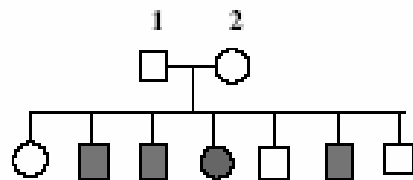
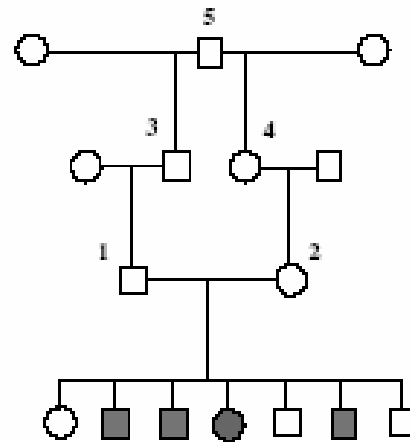


Adjusting for Relatedness among Founders in Linkage Analysis

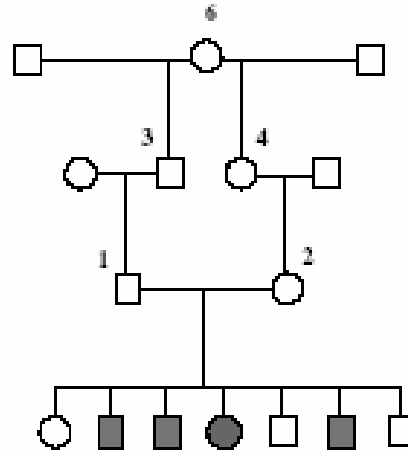
Thore Egeland
Dept of Medical Genetics,
Ullevål University Hospital
Based on Sheehan&Egeland, submitted
ICMS, Edinburgh May 7-11, 2007



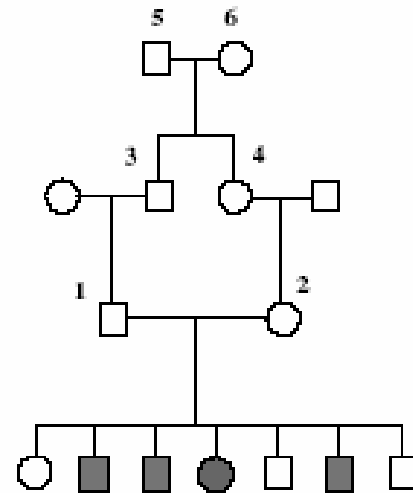
(a)



(b)



(c)



(d)

Problem, purpose summarised

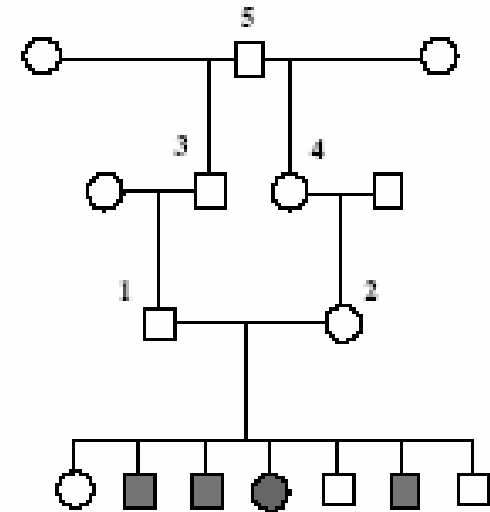
- Normally pedigree considered certain and likelihood modelled as a function of *population parameters, transmission parameters and penetrance parameters (Thompson, 2000)*.
- Ignoring pedigree uncertainty can lead to false positives or false negatives. Goal:

Model and study impact of founder relatedness.

- *Recent reference, different approach: Hössjer (2006).*

Simulation experiment

- 10 markers on each of 10 chromosomes.
- 10cMs apart.
- 5 equiprequent alleles.
- Autosomal recessive disease, penetrance 0.98.
- Disease allele freq. 0.0001.
- No phenocopies.
- Disease locus at 49cM.
- Simulations from pedigree b.
- Data for 1,2 and children.
- Implemented in Allegro.



