Type 1 Error and Power Calculation for Association Analysis

Pak Sham & Shaun Purcell

Advanced Workshop

Boulder, CO, 2005

Statistical Tests

Standard test theory

- Type 1: Rejecting the null hypothesis when it is true (α).
- Type 2: Not rejecting the null hypothesis when it is false (β).
- Fix α (e.g. genome wide α of 0.05 for linkage).
- Optimise 1-β

Gold standard: REPLICATION

Problem: Low Replication Rate

Hirschhorn et al. 2002: Reviewed 166 putative single allelic association with 2 or more replication attempts:

- 6 reliably replicated (≥75% positive replications)
- 97 with at least 1 replication
- 63 with no subsequent replications

Other such surveys have similar findings (Ioannidis 2003; Ioannidis et al. 2003; Lohmueller et al. 2003)

Reasons for Non-Replication

The original finding is false positive

- Systematic bias (e.g. artefacts, confounding)
- Chance (type 1 error)

The attempted replication is false negative

- Systematic bias (e.g. artifacts, confounding)
- Heterogeneity (population, phenotypic)
- Chance (inadequate power)

Type 1 Error Rate vs False Positive Rate

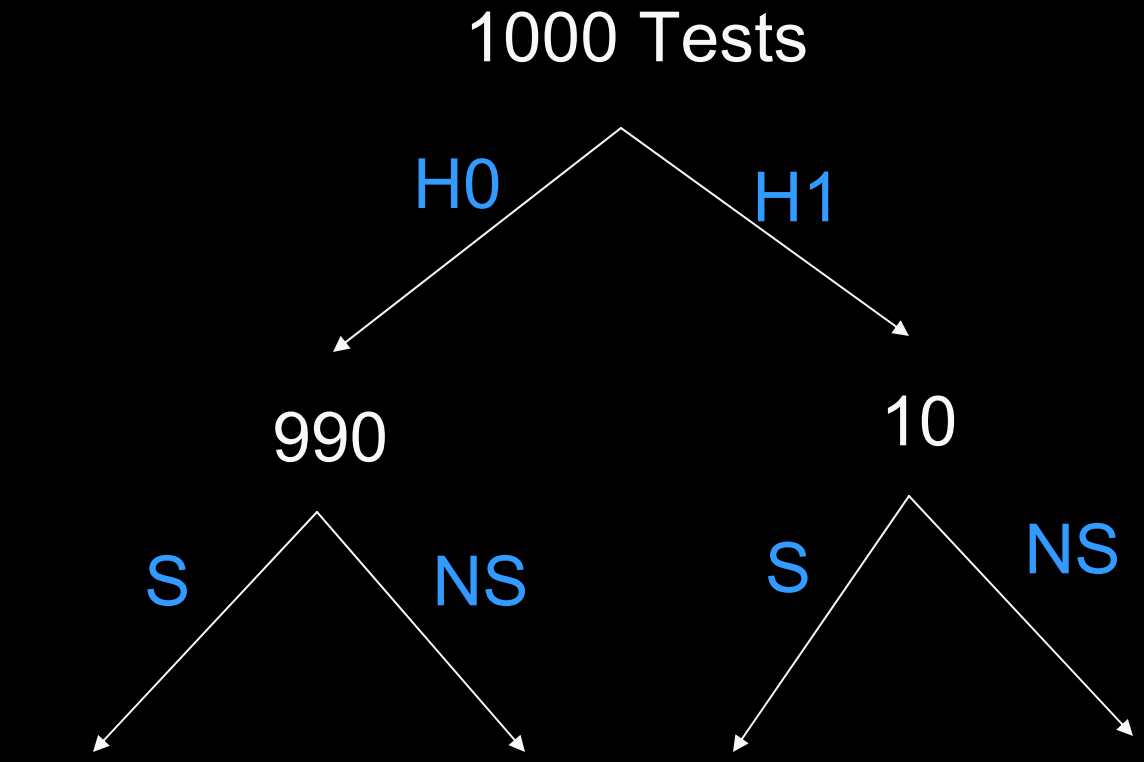
Type 1 error rate = probability of significant result when there is no association

False positive rate = probability of no association among significant results

Why so many false positives?

- Multiple testing
  - Multiple studies
  - Multiple phenotypes
  - Multiple polymorphisms
  - Multiple test statistics
- Not setting a sufficiently small critical p-value
- Inadequate Power
  - Small sample size
  - Small effect size
  - $\rightarrow$ High false positive rate

Both error rates affect false positive rate



α	S
0.05	19.5
0.01	9.9

1-β	S
0.8	8
0.2	2

Multiple testing correction

Bonferroni correction: Probability of a type 1 error among k independent tests each with type 1 error rate of α

$$\alpha^* = 1-(1-\alpha)^k \approx k\alpha$$

Permutation Procedures

Permute case-control status, obtain empirical distribution of maximum test statistic under null hypothesis

False Discovery Rate (FDR)

Under H0: P-values should be distributed uniformly between 0 and 1.

Under H1: P-values should be distributed near 0.

