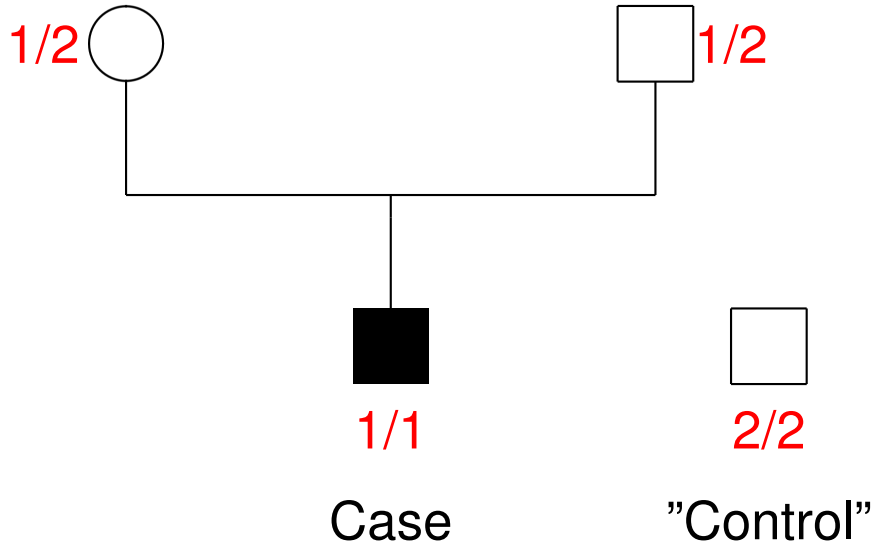
Family-based association (FBA) studies



Sample: Case-parent triads (affected child with both parents)

Method: The two parental alleles that have not been transmitted to the affected child are combined to form the genotype of a “control” individual

Advantage: Genetic background of the “control” individual is identical to that of the case individual

Unmatched analysis of FBA studies

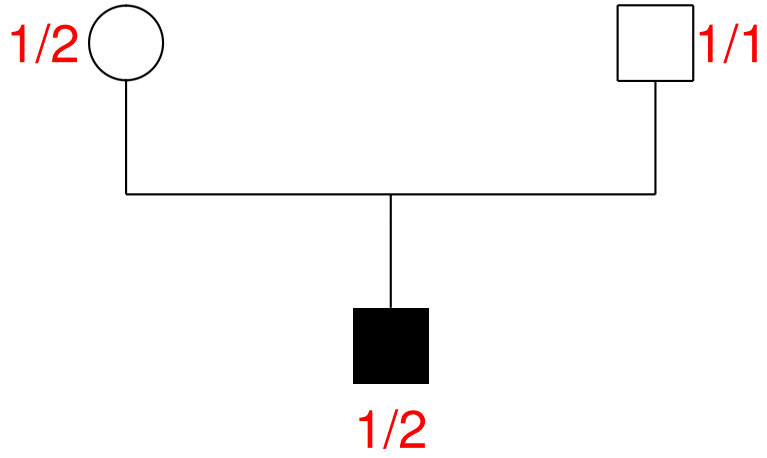
n nuclear families with a single affected child provide genotypes of n cases and n "controls". The statistical analysis can proceed exactly as for case-control studies, i.e., by comparing the genotype distribution (c.f. CC/8), by comparing the allele distribution (c.f. CC/9) between cases and controls, or by Armitage's trend test (c.f. CC/10). However, these kinds of analyses do not take into account the relationship between the cases and controls obtained from the same family (i.e., the one-to-one *matching* of cases and controls). It can be shown that

- for a "perfect" (i.e., randomly mating) population, unmatched analysis is slightly more powerful than matched analysis in most cases.
- for a stratified population, unmatched analysis of FBA studies can be conservative and less powerful than matched analysis.

Matched analysis of FBA studies

Each parent provides one pair of transmitted/non-transmitted alleles.

Example:



	Transmitted allele	Non-transmitted allele
Mother	2	1
Father	1	1

Matched analysis of FBA studies

Diallelic marker locus $\{1, 2\}$, n nuclear families

	Non-transmitted allele	
Transmitted allele	1	2
1	a	b
2	c	d

b: number of parents with genotype $1/2$ who transmit allele 1

c: number of parents with genotype $1/2$ who transmit allele 2

In case that there is no association between the alleles at the marker locus and the disease, b and c should be similar numbers.

Matched analysis of FBA studies: TDT

Test statistic: $$\text{TDT} = \frac{(b-c)^2}{b+c}$$

Null distribution: In case of no association, the distribution of the TDT-statistic can be approximated by a χ^2_1 distribution.

