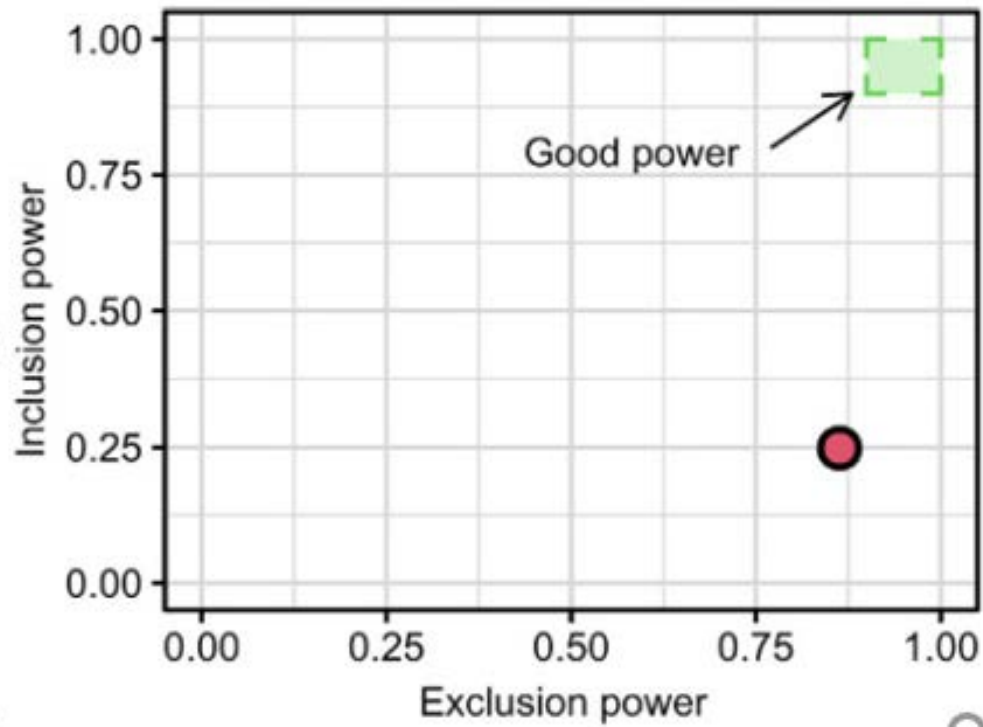
$$P(\text{LR} \geq 10000) = 0.248$$



Inclusion Power Simulation



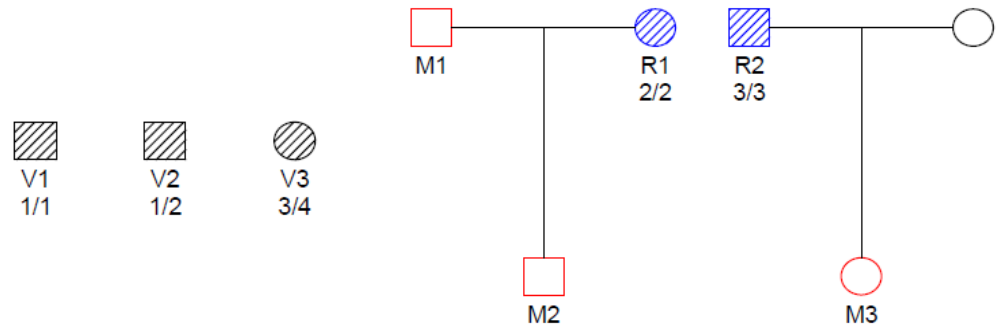
Power summary plot



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
library(dvir)



Disaster Victim Identification in R: dvir package

- Key functions
 - `pairwiseLR()`
 - `sequentialDVI()`
 - `jointDVI()`

Article | [Open Access](#) | Published: 01 July 2021

Joint DNA-based disaster victim identification

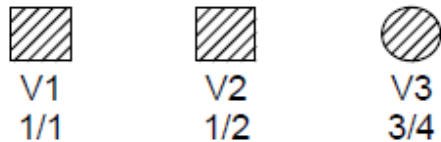
Magnus D. Vigeland  & Thore Egeland

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

Post mortem (PM) and ante mortem (AM) data

- **PM data**

- Victims V1, V2,..., Vs
- Replicated samples merged



Standard marker kits:

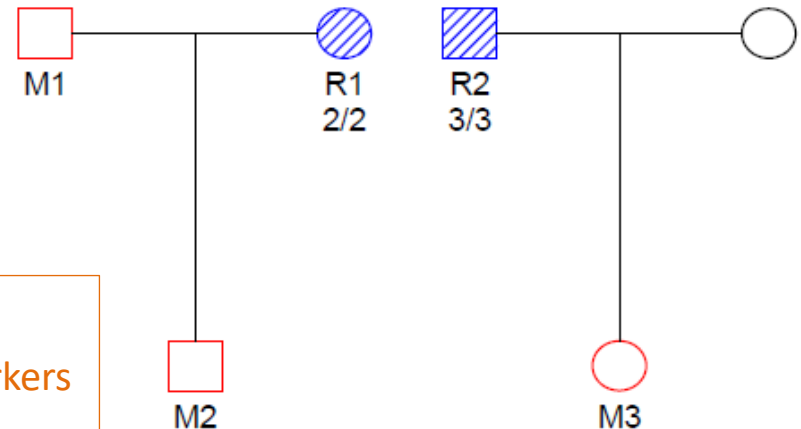
- 15 - 24 autosomal STRs
- up to 50 alleles
- unlinked (mostly)

Sometimes also:

- X, Y, mtDNA markers
- sex
- age

- **AM data**

- Missing persons M1, M2, ..., Mm
- References R1, R2, ...



- **Assignment a**

- $\{V1 = M1, V2 = M2\}$
- compact notation: (M1, M2, *)

The number of possible assignments

m_F	s_F					
	0	1	2	3	4	5
0	19081	3393	529	73	9	1
1	9276	3922	1074	228	40	6
2	4051	3135	1603	559	147	31
3	1546	2004	1768	1054	438	136
4	501	1045	1533	1533	1045	501
5	136	438	1054	1768	2004	1546
6	31	147	559	1603	3135	4051
7	6	40	228	1074	3922	9276
8	1	9	73	529	3393	19081

5 victims and 8 MPs. Table produced by:

```
> ncomb(nVfemales, nMPfemales, nVmales, nMPmales)
```

The model

- Assignments A
- Prior $\pi(a), a \in A$
- Posterior $\pi(a|\text{PM and AM data})$

- Likelihood

$$L(a) = P(\text{PM and AM data} \mid a, \Phi)$$

- Frequentist:

$$- \text{LR} = L(a)/L_0$$

– L_0 = likelihood empty assignment

- From this, we get

$$- P(V_i = M_j \mid \text{PM and AM data})$$

$$- P(V_i = * \mid \text{PM and AM data})$$

Example

	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
11	M_2	M_1	M_3	-Inf	0.00	0.00
12	M_2	M_1	*	-Inf	0.00	0.00
13	M_2	*	M_3	-Inf	0.00	0.00
14	M_2	*	*	-Inf	0.00	0.00

	M_1	M_2	M_3	*
V_1	0.88	0.00	0.00	0.12
V_2	0.02	0.95	0.00	0.03
V_3	0.00	0.00	0.83	0.17

Numerical example

- The posterior with flat prior for the two first assignments:

$$\frac{250}{250 + 50 + \dots + 0} = 0.718$$

$$\frac{50}{250 + 50 + \dots + 0} = 0.144$$

	V ₁	V ₂	V ₃	loglik	LR	posterior
1	M ₁	M ₂	M ₃	-16.118	250.000	0.718
2	M ₁	M ₂	*	-17.728	50.000	0.144
3	*	M ₂	M ₃	-18.421	25.000	0.072
4	M ₁	*	M ₃	-20.030	5.000	0.014
5	*	M ₁	M ₃	-20.030	5.000	0.014
6	*	M ₂	*	-20.030	5.000	0.014
7	*	*	M ₃	-20.030	5.000	0.014
8	M ₁	*	*	-21.640	1.000	0.003
9	*	M ₁	*	-21.640	1.000	0.003
10	*	*	*	-21.640	1.000	0.003

