Linkage Analysis with Ordinal Data: Sex-limitation

Michael Neale, Marleen De Moor & Sarah Medland

Thanks to Fruhling Rijsdijk, Kate Morley et al whose slides we ripped off

Boulder CO International Workshop March 8 2007

Overview

- Background of ordinal trait modeling
- Introduction to sex-limitation theory
- Practical on sex-limited linkage analysis: Dutch twins' exercise participation

Ordinal data

Measuring instrument is able to only discriminate between two or a few ordered categories e.g. absence or presence of a disease. Data take the form of counts, i.e. the number of individuals within each category:

Of 100 individuals:

90 'no'
10 'yes'

	'no'	'yes'
'no'	55	19
'yes'	18	8

Univariate Normal Distribution of Liability

Assumptions:

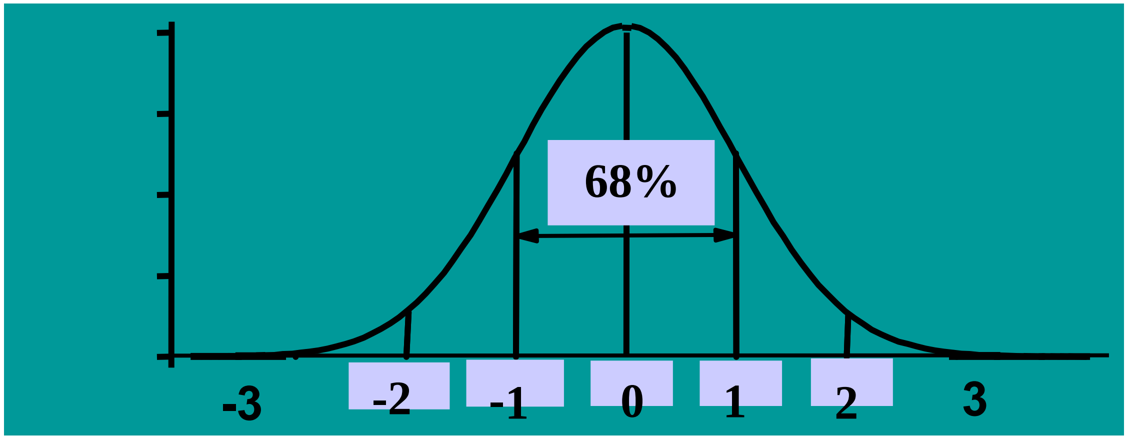
(1) Underlying *normal* distribution of liability

(2) The liability distribution has 1 or more thresholds (cut-offs)

The standard Normal distribution

Liability is a *latent* variable, the scale is arbitrary, distribution is, therefore, assumed to be a *Standard Normal Distribution* (SND) or z-distribution:

- mean (μ) = 0 and SD (σ) = 1
- z-values are the number of SD away from the mean
- area under curve translates directly to probabilities > Normal Probability Density function (Φ)



Two categorical traits:

Data from siblings

In an unselected sample of sib pairs > Contingency Table with 4 observed cells:

cell a:number of pairs concordant for unaffected
cell d: number of pairs concordant for affected
cell b/c: number of pairs discordant for the disorder

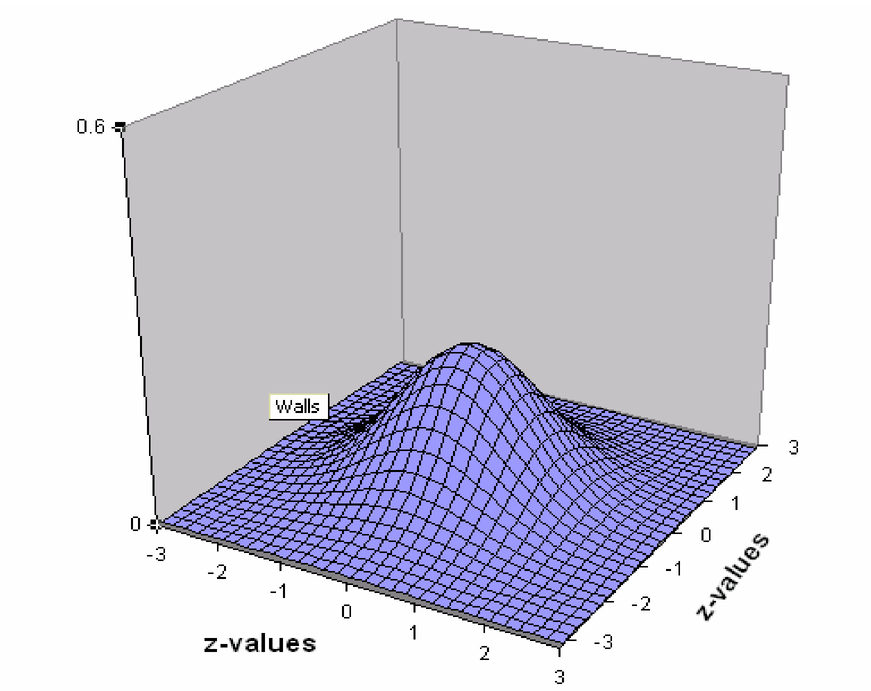
Twin2 \ Twin1	0	1
0	545 (.76)	75 (.11)
1	56 (.08)	40 (.05)

0 = unaffected
1 = affected

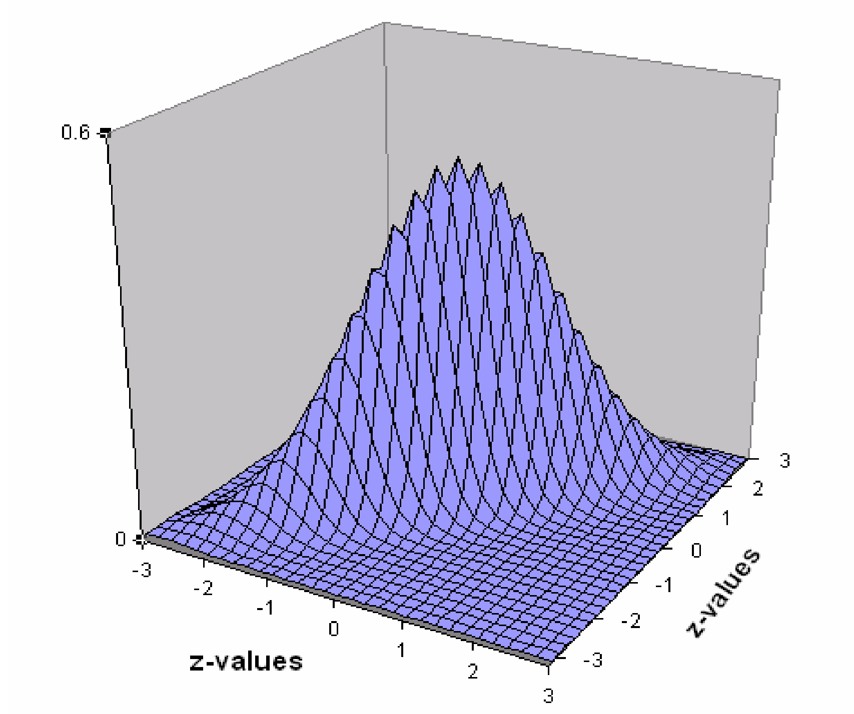
Joint Liability Model for sib/twin pairs