Name: Transmission disequilibrium test (TDT)

TDT: Exercise

Assume that genotyping a sample of 64 families with a single affected child resulted in the following data:

Genotype of mother	father	child	frequency
1/1	1/1	1/1	4
1/1	1/2	1/1	5
1/1	1/2	1/2	1
1/1	2/2	1/2	5
1/2	1/1	1/1	7
1/2	1/1	1/2	3
1/2	1/2	1/1	9
1/2	1/2	1/2	6

Genotype of mother	father	child	frequency
1/2	1/2	2/2	1
1/2	2/2	1/2	7
1/2	2/2	2/2	3
2/2	1/1	1/2	3
2/2	1/2	1/2	5
2/2	1/2	2/2	1
2/2	2/2	2/2	4

Calculate the TDT-statistic for these data.

TDT for a multi-allelic marker

Marker locus $\{a_1, \ldots a_k\}$ with $k \geq 2$, n nuclear families

	Non-transmitted allele				
Transmitted allele	a_1	$\cdots$	a_j	$\cdots$	a_k
a_1	n_{11}	$\cdots$	n_{1j}	$\cdots$	n_{1k}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_i	n_{i1}	$\cdots$	n_{ij}	$\cdots$	n_{ik}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
a_k	n_{k1}	$\cdots$	n_{kj}	$\cdots$	n_{kk}

n_{ij}: number of parents with genotype a_i/a_j who transmit allele a_i

In case that there is no association between the alleles at the marker locus and the disease, n_{ij} and n_{ji} should be similar numbers.

TDT for a multi-allelic marker

Test statistic: $$\mathrm{TDT_{SE}} = \frac{k-1}{k} \sum_{i=1}^{k} \frac{(n_{i.} - n_{.i})^2}{n_{i.} + n_{.i} - 2n_{ii}}$$

with $n_{i.} = \sum_{j=1}^{k} n_{ij}$ and $n_{.i} = \sum_{j=1}^{k} n_{ji}$

Null distribution: In case of no association, the distribution of the $\mathrm{TDT_{SE}}$-statistic can be approximated by a χ^2_{k-1} distribution.

Exercise:

Show that $\mathrm{TDT_{SE}}$=TDT for a diallelic marker.

TDT for a multi-allelic marker

Test statistic: $$\mathrm{TDT}_{\max} = \max_{1 \le i \le k} \frac{(n_{i.} - n_{.i})^2}{n_{i.} + n_{.i} - 2n_{ii}} = \max_{1 \le i \le k} \mathrm{TDT}^{(i)} ,$$

where $\mathrm{TDT}^{(i)}$ is the TDT-statistic for a diallelic marker applied to $a_i, \bar{a}_i = \{a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_k\}$

Null distribution: The (approximate) distribution of the $\mathrm{TDT}_{\max}$ statistic is not known, but can be obtained by simulation.

FBA studies vs. CC studies: Pros and Cons

Cons:

- FBA studies are logistically more demanding. Especially in case of a disease with a late age of onset, it is difficult/impossible to sample a sufficiently large number of affected children with both parents available.
- FBA studies are more expensive in regard to genotyping costs: Three individuals have to be genotyped for obtaining one case and one "control".

Pros:

- FBA studies circumvent the problem of population stratification.
- Often, a family sample adequate for FBA analysis is already available from a previous linkage analysis.

Families with more than a single affected child

Problem:

Some or all of the families in the sample may contain two or more affected children. How to deal with these families?

Possibilities:

1. Randomly select a single child from each family and discard additional affected children in the family.

   Drawback: Result of the analysis cannot be verified. Further, discarding additional affected children may waste information.

Families with more than a single affected child

2. Count, for both parents and *all* affected children, the pair of alleles transmitted/non-transmitted by the parent to each child.

