Workshop

Biostatistics in relationship testing
ESWG-meeting 2014

Daniel Kling
Norwegian Institute of Public Health and Norwegian University of Life Sciences, Norway

Andreas Tillmar
National Board of Forensic Medicine, Sweden
STATISTICAL METHODS IN RELATIONSHIP TESTING
PhD student Daniel Kling\textsuperscript{1} and forensic geneticist Andreas Tillmar\textsuperscript{2}, PhD
\textsuperscript{1}Department of Family Genetics, Norwegian Institute of Public Health, Oslo, Norway
\textsuperscript{2}Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden

09.00-10.30 **Session 1**
Introduction on how to calculate the LR in general
How to calculate LR for a simple paternity case
How to account for mutations, silent alleles, linkage and co-ancestry

10.30-11.00 **Coffee Break**

11.00-12.30 **Session 2**
Training examples under supervision

12.30-13.45 **Lunch Break**


15.15-15.45 **Coffee Break**

15.45-17.00 **Session 3**
How to calculate LR for X-chromosomal markers
FamLinkX

\url{http://www.famlink.se/BiostatWorkshopESWG2014.html}
Outline of this first session

- Likelihood ratio principle
- Theoretical: paternity issues
  - Calculation of LR “by hand”.
  - Accounting for:
    - Mutations
    - Null/silent alleles
    - Linkage and linkage disequilibrium (allelic association)
    - Co-ancestry
DNA-testing to solve relationship issues

Is Mike the father of Brian?

DNA-data

<table>
<thead>
<tr>
<th></th>
<th>vWA</th>
<th>D12S391</th>
<th>D21S11</th>
<th>....</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike:</td>
<td>12,13</td>
<td>22,22</td>
<td>31,32.2</td>
<td>....</td>
</tr>
<tr>
<td>Mother of Brian:</td>
<td>12,14</td>
<td>23,23.2</td>
<td>30,30</td>
<td>....</td>
</tr>
<tr>
<td>Brian:</td>
<td>13,14</td>
<td>22,23</td>
<td>30,31</td>
<td>....</td>
</tr>
</tbody>
</table>

Can we use the information from the DNA-data to answer the question?

**YES!**

We estimate the “weight of evidence” by biostatistical calculations.

We will follow the recommendations given by Gjertson et al., 2007

“ISFG: Recommendations on biostatistics in paternity testing.”
Calculating the “Weight of evidence” in a case with DNA-data

H₁: The alleged father is the father of the child
H₂: The alleged father is unrelated to the child

We wish to find the probability for H₁ to be true, given the evidence

= Pr(H₁|E)

Via Baye’s theorem:

\[
\frac{\text{Pr}(H₁ | E)}{\text{Pr}(H₂ | E)} = \frac{\text{Pr}(E | H₁)}{\text{Pr}(E | H₂)} \cdot \frac{\text{Pr}(H₁)}{\text{Pr}(H₂)}
\]

Posterior odds LR (or PI) Prior odds

Today we focus on the likelihood ratio (LR)

\[
\frac{\text{Pr}(DNA | H₁)}{\text{Pr}(DNA | H₂)}
\]
\[ LR = \frac{Pr(DNA \mid H_1)}{Pr(DNA \mid H_2)} \]

\[ Pr(DNA \mid H_1) = Pr(G_C, G_{AF}, G_M \mid H_1) = Pr(G_C \mid G_{AF}, G_M, H_1) \cdot Pr(G_{AF}, G_M \mid H_1) \]
Paternity trio – An example

\[
\Pr(DNA \mid H_1) = \Pr(G_C, G_{AF}, G_M \mid H_1) = \Pr(G_{AF}, G_M \mid H_1) \cdot \Pr(G_C \mid G_{AF}, G_M, H_1)
\]

\[
\Pr(DNA \mid H_1) = 2 \cdot p_a \cdot p_b \cdot 2 \cdot p_c \cdot p_d \cdot 0.5 \cdot 0.5
\]

\[
\Pr(DNA \mid H_2) = 2 \cdot p_a \cdot p_b \cdot 2 \cdot p_c \cdot p_d \cdot 0.5 \cdot p_c
\]
Paternity trio – An example

\[
LR = \frac{\Pr(DNA|H_1) = 2 \cdot p_a \cdot p_b \cdot 0.5 \cdot 0.5}{\Pr(DNA|H_2) = 2 \cdot p_a \cdot p_b \cdot 0.5 \cdot p_c} = \frac{0.5}{p_c}
\]