(b)

Outline of method

1. Specify pedigree sample space $G=\{g_1, \dots, g_k\}$.
2. Specify prior on G .
3. Calculate posteriors $P(g_i|\text{data})$.
4. Calculate lod-scores weighted wrt to posteriors:

$$lod_i = \log_{10}(L_{A,i} / L_0).$$

1. Pedigree sample space

- *Hard* prior information used to restrict sample size.
- Extremely rare recessive disease:
 - Include pedigree with inbreeding.
 - Helpfull even if correct pedigree is not in sample space?

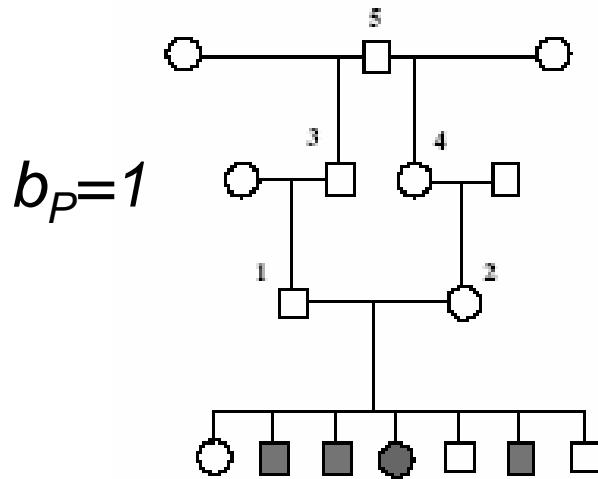
2. Prior

$$\Pr(g) = c \prod_{i=1}^s M_i^{b_i(g)} \prod_{j \neq k=1}^n R_{jk}^{o_{jk}(g)} .$$

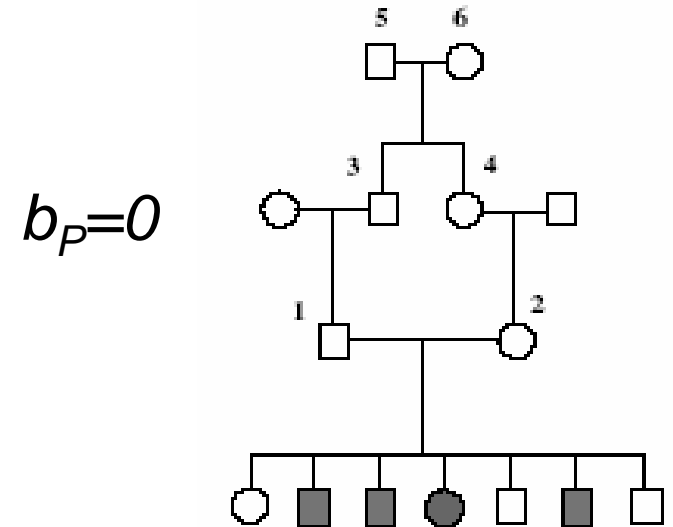
- M denotes *global* user parameters relating to inbreeding, promiscuity,...
- R denotes *local* user parameters relating to parent-child relations.
- $b(g)$ and $o(g)$ calculated from pedigree.
- Flat prior for parameter values of 1.
- Sheehan&Egeland, Ann Hum Gen (2007)
Egeland et al., (2000). For Sci Int.

Prior. *Global* parameter

b_P = number of pairs with precisely one parent in common.



