

Combined Association and
Linkage Analysis
for General Pedigrees
and Genetic Models

Ola Hössjer
Dept. of Mathematics
Stockholm University

Association

Data:	Phenotypes and marker data from several loci, for a number of families.
Principle:	Test association between phenotypes and marker genotypes at all marker loci
Case-control:	Devlin and Risch (1995).
Family-based with parental control, (FBATs):	Falk and Rubinstein (1987), Terwilliger and Ott (1992), Spielman et al. (1993). Laird and Rabinowitz (2000).
Modelling ancestral founder mutation, e.g. with likelihood:	Kaplan et al. (1995), Terwilliger (1995), Lazzeroni (1998), Service et al. (1999), McPeck and Strahs, (1999), Morris et al. (2000, 2002).
Likelihood-based score tests:	Schaid (1996), Clayton (1999), Tu et al. (2000) Whittemore and Tu (2000), Shih and Whittemore (2002).
Spatial/clustering:	Molitor et al. (2003a,b).

Linkage

- Data: Phenotypes and marker data from several loci, for a number of families.
- Principle: Test *coinheritance* of phenotypes and marker data at a number of loci (not necessarily marker loci).
- Parametric: Morton (1955), Kruglyak et al. (1996)
- Score tests: Whittemore (1996), McPeck (1999), Hössjer, (2003,2005).
- Nonparametric: Penrose (1935), Risch (1990), Whittemore and Halpern (1994), Kruglyak et al. (1996)
- Profiling: Risch (1984), Amos (1994), Almasy and Blangero (1998).
- Regression: Haseman and Elston (1972), Sham et al. (2002).
- Est. equations: Liang et al. (2001), Chen et al. (2002).

Combined Association and Linkage

Approximate joint
likelihood (pseudomarkers):

Terwilliger and Göring (2000),
Göring and Terwilliger (2000).

QTL variance components:

Fulker et al. (1999), Sham et al. (2000).
Abecasis et al. (2000).

Goals:

- Combined score test for association and linkage
- 'Biological' parameters (dis locus penetrances, ass between marker and dis gene, marker allele frequencies)
- Arbitrary (outbred) pedigrees
- Wide class of genetic models
- Multipoint approach
- Efficiency/power comparison between association, linkage and combined score tests.

Setup:

Let

$$\begin{aligned}K &= \text{nr. of markers} \\x_i &= \text{marker locus } i, i = 1, \dots, K, \\ \tau &= \text{disease locus}\end{aligned}$$

and, for one family, define

$$\begin{aligned}n &= \text{nr. of individuals of the family} \\Y &= (Y_1, \dots, Y_n), \text{ phenotypes of family members} \\G &= (G_1, \dots, G_n), \text{ disease genotypes of} \\ &\quad \text{family members} \\H_i &= (H_{i1}, \dots, H_{in}), \text{ marker genotypes at } x_i, \\ \mathcal{T} &\subset \{1, \dots, n\}, \text{ set of genotyped individuals} \\M_i &= \{H_{ik}, k \in \mathcal{T}\}, \text{ marker data at } x_i \\M &= (M_1, \dots, M_K), \text{ marker data at all loci.}\end{aligned}$$

Data: (Y, M) for a number (N) of outbred families of arbitrary (and possibly different) form

Hypothesis Test

Consider a fixed marker locus x_i . Let

$$\begin{aligned}\Delta &= \text{association between } H_i \text{ and } G, \\ \varepsilon &= \text{penetrance between } Y \text{ and } G,\end{aligned}$$

Hypothesis test at x_i :

$$\begin{aligned}\text{Null}(x_i): & \quad x_i = \tau, \varepsilon = \Delta = 0, \\ \text{Alternative}(x_i): & \quad x_i = \tau, \varepsilon \neq 0.\end{aligned}$$

where

$$\begin{aligned}\Delta = 0 &: \text{No association between } H_i \text{ and } G. \\ \varepsilon = 0 &: \text{No penetrance between } Y \text{ and } G.\end{aligned}$$

Ex: Binary Phenotypes

Individual phenotype:

$$Y_k = \begin{cases} 1, & k \text{ is affected,} \\ 0, & k \text{ is unaffected.} \end{cases}$$

Penetrance:

$$\begin{aligned} \psi(j) &= P(Y_k = 1 \mid |G_k| = j), \\ &= K_p + \varepsilon u(j) \end{aligned}$$

for $j = 0, 1, 2, \dots$, where

$$\begin{aligned} |G_k| &= \text{number of disease alleles of } G_k, \\ K_p &= P(Y_k = 1 \mid \varepsilon = 0) \\ &= \text{prevalence under null model.} \end{aligned}$$

For entire family,

$$P(Y|G) = \prod_{k=1}^n P(Y_k|G_k)$$

if no polygenic or shared environmental effects.

Ex: Gaussian Phenotypes

For each individual, assume

$$Y_k | |G_k| = j \in N(\psi(j), \sigma^2),$$

where

$$\psi(j) = m + \varepsilon u(j), \quad j = 0, 1, 2,$$

and

$$\begin{aligned} m &= E(Y_k) = \text{mean under null model.} \\ \sigma^2 &= \text{environmental variance.} \end{aligned}$$

For entire family,

$$Y|G \in N(\mu, \sigma^2 \Sigma),$$

where

$$\begin{aligned} \mu &= (\psi(|G_1|), \dots, \psi(|G_n|)), \\ \Sigma &= n \times n \text{ correlation matrix, possibly including} \\ &\text{shared environmental and polygenic effects.} \end{aligned}$$

Gaussian Liability

Penetrance

$$P(Y|G) = \int P(Y|X)P(X|G)dX,$$

for entire family, where

$X = (X_1, \dots, X_n)$ = vector of individual liabilities

with distribution

$$X|G \in N(\mu, \sigma^2 \Sigma)$$

and μ , σ^2 and Σ are as before.

