Genetic

Epidemiology and Human Genetics

David Duffy

Queensland Institute of Medical Research
Brisbane, Australia

Genetic Epidemiology

A science that deals with aetiology, distribution, and control of disease in groups of relatives and with inherited causes of disease in populations (Morton, 1982).

This is double barrelled, but information about populations is inferred via correlations observed among samples of relatives (and *vice versa*). Inheritance may be genetic or nongenetic.

Related/salient disciplines include:

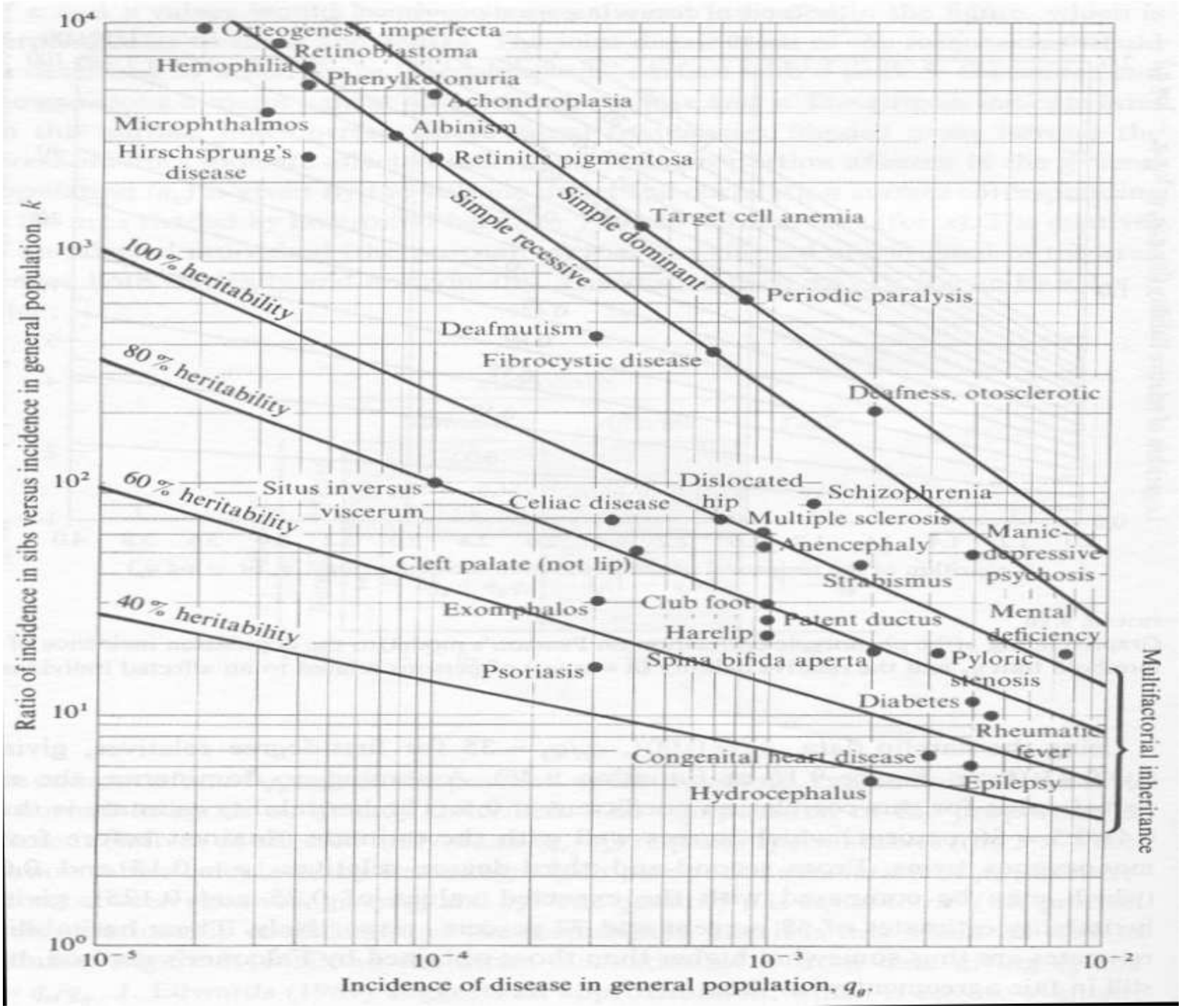
Molecular epidemiology	Human genetics
Clinical genetics	Cancer genetics
Biochemical genetics	Molecular biology
Genomics	Bioinformatics
Statistical genetics	Evolutionary and population genetics
Chronic disease epidemiology	Physical anthropology