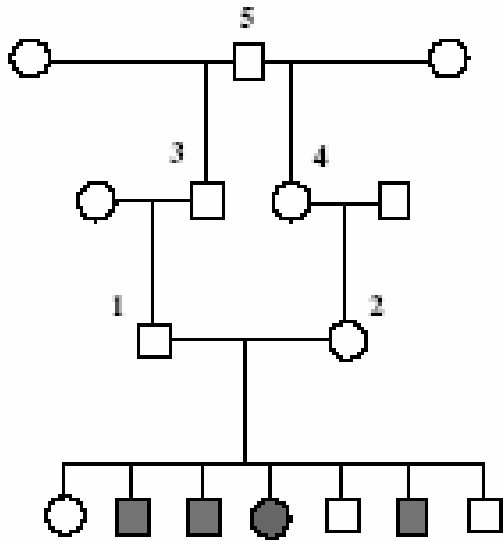
(b)



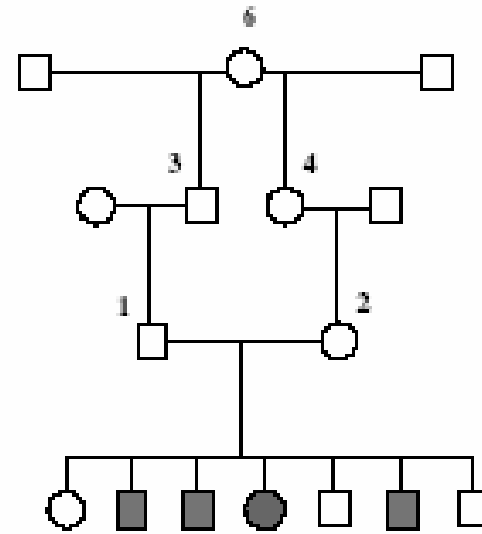
(d)

$$\frac{\text{prior ped. (b)}}{\text{prior ped. (c)}} = K = \frac{M_P^1}{M_P^0} = M_P = 0.5.$$

Prior. *Local* parameters



(b)



(c)

$$\frac{\text{prior ped. (b)}}{\text{prior ped. (c)}} = \mathbf{K} = \frac{R_{53}R_{54}}{R_{63}R_{64}} = \frac{10 * 10}{1 * 1} = 100$$

3. Calculate posteriors

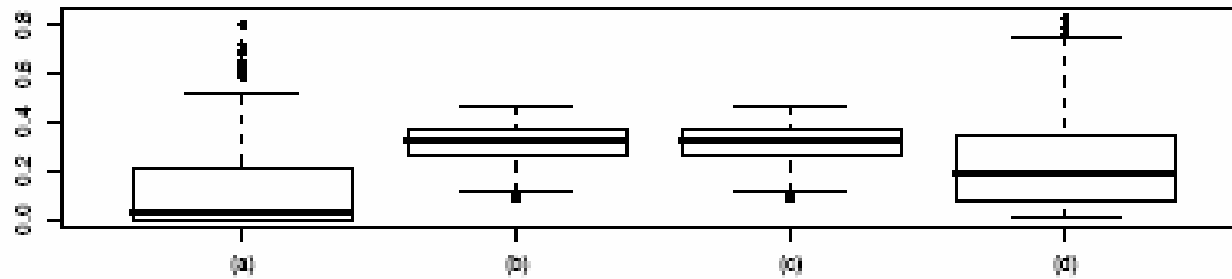
- Calculate posteriors $P(g_i|\text{data})$
 - *based on markers not linked to disease.*
 - Bayes' theorem. Likelihoods calculated using Merlin.

Posteriors. Results

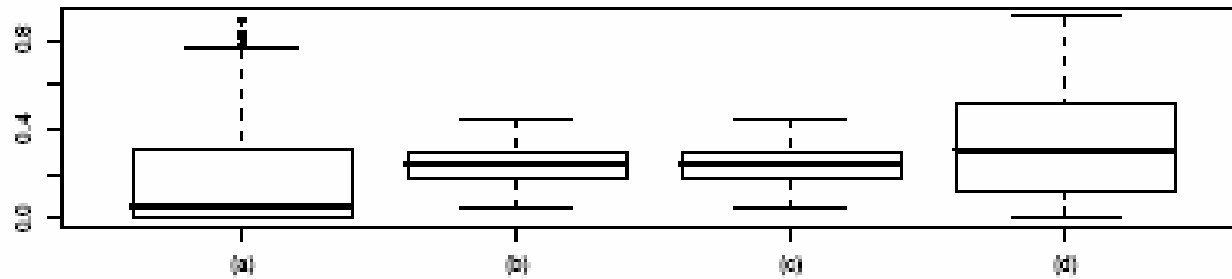
pedigree	prior1	posterior1	prior2	posterior2	prior3	posterior3
unrelated	0,250	0,137	0,333	0,184	0,005	0,005
same gf	0,250	0,312	0,167	0,238	0,495	0,618
same gm	0,250	0,312	0,167	0,238	0,005	0,006
cousins	0,250	0,239	0,333	0,340	0,495	0,371

