European Journal of Human Genetics (2001) 9, 301 – 306


www.nature.com/ejhg

ARTICLE

The family based association test method: strategies for studying general genotype – phenotype associations

Steve Horvath*[,1,2,3], Xin Xu[4] and Nan M Laird[5]

[1]*Institute for Medical Statistics & Genetic Epidemiology, University of Bonn, Bonn, Germany;* [2]*Department of Human Genetics, University of California at Los Angeles, Los Angeles, California, USA;* [3]*Department of Biostatistics, University of California at Los Angeles, Los Angeles, California, USA;* [4]*Program for Population Genetics, Harvard School of Public Health, Boston, Masachusetts, USA;* [5]*Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA*

With possibly incomplete nuclear families, the family based association test (FBAT) method allows one to evaluate any test statistic that can be expressed as the sum of products (covariance) between an arbitrary function of an offspring's genotype with an arbitrary function of the offspring's phenotype. We derive expressions needed to calculate the mean and variance of these test statistics under the null hypothesis of no linkage. To give some guidance on using the FBAT method, we present three simple data analysis strategies for different phenotypes: dichotomous (affection status), quantitative and censored (eg, the age of onset). We illustrate the approach by applying it to candidate gene data of the NIMH Alzheimer Disease Initiative. We show that the RC-TDT is equivalent to a special case of the FBAT method. This result allows us to generalise the RC-TDT to dominant, recessive and multi-allelic marker codings. Simulations compare the resulting FBAT tests to the RC-TDT and the S-TDT. The FBAT software is freely available. *European Journal of Human Genetics* (2001) **9,** 301 – 306.

Keywords: TDT; RC-TDT; linkage disequilibrium; missing genotypes; censored traits; FBAT

Introduction

The transmission/disequilibrium test (TDT),[1,2] has been generalised by altering the TDT statistic, eg, to include multiple alleles, different mode of inheritance models or quantitative phenotypes. The study designs for the TDT have also been altered; for example, by collecting unaffected offspring to define family-based association tests that can deal with missing parental genotype information. Recently, Rabinowitz and Laird[3] introduced a general approach to family-based association tests that allows evaluation of many of the generalised TDT statistics – even when there are missing parental genotypes – by conceptualising transmission/disequilibrium type tests as a two stage procedure. In the first stage, one specifies a statistic (of a general form described below) to test for association between the trait locus and the marker locus. In the second stage, one computes the distribution of the genotyped marker data by treating the offspring genotype data as random, and conditioning on parts of the data as is described below. The first stage allows flexibility in modeling – we model test statistics which can be expressed as the sum of products between an arbitrary function of a subject's genotype with an arbitrary function of the subject's trait. The second stage ensures correct false positive rates regardless of population admixture, genetic model misspecification or the ascertainment strategy. In this paper we focus on the null hypothesis of no linkage between the marker and a trait influencing locus.

*Correspondence: S Horvath, UCLA Department of Human Genetics, Gonda Neuroscience & Genetics Research Center, 695 Charles E Young Dr. South, Suite 6506, Los Angeles, CA 90095-7088, USA.
Tel: +1 310 825 9299; Fax: +1 810 277 7453;
E-mail: shorvath@mednet.ucla.edu
Received 6 December 2000; accepted 18 December 2000

Methods

The null distribution for complete parental data

In nuclear families with complete parental genotypes, Rabinowitz and Laird[3] define the null distribution of the offspring marker data by conditioning on the observed traits in all family members and the parental marker genotypes.[4,5]

Let $P_i(g_j=g)$ be the conditional probability under the null hypothesis that the genotype of the *j*th offspring in the *i*th family equals g (conditional on the observed traits in all family members, and conditional on the parental genotypes). For example, in case of a simplex family (labelled by 3) with both parental genotypes known and one affected child (labelled by 1), the conditional distribution of an $AB \times AB$ mating is given by the usual Mendelian transmission probabilities: $P_3(g_1=AA)=1/4$, $P_3(g_1=AB)=1/2$, $P_3(g_1=BB)=1/4$. Let $P_i(g_j=g\ g_k=\tilde{g})$ be the conditional probability under the null hypothesis that in family *i*, the *j*th child has genotype g and the *k*th ($j \neq k$ child) has genotype $\tilde{g}$. We simplify our notation by dropping indices, eg, $P(g\,\tilde{g})=P_i(g_j=g\ g_k=\tilde{g})$. When both parents are known, one can easily show that $P(g\,\tilde{g})=P(g)P(\tilde{g})$ under the null hypothesis of no linkage.

The general offspring genotype distribution

In nuclear families with incomplete parental genotypes, Rabinowitz and Laird[3] define the offspring genotype null distribution by conditioning on the observed traits, the partially observed parental genotypes and on the offspring genotype configuration (which is defined as the set of offspring genotypes). Technically speaking, the partially observed parental genotypes and the offspring genotype configuration are sufficient statistics for the missing parental genotypes. In the technical report accompanying the FBAT software, we present explicit formulas for these conditional probabilities.