Familial aggregation

For a dichotomous trait, such as a disease, if the probability an individual expresses a phenotype varies according the phenotype of relatives, we say it *aggregates within families.*

For many common diseases, the risk to an individual is doubled if a first degree relative is affected. For rare Mendelian disorders, this risk may increase 1000 or 1000000 fold compared to the baseline population risk.

Familial aggregation can be due either to **genes** or **family environment**.


Ratio of incidence in sibs versus incidence in general population, k
Incidence of disease in general population, q_0
10^4
10^3
10^2
10^1
10^0
10^{-5}
10^{-4}
10^{-3}
10^{-2}
Osteogenesis imperfecta
Retinoblastoma
Hemophilia
Phenylketonuria
Achondroplasia
Microphthalmos
Albinism
Hirschsprung's disease
Retinitis pigmentosa
Simple dominant
Simple recessive
Target cell anemia
100 % heritability
Periodic paralysis
Deafmutism
Fibrocystic disease
80 % heritability
Deafness, otosclerotic
Situs inversus viscerum
Dislocated hip
Schizophrenia
60 % heritability
Celiac disease
Multiple sclerosis
Manic-depressive psychosis
Anencephaly
Cleft palate (not lip)
Strabismus
Exomphalos
Club foot
Patent ductus
Mental deficiency
Harelip
40 % heritability
Spina bifida aperta
Pyloric stenosis
Psoriasis
Diabetes
Rheumatic fever
Congenital heart disease
Epilepsy
Hydrocephalus
Multifactorial inheritance

Familial correlation

For a quantitative (metric) trait, such as blood pressure or plasma glucose level, we can detect familial causes by measuring the correlation of trait values among family members.

TABLE I.

NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES.

(All Female heights have been multiplied by 1·08).

Heights of the Mid-parents in inches.	Heights of the Adult Children.													Total Number of		Medians.
	Below	62·2	63·2	64·2	65·2	66·2	67·2	68·2	69·2	70·2	71·2	72·2	73·2	Above	Adult Children.	Mid-parents.	
Above ..	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	..
72·5	..	..	..	..	..	..	..	1	2	1	2	7	2	4	19	6	72·2
71·5	..	..	..	..	1	3	4	3	5	10	4	9	2	2	43	11	69·9
70·5	1	..	1	..	1	1	3	12	18	14	7	4	3	3	68	22	69·5
69·5	..	..	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68·9
68·5	1	..	7	11	16	25	31	34	48	21	18	4	3	..	219	49	68·2
67·5	..	3	5	14	15	36	38	28	38	19	11	4	..	..	211	33	67·6
66·5	..	3	3	5	2	17	17	14	13	4	..	..	..	..	78	20	67·2
65·5	1	..	0	5	7	11	11	7	7	5	2	1	..	..	66	12	66·7
64·5	1	1	4	4	1	5	5	..	2	..	..	..	..	..	23	5	65·8
Below ..	1	..	2	4	1	2	2	1	1	..	..	..	..	..	14	1	..
Totals ..	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	..
Medians ..	..	..	66·3	67·8	67·9	67·7	67·9	68·3	68·5	69·0	69·0	70·0	..	..	..	..	..

NOTE.—In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62·2, 63·2, &c., instead of 62·5, 63·5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents.