- Assumed to follow a **bivariate normal** distribution, where both traits have a mean of 0 and standard deviation of 1, but the **correlation** between them is unknown.
- The **shape** of a bivariate normal distribution is determined by the **correlation** between the traits

Bivariate Normal

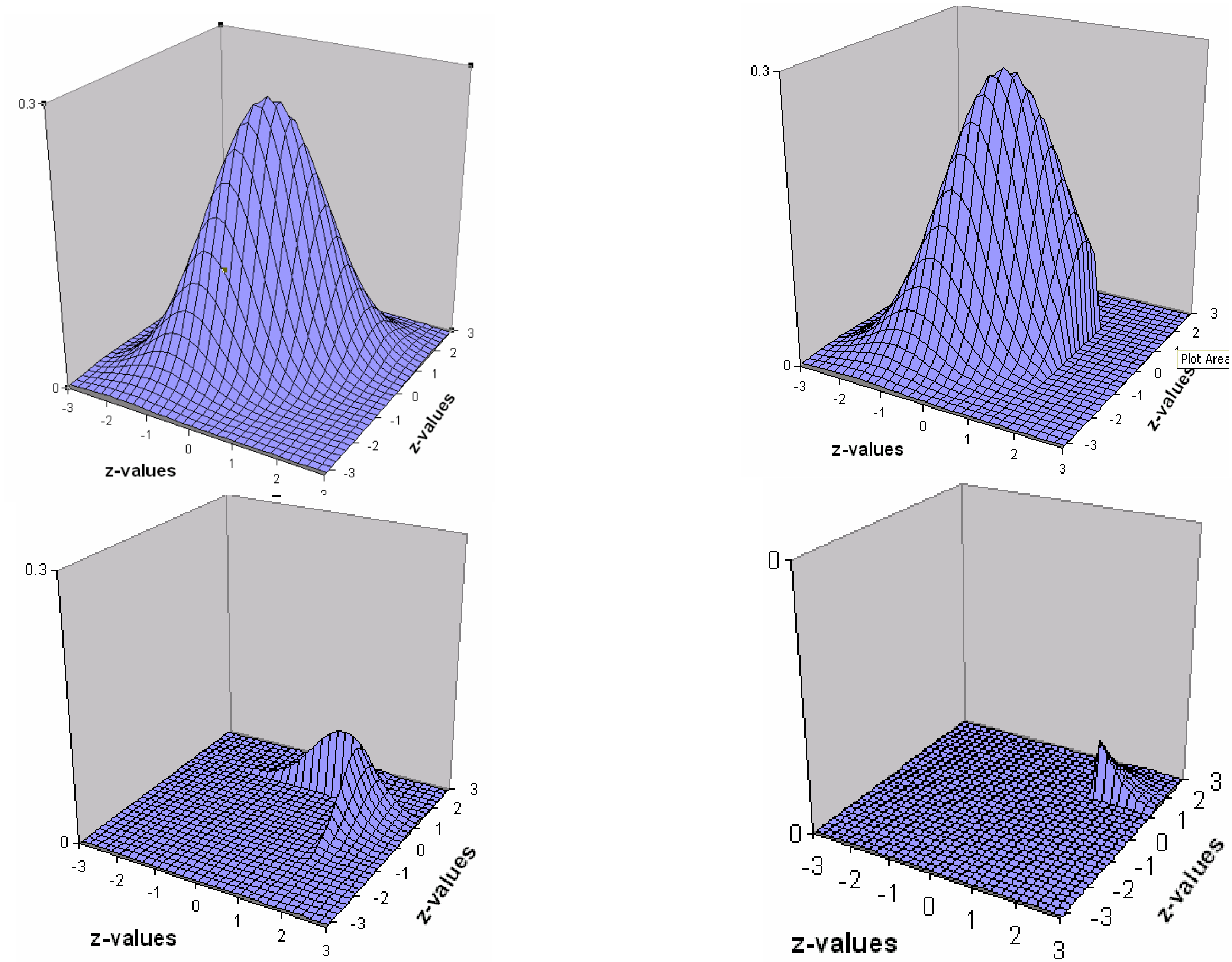


r = .00



r = .90

Bivariate Normal (R=0.6) partitioned at threshold 1.4 (z-value) on both liabilities



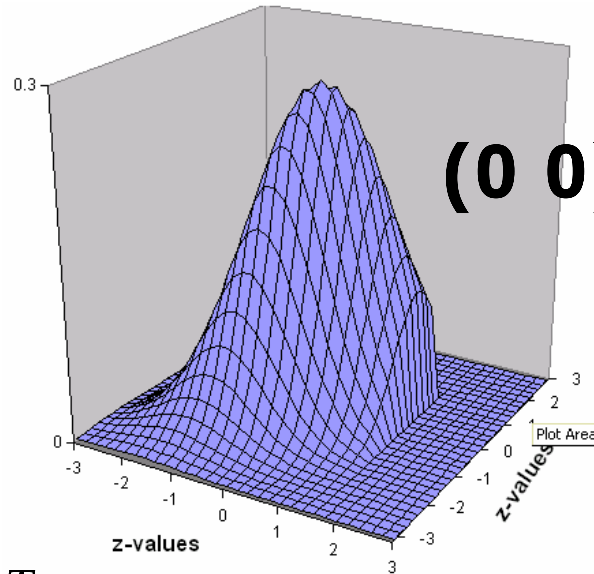
How are expected proportions calculated?

By **numerical integration** of the bivariate normal over two dimensions: the liabilities for twin1 and twin2

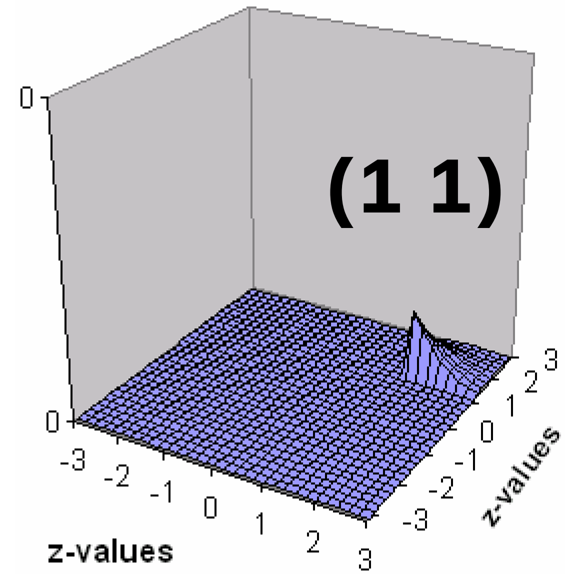
e.g. the probability that both twins are affected :

$$\int_{T_1}^{\infty}\int_{T_2}^{\infty} \Phi(L_1, L_2; \mu, \Sigma)\, dL_1 dL_2$$