Examples:

$$\begin{aligned} Y_k &= 1_{\{X_k \geq T\}}, \text{ (liab. threshold):} \\ P(Y_k \geq T|X_k) &= \exp\left(-\int_0^T \lambda(t)\right), \text{ where} \\ &\lambda(t) = \lambda_0(t) \exp(\beta X_k) \text{ (frailty)} \\ h(E(Y_k|X_k)) &= m + \beta X_k, \text{ where } h \text{ is link} \\ &\text{function (GLM)} \end{aligned}$$

Likelihood

Retrospective likelihood (Prentice and Pyke, 1979)

$$L(x, \theta) = \prod_{N\text{families}} P_{x,\theta}(M|Y),$$

where subscript x means ' $x = \tau$ ' and

$$\begin{aligned} x &= x_i \text{ for some } i = 1, \dots, K, \\ \theta &= (q, \Delta, \varepsilon), \\ q &= (q_1, \dots, q_d) = \text{vector of marker allele} \\ &\quad \text{frequencies at } x_i \text{ (nuisance parameters)}. \end{aligned}$$

One can show

$$\partial \log L / \partial \Delta |_{\theta=\theta_0} = \partial \log L / \partial \varepsilon |_{\theta=\theta_0} = 0,$$

for outbred pedigrees at null model

$$\theta_0 = (q, 0, 0).$$

Hence reparametrize

$$\epsilon = (\epsilon_0, \epsilon_1, \epsilon_2) = (q, \Delta\varepsilon, \varepsilon^2).$$

Expanding Family Likelihood

Retrospective likelihood is

$$L(x, \epsilon) = P_{x, \epsilon}(M|Y),$$

for one family. Let

$$\begin{aligned} n &= \text{nr. of individuals} \\ f &= \text{nr. of founders} \\ m &= 2(n - f) = \text{nr. of meioses} \\ v &= (v_1, \dots, v_m) = \text{inheritance vector at } x \\ b &= (b_1, \dots, b_{2f}) = \text{founder marker alleles at } x \\ a &= (a_1, \dots, a_{2f}) = \text{founder disease alleles at } x. \end{aligned}$$

Assume x captures all association with τ :

$$L(x, \epsilon) = \sum_{b, v} P(M|b, v) P_{x, \epsilon}(b, v|Y),$$

where 2nd term is complete marker data likelihood, which is split into association and penetrance terms by summing over a :

$$\begin{aligned} P_{x, \epsilon}(b, v|Y) &\propto \sum_a P_{q, \Delta}(b|a) P_{\epsilon}(a, v|Y) \\ &\propto \sum_a P_{q, \Delta}(a, b) P_{\epsilon}(Y|a, v) \\ &= \sum_a P_{q, \Delta}(a, b) P_{\epsilon}(Y|G). \end{aligned}$$

Score Vector

Let

$$\begin{aligned} S(x) &= \partial \log L(x, \epsilon) / \partial \epsilon |_{\epsilon=(q,0,0)} \\ &= (S_0(x), S_1(x), S_2(x)), \end{aligned}$$

where

$$\begin{aligned} S_0(x) &= \partial \log L(x, \epsilon) / \partial \epsilon_0 |_{\epsilon=(q,0,0)} \\ &= \text{marker allele freq. score} \\ S_1(x) &= \partial \log L(x, \epsilon) / \partial \epsilon_1 |_{\epsilon=(q,0,0)} \\ &= \text{association score} \\ S_2(x) &= \partial \log L(x, \epsilon) / \partial \epsilon_2 |_{\epsilon=(q,0,0)} \cdot \\ &= \text{linkage score} \end{aligned}$$

One shows

$$S_i(x) = \sum_{b,v} P_q(b, v | M) S_i(b, v), \quad i = 0, 1, 2,$$

where

$$\begin{aligned} P_q(b, v | M) &= \text{multipoint probability} \\ S_i(b, v) &= \text{score component } i \text{ for complete} \\ &\quad \text{marker data.} \end{aligned}$$

Score Vector, Biallelic Marker ($d = 2$)

Assume

$$q = (q_0, q_1) \Rightarrow a_j \text{ and } b_j \text{ binary,}$$

and

$$\begin{aligned} P_{q,\Delta}(a, b) &\stackrel{\text{HW-eq.}}{=} \prod_{j=1}^{2f} P(a_j, b_j), \\ \Delta &= \text{correlation coefficient between } a_j \text{ and } b_j. \end{aligned}$$

Then

$$\begin{aligned} S_0(b, v) &= n_1/q_1 - \text{constant} \\ S_1(b, v) &= \sum_{k=1}^n \omega_k (b_{2k-1} + b_{2k}) - \text{constant} \\ S_2(b, v) &= \sum_{k < l} \omega_{kl} \text{IBD}_{kl} - \text{constant} \end{aligned}$$

where

$$\begin{aligned} n_1 &= \text{nr. of founder marker alleles } b_j = 1 \\ \omega_k &= \text{weight assigned to individual } k \\ \omega_{kl} &= \text{weight assigned to pair } k, l \\ \text{IBD}_{kl} &= \text{nr. of alleles shared IBD by } k \text{ and } l. \end{aligned}$$

Example of Weights

1) Binary phenotypes:

$$\omega_k = Y_k - E(Y_k|\varepsilon = 0) = \begin{cases} -K_p, & Y_k = 0, \\ 1 - K_p, & Y_k = 1. \end{cases}$$

2) Gaussian (quantitative) phenotypes

$$\omega_k = Y_k - E(Y_k|\varepsilon = 0) = Y_k - m,$$

if no polygenic or shared environmental effects.

In both cases¹

$$\omega_{kl} = \omega_k \omega_l$$

¹This is *not* true in presence of polygenic and sh. env. effects.

Previous Work Association

$S_1(b, v)$:

Binary: Schaid (1996), Clayton (1999), Whittemore and Tu (2000), Tu et al. (2000).

General: Shih and Whittemore (2002).

with penetrance modelled directly as $P(Y|H_i)$ at $x = x_i$.