Simple versus Complex Inheritance

Aggregation of a trait in a family may be due to a single strong causative agent (that is both sufficient and necessary). The best example of these are **Mendelian disorders**, where the penetrance of the trait locus genotypes is high, and the risk alleles rare. This is **simple inheritance**.

Complex inheritance is characterized by the presence of multiple causes of familial aggregation, which are sufficiently common in the population that they co-occur and interact. These factors may be genetic or environental, and might be neither sufficient or necessary for the development of the condition under their control.

Most common chronic diseases are regarded as complex. An interesting argument at the moment is whether common diseases are due to the interaction of large numbers of less common risk alleles (**polygenic**), or to a smaller number of common risk alleles (**oligogenic**). Regardless of this, we know of many important environmental risk factors that also affect these diseases.

Family Environment

- Geographical Location
- Infection
- Exposure to toxins etc
- Diet
- Education
- Health-related Behaviours

Fatal pleural mesothelioma diseases caused by familial household contacts with asbestos fiber dust

Schneider J, Grossgarten K, Woitowitz HJ.

The case histories of a family are described where 3 out of 4 developed asbestos-related diseases. Only the husband had direct occupational exposure handling blue-asbestos materials while working in a producing insulating factory in 1950-59. He died of pulmonary asbestosis as an occupational disease. His wife and his son died of asbestos related mesothelioma. Detailed exposure history revealed exposure to asbestos by laundering her husband's contaminated working clothes. His son was exposed to asbestos during childhood by helping his mother laundering the father's working clothes and in addition by visiting his father's working place regularly. The significance of nonoccupational exposure to asbestos is emphasized as a causative factor in the development of malignant mesothelioma.

Am. J. Hum. Genet. 46:994–999, 1990

Simulation of Mendelism Revisited: The Recessive Gene for Attending Medical School

Peter McGuffin and Philip Huckle

Department of Psychological Medicine, University of Wales College of Medicine, Cardiff

Summary

Much of the recent confusion concerning studies of complex phenotypes such as neuropsychiatric disorders may derive from the inappropriate assumption of simple Mendelian transmission. This has sometimes led to unrealistic expectations regarding the potential benefits of linkage studies. To investigate how Mendelism may be simulated, we collected data on a common familial behavioral trait, attendance at medical school, among the relatives of 249 preclinical medical students. The "risk" of first-degree relatives going to medical school was approximately 61 times that of the general population. Complex segregation analysis carried out under a unified model provided strong evidence of vertical transmission. The results were compatible with transmission of a major effect, and a recessive model provided as satisfactory a fit as a general single-locus model. Moreover, a commonly applied test, allowing the transmission probability parameter (τ_2) to deviate from its Mendelian value, did not give a significant improvement of fit. Only a more general model where all three transmission probabilities (τ_1, τ_2, and τ_3) were unrestricted resulted in a significantly better fit than did the recessive model.

Table 1

Frequency of Attending Medical School in Adult Relatives of Medical Students

Category of Adult Relatives (*N*)	Attending Medical School (%)
Fathers (249)	16.1
Mothers (249)	6.0
Siblings (137)	21.9
Grandparents (598)	2.8
Uncles/aunts (1,313)	2.1

Three Phases of Genetic Epidemiology

Understanding	**Discovery**	**Characterization**
AIM:		
Is there a genetic aetiology?	Where are the genes?	What do they mean?
How strong is it?	What are the variants?	Penetrance (risk)
Mode of inheritance, etc.		Prevalence (how much)
GENES:		
Unmeasured	**Inferred from markers**	**Measured**
OPTIMAL DESIGN:		
Population-based	**Opportunistic**	**Population-based**
Families, populations	Families	Individuals and families
STATISTICAL ANALYSIS:		
Pedigree	**Linkage**	**Epidemiological**

Who to study: Populations

Closest to population genetics:

- “General” population surveys eg gene frequency studies.
- Population isolates eg Amish, Aland Islanders, Bedouin.
- Interaction between populations ie Caucasian/Japanese intermarriage in Hawaii.
- Migration of populations into different environments eg Tokelauan islanders and hypertension; American blacks and sickle cell anaemia.