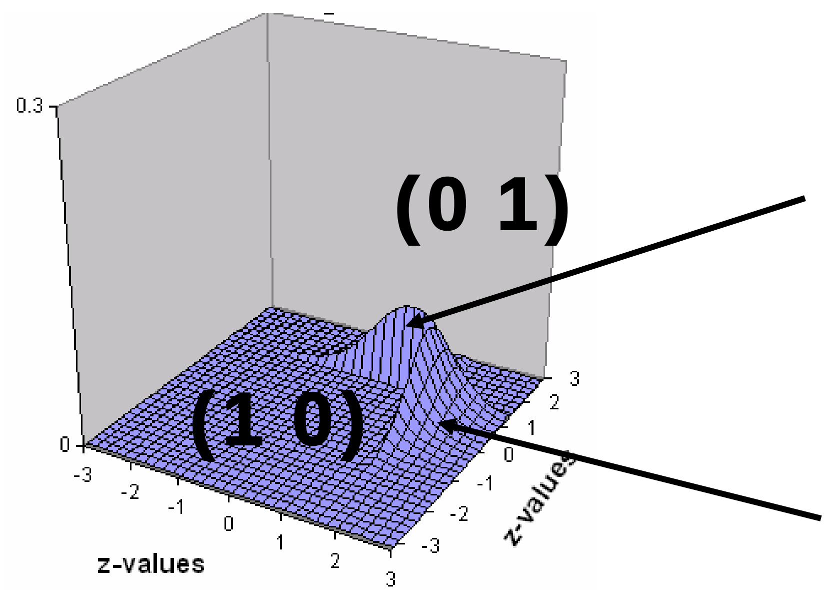
Φ is the bivariate normal probability density function,
L_1 and L_2 are the liabilities of twin1 and twin2, with means 0,
and Σ is the correlation matrix of the two liabilities
T_1 is threshold (z-value) on L_1, T_2 is threshold (z-value) on L_2



$$\int_{-\infty}^{T_1}\int_{-\infty}^{T_2}\Phi(L_1,L_2;\mu,\Sigma)\,dL_1dL_2$$



$$\int_{T_1}^{\infty}\int_{T_2}^{\infty}\Phi(L_1,L_2;\mu,\Sigma)\,dL_1dL_2$$



$$\int_{-\infty}^{T_1}\int_{T_2}^{\infty}\Phi(L_1,L_2;\mu,\Sigma)\,dL_1dL_2$$

$$\int_{T_1}^{\infty}\int_{-\infty}^{T_2}\Phi(L_1,L_2;\mu,\Sigma)\,dL_1dL_2$$

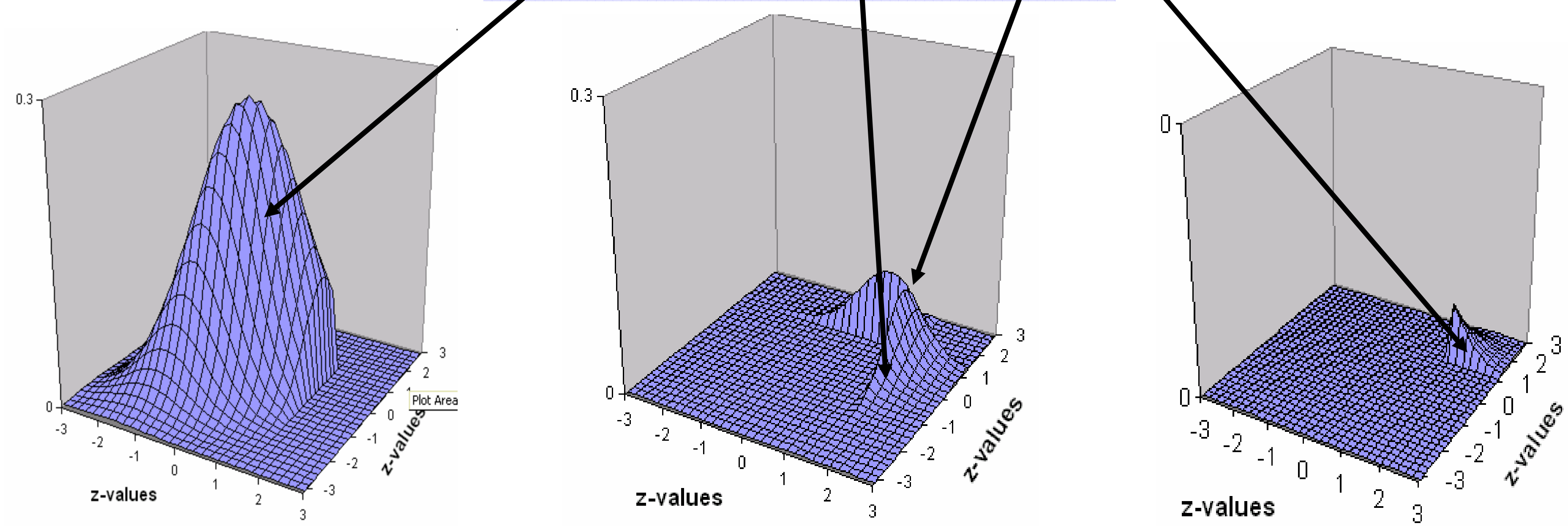
How is numerical integration performed?

There are programmed mathematical subroutines that can do these calculations

Mx uses one written by Alan Genz

Expected Proportions of the BN, for R=0.6, Th1=1.4, Th2=1.4

Liab 2 / Liab 1	0	1
0	.87	.05
1	.05	.03

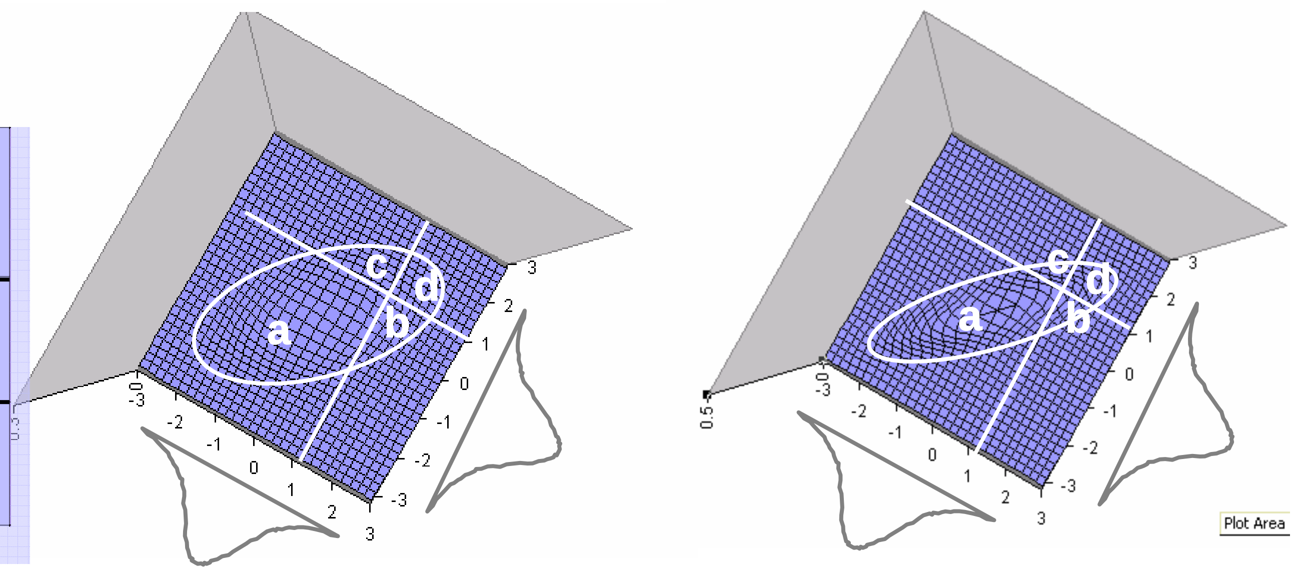


How can we estimate correlations from CT?

