Statistical methods in forensic genetics

Thore.Egeland
Daniel Kling
Oskar Hansson
Guro Dørum
(In order of appearance)
Copenhagen Oct 7- Oct 10 2013

folkehelseinstituttet

NORWEGIAN UNIVERSITY OF LIFE SCIENCES
MDCCCLIX

EUROFORGEN
Network of Excellence

Program. Exercises/demos for all sessions

http://arken.umb.no/~theg/Copenhagen2013/

Day 1: Monday 7 Oct *14:00-17:00* Preliminaries (Thore Egeland)

Principles, paternity index, LR, W, Bayes theorem on odds form.

Day 2: *09:00-12:30* Kinship cases (Thore Egeland)

Theta correction, mutation, silent alleles, complex pedigrees. Familias 2.0.

13:30-17:00 Familias extensions (Daniel Kling)

Simulations, DVI module and new mutation model. Familias 3.0.

Day 3: *09:00-12:30* Linkage and linkage equilibrium (Daniel Kling)

Linked markers, X chromosomal markers. FamLink and FamLinkX.

13:30-17:00 DNA mixtures. (Guro Dørum and Oskar Hansson)

Introduction to R, forensim and LRmix

Day 4: *09:00-12:00* More on forensim and LRmix.

(Oskar Hansson and Guro Doerum)

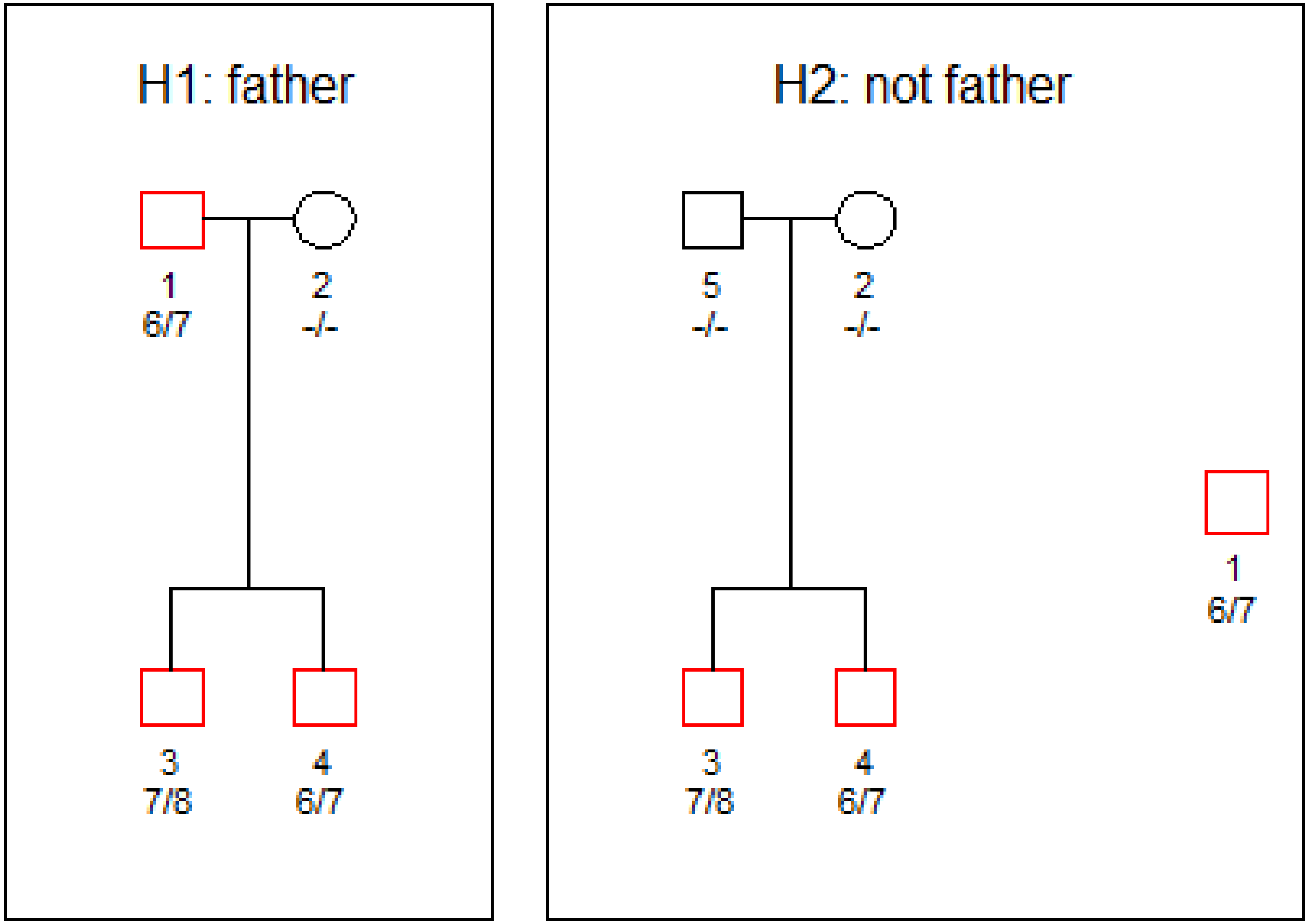
12:00-12:30 Final discussion and summary (Thore Egeland)

12:30-14:00 Lunch and departure.

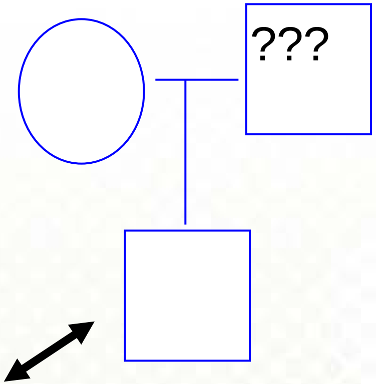
Oct 7. 14:00-17:00 Preliminaries
Oct 8. 09:00-12:30 Kinship cases

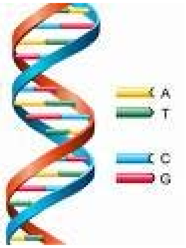
- Objectives: Learn, discuss
  - Statistical evaluation of evidence. Principles
  - Paternity cases
    - Complex cases
  - practical approach based on Familias 2 exercises:

Motivating example 1

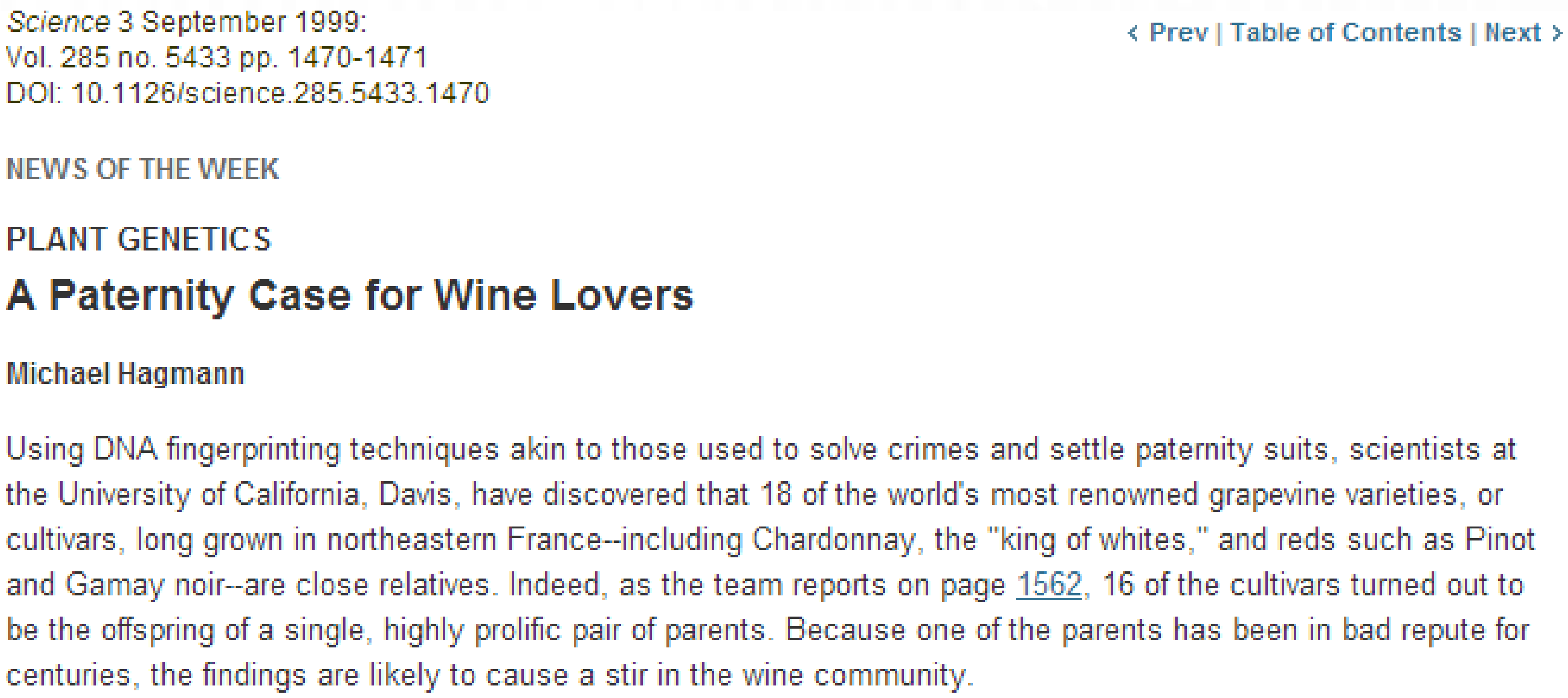


Motivating example 2





‘Red wine, paternity and disaster’



Science 3 September 1999:
Vol. 285 no. 5433 pp. 1470-1471
DOI: 10.1126/science.285.5433.1470

‹ Prev | Table of Contents | Next ›

NEWS OF THE WEEK

PLANT GENETICS

A Paternity Case for Wine Lovers

Michael Hagmann

Using DNA fingerprinting techniques akin to those used to solve crimes and settle paternity suits, scientists at the University of California, Davis, have discovered that 18 of the world's most renowned grapevine varieties, or cultivars, long grown in northeastern France--including Chardonnay, the "king of whites," and reds such as Pinot and Gamay noir--are close relatives. Indeed, as the team reports on page 1562, 16 of the cultivars turned out to be the offspring of a single, highly prolific pair of parents. Because one of the parents has been in bad repute for centuries, the findings are likely to cause a stir in the wine community.

Contents

- Part I. Review of basic forensic genetics
- Part II. Introduction to Familias 2. Demo.
- Part III. Non-standard cases.
  - Complex pedigrees
  - Exercises Familias 2
  - Theta correction, mutation, silent alleles
- Part IV.
  - Exercises Familias 2
  - Discussion

Part I

- Review of basic forensic genetics

- *“DNA typing as a model for a scientifically defensible approach to questions of shared identity, are driving the older forensic sciences toward a new scientific paradigm.”*

Saks, Koehler, Science, 2005

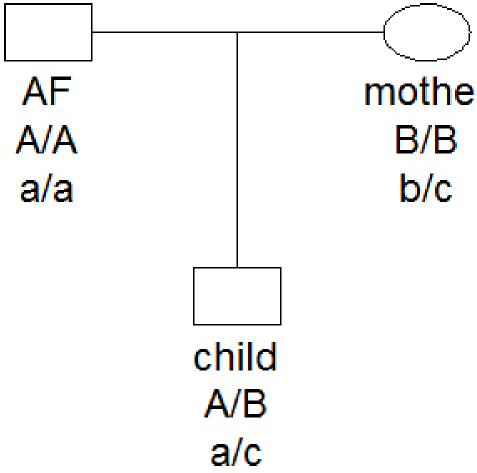
General principles

1. To evaluate the uncertainty of any given proposition it is necessary to consider at least one alternative proposition.
2. Scientific interpretation is based on questions of the kind: “What is the probability of the evidence given the proposition?”
3. Scientific evidence is conditioned not only by the competing propositions, but also by the framework of circumstances within which they are to be evaluated

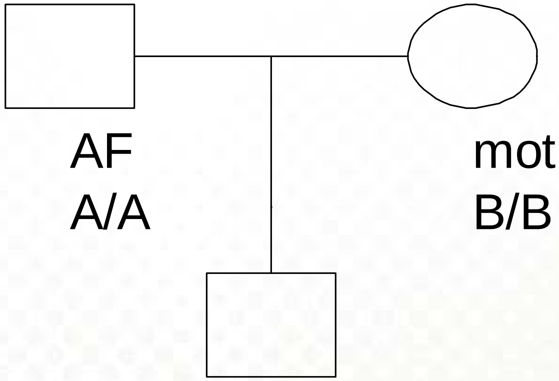
Exercise S1. Standard paternity case

- H_1: The alleged father (AF) is the real father
- H_2: AF and the child are unrelated.