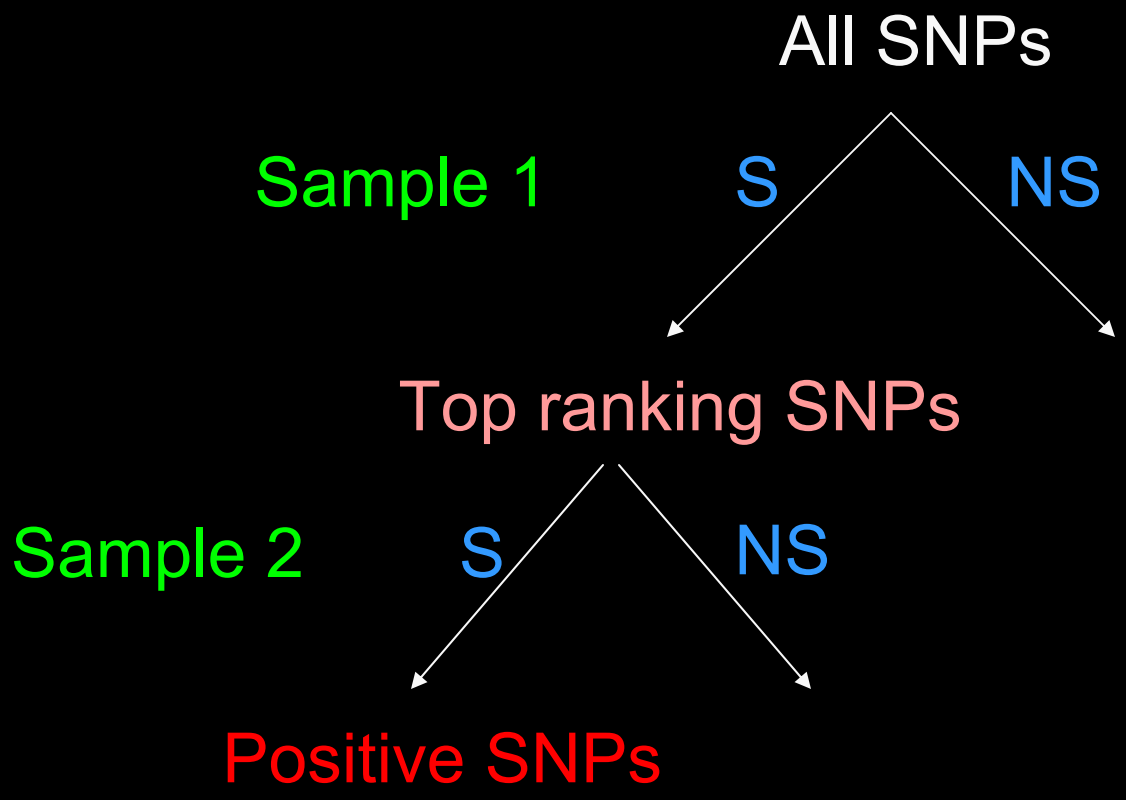
Observed distribution of P-values is a mixture of these two distributions.

FDR method finds a cut-off P-value, such that results with smaller P-values will likely (e.g. 95%) to belong to the H1 distribution.

False Discovery Rate (FDR)

Ranked P-value	FDR	Rank	FDR*Rank
0.001	0.05	1/7	0.007143
0.006	0.05	2/7	0.014286
0.01	0.05	3/7	0.021429
0.05	0.05	4/7	0.028571
0.2	0.05	5/7	0.035714
0.5	0.05	6/7	0.042857
0.8	0.05	7/7	0.05

Multi-stage strategies



Meta-Analysis

Combine results from multiple published studies to:

- enhance power
- obtain more accurate effect size estimates
- assess evidence for publication bias
- assess evidence for heterogeneity
- explore predictors of effect size

Discrete

Case-control

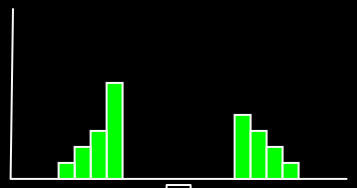
	Aff	UnAff
A	n_1	n_2
a	n_3	n_4

TDT

	Tr	UnTr
A	n_1	n_2
a	n_3	n_4

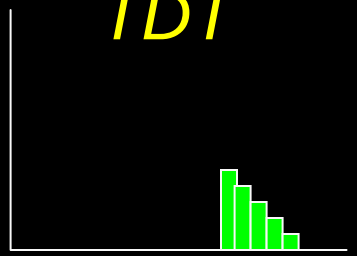
Threshold

Case-control



	High	Low
A	n_1	n_2
a	n_3	n_4

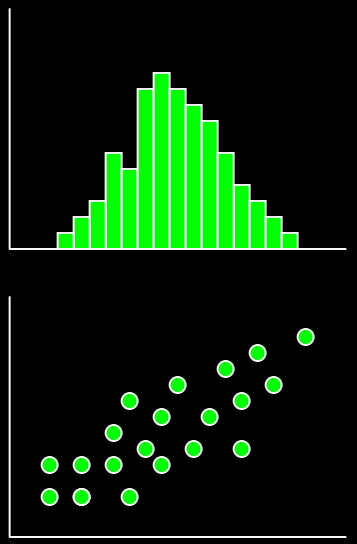
TDT



	Tr	UnTr
A	n_1	n_2
a	n_3	n_4

Quantitative

Variance components



Discrete trait calculation

p	Frequency of high-risk allele
K	Prevalence of disease
R_{AA}	Genotypic relative risk for AA genotype
R_{Aa}	Genotypic relative risk for Aa genotype
N, α, β	Sample size, Type I & II error rate

Risk is P(D|G)

$g_{AA} = R_{AA}\, g_{aa}$ $\qquad$ $g_{Aa} = R_{Aa}\, g_{aa}$

$K = p^2\, g_{AA} + 2pq\, g_{Aa} + q^2\, g_{aa}$

$g_{aa} = K / (p^2\, R_{AA} + 2pq\, R_{Aa} + q^2)$

Odds ratios (e.g. for AA genotype) = $\dfrac{g_{AA} / (1- g_{AA})}{g_{aa} / (1- g_{aa})}$

Need to calculate P(G|D)

Expected proportion *d* of genotypes in cases

$d_{AA} = g_{AA}\, p^2 / (g_{AA}p^2 + g_{Aa}2pq + g_{aa}q^2)$

$d_{Aa} = g_{Aa}\, 2pq / (g_{AA}p^2 + g_{Aa}2pq + g_{aa}q^2)$

$d_{aa} = g_{aa}\, q^2 / (g_{AA}p^2 + g_{Aa}2pq + g_{aa}q^2)$

$$P(G \mid D) = \frac{P(D \mid G)P(G)}{\sum_G P(D \mid G)P(G)}$$

Expected number of A alleles for cases

$2N_{Case} (d_{AA} + d_{Aa} / 2)$

Expected proportion *c* of genotypes in controls

$c_{AA} = (1-g_{AA})\, p^2 / ((1-g_{AA})\, p^2 + (1-g_{Aa})\, 2pq + (1-g_{aa})\, q^2)$