prior1: 'flat'
prior2: 'global'
prior3: 'local'

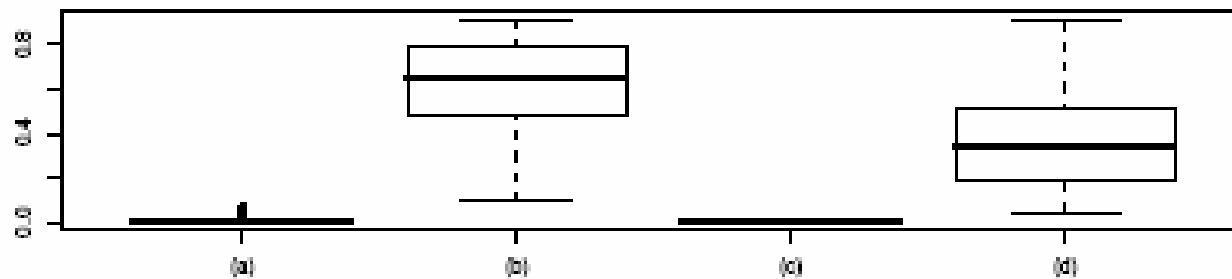
Posterior: Flat prior



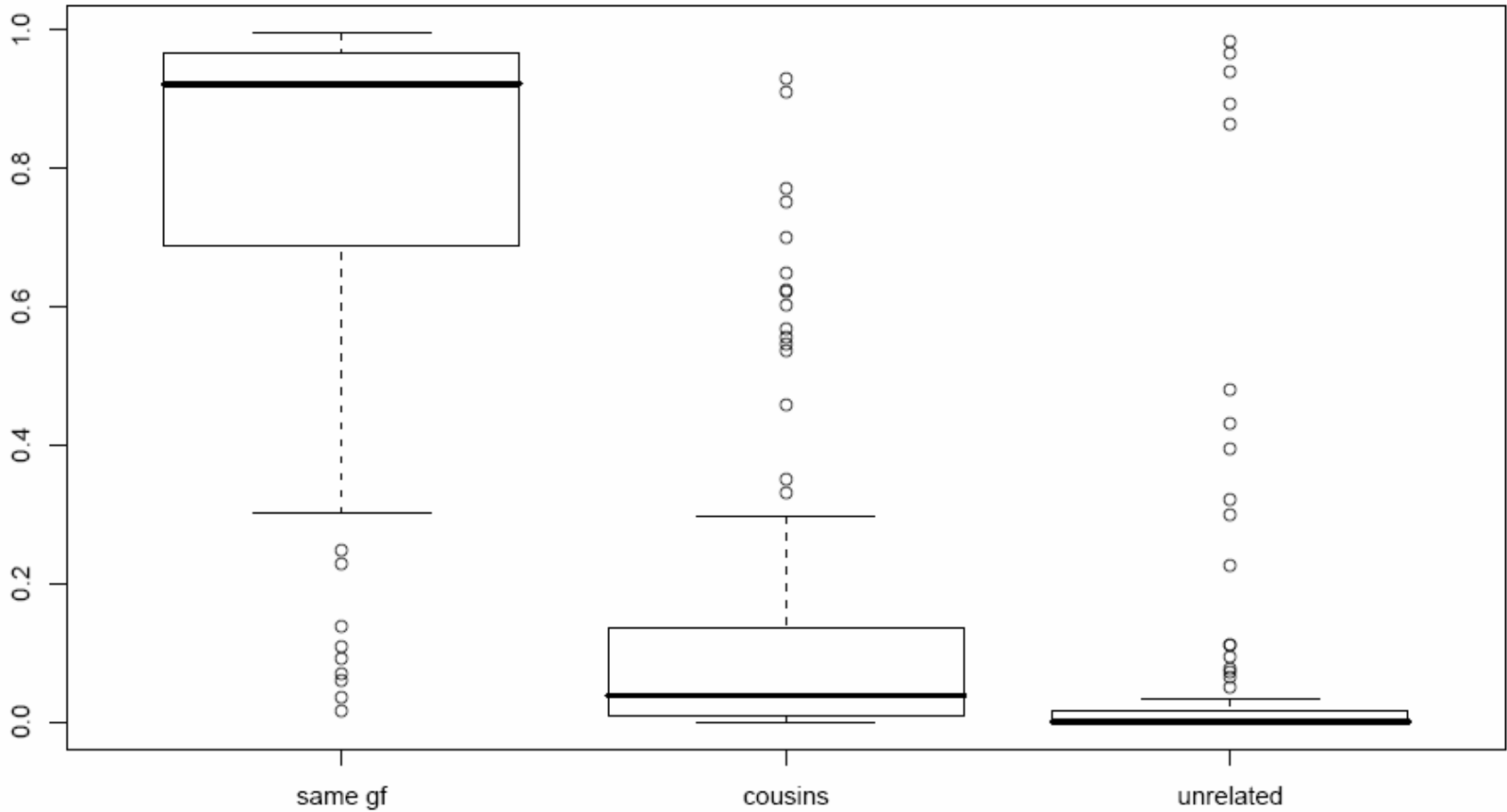
Posterior: Multiple marriages downweighted