Figure1. Standard paternity case.



Standard paternity



AF
A/A

mother
B/B

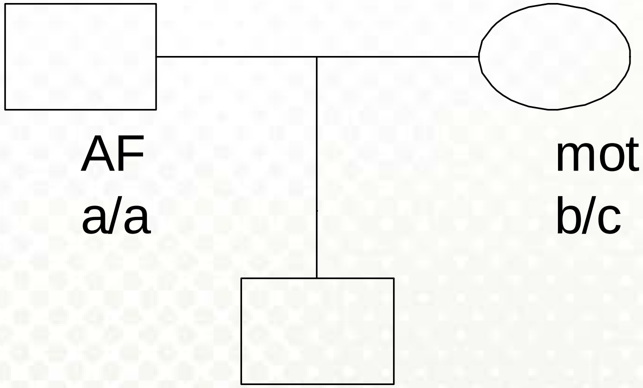
child
A/B

$$LR = \frac{\text{probability of data } given\ AF \text{ father}}{\text{probability of data } given\ AF \text{ unrelated}}$$

$$= \frac{P(child \mid mother, AF)}{P(child \mid mother)} = \frac{1}{p_A} = \frac{1}{0.05} = 20.$$

- Interpretation: The data is 20 times more likely assuming AF to be the father compared to the alternative that some unknown man is the father.

Standard paternity c



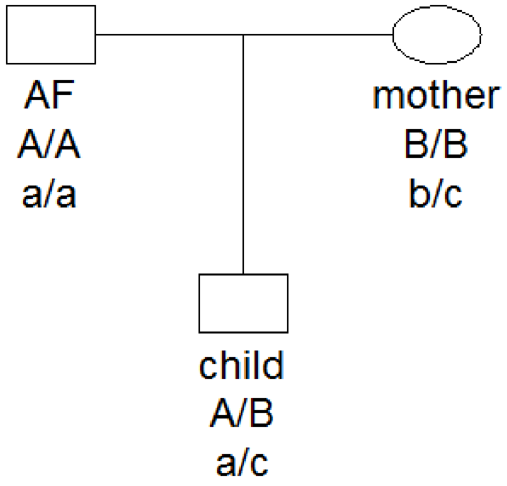
AF
a/a

mother
b/c

child
a/c

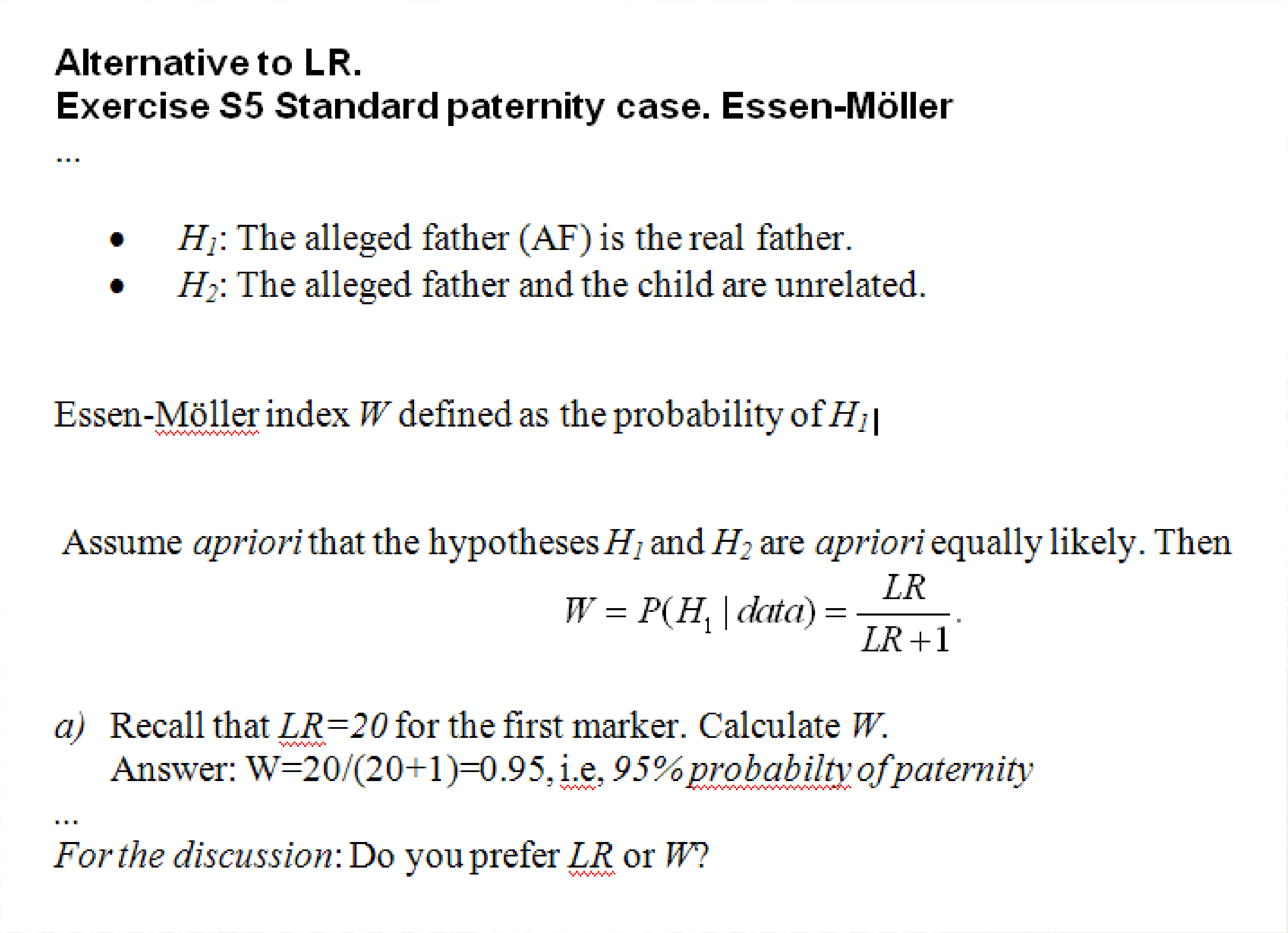
$$LR = \frac{P(child\,|\,mother, AF)}{P(child\,|\,mother)} = \frac{0.5}{0.5p_a} = \frac{0.5}{0.5*0.1} = 10.$$

Figure1. Standard paternity case.



$$LR = 20 * 10 = 200.$$

- Interpretation: The data is 200 times more likely assuming AF to be the father compared to the alternative that some unknown man is the father”.



Alternative to LR.
Exercise S5 Standard paternity case. Essen-Möller

…

- H_1: The alleged father (AF) is the real father.
- H_2: The alleged father and the child are unrelated.

Essen-Möller index W defined as the probability of H_1

Assume *apriori* that the hypotheses H_1 and H_2 are *apriori* equally likely. Then

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

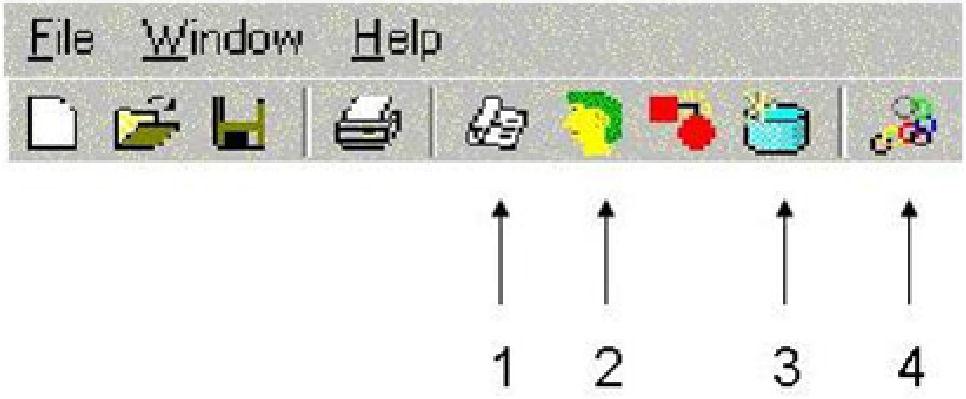
a) Recall that $LR=20$ for the first marker. Calculate W.
Answer: W=20/(20+1)=0.95, i.e, *95% probabilty of paternity*

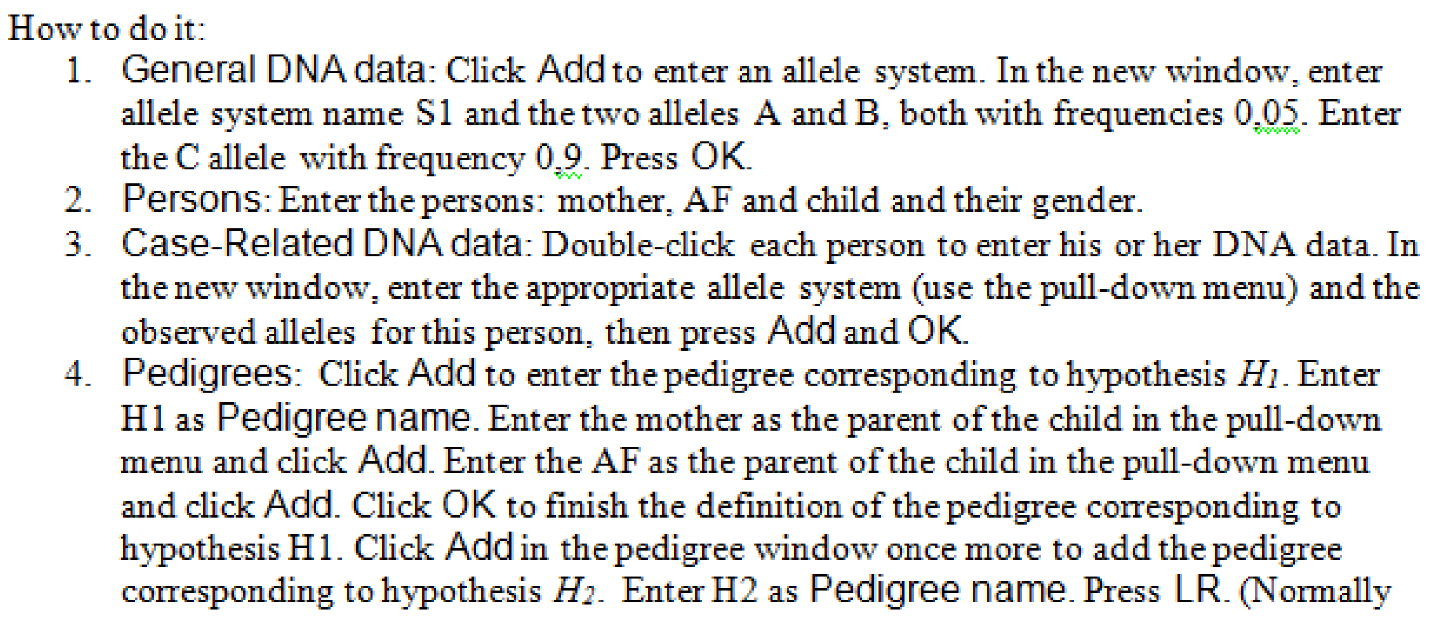
…

For the discussion: Do you prefer LR or W?

Part II.