Clayton (1999) and Whittemore and Tu (2000) defined

$$S_1(b, v) = S_1^F(b) + S_1^{NF}(b, v)$$

where

$$\begin{aligned} S_1^F(b) &= E(S_1(b, v)|b) \\ &= \text{founder statistic} \\ S_1^{NF}(b, v) &= S_1(b, v) - E(S_1(b, v)|b) \\ &= \text{nonfounder statistic} \end{aligned}$$

and the latter includes TDT test as special case.

Previous Work Linkage

$$S_2(b, v) = S_2(v):$$

- Binary: Whittemore and Halpern (1994), McPeck (1999).
Quantitative: Commenges (1994), Tang and Siegmund (2001),
 Putter et al. (2002), Wang and Huang (2002).
General: Hössjer (2003, 2005).

Test Statistics:

Fisher information entries

$$I_{ij} = E(S_i S_j), \quad i, j = 0, 1, 2.$$

and matrix

$$J = \begin{cases} I = (I_{ij})_{i,j=1}^2, & q \text{ known} \\ I - (I_{01}, I_{02}) I_{00}^{-1} (I_{01}, I_{02})^T, & q \text{ estimated.} \end{cases}$$

Two-sided association test:

$$T_1(x) = S_1^2(x) / J_{11}.$$

One-sided linkage test:

$$T_2(x) = S_2(x) / \sqrt{J_{22}}.$$

Combined test:

$$T_{\text{combined}}(x) = \begin{cases} (S_1, S_2) J^{-1} (S_1, S_2)^T, & \text{if } X \geq 0, \\ S_1^2 / J_{11}, & \text{if } X < 0, \end{cases}$$

where $X = (S_1, S_2) J^{-1/2} c^T$ and $c = ((0, 1) J^{1/2})^\perp$.

Remark: If J is diagonal, as for complete marker data, then $X = S_2$.

Asymptotics

Assume N pedigrees of identical form, with same Y . If N large and η_1 and η_2 small, approximate

$$\begin{aligned} T_1 &\in \chi^2(1, N\eta_1^2) \\ T_2 &\in N(\sqrt{N}\eta_2, 1) \\ T_{\text{combined}} &= (X_1 + \sqrt{N}\eta_1)^2 \\ &\quad + (X_2 + \sqrt{N}\eta_2)^2 \mathbf{1}_{\{X_2 + \sqrt{N}\eta_2 \geq 0\}}, \end{aligned}$$

where X_1 and X_2 are independent $N(0, 1)^2$.

Noncentrality parameters, complete marker data:

$$\begin{aligned} \eta_1^2 &= J_{11} \cdot (\Delta\varepsilon)^2, \\ \eta_2^2 &= J_{22} \cdot \varepsilon^4 \end{aligned}$$

²This is true for complete marker data.

Power

Let

$N_i(\alpha, \beta)$ = sample size required for level α
test T_i over region(s) Ω
to achieve power β .

for $i = 1, 2$. Then

$$N_i(\alpha, \beta) = C_i / \eta_i^2,$$

where $C_i = C_i(\alpha, \beta, \Omega)$ is constant correcting for multiple testing over region Ω .

Required sample size for T_{combined} more complicated, but

$$N_{\text{combined}}(\alpha, \beta) \approx C_{\text{combined}} / (\eta_1^2 + \eta_2^2),$$

where $C_{\text{combined}} = C_{\text{combined}}(\alpha, \beta, \Omega)$ is constant correcting for multiple testing over region Ω .

Ex: Nuclear Families

Parents : $k = 1, 2$
Children : $k = 3, \dots, n$

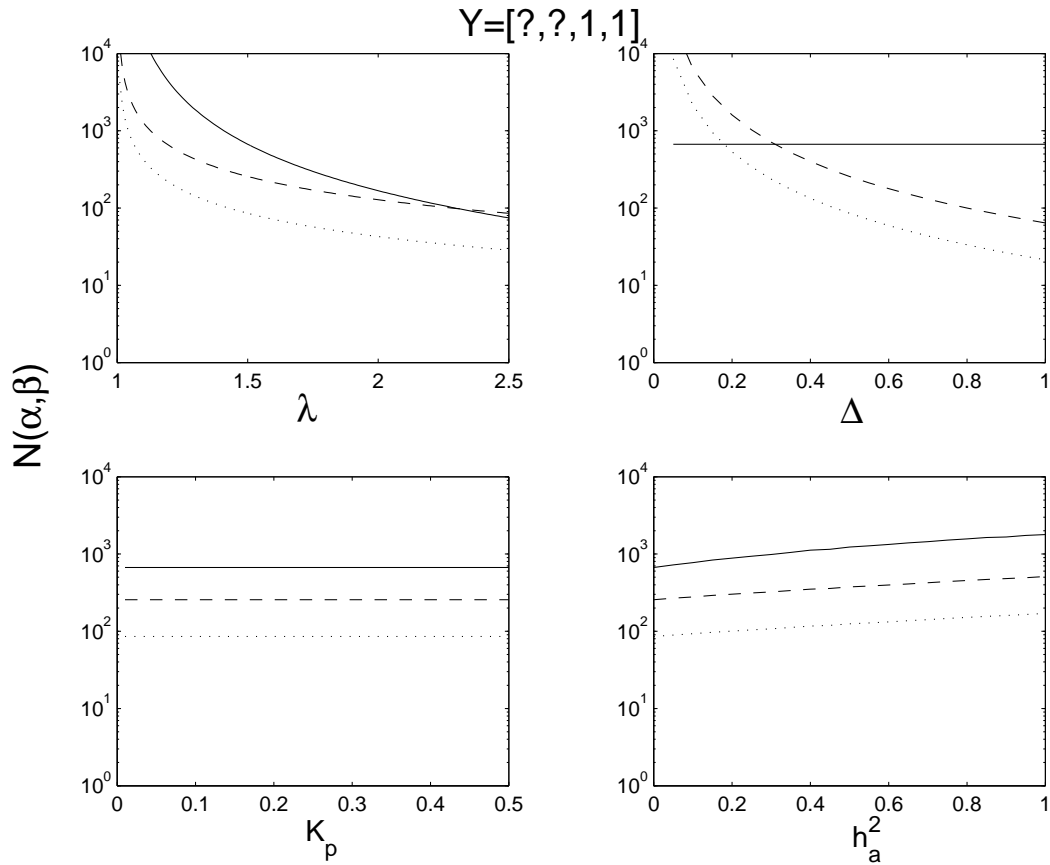
For additive model and complete marker data:

$$J_{11} = 0.5 \cdot \begin{cases} (\sum_{k=1}^n \omega_k)^2 + (\omega_1 - \omega_2)^2 + \sum_{k=3}^n \omega_k^2, & q \text{ known,} \\ (\omega_1 - \omega_2)^2 + \sum_{k=3}^n \omega_k^2, & q \text{ estimated,} \\ \sum_{k=3}^n \omega_k^2, & \text{nonfounder stat.} \end{cases}$$

and

$$J_{22} = 0.125 \sum_{3 \leq k < l \leq n} \omega_{kl}^2.$$

Affected Sib Pair



T_1 : known q (dotted), estimated q and nonfounder statistic (dashed)

T_2 : solid

Genomewide scan (22 chromosomes, length 3575 cM), $\alpha = 0.05$, $\beta = 0.8$

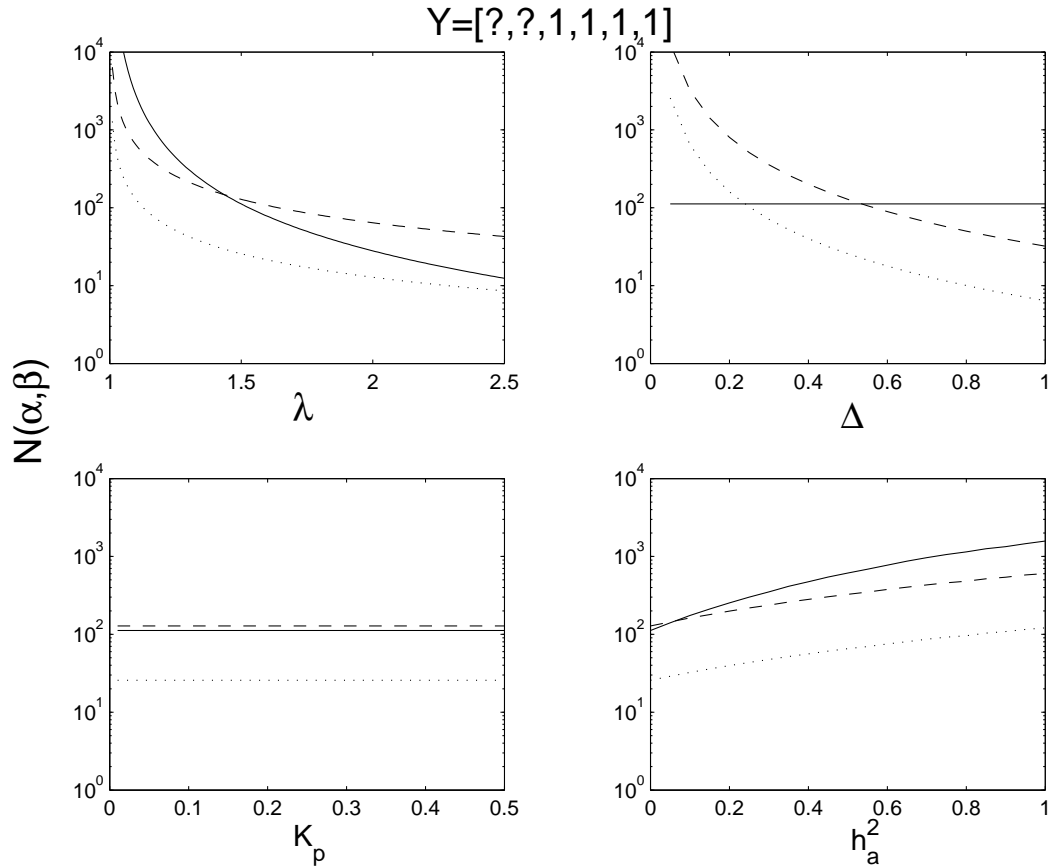
$K_p = 0.1$

$\Delta = 0.5$

Relative risk for MZ pair: $\lambda = 1.5 = 1 + (1 - K_p)^2 \varepsilon^2$.

Dense marker map for linkage. 0.1 cM between independent association tests.

Affected Sib Quartet



T_1 : known q (dotted), estimated q and nonfounder statistic (dashed)