Probability to observe a certain allele in the population (e.g. population frequency)

\[
p_c \approx \frac{x_c + 1}{N + 1}
\]
Paternity trio – real data

\[ \Pr(DNA \mid H_1) = 2 \cdot p_{10} \cdot p_{11} \cdot 12 \cdot p_{13} \cdot p_{14.2} \cdot 0.5 \cdot 0.5 \]

\[ \Pr(DNA \mid H_2) = 2 \cdot p_{10} \cdot p_{11} \cdot 12 \cdot p_{13} \cdot p_{14.2} \cdot 0.5 \cdot p_{14.2} \]

LR = \frac{0.5}{p_{14.2}}
Paternity trio - Mutation

\[
Pr(DNA \mid H_1) = 2 \cdot p_{10} \cdot p_{11} \cdot 2 \cdot p_{13} \cdot p_{15.2} \cdot 0.5 \cdot \]

Mutation?
Mutations

“The possibility of mutation shall be taken into account whenever a genetic inconsistency is observed”

Brinkmann et al (1998) found 23 mutations in 10,844 parent/child offsprings. Out of these 22 were single step and 1 were two-step mutations.

AABB summary report covering ~4000 mutations showed approx 90-95% single step, 5-10% two-step and 1% or less more than two steps.

Note that these are the observed mutation rate, not the actual mutation rate (hidden mutations, assumption of mutation with the smallest step.

Mutation rate depend on marker, sex (female/male), age of individual, allele size
Mutations cont

Allele: 9 10 11 12 13 14 15
-3 -2 -1 0 +1 +2 +3

1- \( \mu_{tot} \)

\( \mu_{-3} \) \( \mu_{-2} \) \( \mu_{-1} \) \( \mu_{+1} \) \( \mu_{+2} \) \( \mu_{+3} \)

\[ \mu_{tot} = \mu_{+1} + \mu_{-1} + \mu_{+2} + \mu_{-2} + \ldots. \]

Approaches to calculate LR accounting for mutations:

\[ LR \sim (\mu_{tot} \cdot \text{adj\_steps})/p(\text{paternal allele}) \text{ e.g “stepwise mutation model”} \]

\[ LR \sim \mu/A \] (A=average PE)

\[ LR \sim \mu \]

\[ LR \sim \mu/p(\text{paternal allele}) \]
Paternity trio - Mutation

\[ \text{mut}_{13 \rightarrow 14.2} + \text{mut}_{15.2 \rightarrow 14.2} \]

Mutation model decreasing with range

\[ \text{mut}_{15.2 \rightarrow 14.2} \]

Total mutation rate

50% for 15.2 to be transmitted

90% of mutations are 1 step

\[ 0.5 \cdot \mu_{\text{Tot}} \cdot 0.9 \cdot 0.5 \]

50% are loss of fragment size
Paternity trio - Mutation

\[
\Pr(DNA \mid H_1) = 2 \cdot p_{10} \cdot p_{11} \cdot 2 \cdot p_{13} \cdot p_{15} \cdot 0.5 \cdot (\text{mut}_{13->14.2} + \text{mut}_{15.2->14.2})
\]

\[
\Pr(DNA \mid H_2) = 2 \cdot p_{10} \cdot p_{11} \cdot 2 \cdot p_{13} \cdot p_{15} \cdot 0.5 \cdot p_{14.2}
\]

\[
\text{LR} = \frac{0.5 \cdot \mu_{\text{Tot}} \cdot 0.9 \cdot 0.5}{p_{14.2}}
\]
Paternity trio – Silent allele

- $G_M = 10,11$
- $G_{AF} = 17.2,17.2$
- $G_C = 10,10$
- $G_C = 10, s$
- $G_{AF} = 17.2, s$

M → AF → C

C

$G_C = 10, s$
Null/silent allele

“STR null alleles mainly occur when a primer in the PCR reaction fails to hybridize to the template DNA...”