Full contingency table

	"A" allele	"a" allele
Case	$2N_{Case}\ (\ d_{AA} + d_{Aa} / 2\ )$	$2N_{Case}\ (\ d_{aa} + d_{Aa} / 2\ )$
Control	$2N_{Control}\ (\ c_{AA} + c_{Aa} / 2\ )$	$2N_{Control}\ (\ c_{aa} + c_{Aa} / 2\ )$

$$\chi^2 = \frac{(O-E)^2}{E}$$

Incomplete LD

Effect of incomplete LD between QTL and marker

	A	a
M	$pm_1 + \delta$	$qm_1 - \delta$
m	$pm_2 - \delta$	$qm_2 + \delta$

$\delta = D' \times D_{MAX}$ $\quad D_{MAX} = \min\{pm_2 , qm_1\}$

Note that linkage disequilibrium will depend on both

D' and QTL & marker allele frequencies

Incomplete LD

Consider genotypic risks at marker:

$$P(D|MM) = [\ (pm_1 + \delta)^2\ P(D|AA)$$
$$+\ 2(pm_1 + \delta)(qm_1 - \delta)\ P(D|Aa)$$
$$+\ (qm_1 - \delta)^2\ P(D|aa)\ ]$$
$$/\ m_1^2$$

Geno.	Haplo.
AAMM	AM/AM
AaMM	AM/aM or aM/AM
aaMM	aM/aM
MM	

Calculation proceeds as before, but at the marker

Fulker association model

The genotypic score (1,0,-1) for sibling *i* is decomposed into between and within components:

$$[A]_i = \left(\frac{\sum_{j=1}^{s} A_j}{s} \right) + \left(A_i - \frac{\sum_{j=1}^{s} A_j}{s} \right) = [A_B]_i + [A_W]_i$$

sibship genotypic mean

deviation from sibship genotypic mean

NCPs of B and W tests

Approximation for between test

$$\lambda_B \approx \frac{\frac{s+1}{2}V_A + \frac{s+3}{4}V_D}{V_N + sV_S}$$

Approximation for within test

$$\lambda_W \approx (s-1)\left[\frac{\frac{1}{2}V_A + \frac{3}{4}V_D}{V_N}\right]$$

Sham et al (2000) AJHG 66

GPC

Usual URL for GPC

http://statgen.iop.kcl.ac.uk/gpc/

Purcell S, Cherny SS, Sham PC. (2003)
Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits.
Bioinformatics, 19(1):149-50

Exercise 1:

Candidate gene case-control study

Disease prevalence 2%

Multiplicative model

genotype risk ratio Aa = 2

genotype risk ratio AA = 4

Frequency of high risk disease allele = 0.05

Frequency of associated marker allele = 0.1

Linkage disequilibrium D-Prime = 0.8

Sample size: 500 cases, 500 controls

Type 1 error rate: 0.01

Calculate

Parker allele frequencies in cases and controls

NCP, Power

Exercise 2

For a discrete trait TDT study

- Assumptions same models as in Exercise 1
- Sample size: 500 parent-offspring trios
- Type 1 error rate: 0.01

Calculate:

- Ratio of transmission of marker alleles from heterozygous parents
- NCP, Power

Exercise 3:

Candidate gene TDT study of a threshold trait

- 200 affected offspring trios
- “Affection” = scoring > 2 SD above mean
- Candidate allele, frequency 0.05, assumed additive
- Type 1 error rate: 0.01
- Desired power: 0.8

What is the minimum detectable QTL variance?

Exercise 4:

An association study of a quantitative trait

- QTL additive variance 0.05, no dominance
- QTL allele frequency 0.1
- Marker allele frequency 0.2
- D-Prime 0.8
- Sib correlation: 0.4
- Type 1 error rate = 0.005
- Sample: 500 sib-pairs

Find NCP and power for between-sibship, within-sibship and overall association tests.

What is the impact of adding 100 sibships of size 3 on the NCP and power of the overall association test?

Exercise 5:

Using GPC for case-control design

- Disease prevalence: 0.02
- Assume multiplicative model
  - genotype risk ratio Aa = 2
  - genotype risk ratio AA = 4
- Frequency of high risk allele = 0.05
- Frequency of marker allele = 0.05, D-prime =1

Find the type 1 error rates that correspond to 80% power

- 500 cases, 500 controls
- 1000 cases, 1000 controls
- 2000 cases, 2000 controls

.

Exploring power of association using GPC

Linkage versus association

- difference in required sample sizes for specific QTL size

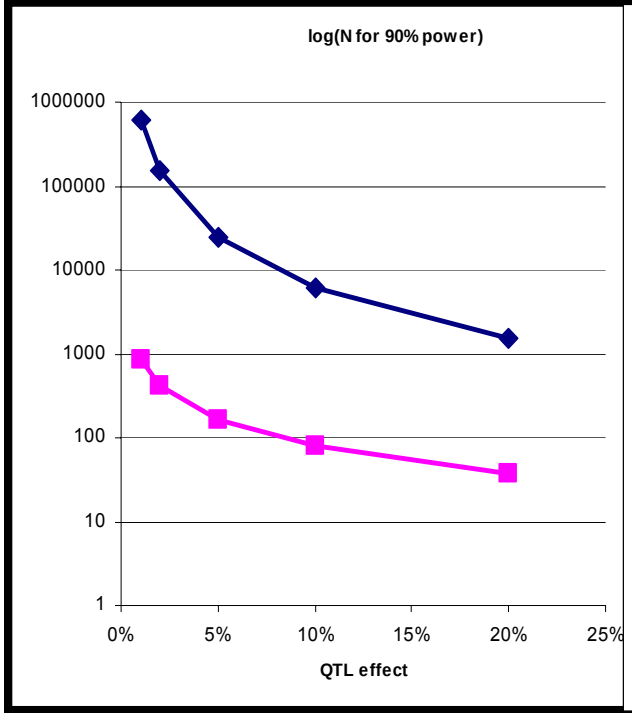
TDT versus case-control

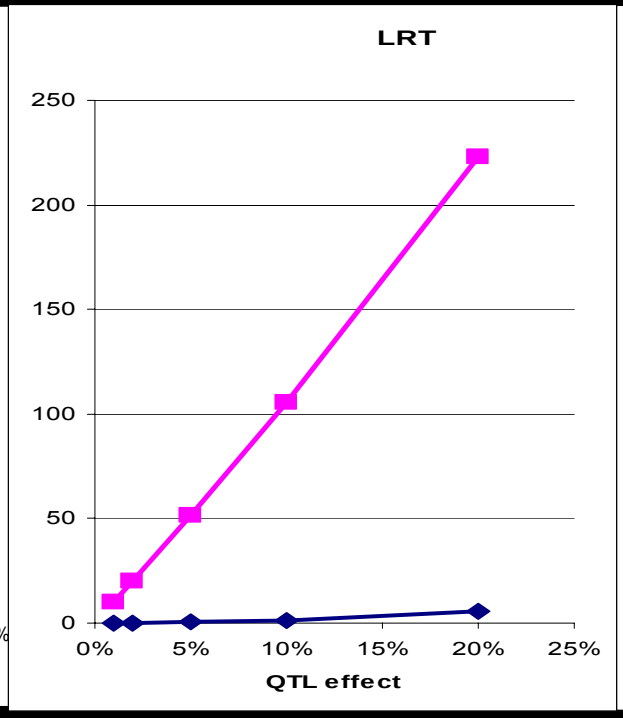
- difference in efficiency?