T_2 : solid

Genomewide scan (22 chromosomes, length 3575 cM), $\alpha = 0.05$, $\beta = 0.8$

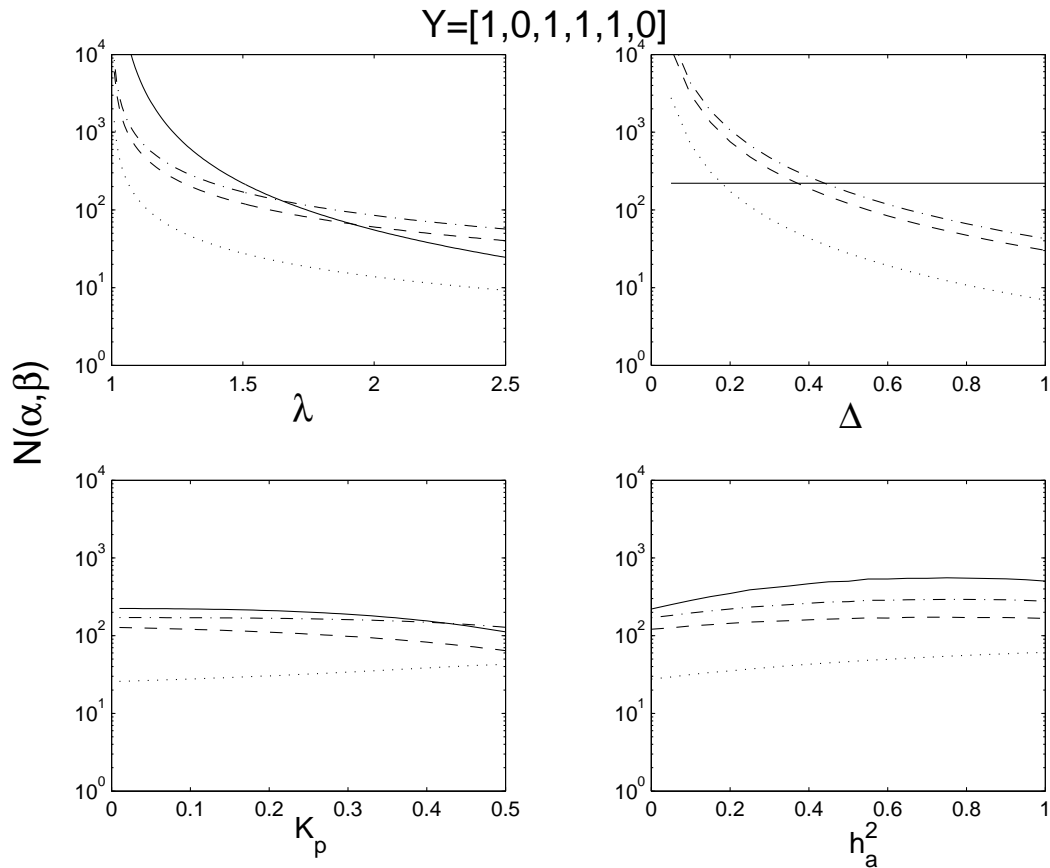
$K_p = 0.1$

$\Delta = 0.5$

Relative risk for MZ pair: $\lambda = 1.5 = 1 + (1 - K_p)^2 \varepsilon^2$

Dense marker map for linkage. 0.1 cM between independent association tests.

Aff/Unaff Family, 4 Children



T_1 : known q (dotted), estimated q (dashed) and nonfounder stat. (dash-dotted)

T_2 : solid

Genomewide scan (22 chromosomes, length 3575 cM), $\alpha = 0.05$, $\beta = 0.8$

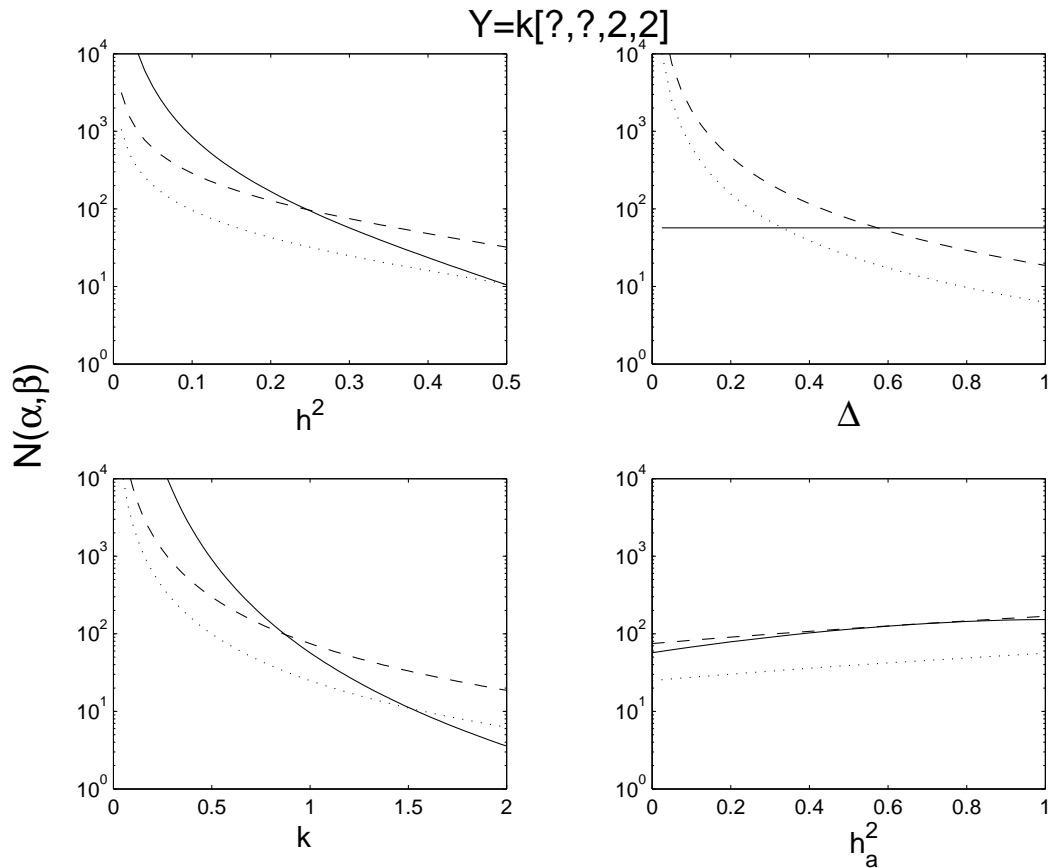
$K_p = 0.1$

$\Delta = 0.5$

Relative risk for MZ pair: $\lambda = 1.5 = 1 + (1 - K_p)^2 \varepsilon^2$

Dense marker map for linkage. 0.1 cM between independent association tests.