The correlation (shape) of the bivariate normal and the two thresholds determine the relative proportions of observations in the 4 cells of the contingency table.

Conversely, the sample proportions in the 4 cells can be used to estimate the correlation and the thresholds.

Twin2 / Twin1	0	1
0	a	b
1	c	d



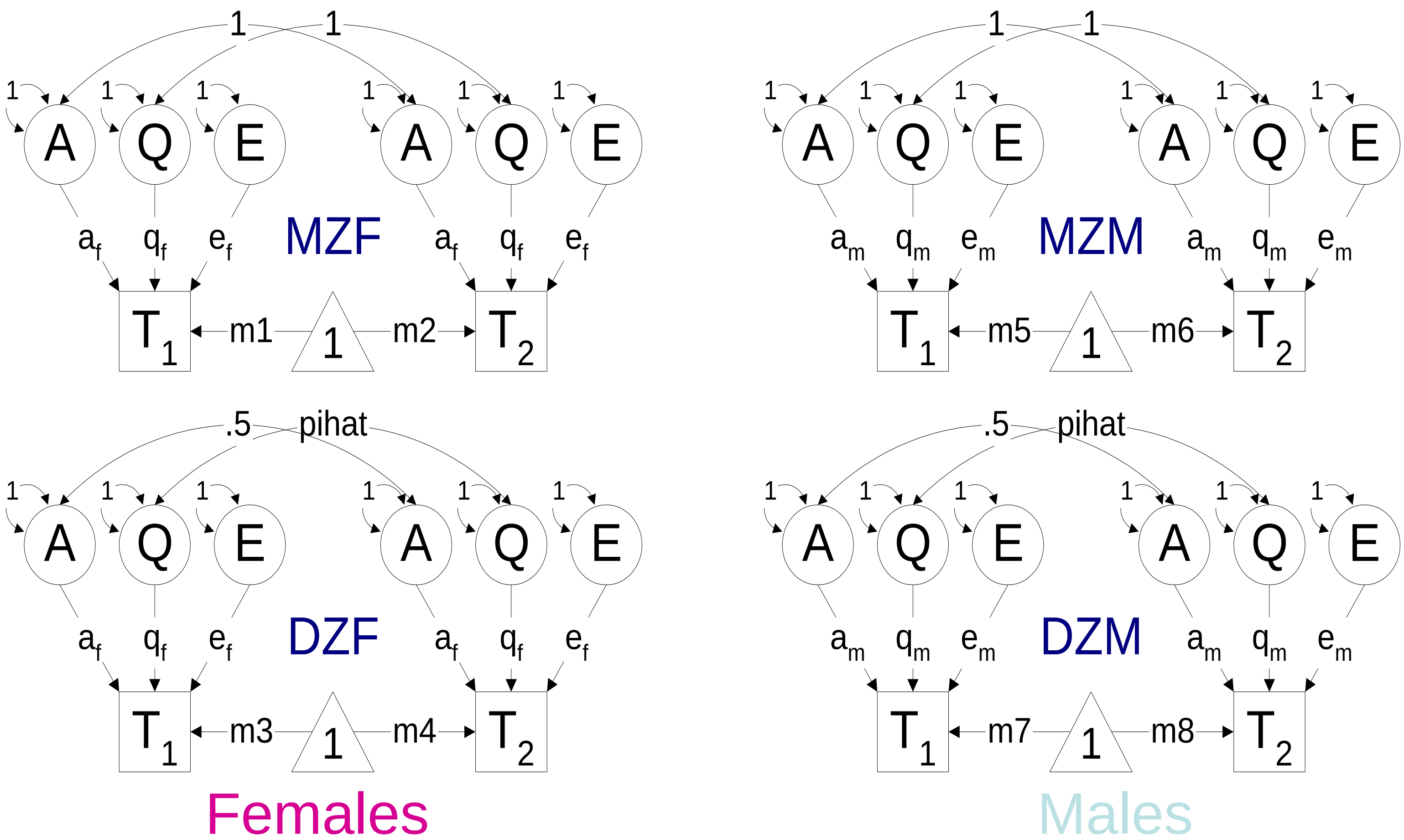
Summary

It is possible to estimate a *tetrachoric* correlation between categorical traits from simple counts because we assume that the underlying joint distribution is *bivariate normal*

The relative sample proportions in the 4 cells are translated to proportions under the bivariate normal so that the most likely correlation and the thresholds are derived

Next: use correlations in a linkage analysis

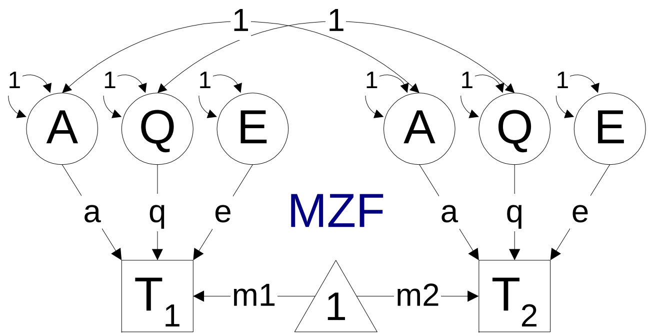
Heterogeneity

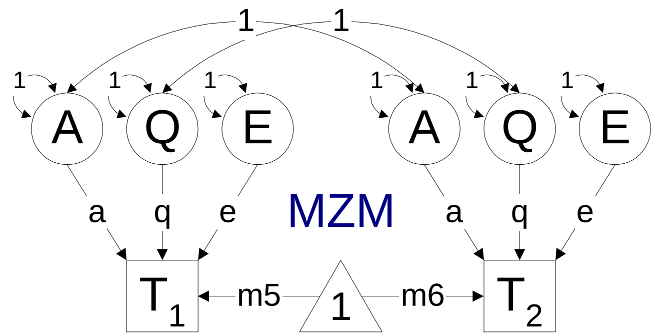


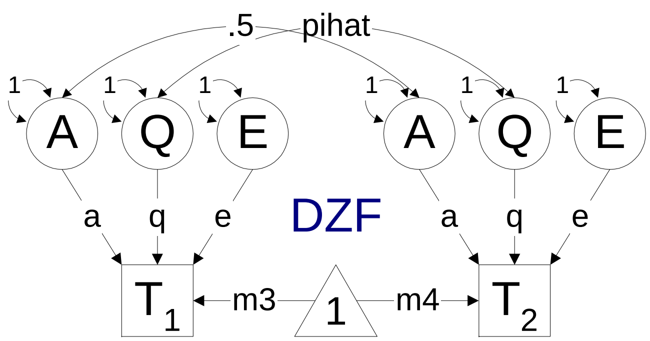
What about DZO?