Same grand father more likely ($R=(10,10,1,1)$)



400 markers, 20 chromosomes, simulated assuming same grand father. Flat prior



More data, 'same grand mother pedigree' disregarded

Outline of method

1. Specify pedigree sample space $G=\{g_1, \dots, g_k\}$.
2. Specify prior on G .
3. Calculate posteriors $P(g_i|\text{data})$ based on markers not linked to disease.
4. **Calculate lod-scores weighted wrt to posteriors:**

$$lod_i = \log_{10}(L_{A,i} / L_0).$$

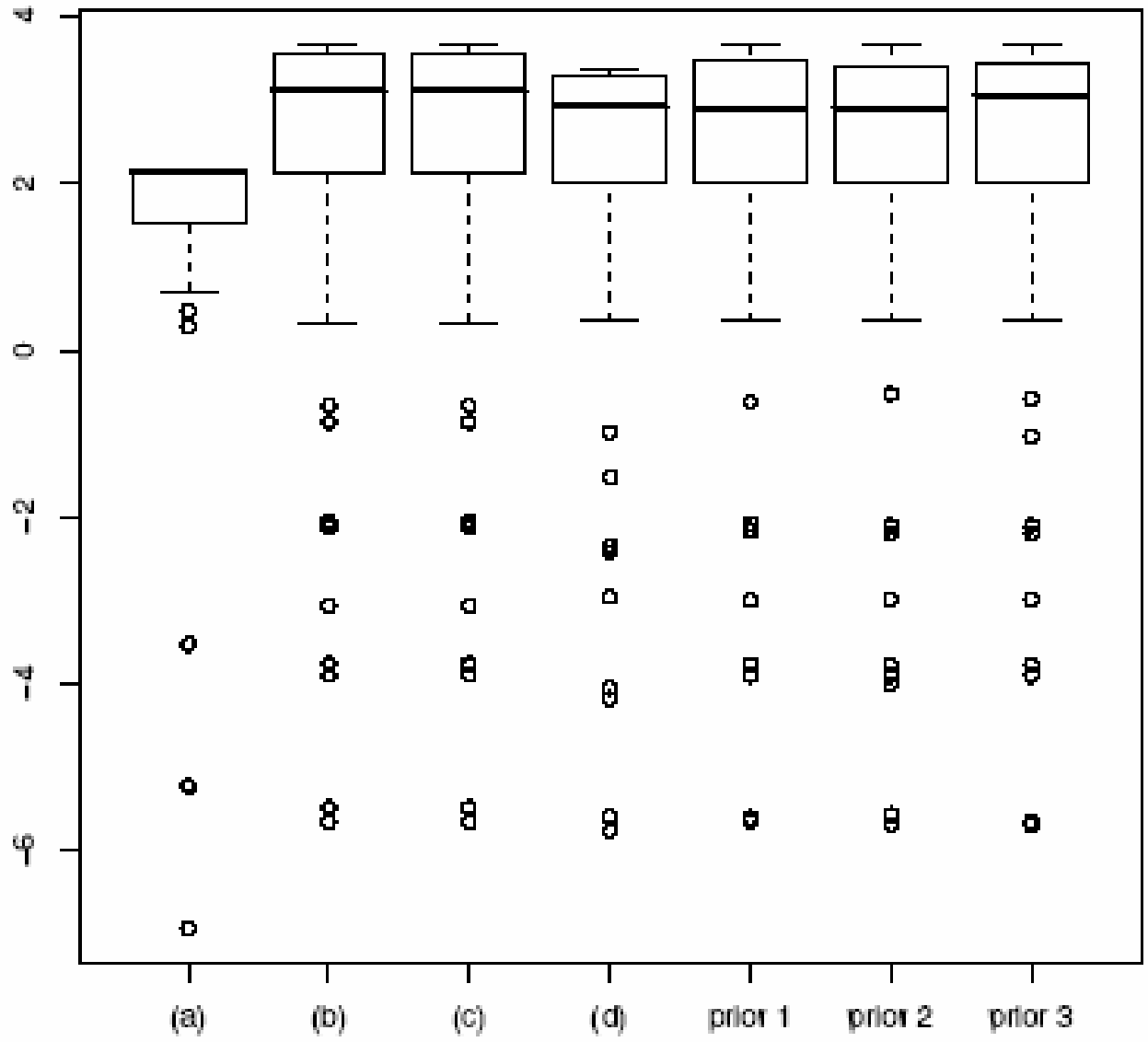
4. Calculate lod-scores weighted wrt to posteriors