Quantitative Concordant Sib Pair



T_1 : known q (dotted), estimated q and nonfounder statistic (dashed)

T_2 : solid

Genomewide scan (22 chromosomes, length 3575 cM), $\alpha = 0.05$, $\beta = 0.8$

$m = 0$, $\sigma = 1$, $k = 1$

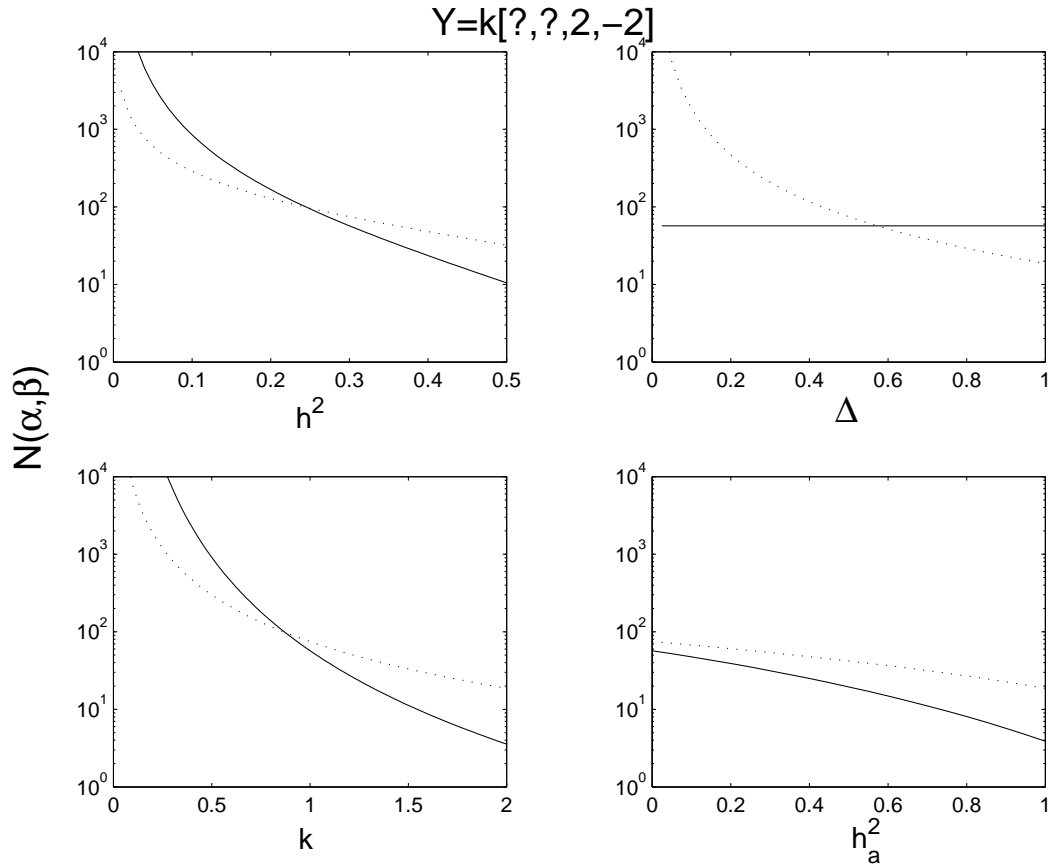
$\Delta = 0.5$

Heritability at main locus $h^2 = 0.3 = \text{Var}(\psi_{|G_k|}) / \text{Var}(Y_k)$

$h_a^2 = 0$

Dense marker map for linkage. 0.1 cM between independent association tests.

Quantitative Discordant Sib Pair



T_1 , known q , estimated q and nonfounder statistic: dotted

T_2 : solid

Genomewide scan (22 chromosomes, length 3575 cM), $\alpha = 0.05$, $\beta = 0.8$

$m = 0$, $\sigma = 1$, $k = 1$

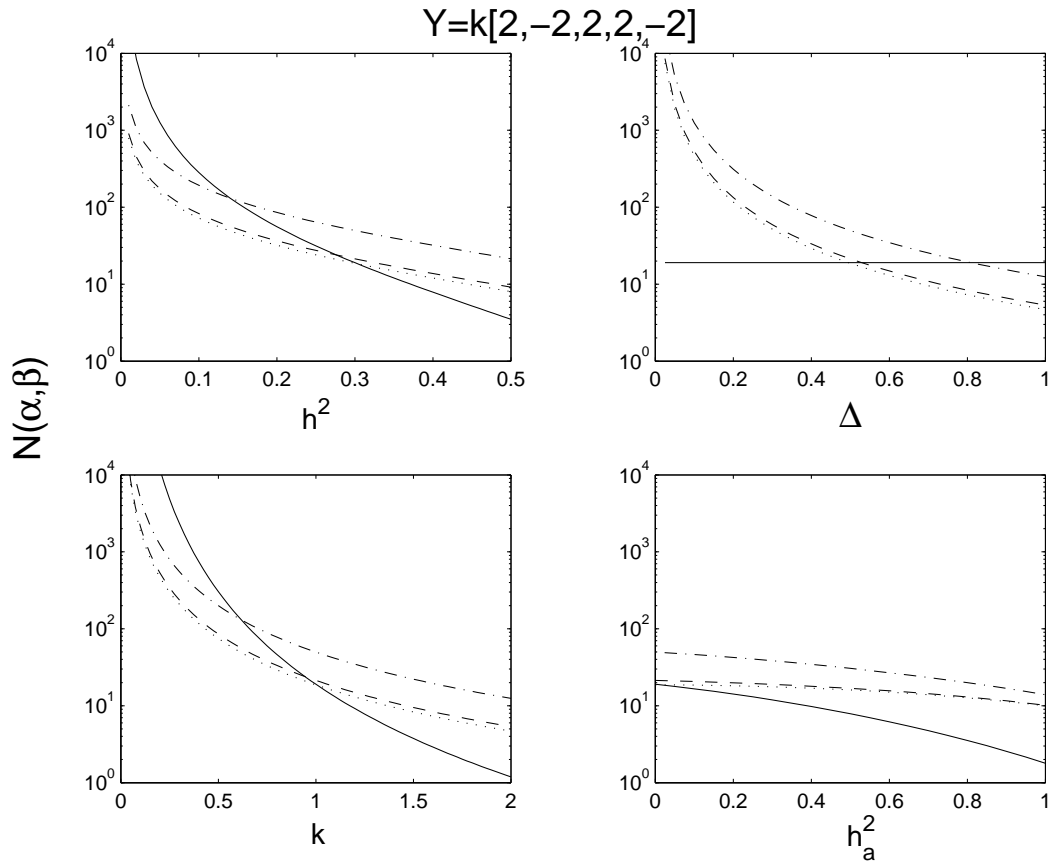
$\Delta = 0.5$

Heritability at main locus $h^2 = 0.3 = \text{Var}(\psi_{|G_k|}) / \text{Var}(Y_k)$

$h_a^2 = 0$

Dense marker map for linkage. 0.1 cM between independent association tests.