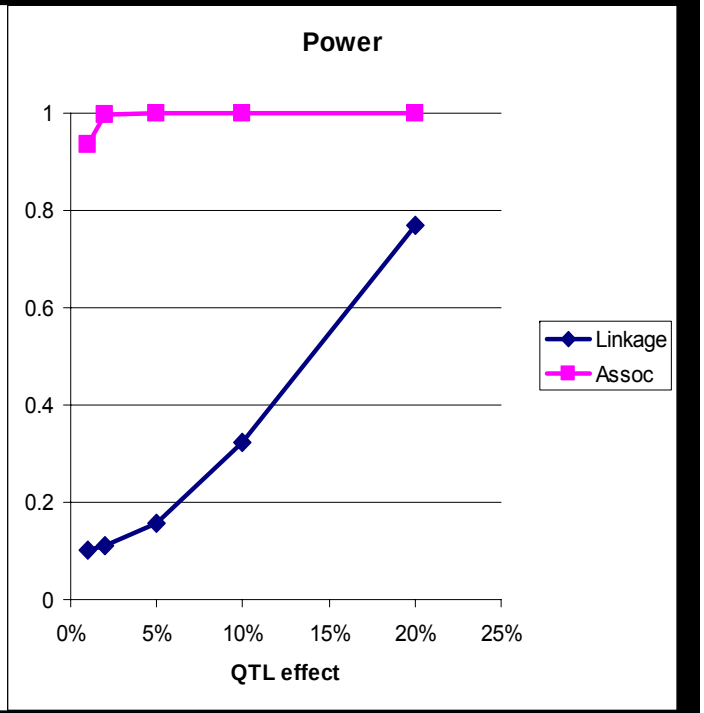
Quantitative versus binary traits

- loss of power from artificial dichotomisation?