- Introduction to Familias. Demo.
- *Videos in English and Spanish available from course page* http://arken.umb.no/~theg/Copenhagen2013/
- http:// familias.name
  - Reference: Egeland, Mostad et al. (2000)
  - Validation: J. Drábek (2008)





How to do it:

1. General DNA data: Click Add to enter an allele system. In the new window, enter allele system name S1 and the two alleles A and B, both with frequencies 0,05. Enter the C allele with frequency 0,9. Press OK.
2. Persons: Enter the persons: mother, AF and child and their gender.
3. Case-Related DNA data: Double-click each person to enter his or her DNA data. In the new window, enter the appropriate allele system (use the pull-down menu) and the observed alleles for this person, then press Add and OK.
4. Pedigrees: Click Add to enter the pedigree corresponding to hypothesis H_1. Enter H1 as Pedigree name. Enter the mother as the parent of the child in the pull-down menu and click Add. Enter the AF as the parent of the child in the pull-down menu and click Add. Click OK to finish the definition of the pedigree corresponding to hypothesis H1. Click Add in the pedigree window once more to add the pedigree corresponding to hypothesis H_2. Enter H2 as Pedigree name. Press LR. (Normally

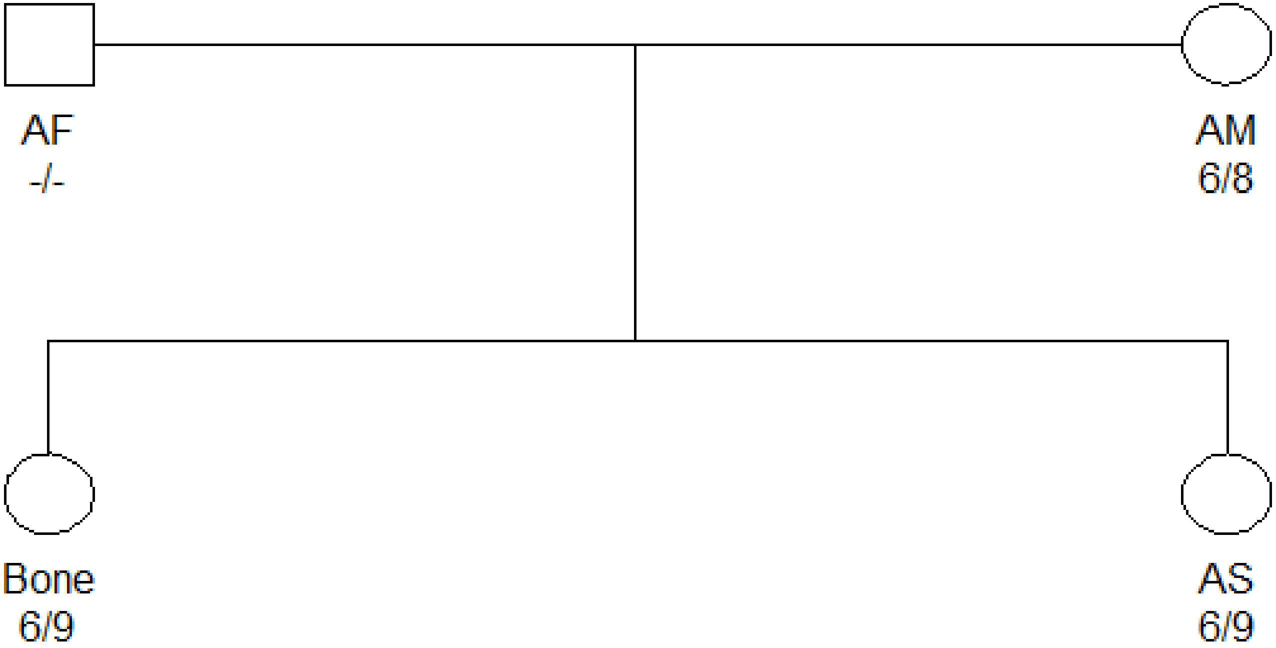
Part III.

- Non-standard cases.
  - Complex pedigrees
- Demo, exercise:

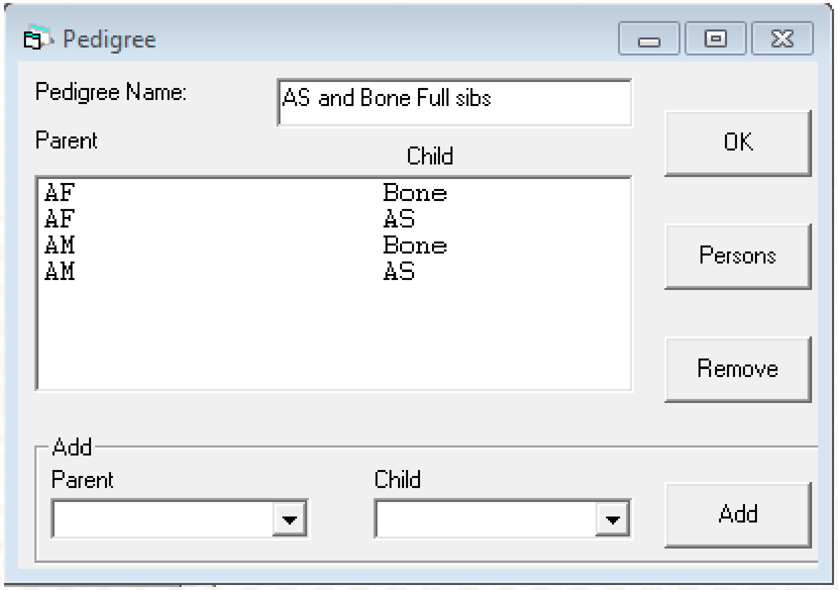
Exercise S3 Cap paper challenge (“bone”)

- H_1: The individual providing the bone is the daughter of AM and sister of AS.
- H_2: The individual providing the bone is not related to the tested individuals (AM,AS) of Figure 3

Figure 3. CAP paper challenge. Marker F13B.
Frequencies: 6:0.086, 8:0.152, 9:0.328, rest:0.434



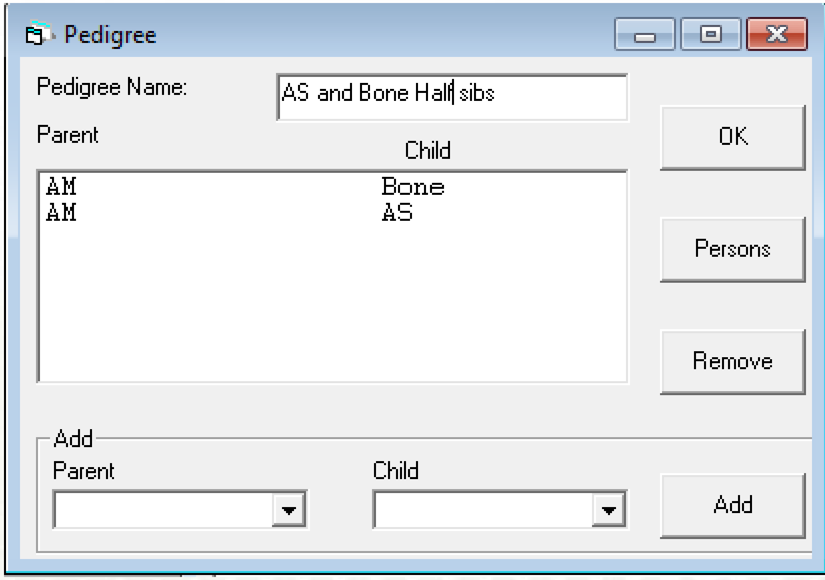
Full sibs



Pedigree Name: AS and Bone Full sibs

Parent	Child
AF	Bone
AF	AS
AM	Bone
AM	AS

Half sibs



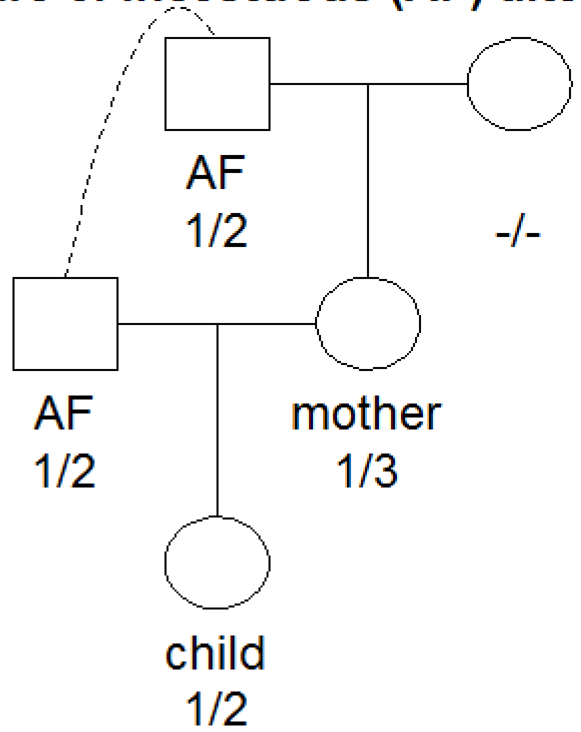
Pedigree Name: AS and Bone Half sibs

Parent	Child
AM	Bone
AM	AS

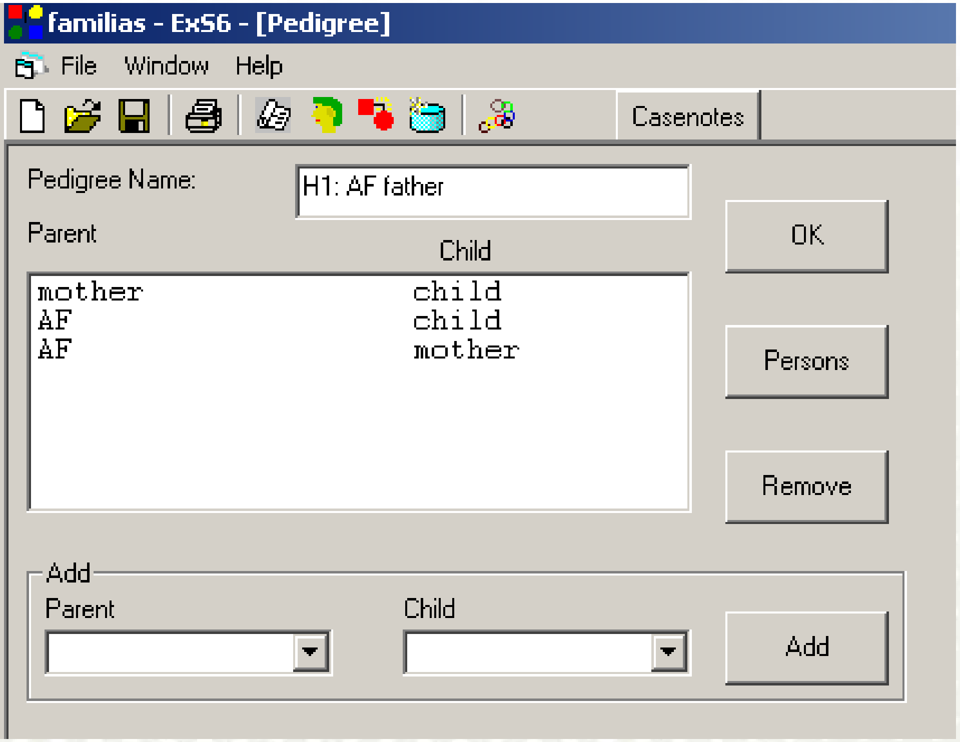
The father AF is not typed but included in the full sibs alternative to the left to define Bone and AS to be full sisters.