- Posterior pairing probability:

$$P(V_1 = M_1 \mid \text{PM and AM data}) =$$

$$0.718 + 0.144 + 0.0144 + 0.003 = 0.88$$

Sum posterior for rows $V_1 = M_1$

	M ₁	M ₂	M ₃	*
V ₁	0.88	0.00	0.00	0.12
V ₂	0.02	0.95	0.00	0.03
V ₃	0.00	0.00	0.83	0.17

Example: ICMP ex Yugoslavia. Joint solution

PM data



AM data

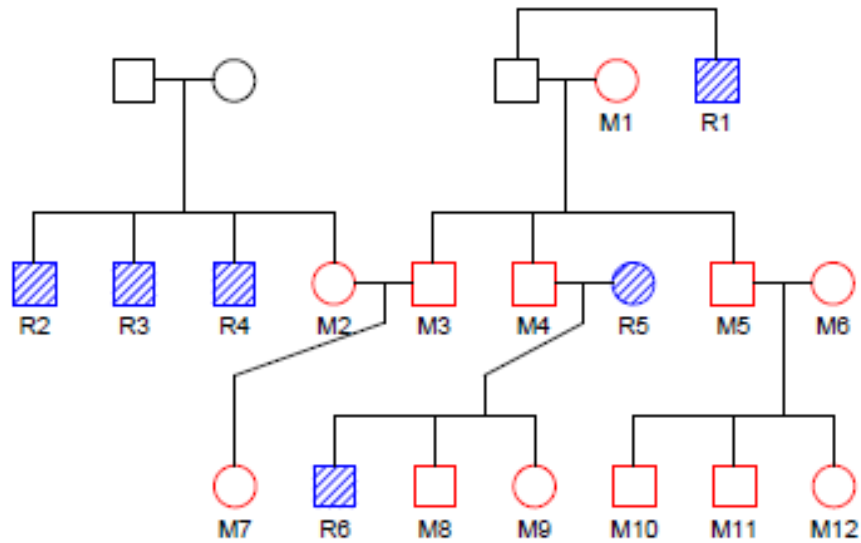


Figure 5. A large reference family with 12 missing individuals.

```
> library(dvir)
> pm = icmp$pm
> am = icmp$am
> missing = icmp$missing
> res = 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

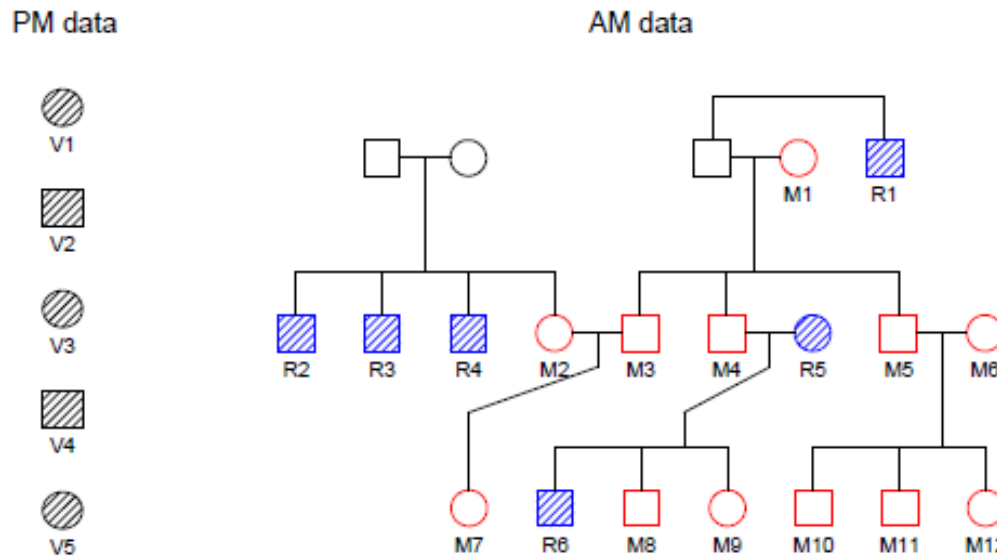