Who to study: Families

Examining aggregation of disease in families which may be due to heredity or shared exposure to environmental risk factors.

- Population based sample
- *Ascertained* through a proband (affected or unaffected)

Could be:

- Large complete multigenerational pedigree
- Nuclear families
- Pairs of relatives: distantly related cases, siblings, twins

Phenometric and Genometric studies

Newton Morton coined two terms to divide up genetic studies:

Phenometric refers to studying *just the trait* in families or populations.

Genometric refers to studies where genotypes as well as the trait have been measured: **genetic linkage** and **genetic association**.

Phenometric studies: Is it genetic? 1

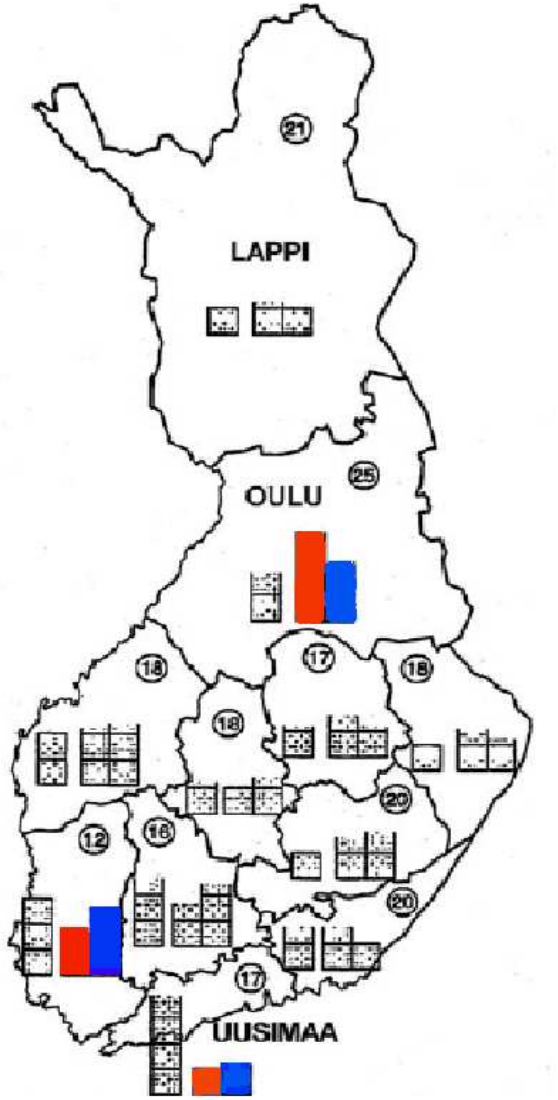
- **Comparison of ethnic groups**: does risk of disease vary by ethnicity
- **Comparison of migrants** to aboriginal population: does risk vary by group
- **Admixture** of population: how does risk alter in offspring
- **Comparison of relatedness of disease cases** compared to that of pairs of controls
  - Cancer in Utah versus relatedness estimated from genealogy
  - Cancer in UK versus sharing of surnames

- **Weinberg proband-control** studies: family history in cases compared to controls
  - CASH study of breast cancer
  - Cooke's studies of asthma

Phenometric studies: Is it genetic? 2

- **Path analysis** of family material: measuring family correlations
- **Twin studies**: are monozygotic (MZ) twins more similar than dizygotic (DZ) twins
- **Adoption studies**: are adopted children at same or different risk
- **Segregation analysis** of family material: fitting major gene models