“Laboratories should try to resolve this situation with different pairs of STR primers”

“...no data from which to estimate the null allele frequency exists, consider a generous bracket of plausible values for the frequency and compute a corresponding range of values for the PI. If the resulting range of case-specific PI values are all large (as large as a laboratory’s threshold value for issuing non-exclusion paternity reports), then report that the PI is “at least greater than the smallest value” in the range”

Impact of null/silent allele may vary on the pedigree and the genotype constellations
Paternity trio – Silent allele

\[ \Pr(DNA | H_1) = 2 \cdot p_{10} \cdot p_{11} \cdot 2 \cdot p_{17.2} \cdot p_s \cdot 0.5 \cdot 0.5 \]

\[ \Pr(DNA | H_2) = 2 \cdot p_{10} \cdot p_{11} \cdot (2 \cdot p_{17.2} \cdot p_s + p_{17.2} \cdot p_{17.2}) \cdot 0.5 \cdot (p_{10} + p_s) \]

\[ LR = \frac{p_{17.2} \cdot p_s}{(2 \cdot p_{17.2} \cdot p_s + p_{17.2} \cdot p_{17.2}) \cdot (p_{10} + p_s)} \]
**Linkage and Linkage disequilibrium**

- **Linkage**
  - Can be described as the co-segregation of closely located loci within a family or pedigree.
  - **Effects the transmission probabilities!**

- **Linkage disequilibrium (LD)**
  - Allelic association.
  - Two alleles (at two different markers) which is observed more often/less often than can be expected.
  - **Effects the founder genotype probabilities, not the transmission probabilities!**
Marker 1  Marker 2
Mother:  16,18  21,24

If LD, different probabilities for each phase

If linkage, different probability of transmission

Distance R, Theta
Mother: 16,18 21,24
Child1: 16,25 21,29
Child2: 16,22 24,30
Linkage and LD when it comes to forensic DNA-markers?

- 14 pair of autosomal STRs < 50cM
  - Phillips et al 2012
  - vWA-D12S391 (around 10cM, i.e. 10% recomb rate)

- Linkage will have an effect on LR for some pedigrees
  - Incest.
  - Sibling, uncle-nephew.
  - Impact of linkage will depend on
    - Pedigree
    - Markers typed
    - Individuals’ DNA data

- LE for vWA-D12S391!
Linkage cont.

• Assuming LE (no allelic association)
  – “linkage have an impact…”

• Thus...
  – For standard paternity cases (father vs unrelated), linkage has no impact on LR
  – BUT for paternity vs uncle (or other close related alternative), linkage has impact on LR
Presenting the evidence

• Likelihood ratio (LR) or Paternity index (PI)
• Posterior probability (must set priors)

$$W_i = \frac{\pi_i L_i}{\sum_{j=1}^{n} \pi_j L_j}$$

$L_i$=Likelihood under hypothesis $i$
$\pi_i$=Prior probability for hypothesis $i$

Simplifies to $LR/LR+1$,
assuming two hypotheses with equal priors
Subpopulation correction

- What is subpopulation effects?
  - First developed by Wright in (1965)
  - Balding (1994)
- Hardy Weinberg equilibrium
- How to estimate it?
  - Reasonable values (0.01-0.05)
  - Sampling formula \( (Fst = \theta) \)

\[
p'(a_i) = \frac{Fst \ast n_{a_i} + (1 - Fst) \ast p(a_i)}{1 + (N_{Obs} - 1) \ast Fst}
\]
Example 1. Genotype probabilities

- **HWE**
  - **Homozygous**: \( p(A)p(A) \)
  - **Heterozygous**: \( 2p(A)p(B) \)

- **HWD**
  - **Homozygous**: \( \text{Fst}^*p(A) + (1-\text{Fst})^*p(A)p(A) \)
  - **Heterozygous**: \( (1-\text{Fst})^*p(A)p(B) \)

\[
P(A, A) = P(\text{Sampling two } A:s) = \frac{\theta^*0 + (1-\theta)p(A)}{1+(0-1)\theta} \cdot \frac{\theta^*1 + (1-\theta)p(A)}{1+(1-1)\theta} = \theta p(A) + (1-\theta)p(A)^2
\]

\[
P(A, B) = P(\text{Sampling one } A \text{ and one } B) = \frac{\theta^*0 + (1-\theta)p(A)}{1+(0-1)\theta} \cdot \frac{\theta^*0 + (1-\theta)p(B)}{1+(1-1)\theta} = (1-\theta)p(A)p(B)
\]
Example 2. Random Match Prob.