Example: Complex pedigrees I. Exercise S6

Figure 5. Incestuous (AF) alternative



Pedigree in Familias

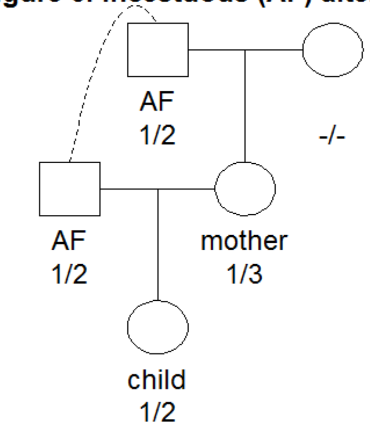


Pedigree Name: H1: AF father

Parent	Child
mother	child
AF	child
AF	mother

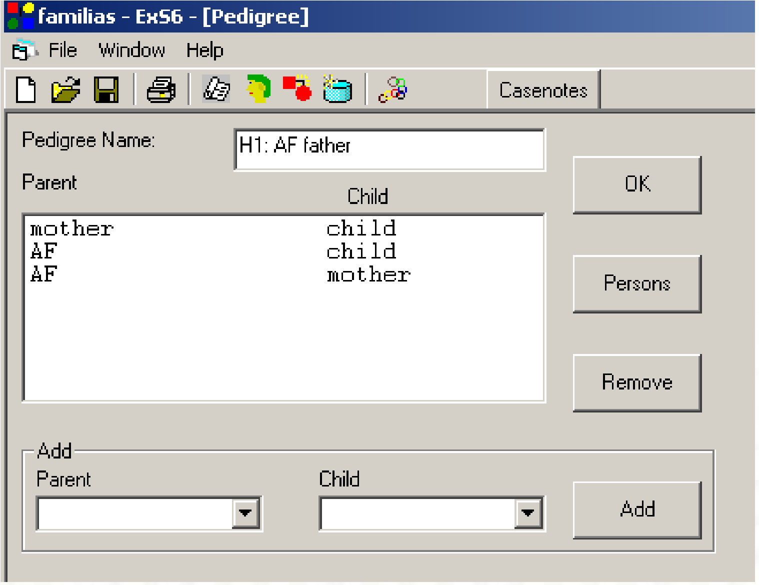
Add
Parent
Child

Figure 5. Incestuous (AF) alternative



Exercises Day 1

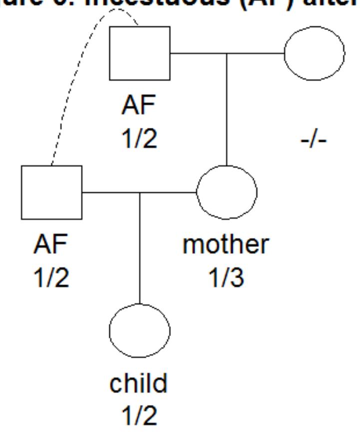
- Exercises, input files and solutions available from course homepage, Day 1 http://arken.umb.no/~theg/Copenhagen2013
- Do S1 a-g, S3, S4, S5, S6 (as time permits)
- Exercise S6:

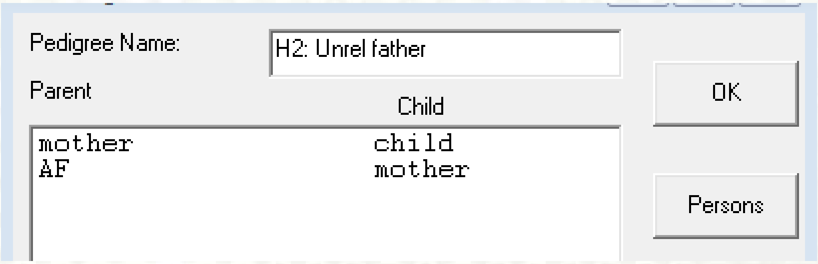


Pedigree Name: H1: AF father

Parent	Child
mother	child
AF	child
AF	mother

Figure 5. Incestuous (AF) alternative

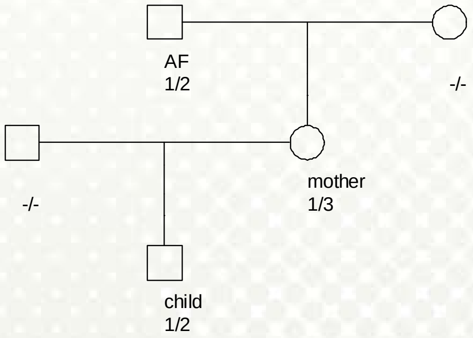


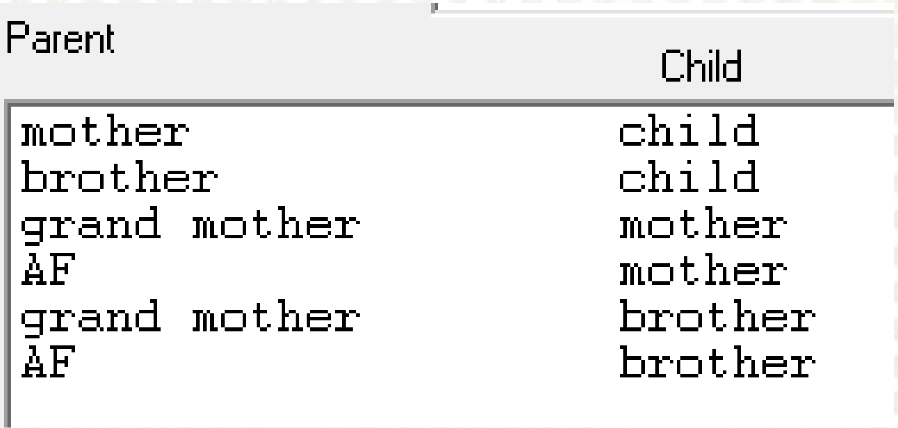


Pedigree Name: H2: Unrel father

Parent	Child
mother	child
AF	mother

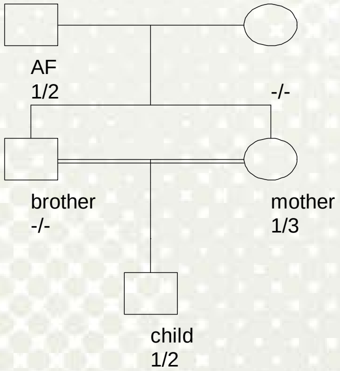
H2:Unrelated

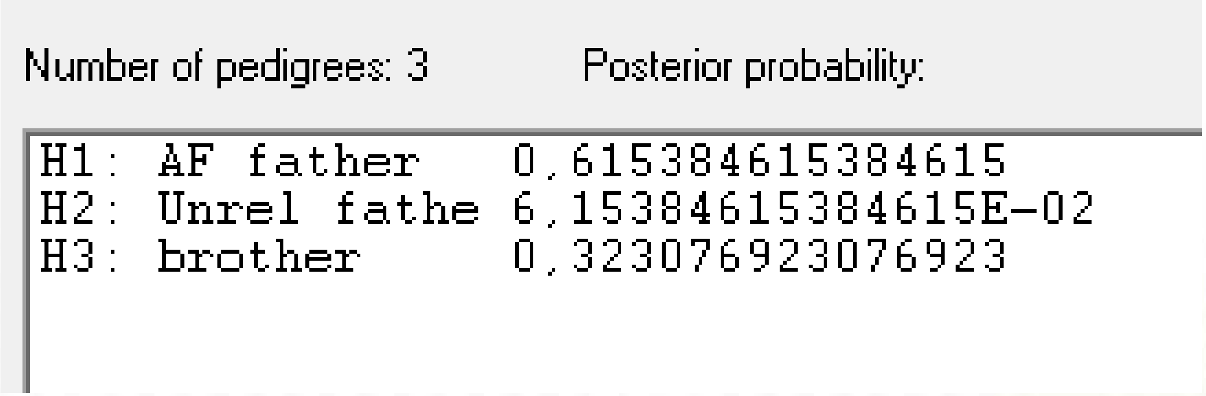




Parent	Child
mother	child
brother	child
grand mother	mother
AF	mother
grand mother	brother
AF	brother

H3: Brother

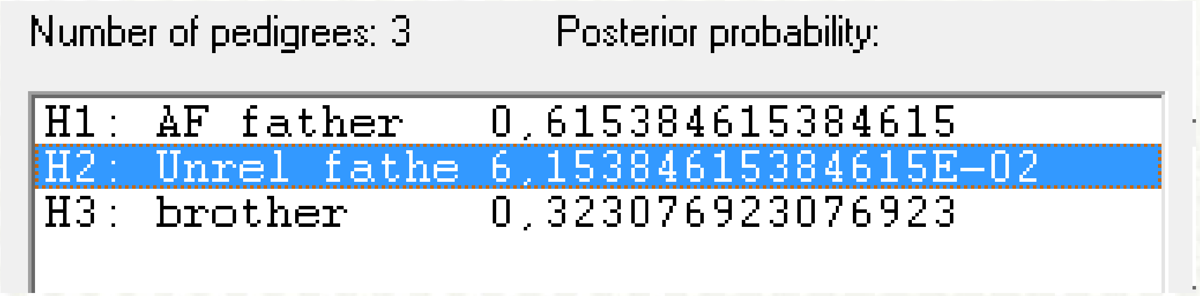




Number of pedigrees: 3   Posterior probability:

```
H1:  AF father    0,615384615384615
H2:  Unrel fathe 6,15384615384615E-02
H3:  brother      0,323076923076923
```

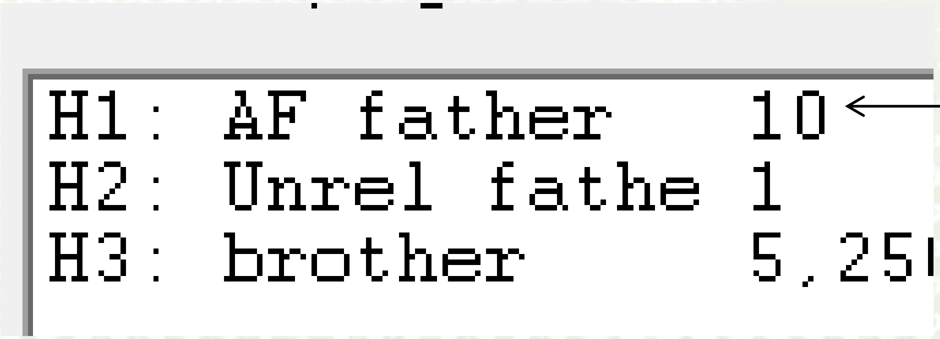
Highlight denominator



Number of pedigrees: 3   Posterior probability:

```
H1:  AF father    0,615384615384615
H2:  Unrel fathe 6,15384615384615E-02
H3:  brother      0,323076923076923
```

Click: “Relative or ...”



```
H1:  AF father    10
H2:  Unrel fathe 1
H3:  brother      5,25
```

$$\frac{0.615}{0.0615} = 10$$

Complications

- Non-standard cases.
  - Complications:
    1. Theta-correction
    2. Mutations
    3. Silent alleles

NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Problem: Unrelated people not unrelated!
Solution: Theta-correction

commentary

Nature **371**, 735 - 738 (27 October 2002); doi:10.1038/371735a0

DNA fingerprinting dispute laid to rest

ERIC S. LANDER & BRUCE BUDOWLE

Whitehead Institute for Biomedical Research, Nine Cambridge Center, Cambridge, Massachu
Forensic Science Research and Training Center, FBI Laboratory, FBI Academy, Quantico, Vi

Hardy-Weinberg (HW) equilibrium

Two alleles A, B.