Interethnic differences in disease rate



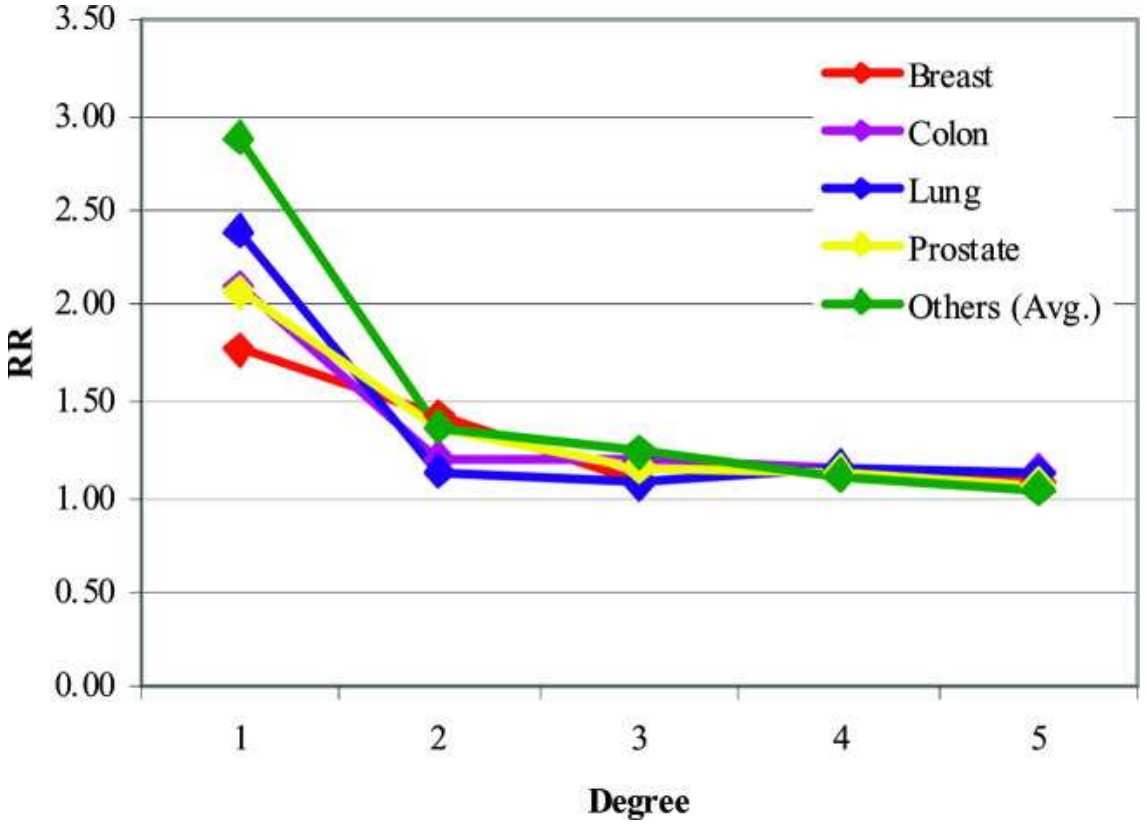
The **Finnish Disease Heritage** comprises 36 rare diseases (usually of childhood onset) affecting (in total) approximately 1 per 1000 individuals, and occur rarely in other populations.

They are especially common in “Sparse Finland”, where the original settler population was small, but has expanded relatively quickly.

Relatedness of cases

Weinberg proband-control study

This refers to the classic geneticist's version of the case-control study, where one compares the risk of disease in relatives of a case to the risk to relatives of a control (Kerber and O'Brien, Cancer 2005; 103: 1906-15).



More on Kerber and O'Brien 2005

Used the Utah genealogical and cancer registries to construct a cohort of 662515 individuals (born 1870-1984) with cancer incidence data 1966–1996.

Cancer	No. Cases	No. Controls	1st degree Relative Risk	PAR (age 65-84 years)
Breast	7919	7918	1.8 (1.6–1.9)	39%
Lung	4632	4632	2.4 (1.9–3.0)	23%
Colon	3180	3180	2.1 (1.8–2.5)	35%
Prostate	11573	11573	2.1 (1.9–2.2)	57%

The relationship between risk and degree of relatedness

As degree of relatedness to the **proband** falls, the probability that a relative carries the same genotype as the proband decreases.