$$\text{lod}_i = \log_{10}(L_{A,i} / L_0),$$

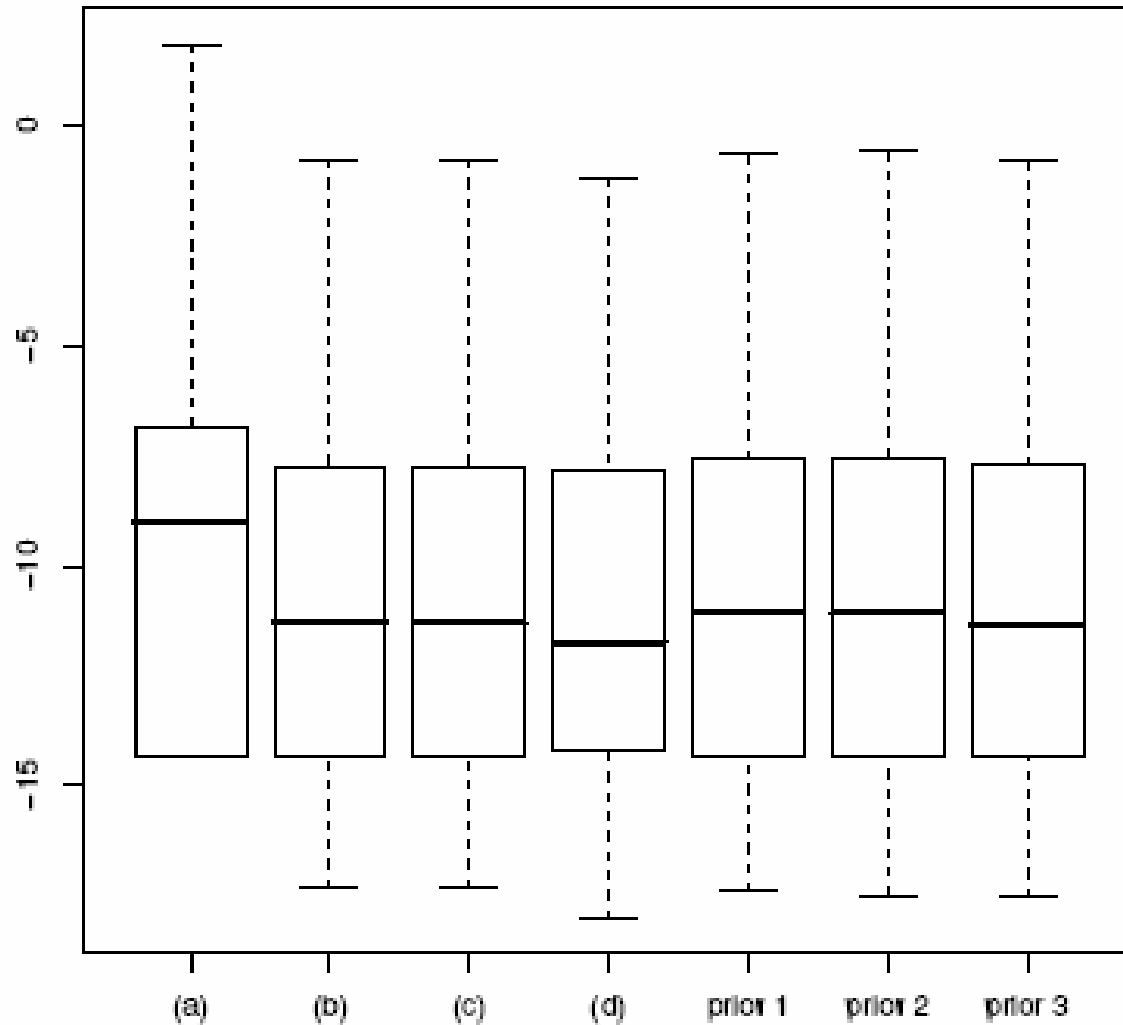
$$L_{A,i} = L_{A,i,1}P(g_1 | data) + K + L_{A,i,k}P(g_k | data)$$

$$L_0 = L_{0,1}P(g_1 | data) + K + L_{0,k}P(g_k | data)$$

Implemented in Merlin.



Lod scores for marker closest to disease locus



Lod scores for marker unlinked to disease locus

Discussion

- Alternative. Expected lod score:

$$elod_i = \log_{10} \left(\frac{L_{A,i,1}}{L_{o,1}} \right) P(g_1 | data) + K + \log_{10} \left(\frac{L_{A,i,k}}{L_{o,k}} \right) P(g_k | data)$$

- Ignore markers suspected to be linked to disease for posteriors. Practical? Necessary?
- Can relations be determined unambiguously?
- What about LD?
- Prior practical?
- Full Bayesian approach?
- Genotyping errors (Epstein et al., 2000).
- Alternative approaches? Alternative applications?:
Power calculations. Homozygosity mapping.

References

- *Hössjer, O. (2006). Theor Pop Biol 70, 146-173.*
- *Leutenegger, A.-L et al. (2002). Gen Epidem, 413-425.*
- *Miano et al. (2000). Am J Hum Gen, 1146-1148.*
- *Sheehan, N.A. & Egeland (2007). Ann Hum Gen.*
- *Thompson, E. (2000). NSF-CBMS, vol 6. References therein.*
- *Weir, BS. Nat Rev Genet, 2006.*

- *More complete list of references in Sheehan&Egeland, submitted.*