Quantitative 4 Children Family



T_1 : known q (dotted), estimated q (dashed) and nonfounder stat. (dash-dotted)

T_2 : solid

Genomewide scan (22 chromosomes, length 3575 cM), $\alpha = 0.05$, $\beta = 0.8$

$m = 0$, $\sigma = 1$, $k = 1$

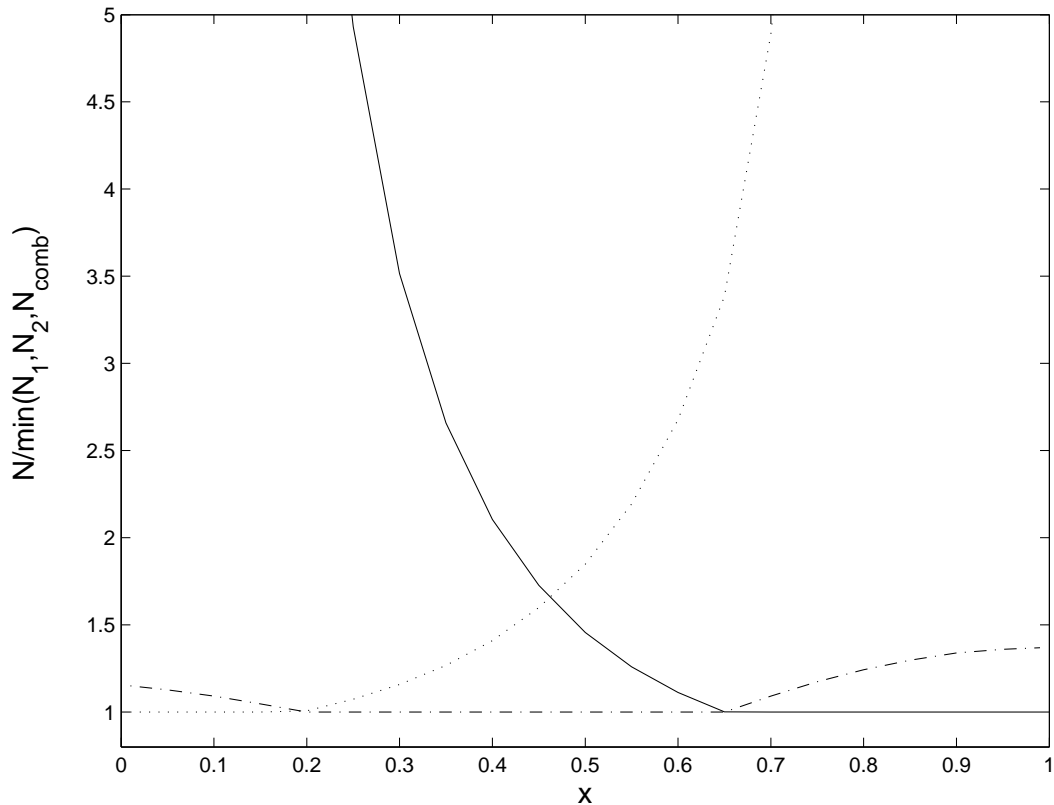
$\Delta = 0.5$

Heritability at main locus $h^2 = 0.3 = \text{Var}(\psi_{|G_k|}) / \text{Var}(Y_k)$

$h_a^2 = 0$

Dense marker map for linkage. 0.1 cM between independent association tests.

$N_1(\alpha, \beta)$, $N_2(\alpha, \beta)$ and $N_{\text{combined}}(\alpha, \beta)$



Pointwise test, $\Omega = \{x\}$, $\alpha = 0.05$, $\beta = 0.8$

N_1 : dotted

N_2 : solid

N_{combined} : dash-dotted

All three curves standardized by $\min(N_1, N_2, N_{\text{combined}})$

$\eta_2/\eta_1 = \tan(\pi x/2)$

100 000 Monte Carlo iterates when computing N_{combined} for each x .

Conclusions

- T_{combined} very robust.
- Never performs much worse than best of T_1 and T_2 .
- Often better than best of T_1 and T_2 .
- With multiple testing ($\Omega \neq \{x\}$), T_2 is favoured slightly over T_1 and T_{combined} , because dependence extends over longer distances for linkage. (Simulation or new analytical formulae required to make this precise.)
- Still, T_{combined} is likely to be the most robust of the three tests for chromosomewide or genomewide scans.

Hence, choose combined association and linkage test, e.g. T_{combined} , when little is known about the genetic model!

References

- Abecasis, G.R. et al. (2000). A general test for association for quantitative traits in nuclear families. *Am. J. Hum. Genet.*, **66**, 279-292.
- Almasy, L. and Blangero, J. (1998). Multipoint quantitative trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.*, **62**, 1198-1211.
- Amos, C.I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.*, **54**, 535-543.
- Chen, W-M. et al. (2004). Quantitative trait linkage analysis by generalized estimating equations: Unification of variance components and Haseman-Elston regression. *Genet. Epid.* **26**, xx-xx.
- Clayton, D. (1999). A generalization of the transmission/disequilibrium test for uncertain haplotype transmission. *Am. J. Hum. Genet.* **65**, 1170-1177.
- Commenges, D. (1994). Robust genetic linkage analysis based on a score test of homogeneity: The weighted pairwise correlation statistic. *Genetic Epidemiol.* **11**, 189-200.
- Devlin, B. and Risch N. (1995). A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**(2), 311-22.
- Falk, C. and Rubinstein, P. (1987). Haplotype relative risks: an easy and reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* **51**, 227-233.
- Fulker, D.W. et al. (1999). Combined linkage and association sib-pair analysis for quantitative traits. *Am. J. Hum. Genet.* **64**, 259-267.