Therefore, if the trait has a genetic cause, the risk to relatives should decrease as the relationship becomes more distant, but the shape of this curve is indicative of the type of inheritance:

> "The striking persistence of increased risk among relatively distant kin of patients, not only for such cancers as breast and colon but for less obviously familial cancers, such as chronic lymphocytic leuke- mia and cancers of the testis and liver, suggests that relatively simple mechanisms of shared susceptibility are at work in these families. If elevated familial re- currence risks beyond close relatives were based on interactions among multiple factors (independent genes and/or specific environmental exposures), then the probability of sharing these with third-degree, fourth-degree, or fifth-degree relatives would be small and indistinguishable from baseline. Autosomal-dominant gene effects are among the possible explanations for such persistence of risk…"

Pattern of recurrence risks versus relatedness

Risch (1990) shows that if a single major locus is acting on a trait, then there is **linear relationship**:

$$RR_1 - 1 = 2(RR_2 - 1) = 4(RR_3 - 1)$$

Where, RR_i is the recurrence risk ratio for an ith degree relative.

If multiple genes are acting multiplicatively on risk, then the risk falls off more steeply, according to **a square root relationship**.

Twin and Adoption Studies

The strong conclusions drawn by Kerber and O'Brien (2005) rely on strong assumptions about family environmental effects. It is quite possible for dietary similarity to fall off with relationship in the same way as genetic similarity.

The **Classical twin study** and the **adoption study** are two designs that try and control the effects of family environment.

The Classical Twin Study

Contrasts the similarity of monozygotic (MZ, identical) twins to the similarity of dizygotic (DZ, fraternal) twins.

If family upbringing acts equally on MZ twins, as it does on DZ twins, then a MZ-DZ difference in trait correlation must be due to genetic factors.

$RR_{MZ} > 4RR_{DZ}$	**Epistasis** - must be polygenic
$RR_{MZ} - 1 > 2(RR_{DZ} - 1)$	Genetic **dominance** (or epistasis)
$RR_{MZ} - 1 = 2(RR_{DZ} - 1)$	**Additive genetic** effect - either monogenic or polygenic
$RR_{MZ} = RR_{DZ} > 1$	No genetic contribution - effects of **family environment**
$RR_{MZ} = RR_{DZ} = 1$	No familial aggregation

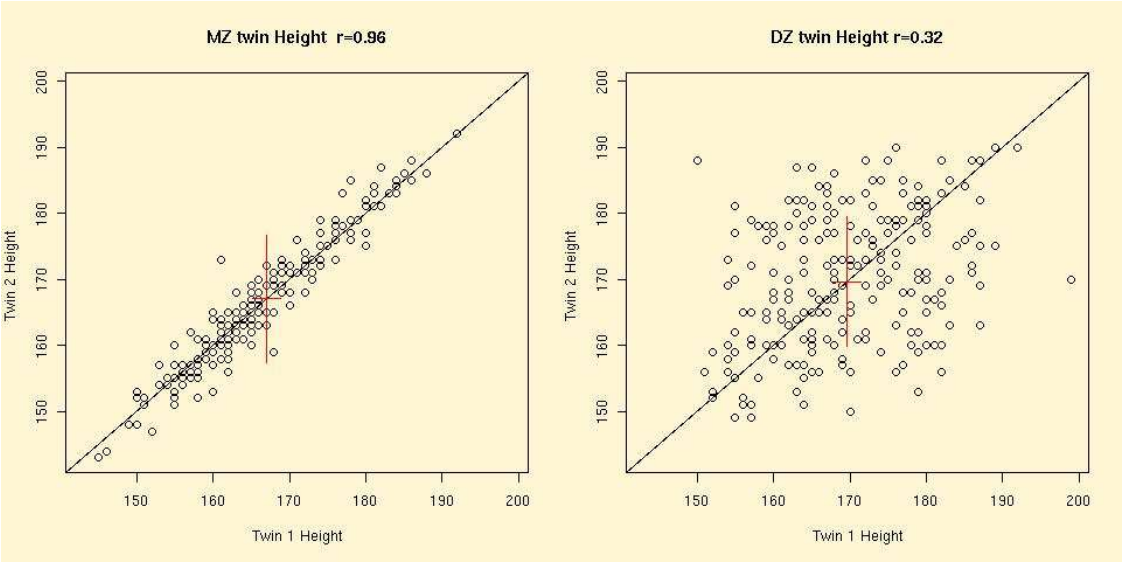
Classical Twin Study of Lichtenstein et al (2000)