Linkage versus association

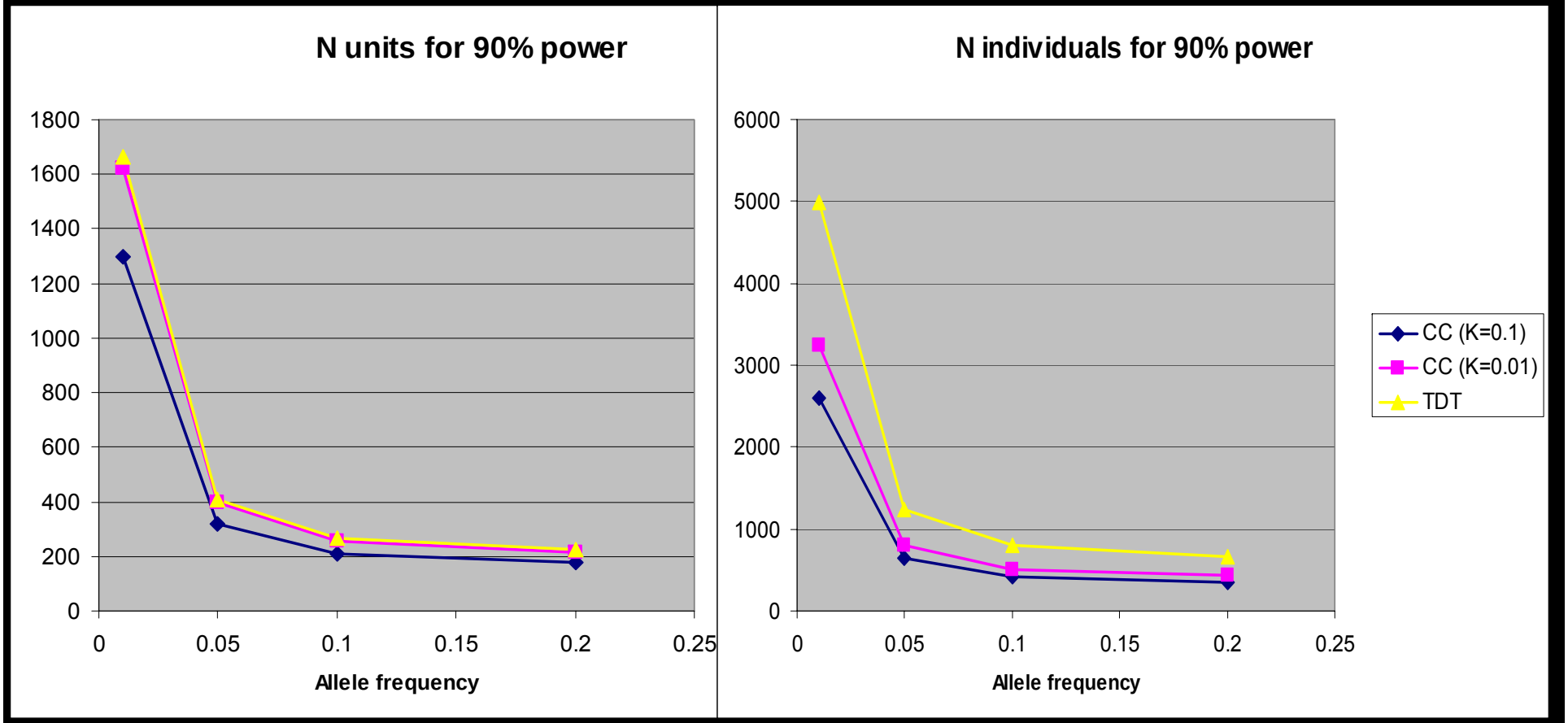






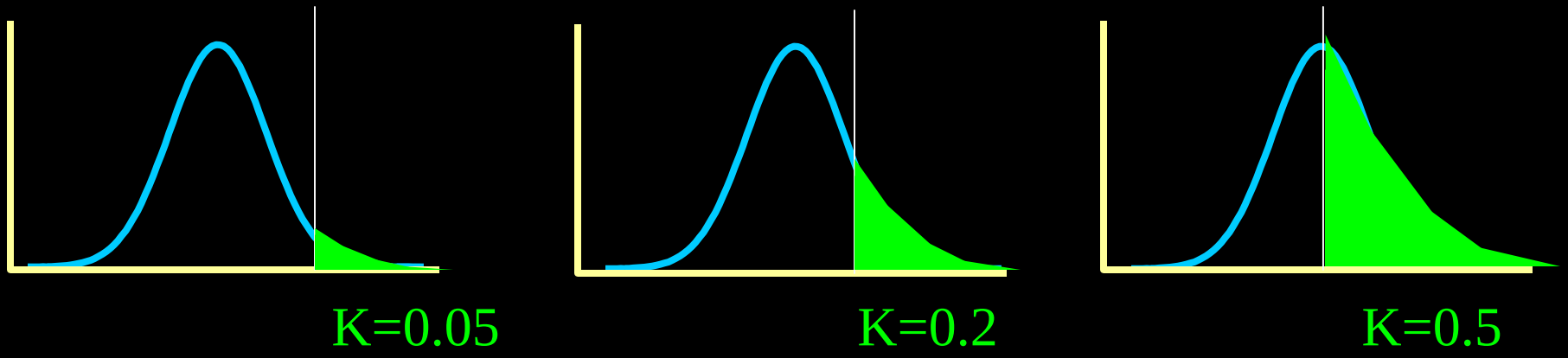
QTL linkage: 500 sib pairs, r=0.5
QTL association: 1000 individuals

Case-control versus TDT



p = 0.1; RAA = RAa = 2

Quantitative versus discrete



To investigate: use threshold-based association

Fixed QTL effect (additive, 5%, *p=0.5*) 500 individuals

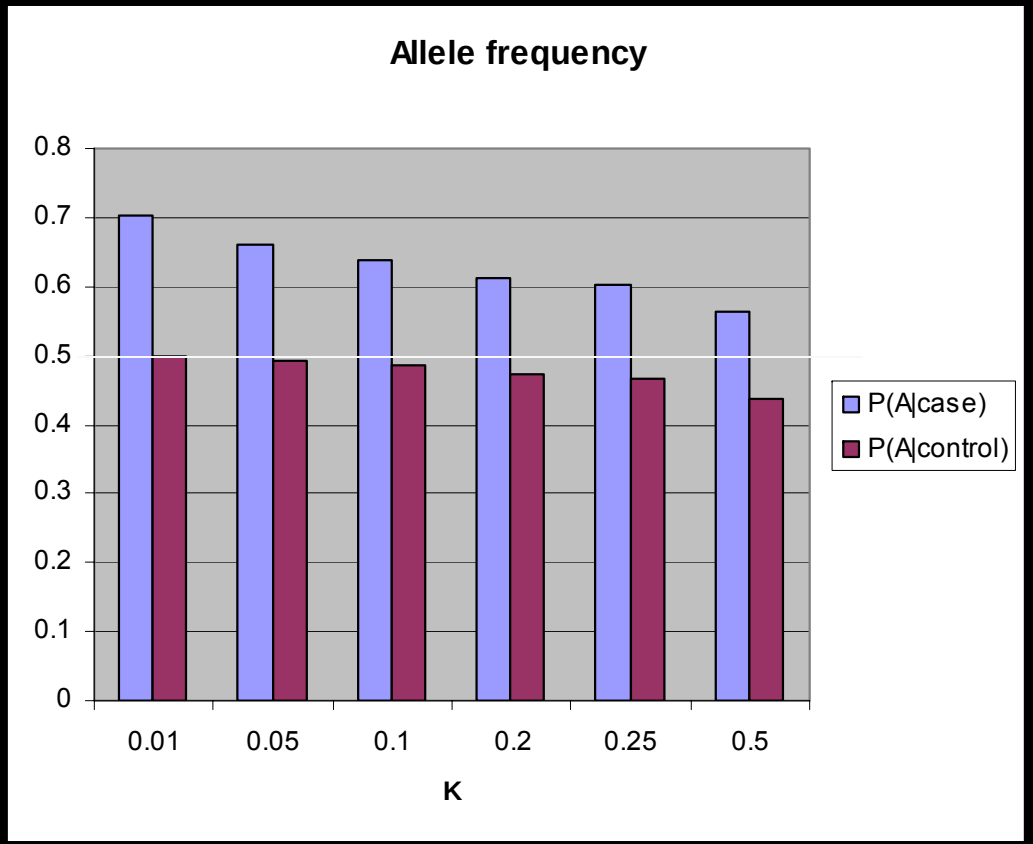
For prevalence *K*

Group 1 has N $500K$ and *T* $-6 \leq X \leq \Phi^{-1}(K)$

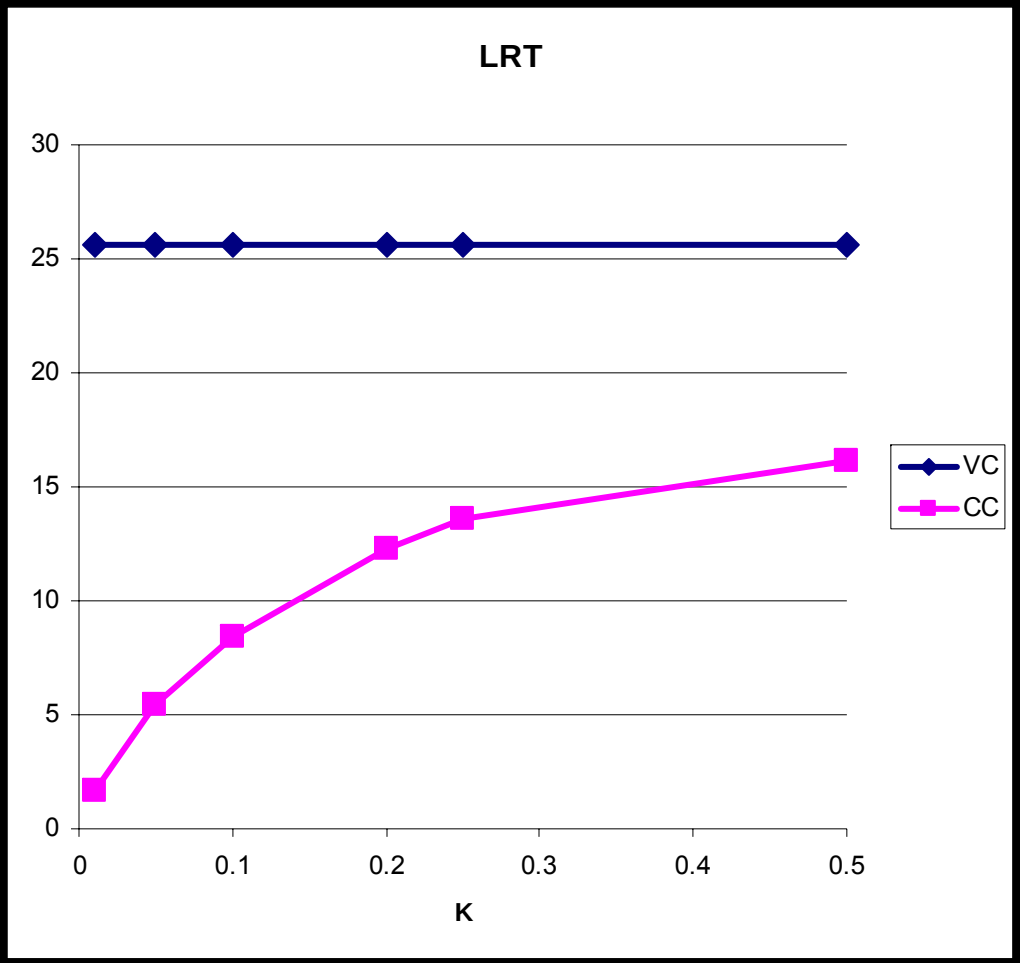
Group 2 has N $500(1-K)$ and *T* $\Phi^{-1}(K) \leq X \leq 6$