$p_A = 0.4,\ p_B = 0.6.$

$$\begin{aligned}
&\text{Fraction A/A: } 0.4^2 &&= 0.160\\
&\text{Fraction A/B: } 2*0.4*0.6 &&= 0.480\\
&\underline{\text{Fraction B/B: } 0.6^2} &&\underline{= 0.360}\\
&\text{Sum} &&= 1.000
\end{aligned}$$

Problem: Above requires HW, not valid if 'unrelated people' are slightly related
Solution: theta-correction

Theta - correction

$$Homozygous \quad A,A: \ \theta p_A + p_A^2(1-\theta),$$

$$Heterozygous \quad A,B: 2p_A p_B(1-\theta).$$

$$\theta = 0.1 \text{ (extreme case)}$$

$$\text{Fraction A/A: } 0.1*0.4+0.4^2*(1\text{-}0.1) \ = 0.184 > 0.160$$

$$\text{Fraction A/B: } 2*0.4*0.6*(1\text{-}0.1) \qquad = 0.432 < 0.480$$

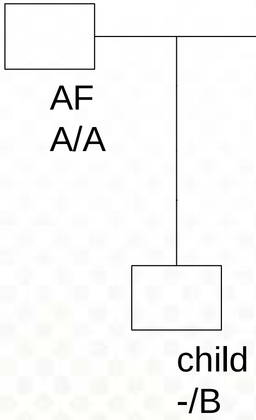
$$\underline{\text{Fraction B/B: } 0.1*0.6+0.6^2*(1-0.1) = 0.384 > 0.360}$$

$$Sum \qquad\qquad\qquad\qquad 1.000 \quad 1.000$$

General case: Sampling formula

- If x alleles are of type A of a total n alleles sampled from the subpopulation, the next will be of type A with probability

$$\frac{x\theta + (1-\theta)p_A}{1+(n-1)\theta}$$

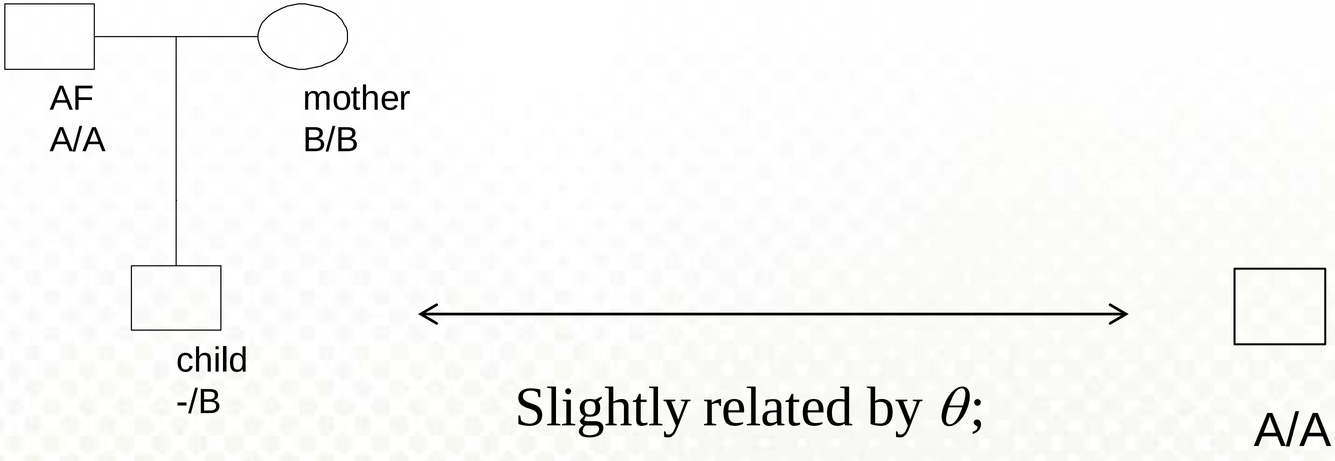


$$LR = \frac{1}{P(\text{child gets } A \mid \text{not father})}$$

$$x = 2, n = 4$$

$$P(\text{child gets } A \mid \text{not father}) = \frac{x\theta + (1-\theta)p_A}{1+(n-1)\theta}$$

$$= \frac{2\theta + (1-\theta)p_A}{1+3\theta} \Rightarrow \underline{\underline{LR = \frac{1+3\theta}{2\theta + (1-\theta)p_A}}} = 11.91$$



AF
A/A

mother
B/B

child
-/B

A/A

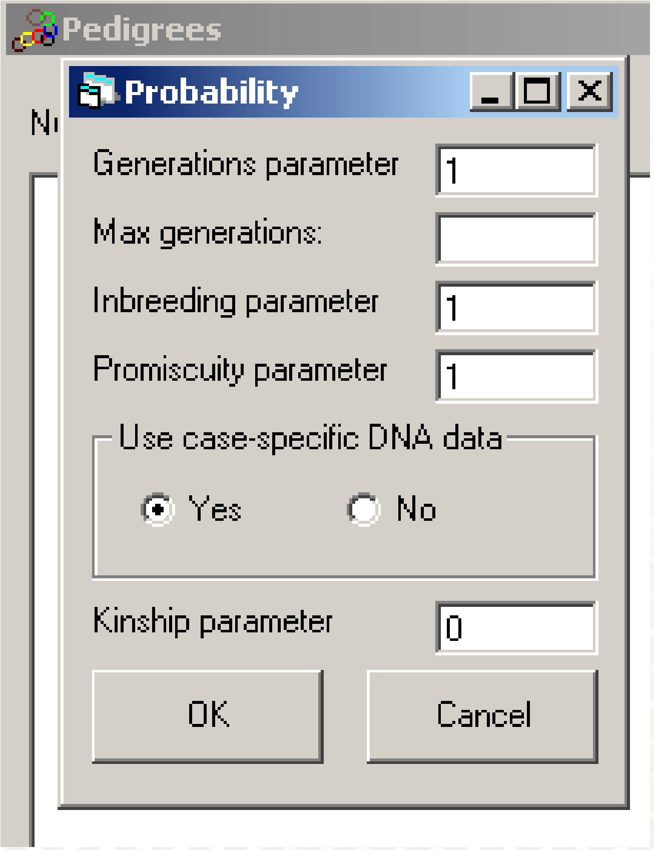
Slightly related by θ;

LR changes

$$x = 4, n = 6$$

$$P(\text{child gets } A \mid \text{not father}) = \frac{x\theta + (1-\theta)p_A}{1+(n-1)\theta}$$

$$= \frac{4\theta + (1-\theta)p_A}{1+5\theta} \Rightarrow \underline{\underline{LR = \frac{1+5\theta}{4\theta + (1-\theta)p_A}}} = 8.53$$



Generations parameter 1

Max generations:

Inbreeding parameter 1

Promiscuity parameter 1

Use case-specific DNA data

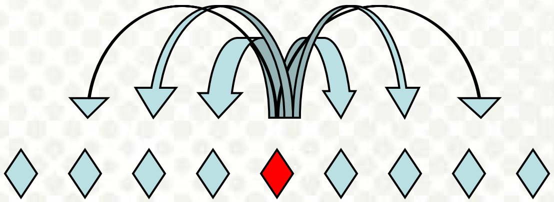
● Yes ○ No

Kinship parameter 0 ← θ

Exercise S1h

Modelling mutations

- Mutation rate varies with
  - Sex of parent and locus.
    Alleles tend to mutate to close alleles:
  - Several models



Key concept: Mutation matrix

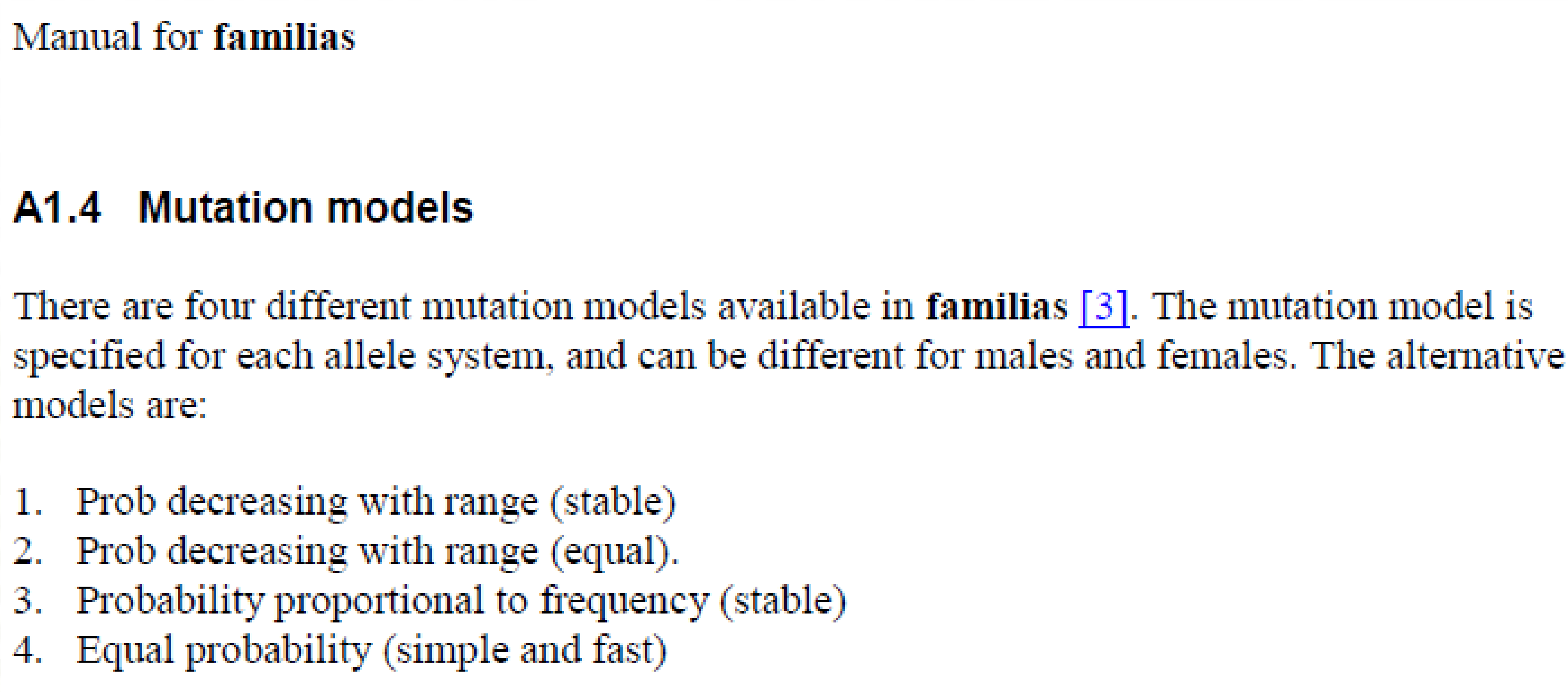
- Two alleles 1 and 2
- Probability of mutating from 1 to 2=0.01
- Probability of mutating from 2 to 1=0.01
- *Resulting mutation matrix:*

$$M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$$

Three alleles: Different models possible

				Sum
	0,990	0,005	0,005	1,000
M1=	0,005	0,990	0,005	1,000
	0,005	0,005	0,990	1,000
	0,990	0,009	0,001	1,000
M2=	0,005	0,990	0,005	1,000
	0,001	0,009	0,990	1,000

What's the best model?



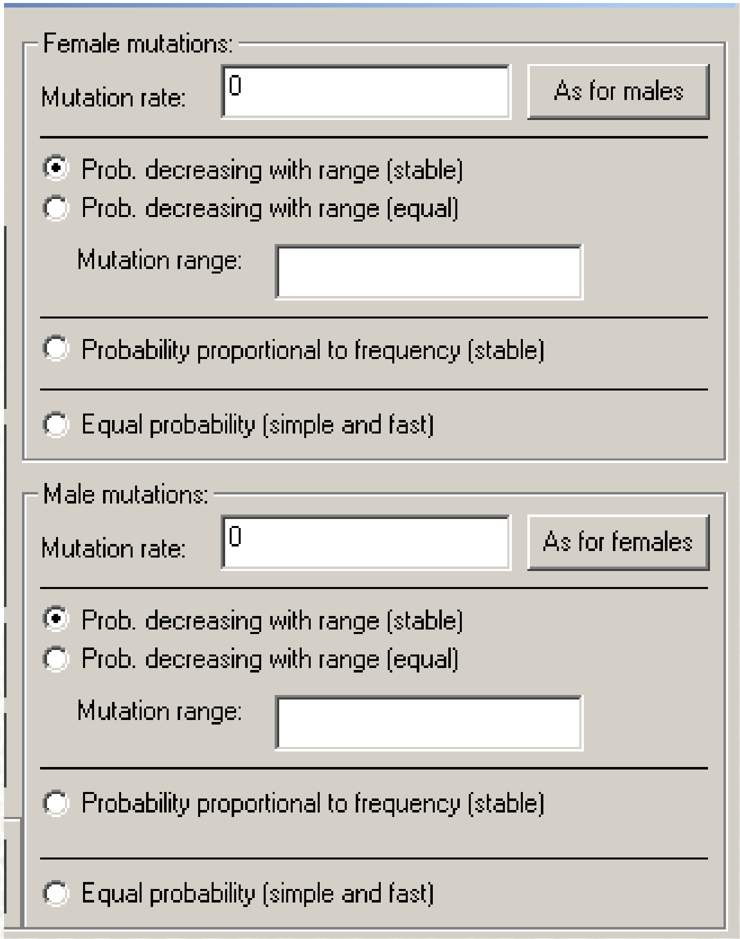
Manual for **familias**

A1.4 Mutation models

There are four different mutation models available in **familias** [3]. The mutation model is specified for each allele system, and can be different for males and females. The alternative models are:

1. Prob decreasing with range (stable)
2. Prob decreasing with range (equal).
3. Probability proportional to frequency (stable)
4. Equal probability (simple and fast)

Mutation models in Familias



Female mutations:

Mutation rate: 0

● Prob. decreasing with range (stable)
○ Prob. decreasing with range (equal)

Mutation range:

○ Probability proportional to frequency (stable)

○ Equal probability (simple and fast)

Male mutations:

Mutation rate: 0

● Prob. decreasing with range (stable)
○ Prob. decreasing with range (equal)

Mutation range:

○ Probability proportional to frequency (stable)

○ Equal probability (simple and fast)

⇦ Next page

Exercise S2, S7

Mutation range

- Mutation range 0.1:
  - mutation probability decreases by one tenth for each additional unit length difference
- Alleles must be named to preserver order!
  - use 09,10,11,12, not
  - 9,10,11,12 as standard (alphabetical) sorting is "10","11","12","9“

Example revisited

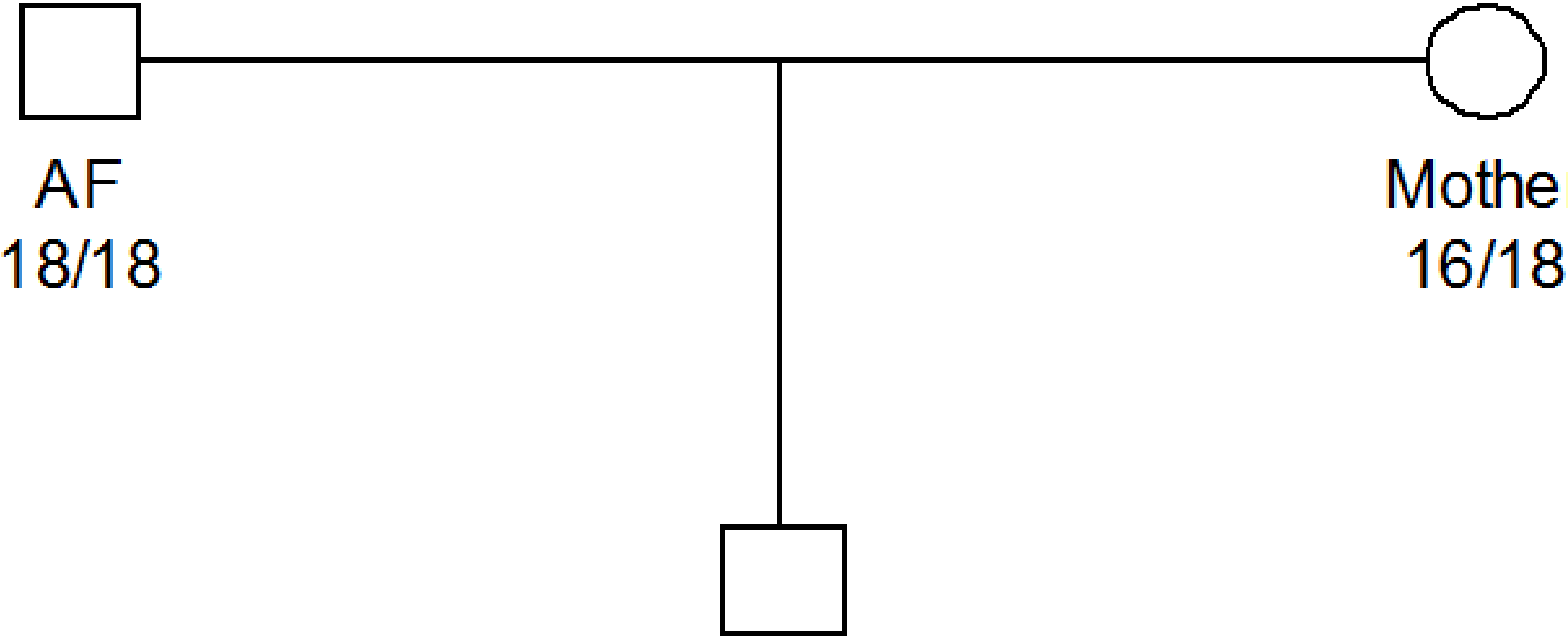
	0,990	0,009	0,001
M2=	0,005	0,990	0,005
	0,001	0,009	0,990

Mutation range =0.001/0.009=0.1111

Example: Three mutations, 21 markers

STRs	Alleged Father	Mother	Daughter
D8S1179	11/13	13/14	13
...	...	...	...
VWA	18	16/18	16/19
TPOX	8/11	7/9	7/8
D18S51	14/17	14/17	16/17
D5S818	11	9/13	9/11
FGA	24/28	19/22	22/27

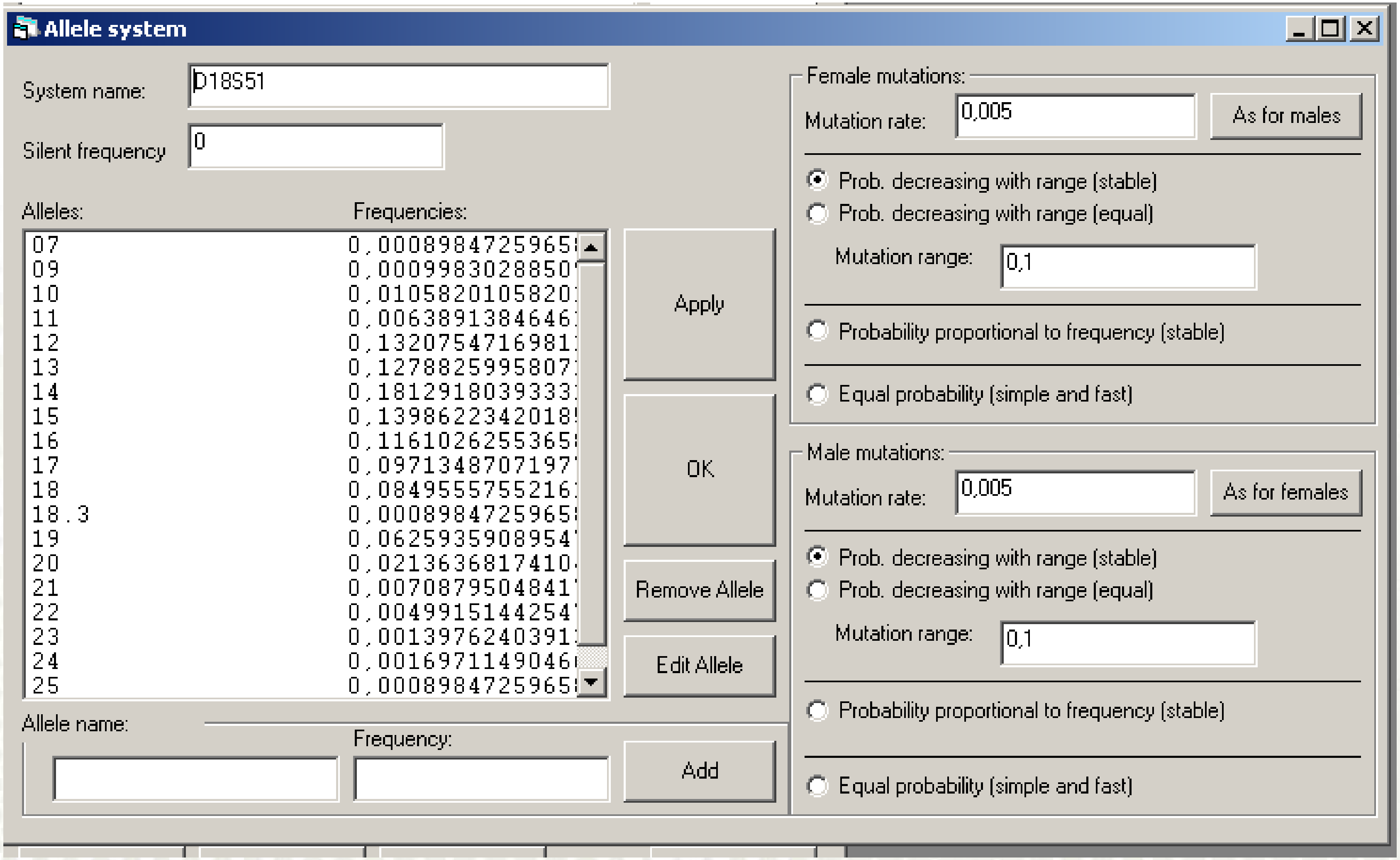
VWA



AF
18/18

Mother
16/18

Daughter
16/19



System name: D18S51

Silent frequency: 0

Alleles:	Frequencies:
07	0,000898472596
09	0,000998302885
10	0,010582010582
11	0,006389138464
12	0,132075471698
13	0,127882599580
14	0,181291803933
15	0,139862234201
16	0,116102625536
17	0,097134870719
18	0,084955575521
18.3	0,000898472596
19	0,062593590895
20	0,021363681741
21	0,007087950484
22	0,004991514425
23	0,001397624039
24	0,001697114904
25	0,000898472596

Female mutations:

Mutation rate: 0,005

● Prob. decreasing with range (stable)
○ Prob. decreasing with range (equal)

Mutation range: 0,1

○ Probability proportional to frequency (stable)

○ Equal probability (simple and fast)

Male mutations:

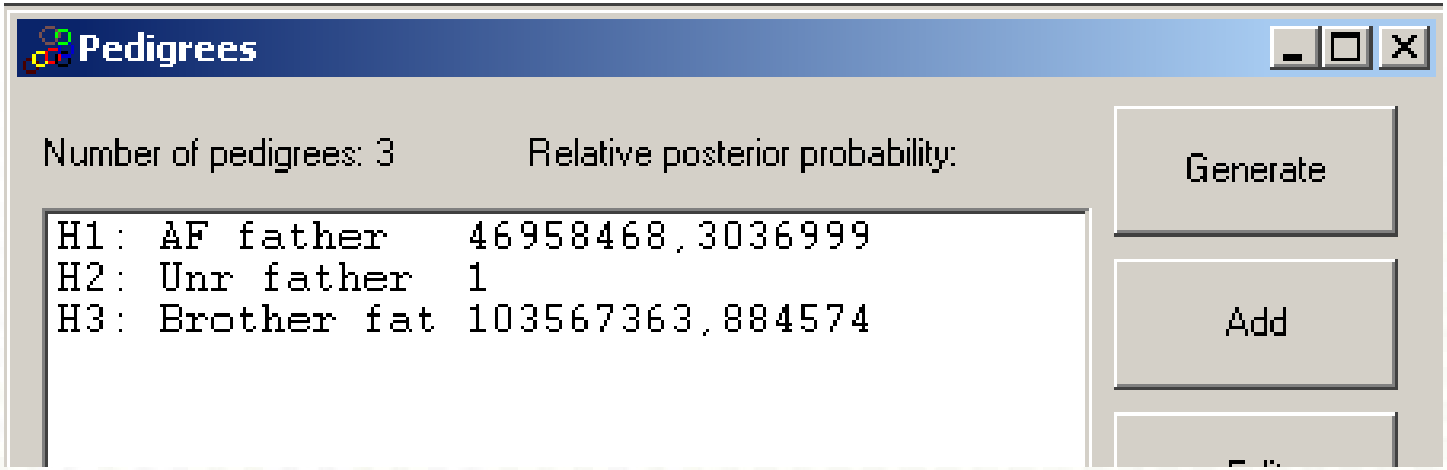
Mutation rate: 0,005

● Prob. decreasing with range (stable)
○ Prob. decreasing with range (equal)