- \( H_1 \): The profiles G1 and G2 are from the same individual
- \( H_2 \): The profiles G1 and G2 are from different individuals

Let G1=A,A and G2=A,A

- **HWE:**
  \[
  \frac{1}{RMP} = \frac{P(\text{Data} \mid H_1)}{P(\text{Data} \mid H_2)} = \frac{P(G1)}{P(G1) \times P(G2)} = \frac{1}{p(A)^2}
  \]

- **HWD:**
  \[
  \frac{1}{RMP} = \frac{P(\text{Data} \mid H_1)}{P(\text{Data} \mid H_2)} = \frac{P(G1)}{P(G1, G2)} = \langle \text{Sampling} \rangle = \frac{P(\text{Sampling two A:s})}{P(\text{Sampling four A:s})}
  \]

  \[
  = \frac{(1 + \theta)(1 + 2\theta)}{[2\theta + (1 - \theta)p(A)][3\theta + (1 - \theta)p(A)]}
  \]
Example 3. Paternity

\( H_1 \): The alleged father (AF) is the real father

\( H_2 \): AF and the child are unrelated.

\[ p(A) = 0.05 \]
Standard paternity case. First marker

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

\[ = \frac{P(\text{child} | \text{mother}, AF)}{P(\text{child} | \text{mother})} = \left\langle \theta = 0 \right\rangle = \frac{1}{p_A} = \frac{1}{0.05} = 20. \]

\[ LR = \frac{P(\text{mother})P(AF)P(\text{child} | \text{mother}, AF)}{P(AF)P(\text{mother})P(\text{child} | \text{mother})} = \left\langle \theta \neq 0 \right\rangle = \frac{P(\text{Sampling two A:s and two B:s})}{P(\text{Sampling three A:s and two B:s})} = \]

\[ = \frac{1}{P(\text{Sampling the third A})} = \frac{1}{2\theta + (1-\theta)p_A} = \frac{1 + 3\theta}{2\theta + (1-\theta)p_A} \]
\[
LR = \frac{P(\text{mother})P(AF)P(\text{child} \mid \text{mother}, AF)}{P(AF)P(\text{mother})P(\text{child} \mid \text{mother})} = \frac{\theta \neq 0}{P(\text{Sampling three A:s and two B:s})} = P(\text{Sampling two A:s and two B:s}) \\
= \frac{1}{P(\text{Sampling the third A})} = \frac{1}{2\theta + (1-\theta)p_A} = \frac{1+3\theta}{2\theta + (1-\theta)p_A}
\]

**LR as a function of Theta**
Example 4. Siblings

H0: Two persons are full siblings

H1: The same two persons are unrelated

● The two persons are homozygous A,A
  \( p(A)=0.05 \)

● We use the “IBD method”

\[
LR = \frac{P(\text{Data} \mid H0)}{P(\text{Data} \mid H1)} = \langle \theta \neq 0 \rangle = \\
\frac{P(\text{Sampling two A:s})*P(\text{IBD}=2|H0)+P(\text{Sampling three A:s})*P(\text{IBD}=1|H0)+P(\text{Sampling four A:s})*P(\text{IBD}=0|H0)}{P(\text{Sampling four A:s})} \\
= 0.25 + 0.5P(\text{Sampling the third A}) + 0.25P(\text{Sampling the last two A:s}) \\
= \frac{0.25 + 0.5 \frac{2\theta + (1-\theta)p_A}{1 + (2-1)\theta} + 0.25 \frac{2\theta + (1-\theta)p_A \ast 3\theta + (1-\theta)p_A}{1 + (2-1)\theta}}{1 + (3-1)\theta} \\
= \frac{2\theta + (1-\theta)p_A \ast 3\theta + (1-\theta)p_A}{1 + (2-1)\theta} \frac{3\theta + (1-\theta)p_A}{1 + (3-1)\theta}
\]
Example 4. Siblings

LR as a function of Theta
Example 5. General relationships

- We have N number of founder genotypes
- If Fst!=0 these are not independent
- We use the sampling formula to calculate updated allele frequencies
- It is more complex to do by hand
  - In simple pairwise relationships use the IBD method (See Example 4)
  - Otherwise, use validated software!