Figure 5. A large reference family with 12 missing individuals.

	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												

Table produced by: `> Bmarginal(res, missing)`

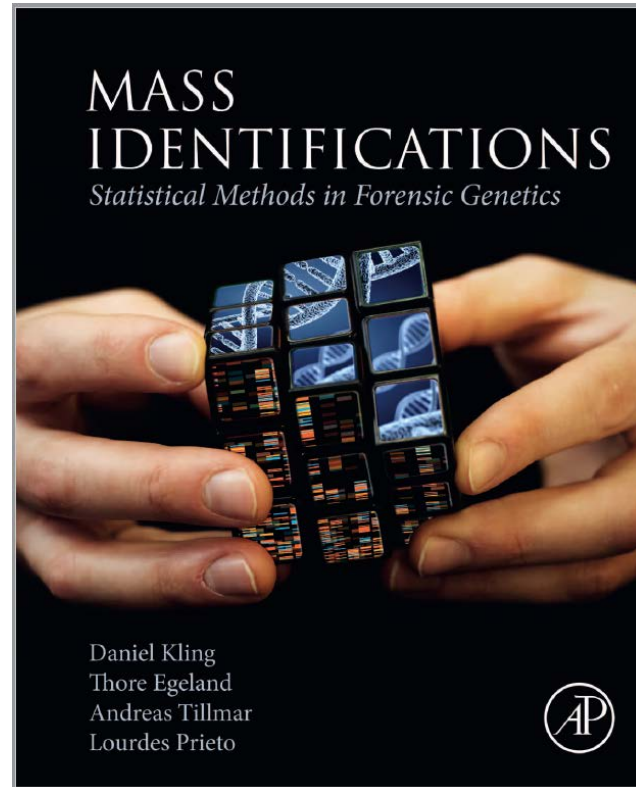
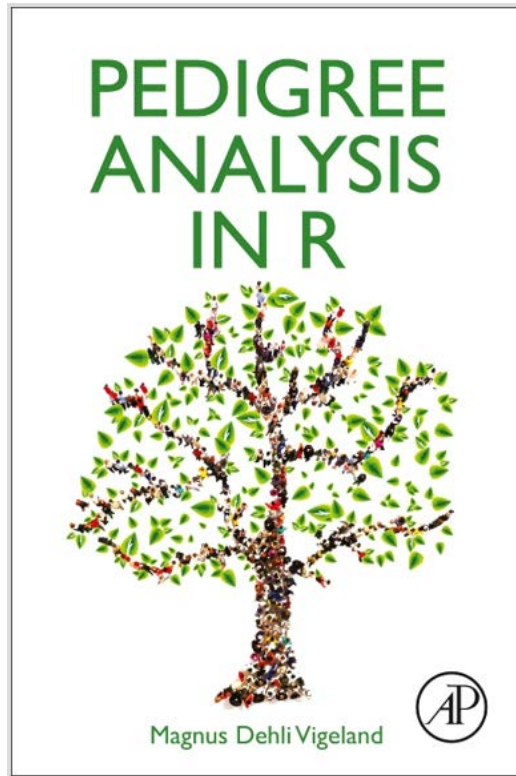
Discussion

- What should be reported?
 - LR and, or posterior?
 - Joint or individual?
 - Use verbal scales?
- Extend model
 - informative priors
 - include non-genetic information in likelihood

$$\begin{aligned} L(a) &= P(PM, AM, age|a, \Phi) \\ &= P(PM, AM|a, \Phi)P(age|a, \Phi) \end{aligned}$$

- beyond conventional forensic markers

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](#)