Mutation range: 0,1

○ Probability proportional to frequency (stable)

○ Equal probability (simple and fast)



Number of pedigrees: 3

Relative posterior probability:

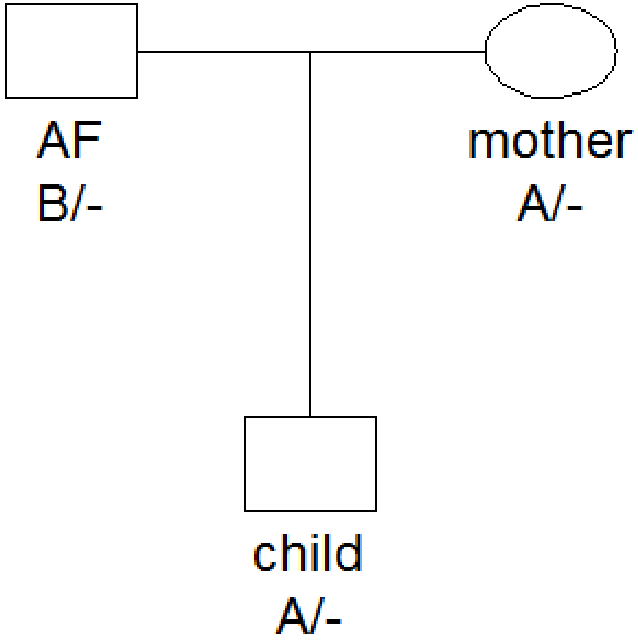
```
H1:  AF father    46958468,3036999
H2:  Unr father   1
H3:  Brother fat 103567363,884574
```

- Caveat. Not relevant database

Several improvements in Familias 3: Daniel’s lecture

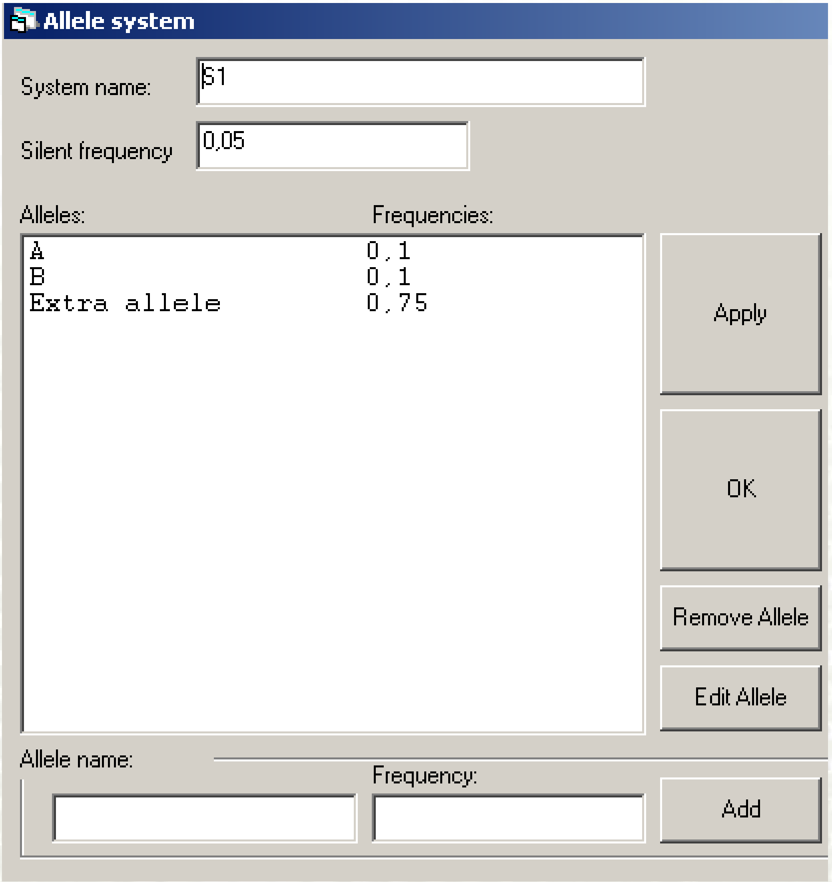
Silent alleles I. Exercise S11

Figure 5. Silent allele?



Silent alleles II. Familias

Silent allele →



System name: S1

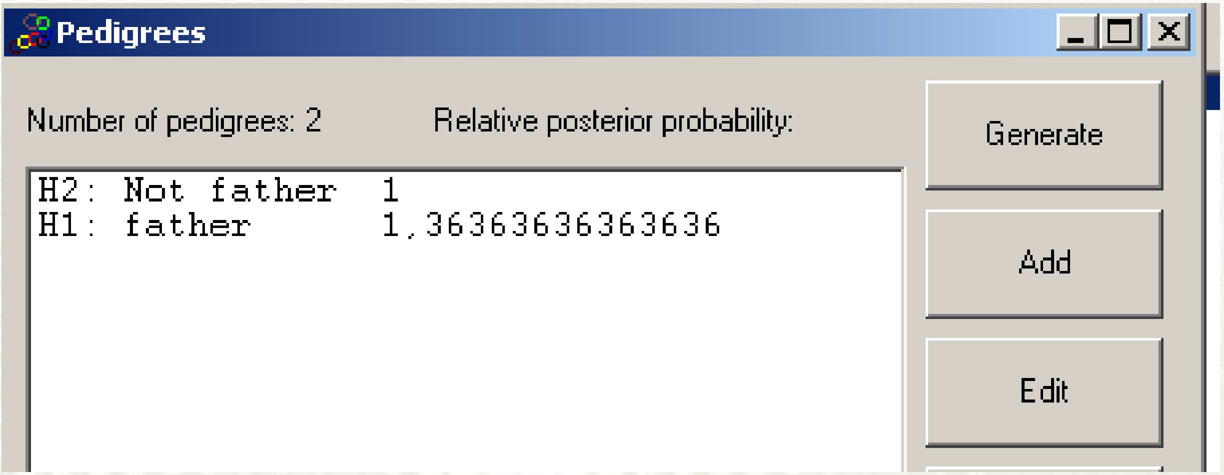
Silent frequency 0,05

Alleles:	Frequencies:
A	0,1
B	0,1
Extra allele	0,75

Silent alleles III. Exercises S11 (S12 theoretical)

$$LR = \frac{p_s(p_A+p_s)}{(p_A+p_s)^2(p_B+2p_s)+p_s p_A(p_B+2p_s)}$$

$$= \frac{0.05\cdot 0.15}{0.15^2\cdot 0.2+0.05\cdot 0.1\cdot 0.2} = 1.36.$$



Number of pedigrees: 2

Relative posterior probability:

```
H2: Not father   1
H1: father       1,36363636363636
```

Part IV.
Exercises as time permits

- Familias exercises: S7, S9, S10, S11
- Theoretical exercises S8, S12

- Simpler input/output
  - Exercise S14
  - GeneMapperToFamilias
    - http://familias.name/examples.html

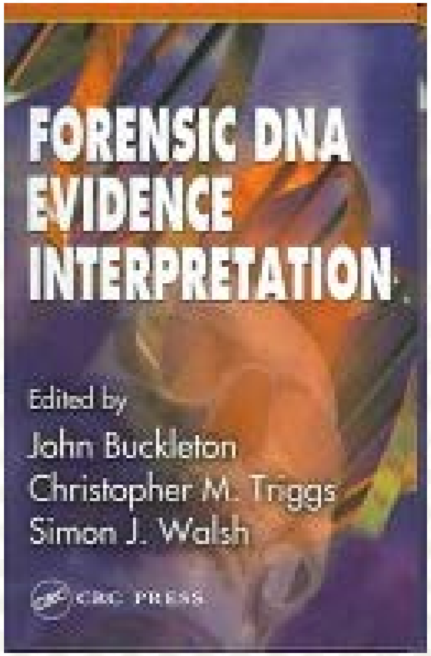
Discussion. Mutation models

- Mutations:
  - Exercise S2,S7,S8 discussed, see solutions
  - Two challenges:
    - Unstable models
    - LR may differ depending on whether a *minimum database* (i.e., only alleles appearing for marker are named) is used or not

Unstable (non-stationary) models

- Exercise S7f Comment:
  - Model 1, Probability decreasing with range is *stable*
  - Model 3, Probability proportional to frequency *stable*
- Stable: introducing a new untyped person, say the father of the alleged father, does not change the LR.
  - This is a reasonable property of a model as introducing irrelevant information should not change the result.
- Other models not stable
- I recommend model 1!
- Note extension in Familias 3 for microvariants

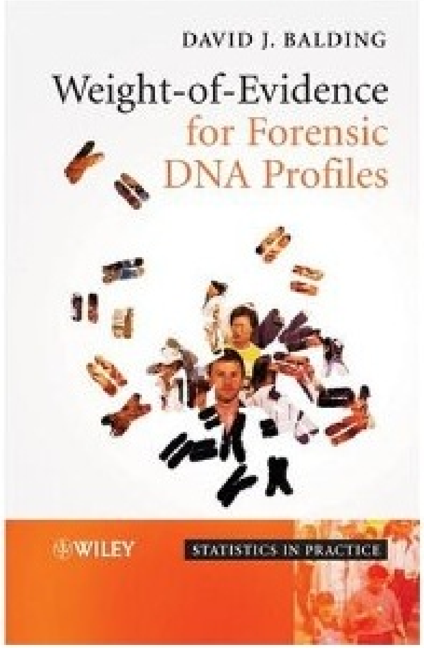
References



FORENSIC DNA EVIDENCE INTERPRETATION

Edited by
John Buckleton
Christopher M. Triggs
Simon J. Walsh

CRC PRESS

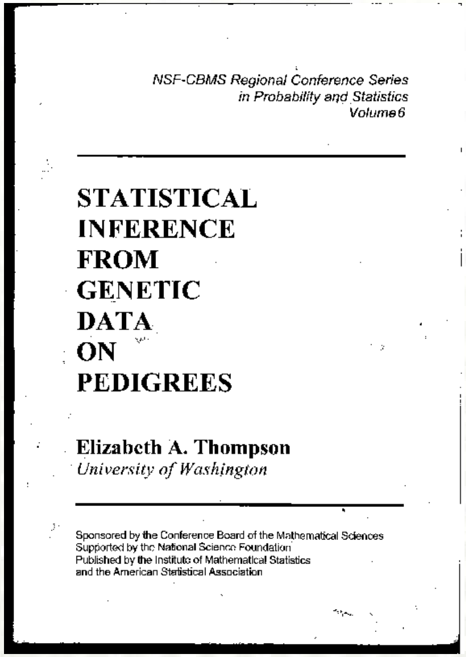


DAVID J. BALDING

Weight-of-Evidence for Forensic DNA Profiles

WILEY

STATISTICS IN PRACTICE



NSF-CBMS Regional Conference Series
in Probability and Statistics
Volume 6

STATISTICAL
INFERENCE
FROM
GENETIC
DATA
ON
PEDIGREES

Elizabeth A. Thompson
University of Washington

Sponsored by the Conference Board of the Mathematical Sciences
Supported by the National Science Foundation
Published by the Institute of Mathematical Statistics
and the American Statistical Association

Our work, projects: arken.umb.no/~theg

Discussion

- ??
- Input/output
  - GeneMapperToFamilias http://familias.name/examples.html
  - Exercise S14 http://arken.umb.no/~theg/Copenhagen2013/familiasExercises.pdf
- Familias on facebook
- Verbal equivalents to LR
- Mutation rates
- (RMNE: Guro)

Verbal equivalents

LR	Verbal Wording	Support for
1000000	Extremely strong	Prosecution
100000	Very strong	Prosecution
10000	Strong	Prosecution
1000	Moderately strong	Prosecution
100	Moderate	Prosecution
1-10	Limited	Prosecution
1	Inconclusive	
0.1	Limited	Defence
0.01	Moderate	Defence
0.001	Moderately strong	Defence
0.0001	Strong	Defence
0.00001	Very strong	Defence
0.000001	Extremely strong	Defence