Risk to cotwins of a cancer case (up to age 75 years) from a sample of 44788 pairs of Scandinavian twins.

Cancer	MZ Twins			DZ Twins		
	Aff-Aff	Aff-UnA	RR	Aff-Aff	Aff-UnA	RR
Breast (F)	42	505	5.2	52	1023	2.8
Lung (M)	15	233	7.7	24	436	6.7
Lung (F)	3	63	25.3	1	185	1.8
Colorectal (M)	10	202	6.9	17	393	5.9
Colorectal (F)	20	214	14.3	15	453	4.4
Prostate	40	299	12.3	20	584	3.1

Classical Twin Analysis of a Quantitative Trait

$r_{MZ} > 4r_{DZ}$	**Epistasis** - must be polygenic
$r_{MZ} > 2r_{DZ}$	Genetic **dominance** (or epistasis)
$r_{MZ} = 2r_{DZ}$	**Additive genetic** effect - either monogenic or polygenic
$r_{MZ} = r_{DZ} > 0$	No genetic contribution - effects of **family environment**
$r_{MZ} = r_{DZ} = 0$	No familial aggregation



Variance Components in a Classical Twin Study

Assuming the absence of **gene by environment interaction**, we can simply partition the phenotype variance into proportions due to:

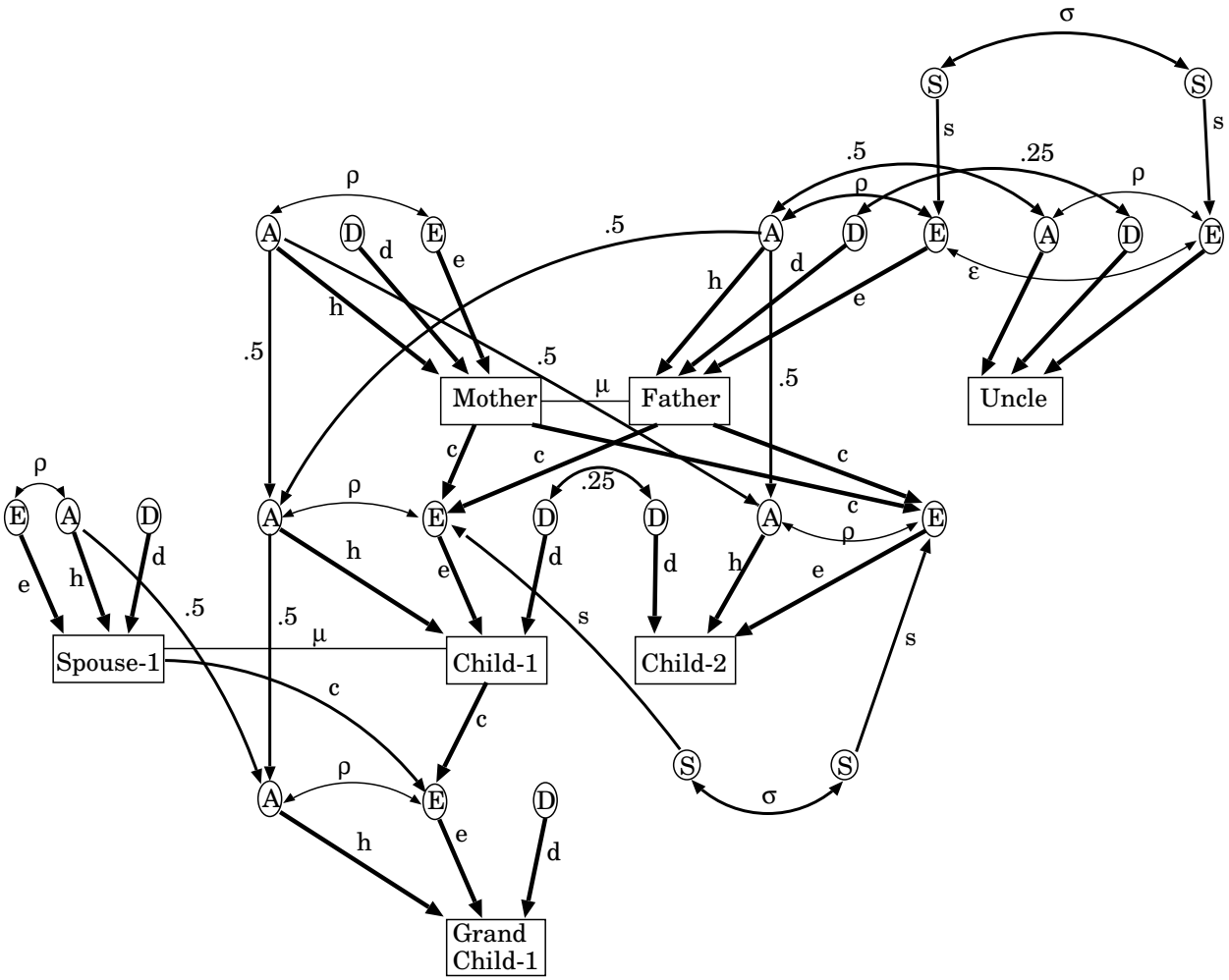
- **Dominant genetic action (D)**: also includes any epistatic variance
- **Additive genetic action (A)**: only transmitted unilineally
- **Family environment (C)**: environmental effects *c*ommon to family members
- **Unique environment (E)**: including measurement error