Quantitative versus discrete

K	T (SD)
0.01	2.326
0.05	1.645
0.10	1.282
0.20	0.842
0.25	0.674
0.50	0.000



Quantitative versus discrete



Incomplete LD

- what is the impact of D’ values less than 1?
- does allele frequency affect the power of the test?
  - (using discrete case-control calculator)

Family-based VC association: between and within tests

- what is the impact of sibship size? sibling correlation?
  - (using QTL VC association calculator)

Incomplete LD

Case-control for discrete traits

Disease K = 0.1

QTL $R_{AA} = R_{Aa} = 2$ p = 0.05

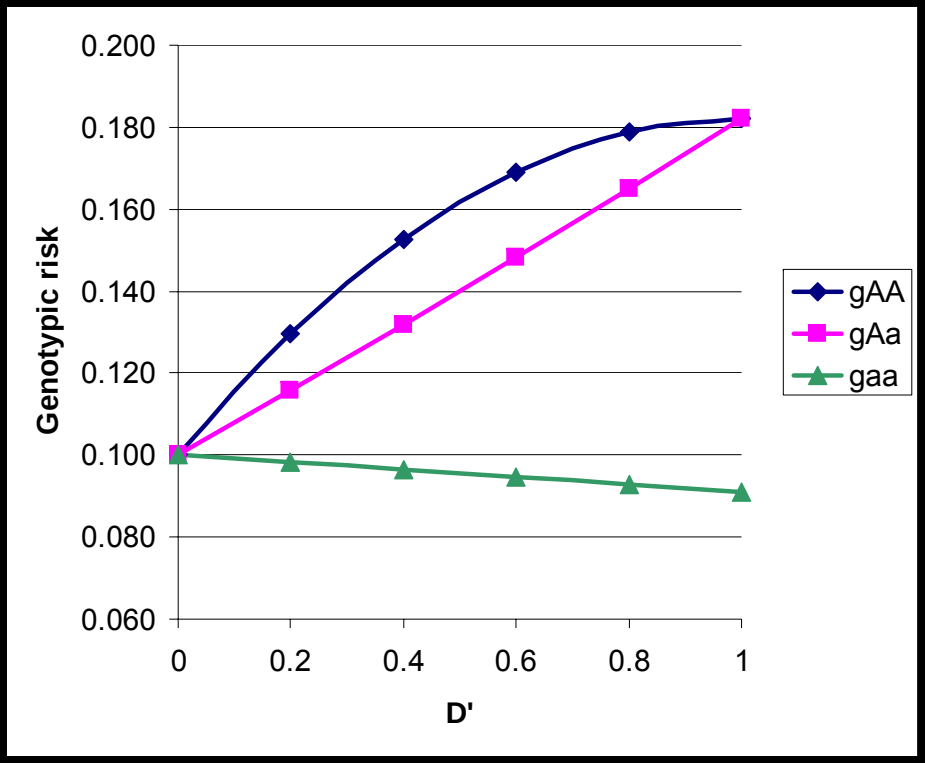
Marker1 m = 0.05 D' = { 1, 0.8, 0.6, 0.4, 0.2, 0}

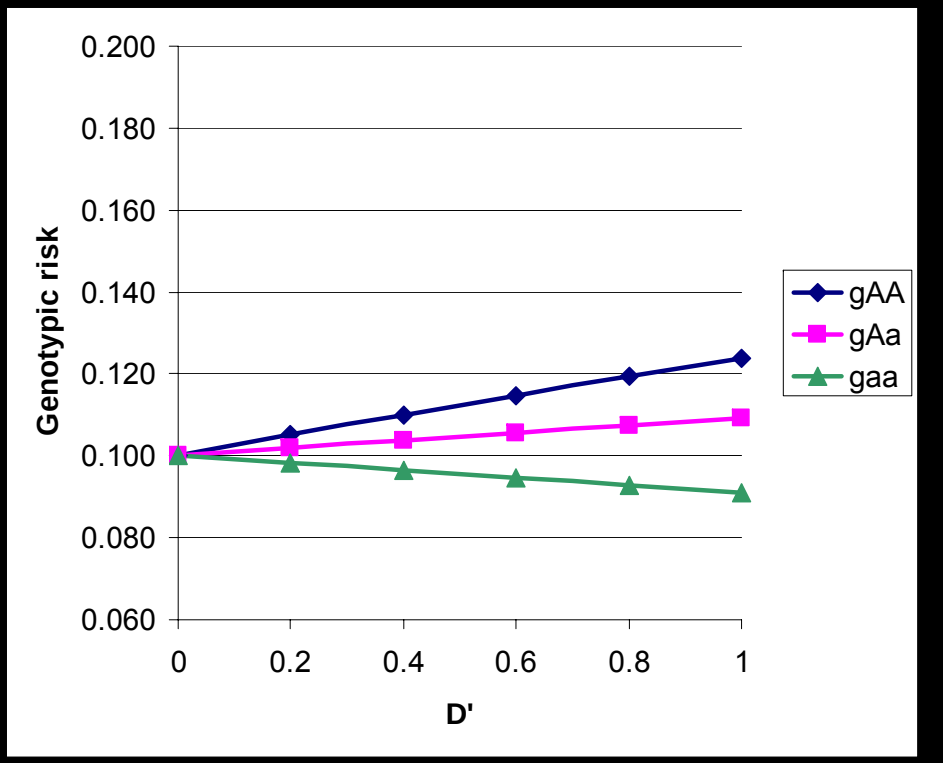
Marker2 m = 0.25 D' = { 1, 0.8, 0.6, 0.4, 0.2, 0}