Göring, H.H.H. and Terwilliger, D. (2000). Linkage analysis in the presence of errors IV: Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am. J. Hum. Genet.* **66**, 1310-1327.

Haseman, J.K. and Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* **2**, 3-19.

Hössjer, O. (2003). Determining Inheritance Distributions via Stochastic Penetrances. *J. Amer. Statist. Assoc.*, **98**, 1035-1051.

Hössjer, O. (2005). Conditional likelihood score functions in linkage analysis. To appear in *Biostatistics*.

Penrose, L.S. (1935). The detection of autosomal linkage in data which consists of brothers and sisters of unspecified parentage. *Annals of Eugenics* **6**, 133-138.

McPeck, S. (1999). Optimal allele-sharing statistics for genetic mapping using affected relatives. *Genetic Epidemiol.* **16**, 225-249.

Morton, N.E. (1955). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7**, 277-318.

Kaplan, N.L. et al. (1995). Likelihood methods for locating disease genes in nonequilibrium populations. *Am. J. Hum. Genet.* **56**, 18-32.

Kruglyak, L. et al. (1996). Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am. J. Hum. Genet.*, **58**, 1347-1363.

Molitor, J. et al. (2003a). Application of Bayesian spatial statistical methods to analysis of haplotype effects and gene mapping. *Genet. Epid.* **25**, 95-105.

Molitor, J. et al. (2003b). Fine-scale mapping of disease genes with multiple mutations via spatial clustering techniques. *Am. J. Hum. Genet.*, **73**, 1368-1384.

Rabinowitz, D. and Laird, N. (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered.* **50**(4), 211-23.

Lazzeroni, L.C. (1998). Linkage disequilibrium and gene mapping: an empirical least-squares approach. *Am. J. Hum. Genet.* **62**, 159-170.

Liang et al. (2001). A robust identity-by-descent procedure using affected relative pairs: Multipoint mapping for complex diseases. *Hum. Hered.* **51**, 64-78.

McPeck, S. and Strahs, A. (1999). Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am. J. Hum. Genet.* **65**, 858-875.

Morris A.P. et al. (2000). Bayesian fine-scale mapping of disease loci, by hidden Markov models. *Am. J. Hum. Genet.* **67**(1), 155-69.

Morris A.P. et al. (2002). Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies. *Am. J. Hum. Genet.* **70**, 686-707.

Prentice, R.L. and Pyke, R. (1979). Logistic disease incidence models and case-control studies. *Biometrika* **66**, 403-411.

Putter, H. et al. (2002). Score test for detecting linkage to quantitative traits. *Genet. Epidemiol.* **22**, 345-355.

Risch, N. (1984). Segregation analysis incorporating genetic markers. I. Single-locus models with an application to type I diabetes. *Am. J. Hum. Genet.*, **36**, 363-386.

Risch, N. (1990). Linkage strategies for genetically complex traits III. The effect of marker polymorphism on analysis of affected relative pairs. *Am. J. Hum. Genet.* **46**, 242-253.

Service et al. (1999). Linkage-disequilibrium mapping of disease genes by reconstruction of ancestral haplotypes in founder populations. *Am. J. Hum. Genet.* **64**, 1728-1738.

Schaid, D.J. (1996). General score tests for association of genetic markers with disease using cases and their parents. *Genet. Epidemiol.* **13**, 423-449.

Sham, P.C., Cherny, S.S., Purcell, S. and Hewitt, J.K. (2000). Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am. J. Hum. Genet.* **66**, 1616-1630.

Sham, P. et al. (2002). Powerful regression-based quantitative-trait linkage analysis for general pedigrees. *Am. J. Hum. Genet.* **71**, 238-253.

Shih, M-C. and Whittemore, A.S. (2002). Tests for genetic association using family data. *Genet. Epidemiol.* **22**, 128-145.

Spielman, R.S, McGinnis, R.E. and Ewens, W.J. (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506-516.

Terwilliger, J.D. (1995). A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am. J. Hum. Genet.* **56**, 777-787.

Terwilliger, J.D. and Göring, H.H.H. (2000). Gene mapping of the 20th and 21st centuries: Statistical methods, data analysis and experimental design. *Hum. Biol.* **1**, 63-132.

Terwilliger, J.D. and Ott, J. (1992). A haplotype-based 'haplotype-relative-risk' approach to detecting allelic associations. *Hum. Hered.* **42**, 337-346.

Tang, H-K. and Siegmund, D. (2001). Mapping quantitative trait loci in oligogenic models. *Biostatistics* **2**, 147-162.

Tu, I-P., Balise, R.R. and Whittemore, A.S. (2000). Detection of disease genes by use of family data. II. Application to nuclear families. *Am. J. Hum. Genet.* **66**, 1341-1350.

Wang, K. and Huang, J. (2002). A score-statistic approach for the mapping of quantitative-trait loci with sibships of arbitrary size. *Am. J. Hum. Genet.* **70**, 412-424.

Whittemore, A. (1996). Genome scanning for linkage: An overview. *Biometrics* **59**, 704-716.

Whittemore, A. and Halpern, J. (1994). A class of tests for linkage using affected pedigree members. *Biometrics* **50**, 118-127.

Whittemore, A.S. and Tu, I-P. (2000). Detection of disease genes theory. *Am. J. Hum. Genet.* **66**, 3128-1340.