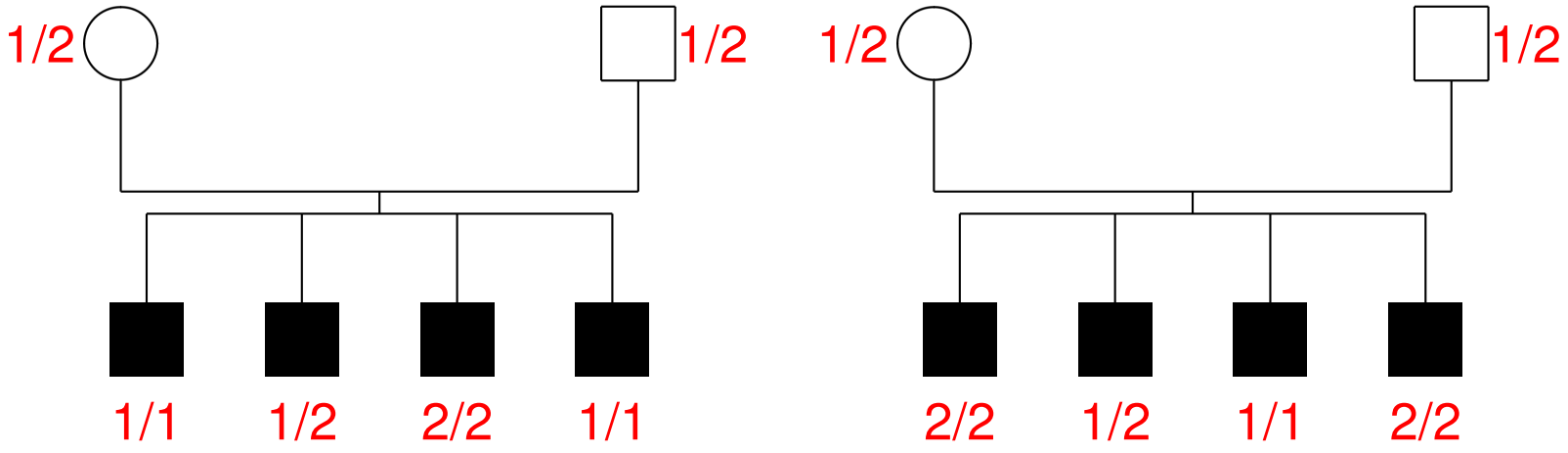
   Drawback: Transmissions by one parent to different children are not independent if there is linkage (but no association) between marker and disease locus. However, the independence of parental transmissions is an essential assumption for approximating the null distribution of the TDT statistic by a χ^2_1 distribution.

   Solution: Obtain the null distribution of the TDT by an appropriate simulation procedure.

Families with more than a single affected child

For $1 \leq i \leq n$, let g_i denote the collection of observed marker genotypes of all individuals (parents and children) in family i and let $\bar{g}_i$ denote the collection of genotypes that is obtained from g_i by replacing the observed genotypes of *all* children with the two non-transmitted parental alleles.

Example:



g_i $\bar{g}_i$

Families with more than a single affected child

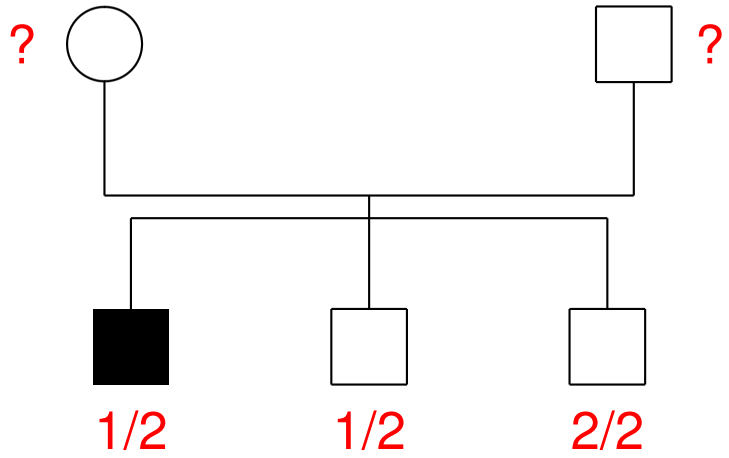
0. Count, for both parents and *all* affected children, the pair of alleles transmitted/non-transmitted by the parent to each child. Sum up these counts over the families and calculate the TDT statistic (Notation: TDT_0).
1. Perform t simulation replicates. In the j-th simulation replicate ($1 \leq j \leq t$), a sample of n families is constructed such that the collection of the observed marker genotypes in family i is either g_i (with probability 1/2) or $\overline{g}_i$ (with probability 1/2). Let TDT_j denote the TDT statistic for the j-th simulation replicate.
2. The P value assigned to TDT_0 is the fraction of simulation replicates resulting in a test statistic greater or equal to TDT_0, i.e.,

$$P = \frac{|\{j : \text{TDT}_j \geq \text{TDT}_0\}|}{t}$$

FBA studies without parents

In FBA studies, missing parents can be compensated by unaffected offspring.

Example:



Let m_A (m_U) be the mean number of "1" alleles among the affected (unaffected) children.

Example: $m_A = 0.5$, $m_U = 0.25$

FBA studies without parents: SDT

Marker: Diallelic locus $\{1, 2\}$

Sample: Sibships with (at least) one affected and (at least) one unaffected child

Method: Let d_+ be the number of sibships for which $m_A > m_U$ and d_- be the number of sibships for which $m_A < m_U$

Test statistic:

$$\text{SDT} = \frac{(d_+ - d_-)^2}{d_+ + d_-}$$

Null distribution: In case of no association, the distribution of the SDT-statistic can be approximated by a χ^2_1 distribution

Name: Sibship disequilibrium test (SDT)