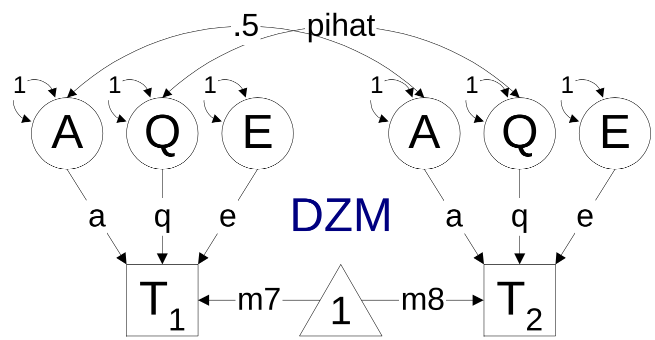
- Var F, Cov MZF, Cov DZF
  - a_f, d_f, e_f
- Var M, Cov MZM, Cov DZM
  - a_m, d_m, e_m
- Var Fdzo = Var F, Var M dzo = Var M
- Cov DZO
  - r_g (but still pihat)

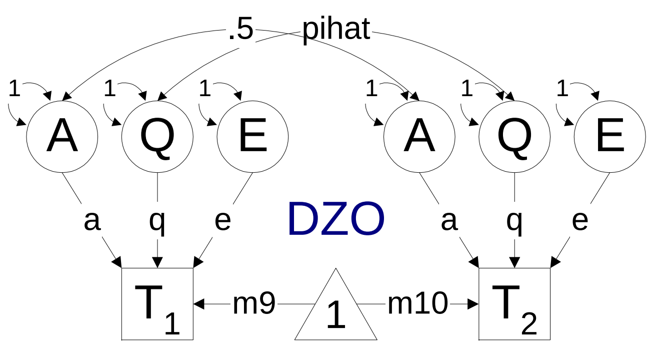
Homogeneity





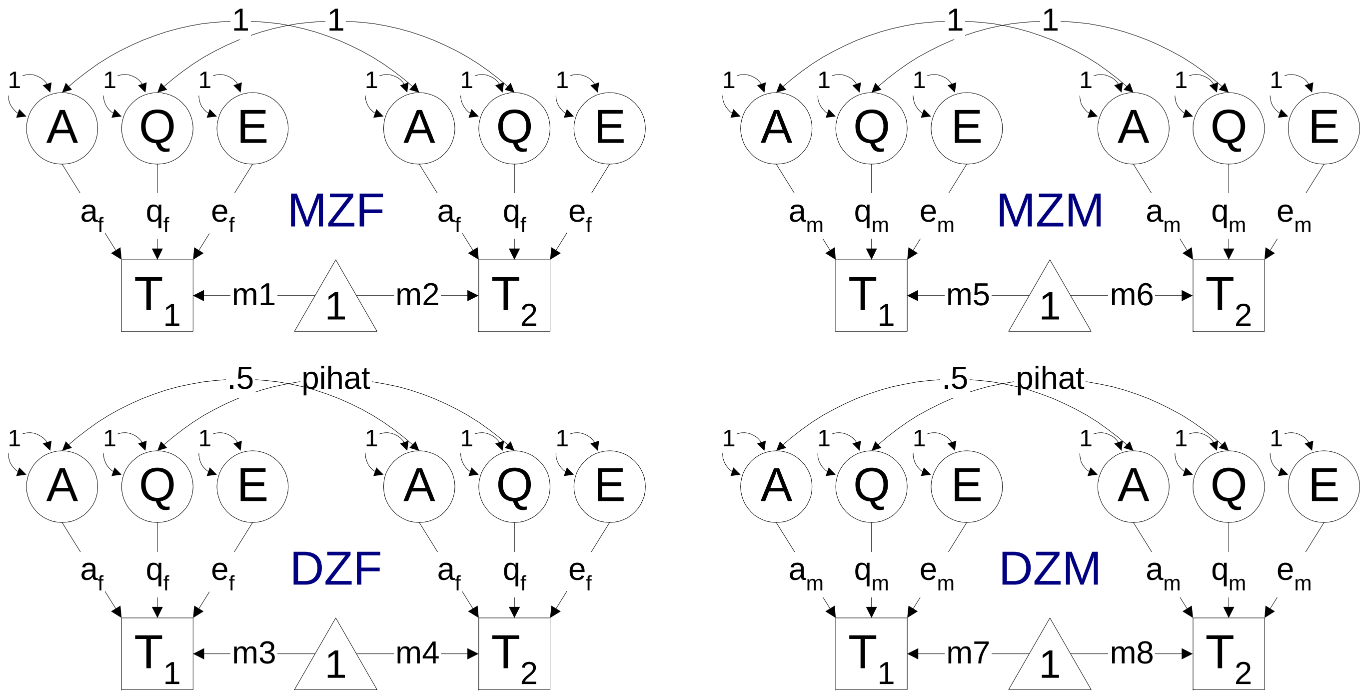


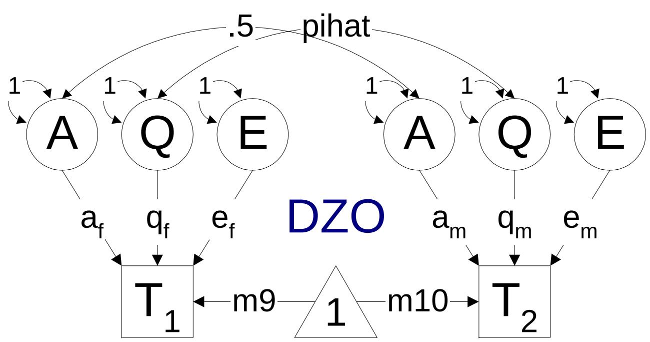




DZO	female	male
female	$a^2+e^2+q^2$	$.5a^2+pihat*q^2$
male	$.5a^2+pihat*q^2$	$a^2+e^2+q^2$