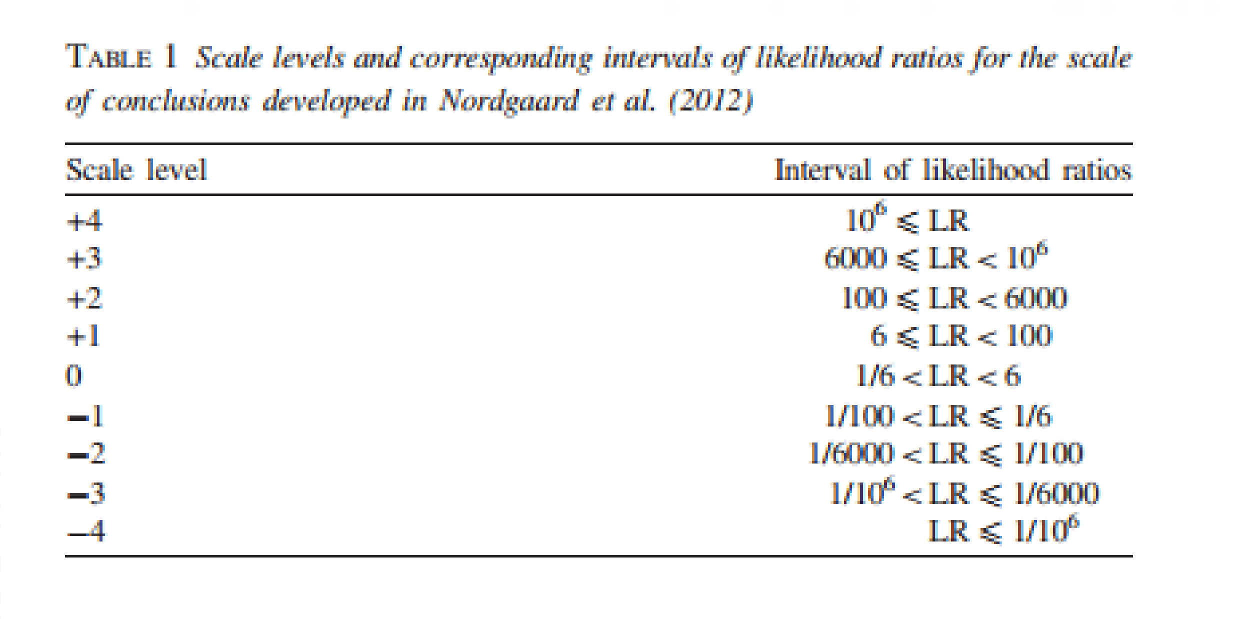
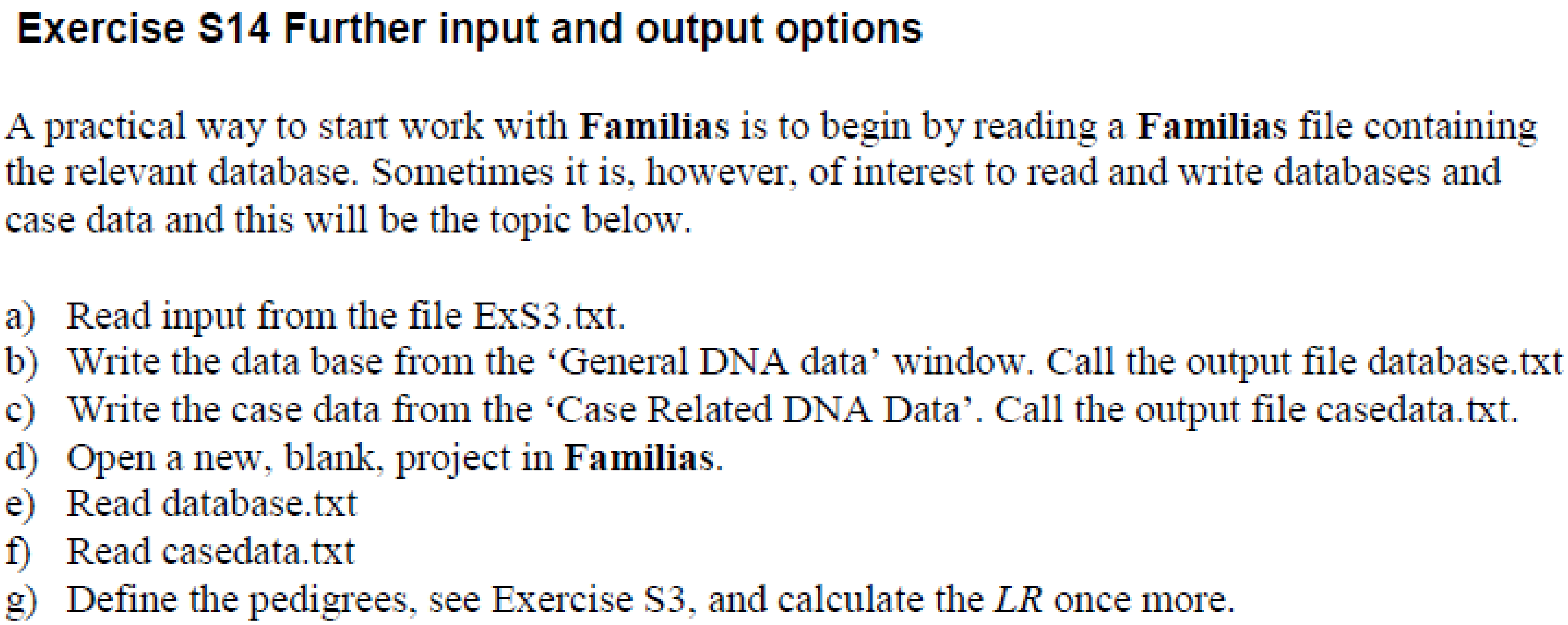


TABLE 1 *Scale levels and corresponding intervals of likelihood ratios for the scale of conclusions developed in Nordgaard et al. (2012)*

Scale level	Interval of likelihood ratios
+4	$10^6 \leqslant LR$
+3	$6000 \leqslant LR < 10^6$
+2	$100 \leqslant LR < 6000$
+1	$6 \leqslant LR < 100$
0	$1/6 < LR < 6$
−1	$1/100 < LR \leqslant 1/6$
−2	$1/6000 < LR \leqslant 1/100$
−3	$1/10^6 < LR \leqslant 1/6000$
−4	$LR \leqslant 1/10^6$

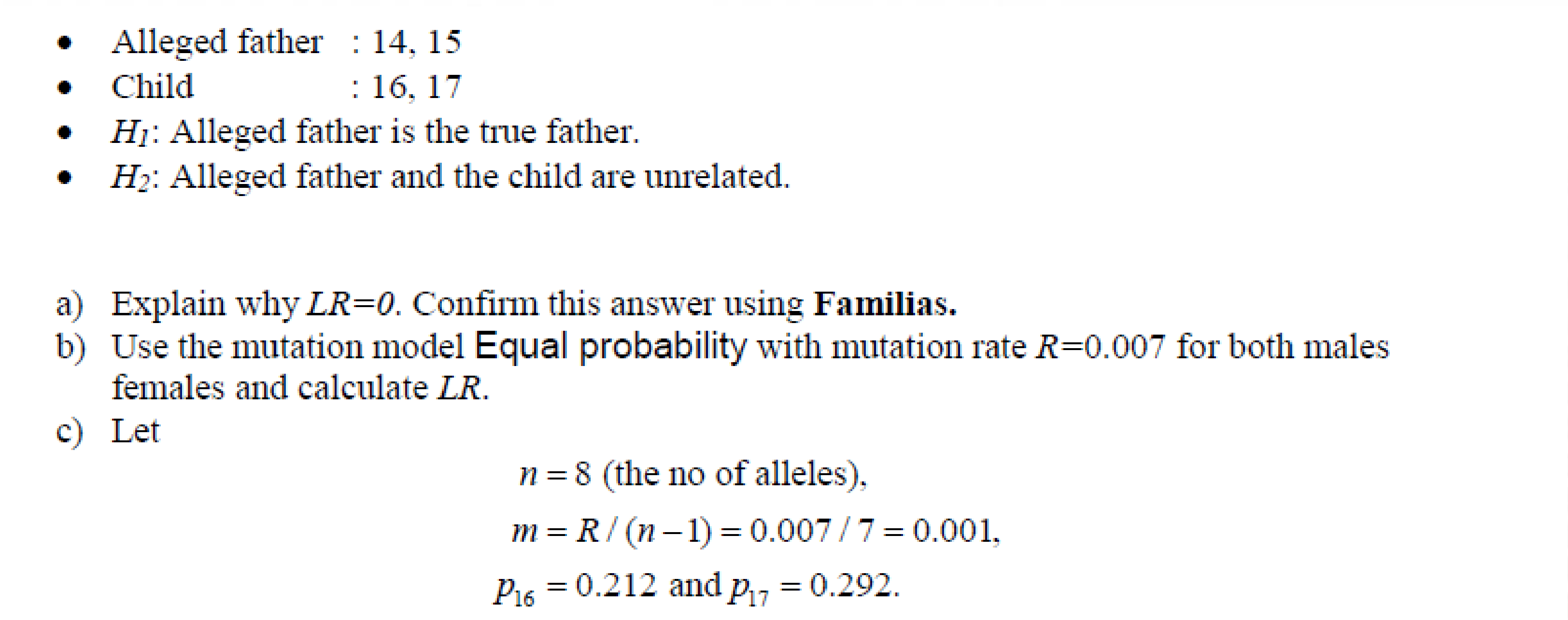


Exercise S14 Further input and output options

A practical way to start work with **Familias** is to begin by reading a **Familias** file containing the relevant database. Sometimes it is, however, of interest to read and write databases and case data and this will be the topic below.

a) Read input from the file ExS3.txt.
b) Write the data base from the 'General DNA data' window. Call the output file database.txt
c) Write the case data from the 'Case Related DNA Data'. Call the output file casedata.txt.
d) Open a new, blank, project in **Familias**.
e) Read database.txt
f) Read casedata.txt
g) Define the pedigrees, see Exercise S3, and calculate the *LR* once more.

Mutation rate



- Alleged father : 14, 15
- Child : 16, 17
- H_1: Alleged father is the true father.
- H_2: Alleged father and the child are unrelated.

a) Explain why $LR=0$. Confirm this answer using **Familias.**

b) Use the mutation model Equal probability with mutation rate R=0.007 for both males females and calculate LR.

c) Let

$$n = 8 \text{ (the no of alleles)},$$

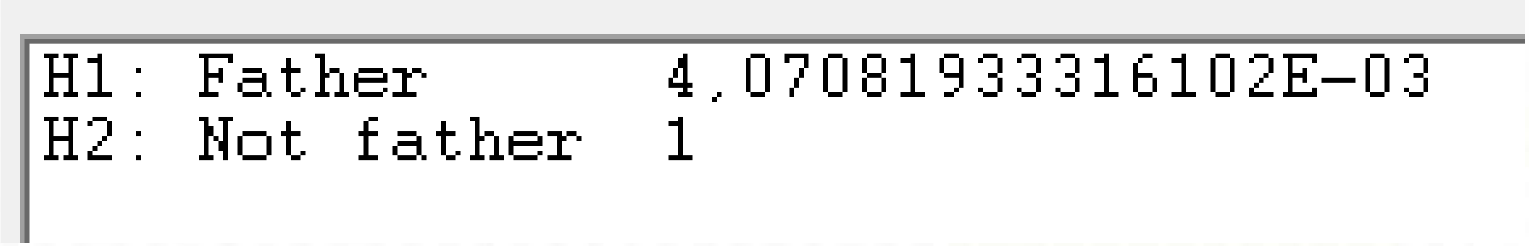
$$m = R/(n-1) = 0.007/7 = 0.001,$$

$$p_{16} = 0.212 \text{ and } p_{17} = 0.292.$$

Probability of a specific mutation 0.001

Mutation rate=7*0.001=0.007

Exercise S2



```
H1: Father        4,0708193316102E-03
H2: Not father    1
```

Formula confirmation:

```
> m=R/7
> p16=0.212
> p17=0.292
> LR=m*(p16+p17)/(2*p16*p17)
> LR
[1] 0.004070819
```