$$r_{MZ} = a^2 + d^2 + c^2 + e^2$$

$$r_{DZ} = \frac{1}{2}a^2 + \frac{1}{4}d^2 + c^2 + e^2$$

Path Analysis

Path analysis is the geneticist's method for analysing familial correlations, and takes its' name from the **path diagrams** that are used to graphically represent the model being fitted.



Familial Blood Pressure Correlations (Tambs et al, 1992)

Relationship	Pairs	Correlation
R=1, K=1		
MZ twins	79	0.521
R=1/2, K=1/4		
Same-sex DZ twins	90	0.154
Brothers	6017	0.220
Sisters	2856	0.211
Brother-sister	9278	0.193
R=1/2, K=0		
Father-son	10674	0.157
Father-daughter	8682	0.145
Mother-son	13262	0.165
Mother-daughter	10692	0.159
Avuncular via MZ twins	89	0.126
R=1/4, K=0		
Grandparent-grandchild	1251	0.065
Avuncular via same-sex sibs	598	0.140
Avuncular via opposite-sex sibs	548	0.083
Cousins via MZ twins	52	0.163
R=1/8, K=0		
Cousins via same-sex sibs	71	-0.047
Cousins via opposite-sex sibs	17	-0.172
R=0, K=0		
Spouse	23936	0.077
Avuncular, spouse of same-sex sibs	420	-0.022
Avuncular, spouse of opposite-sex sibs	381	0.024
Avuncular spouse of MZ twins	66	-0.110

Genometric studies: Finding the genes

- **Parametric linkage analysis** of family material: cosegregation of trait and marker loci
- **Nonparametric linkage analysis** of family material: correlation of trait with marker transmission
- **Association analysis** of unrelateds: genotype-phenotype correlation
- **Familial association analysis**: linkage plus robust association

Parametric Linkage Analysis

Marker-Marker: placing new markers onto a map of previously positioned markers. Seldom necessary for human data now.

Trait-Marker: Using an inheritance model, infer trait locus genotypes and test for linkage to markers.