Sample 250 cases, 250 controls

Incomplete LD

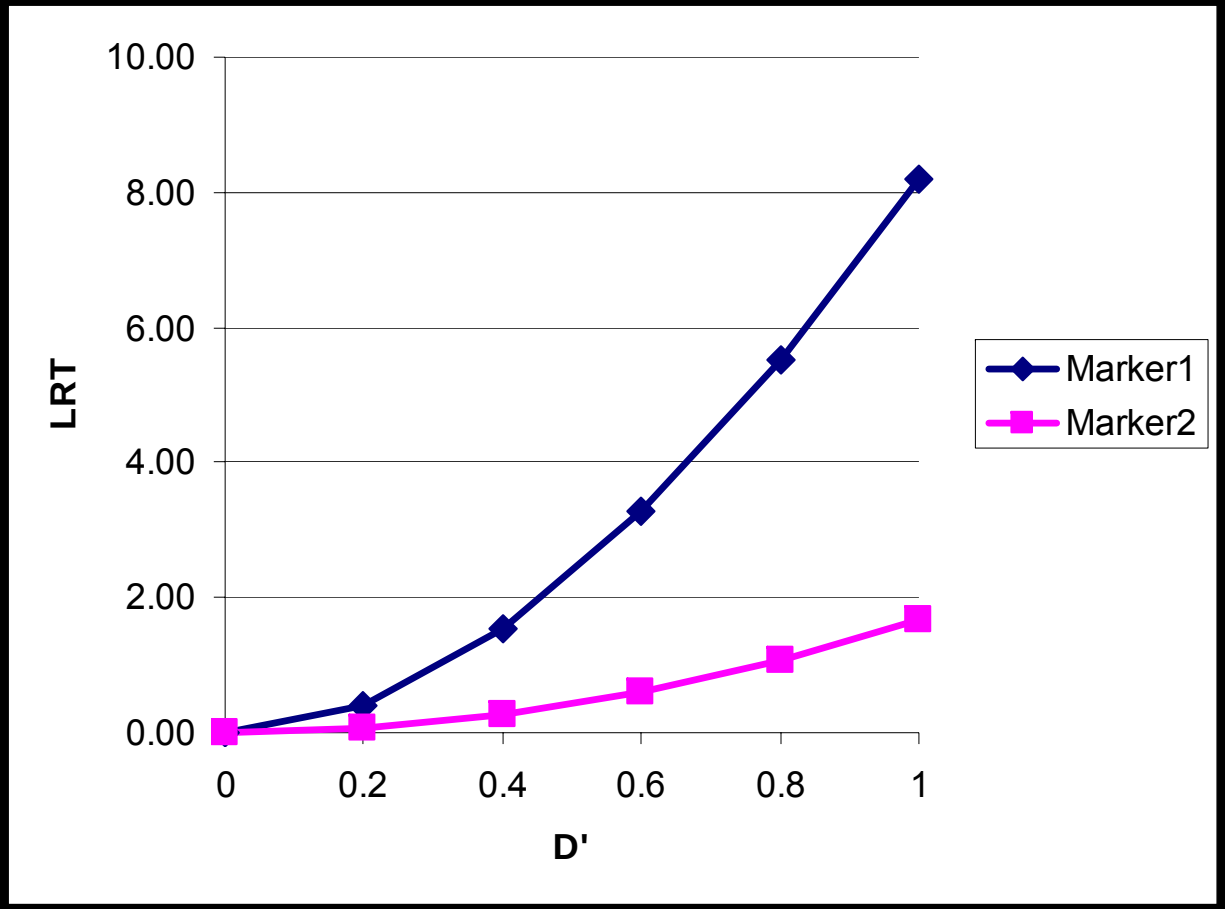
Genotypic risk at marker1 (left) and marker2 (right) as a function of D’





Incomplete LD

Expected likelihood ratio test as a function of D’