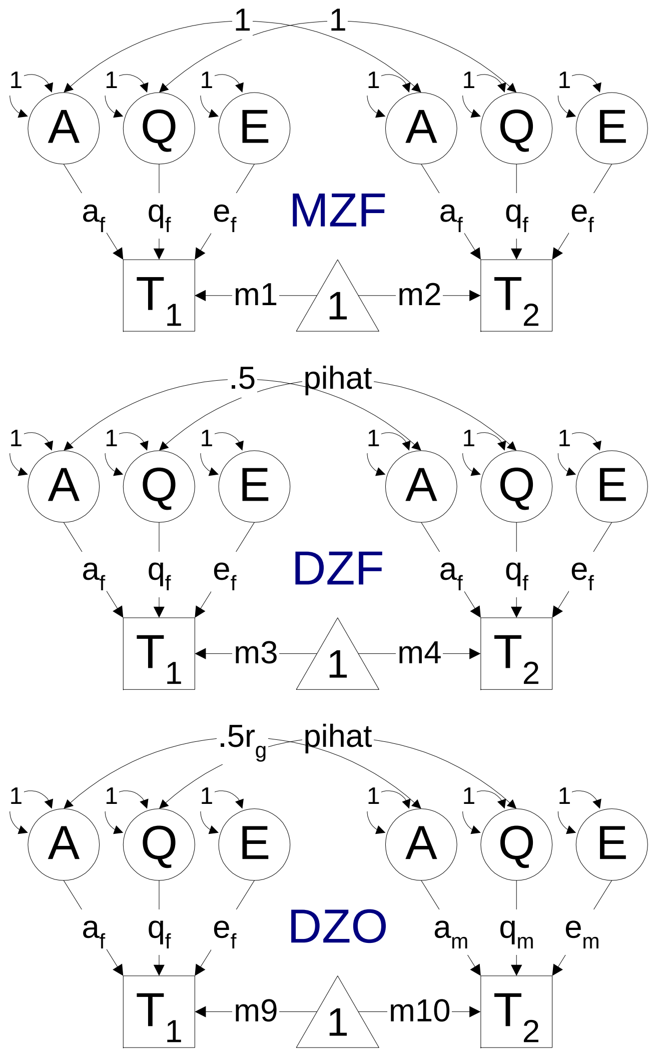
Heterogeneity

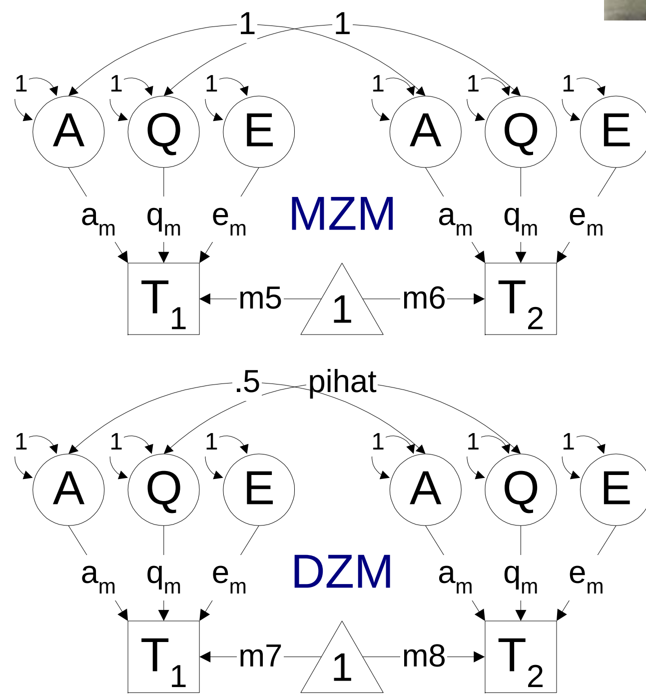




DZO	female	male
female	$a_f^2+e_f^2+q_f^2$	$.5a_fa_m'+piq_fq_m'$
male	$.5a_ma_f'+piq_mq_f'$	$a_m^2+e_m^2+q_m^2$

General Sex Limitation





DZO	female	male
female	$a_f^2+e_f^2+q_f^2$	$.5r_g a_f a_m'+$ pihat$q_f q_m'$
male	$.5r_g a_m a_f'+$ pihat$q_m q_f'$	$a_m^2+e_m^2+q_m^2$

Practical
sex-limited linkage
with ordinal data in Mx

Data: Exercise participation

- Dutch sample of twins and their siblings
- N=9,408 individuals from 4,230 families
- Binary phenotype:

  Exercise participation: Yes/No

  (Criterion: 60 min/week at 4 METs)

Genotyped sub sample

- Sub sample was genotyped
- N=1,432 sibling pairs from 619 families (MZ pairs excluded)
- (266 MM, 525 FF, 328 MF and 313 FM sib pairs)
- Genotypic information:
    - based on 361 markers on average (10.6 cM spacing)
    - IBD probabilities estimated at 1 cM grid in Merlin (multipoint)
    - Pihat calculated in Mx with formula:

      Pihat=0.5*p(IBD=1)+1*p(IBD=2)

Heritability in total sample

Twin/sibling correlations				
MZM	DZM/sibMM	MZF	DZF/sibFF	DOS/sibOS
0.71	0.30	0.55	0.30	0.17

Heritability estimates:

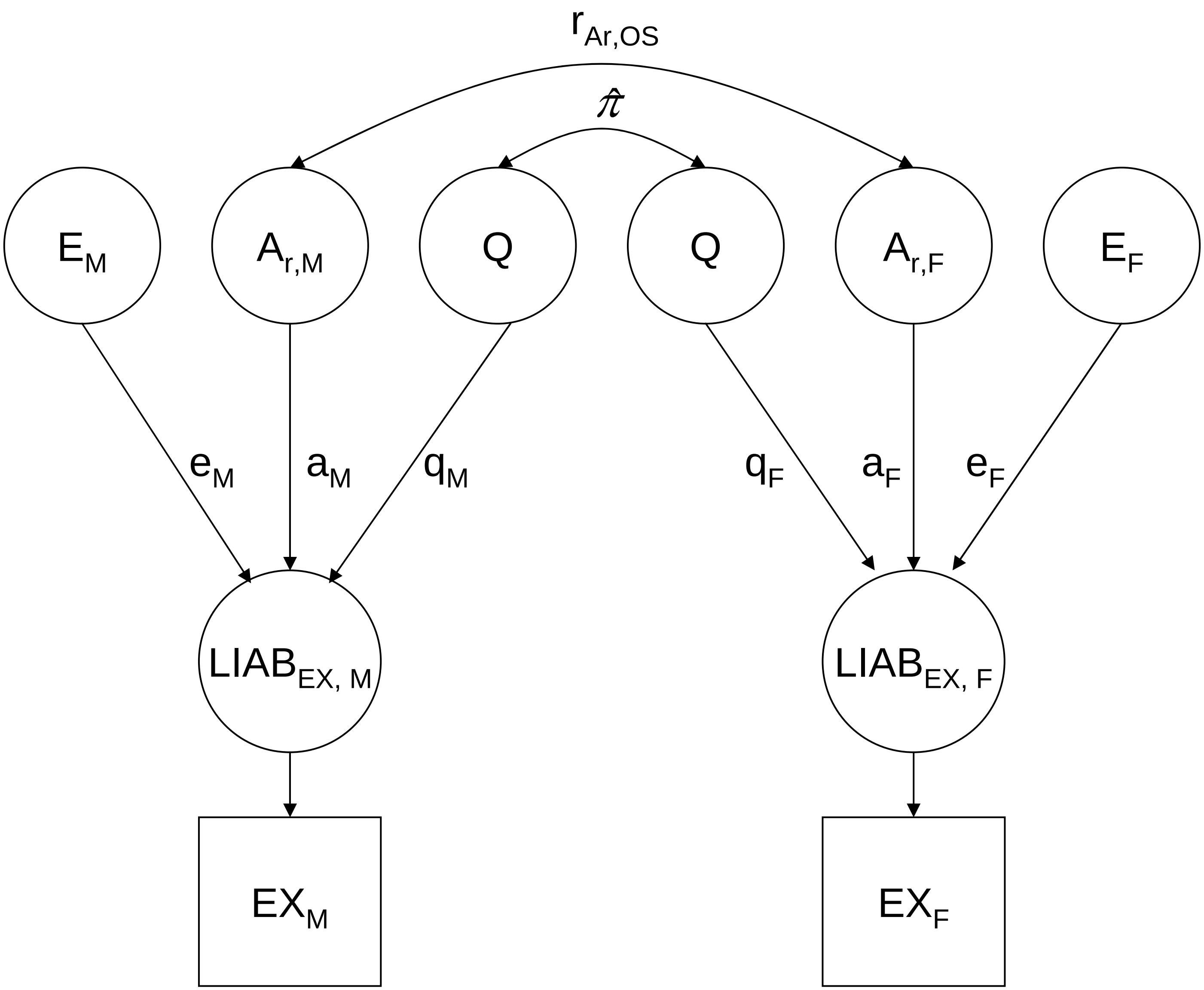
Males: **A** 69.4% **E** 30.6%

Females: **A** 55.7% **E** 44.3%

Genetic correlation OS pairs: 0.27

Thus: partly different genes affect exercise participation in males and females

Path model



r Ar,OS
π̂
EM
Ar,M
Q
Q
Ar,F
EF
eM
aM
qM
qF
aF
eF
LIABEX, M
LIABEX, F
EXM
EXF

Mx script

```
G2: Data from genotyped male-male sibling pairs
 Data NInput=346
 Ord File=c19mm.dat
…
…
Thresholds M +(S|R)*B ;
Covariances A+E+Q | H@A+P@Q _
            H@A+P@Q | A+E+Q ;
…
```

Mx script

```
G1: Calculation group
Data Calc NGroups=7

Begin Matrices ;
X Lower 1 1 Free                ! female genetic structure
Z Lower 1 1 Free                ! female specific environmental structure
G Full 1 1 Free                 ! female qtl

U Lower 1 1 Free                ! male genetic structure
W Lower 1 1 Free                ! male specific environmental structure
F Full 1 1 Free                 ! male qtl
…
Begin Algebra;
A = U*U';                       ! male genetic variance
E = W*W';                       ! male specific environmental variance
Q = F*F';                       ! male qtl variance
V = A+E+Q;                      ! male total variance
P = K*I;                        ! calculates pihat
End Algebra ;
…
```

Mx script

```
G6: constraint males: total variance=1
Constraint
Begin matrices = Group 1;
J unit nvar 1
End matrices;
Constraint V=J;
option no-output
END
…
```

Exercise

- Run the script AEQc19.mx for position 11 on chromosome 19
- Modify the script to test:
  - for sex heterogeneity at QTL
  - significance of QTL males
  - significance of QTL females
- Obtain chi2 in the output and compute LOD scores for females and males with formula:

  LOD=chi2/4.61

- If you have time, repeat this for another position on chromosome 19

Solution

Modify the script:

```
G5
…
Option Multiple Issat
END
…
Save full.mxs

Get full.mxs
!Test for sex heterogeneity
Equate F 1 1 1 G 1 1 1
END

Get full.mxs
!Test for significance female QTL
Drop G 1 1 1
END

Get full.mxs
!Test for significance male QTL
Drop F 1 1 1
END
```

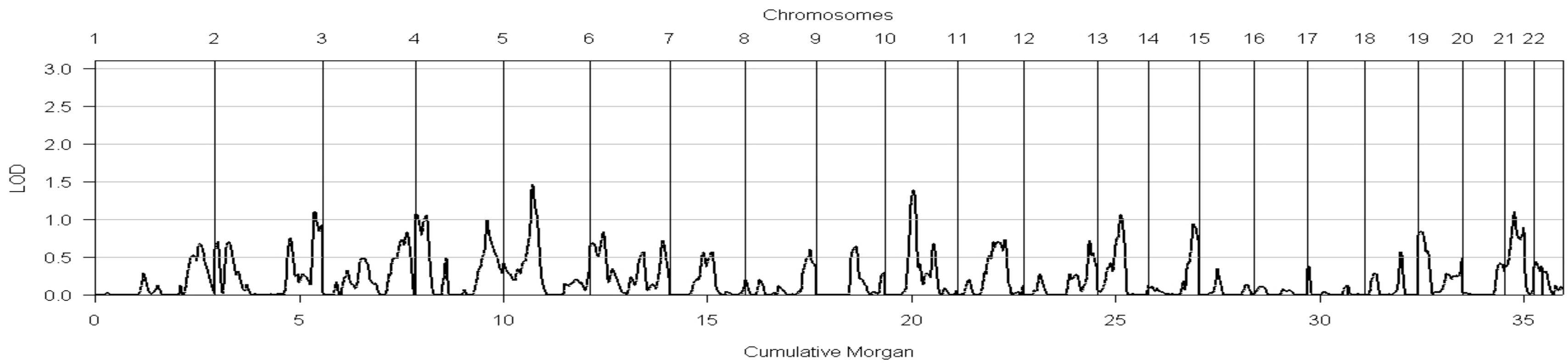
Solution

Results from Mx output:

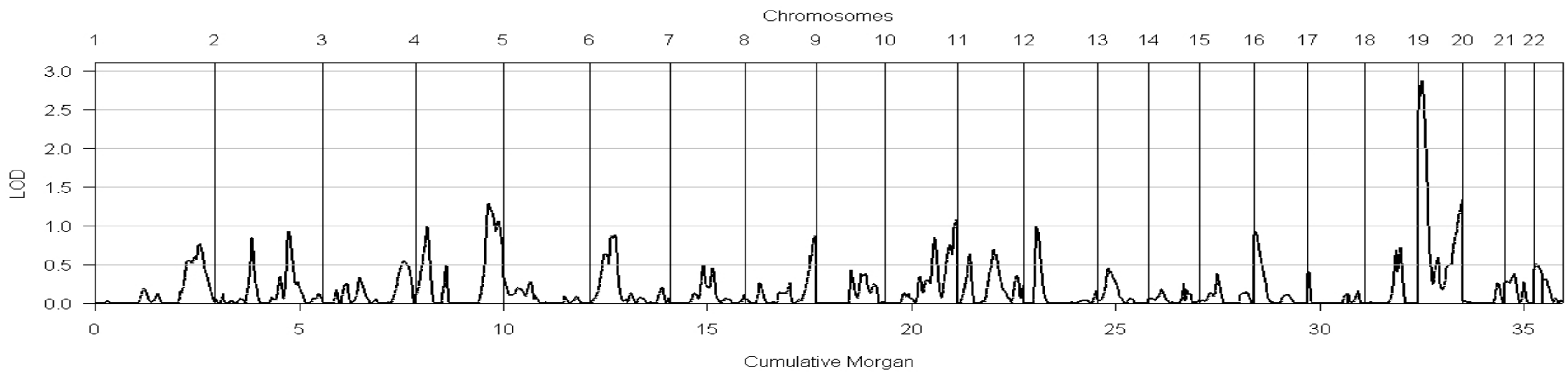
Chromosome 19 position 11	chi2	Δdf	p	**LOD**
Test sex Heterogeneity	2.54	1	0.111	**0.55**
Drop QTL females	12.70	1	<0.001	**2.75**
Drop QTL males	4.17	1	0.041	**0.90**

Results whole genome

Males:



Females:



Issues

- Power to detect linkage (or heritability) with ordinal data is lower than with continuous data
- Power to detect sex heterogeneity at QTL also low
- Unclear what is best way to test sex-specific QTLs
- QTL variance is overestimated, leads to strange estimates in different parts of the model (a_F, a_M, $r_{A,OS}$)

- Sex-limitation only considered here, but model applies to GxE generally.

More advanced scripting

Sarah Medland (2005) TRHG

Parameterization of Sex-Limited Autosomal Linkage Analysis for Mx

Sarah E. Medland[1,2]

[1] *Queensland Institute of Medical Research, Brisbane, Australia*

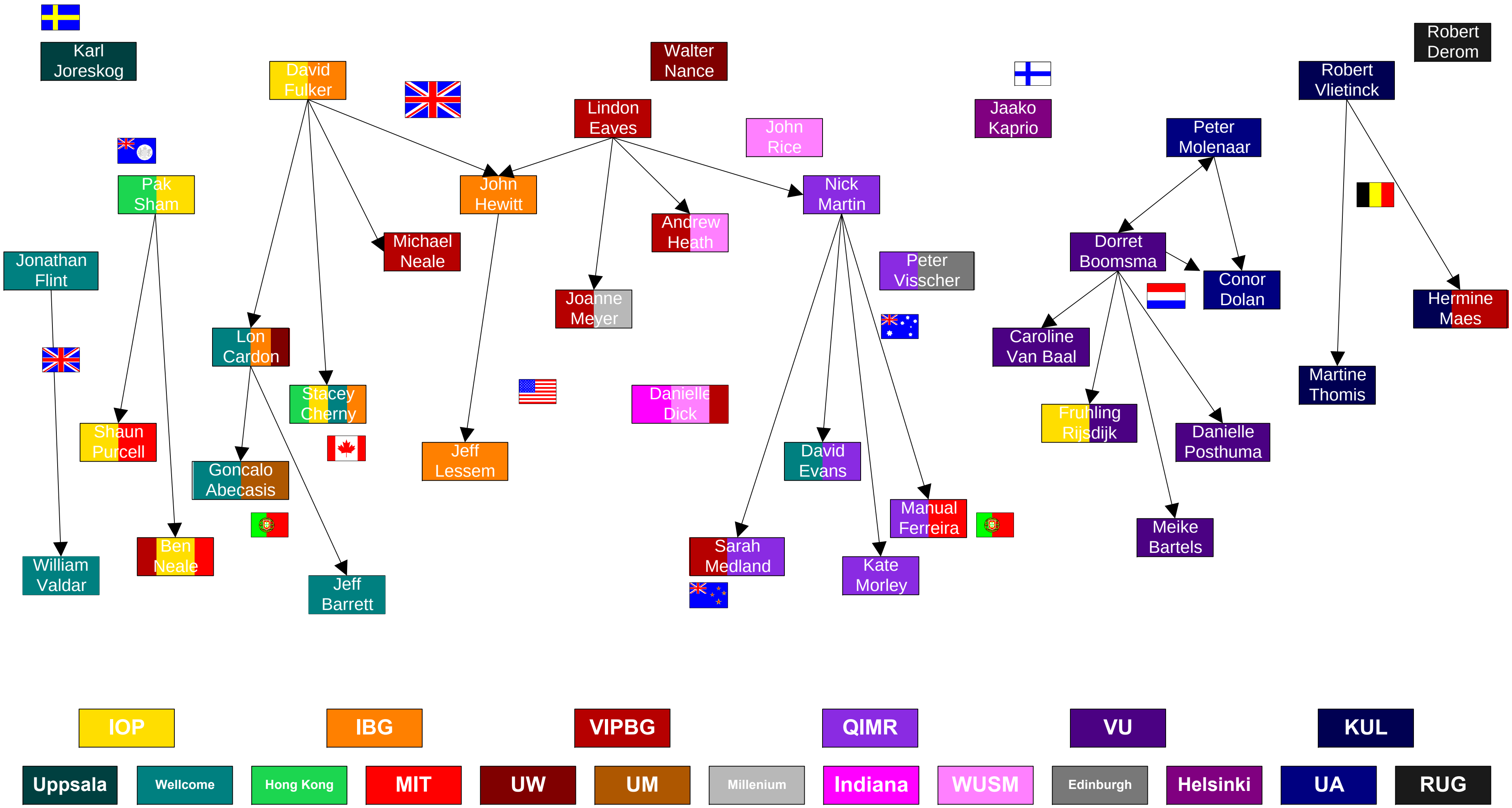
[2] *Cognitive Psychophysiology Laboratory, University of Queensland, Australia*

Incorporation of sex-limitation (genotype-sex interaction) effects into a model of quantitative trait loci (QTL) analysis has been shown to increase the power to detect linkage when analyzing traits in which sex limitation is present (Towne et al., 1997). The present note provides a parameterization of the nonscalar sex-limitation ACE model incorporating autosomal sex-limited QTL effects for use with the Mx matrix algebra program (Neale et al., 2002). An example script designed for use with extended sibships that takes advantage of the versatile treatment of covariates within Mx is included.

structured pedigrees in outbred populations (Seaton et al., 2002).

Within the context of the classical twin design, the parameterization of variance components QTL linkage analysis represents a simple extension of the model through the addition of a QTL-linked component of variance (for a recent description of the methodology see Posthuma et al., 2003). In a similar way the parameterization of a sex-limited nonscalar autosomal QTL model is achieved by extending the variance covariance matrices for a nonscalar sex-limitation script, incorporating the male and female specific QTL parameters. The parameterization adopted in this note

- Efficient script to model sex-limited linkage, only 1 datagroup
- Both continuous and ordinal data
- Especially convenient when sibships are larger than 2


Karl Joreskog
David Fulker
Walter Nance
Lindon Eaves
John Rice
Jaako Kaprio
Robert Derom
Robert Vlietinck
Peter Molenaar
Pak Sham
John Hewitt
Nick Martin
Andrew Heath
Michael Neale
Jonathan Flint
Peter Visscher
Dorret Boomsma
Conor Dolan
Joanne Meyer
Hermine Maes
Caroline Van Baal
Lon Cardon
Stacey Cherny
Danielle Dick
Martine Thomis
Shaun Purcell
Fruhling Rijsdijk
Danielle Posthuma
Jeff Lessem
David Evans
Goncalo Abecasis
Manual Ferreira
Meike Bartels
William Valdar
Ben Neale
Sarah Medland
Kate Morley
Jeff Barrett
IOP
IBG
VIPBG
QIMR
VU
KUL
Uppsala
Wellcome
Hong Kong
MIT
UW
UM
Millenium
Indiana
WUSM
Edinburgh
Helsinki
UA
RUG

THE 20th ANNIVERSARY INTERNATIONAL WORKSHOP ON METHODOLOGY OF TWIN AND FAMILY STUDIES

October 1 - 5, 2007

Leuven, Belgium