Family-based association

Sibship type

1200 individuals, 600 pairs, 400 trios, 300 quads

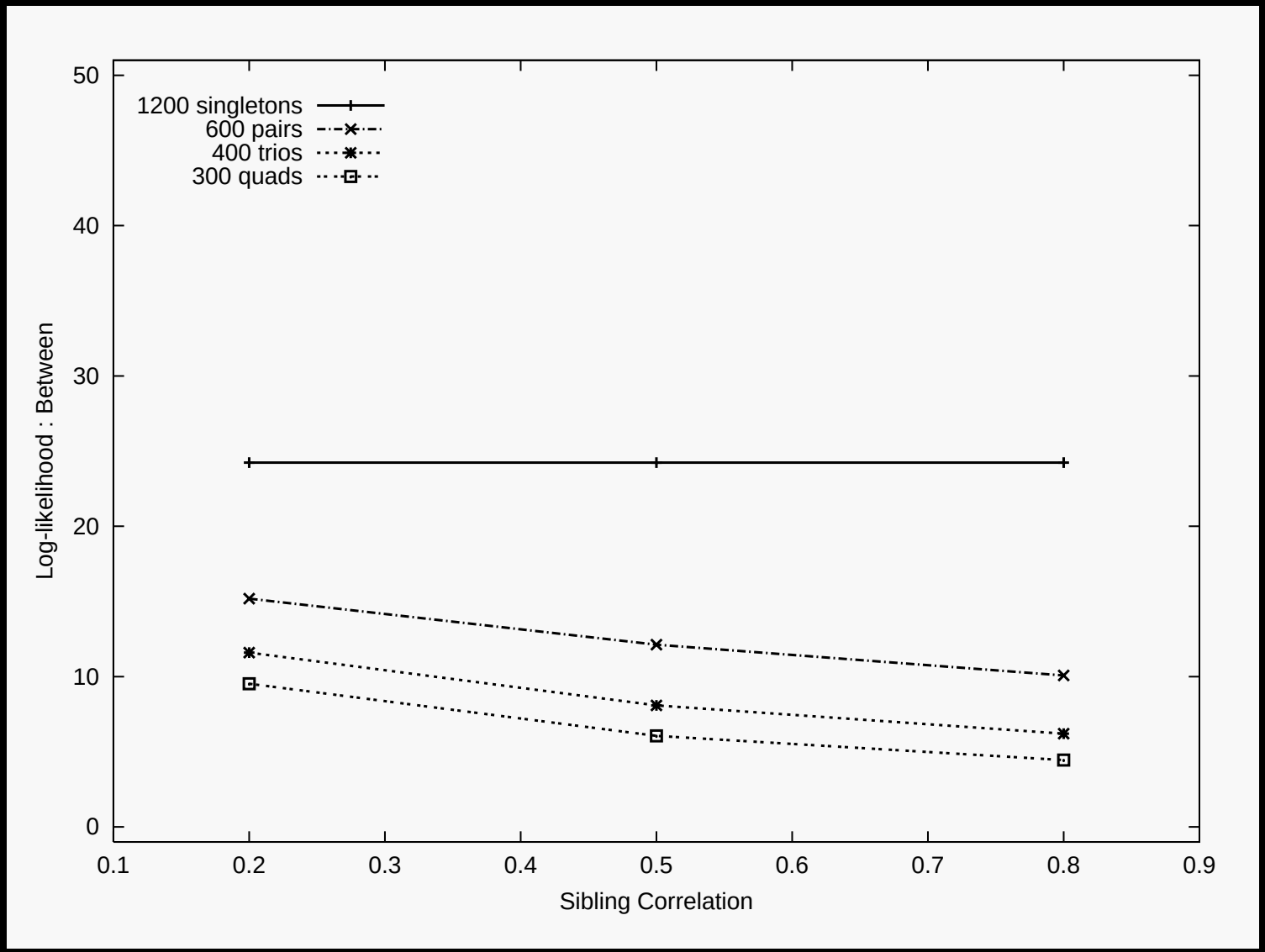
Sibling correlation

r = 0.2, 0.5, 0.8

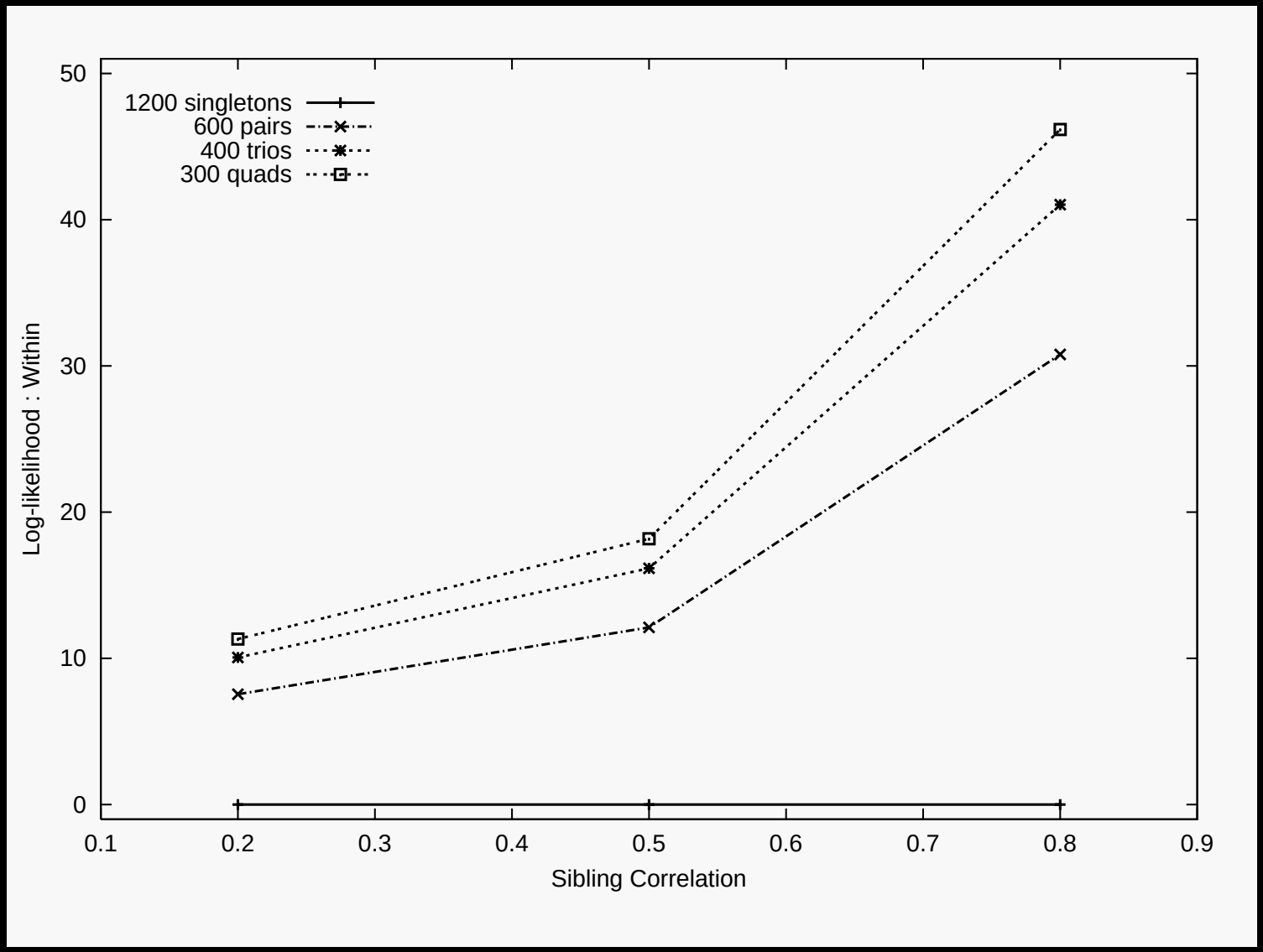
QTL (diallelic, equal allele frequency)

2%, 10% of trait variance

Between-sibship association



Within-sibship association



Total association