The general FBAT statistic

Recall that we assume there are N nuclear families indexed by *i*, each having n_i offspring, indexed by $j=1 \ldots, n_i$. Let X_{ij} denote some function of the *ij*th offspring's marker genotype and let T_{ij} denote some function of the *ij* offspring phenotype. The standard choice which underlies tests such as the TDT,[1] the S-TDT,[6] and the RC-TDT,[7] is to let X_{ij} count the number of *A* alleles (possible values=0,1,2), and let $T_{ij}=1$ denote an affected and $T_{ij}=0$ denote an unaffected or phenotype unknown individual. For the *i*th family, the FBAT method determines the distribution of the following linear combination of offspring genotypes and phenotypes:

$$S_i = \sum_{j=1}^{n_i} X_{ij} T_{ij} \qquad (1)$$

One can easily verify that for the standard choice, S_i counts the number of *A* alleles in the affected offspring.

Let us briefly describe different ways of coding the marker genotypes: a recessive coding is given by setting $X_{ij}=1$ if the *ij*th individual has genotype *AA* and zero otherwise. We have implemented additive, dominant and recessive bi-allelic and multi-allelic marker codings which are described in[8] and in the FBAT manual. In case of a multi-allelic coding, X_{ij} is a vector, eg, if there are three alleles *A*, *B*, *C* then the multi-allelic, additive marker coding is defined by setting $X_{ij}=(2, 0, 0)$ for an individual with genotype *AA*, $X_{ij}=(1, 1, 0)$ for genotype *AB*, $X_{ij}=(0, 0, 2)$ for genotype *CC*, etc. In the definition of our data analysis strategies (see below), we choose additive bi-allelic or multi-allelic marker codings since this choice tends to perform well (see the discussion surrounding Table 2).

Once the distributions of the S_i are known, one can define the score $U=\sum_i\{S_i-E(S_i)\}$, the variance $V=\sum_i V(S_i)$ and define a Mantel-Haenszel type test statistic as $\chi^2=U^TV^-U$. Under the null hypothesis of no linkage between the marker and any gene that influences the trait, χ^2 has a (central) χ^2 distribution where the number of degrees of freedom equals the rank of V. It remains to compute the mean and the variance of S_i. Since we condition on the trait, T_{ij} is fixed. Thus the mean is given by $E(S_i)=\sum_j E(X_{ij})T_{ij}$ where $E(X_{ij})=\sum_g X(g)P_i(g_j=g)=\sum_g X(g)P(g)$; $\sum_g$ denotes the sum over all offspring genotypes that are possible in the *i*th family.

In the technical report accompanying the software we show that the variance of S_i is given by

$$\begin{aligned} Var(S_i) =& \Big(\sum_j T_{ij}\Big)^2 \sum_g \sum_{\tilde{g}} \{X(g)(P(g\tilde{g}) \quad P(g)P(\tilde{g}))X(\tilde{g})^T\} \\ &+ \sum_j T_{ij}^2 \Big(\sum_g X(g)X(g)^T P(g) \quad \sum_g \sum_{\tilde{g}} X(g)P(g\tilde{g})X(\tilde{g})^T\Big). \end{aligned} \qquad (2)$$

Different phenotype codings

We will now address different ways of recoding the phenotype. Let Y_{ij} denote the original phenotype of the *ij*th offspring. For example, when Y_{ij} is a quantitative trait, a quantitative trait TDT[9] can be defined by using the mean centred trait $T_{ij}=Y_{ij}-\mu$ where $\mu=\bar{Y}$ is a (weighted) sample mean of the Y_{ij}.

When dealing with a recoded trait of the form $T_{ij}=Y_{ij}-\mu$, there are different ways of determining the offset μ. Sometimes μ has an intuitive interpretation and can be estimated from the data (see strategies Q1 and C1) or be determined on the basis of prior knowledge about the disease. But if no such prior knowledge exists, μ can be determined as the value that minimises the total null variance $Var(S,\mu)=\sum_i Var(S_i)$; several strategies are possible for dealing with nuisance parameters in family based association test statistics.[10] For strategy 2 presented below, we chose to determine the nuisance parameter by minimising the null variance since this method is fairly standard and straightforward. The FBAT software can determine the minimum value of any nuisance parameter which is a linear offset of the trait value ($T_{ij}=Y_{ij}-\mu$) or which is part of a 'time to onset trait coding' defined by ($T_{ij}=Y_{ij}-\delta_{ij}\mu$) where $\delta_{ij}=1$ for affecteds and 0 for unaffecteds (see below).

Simple strategies for data analysis

The FBAT method allows great flexibility in choosing a test statistic. To give some guidance on how to use it, we present simple strategies (see Table 1) along which a data analysis

Table 1 Different strategies of the FBAT method applied to the NIMH Alzheimer disease data. The bi-allelic tests focus on allele 4 of APOE and allele 2 of A2M

Gene	*Strategy*	*Trait*	*Mode*[b]	*Offset*	*No fams*[a]	χ^2	*df*	P *value*
APOE4	D1	dichot	bi	n	71	13.1	1	0.000298
	D2		bi	y	71	17.0	1	0.000037
	D3		multi	n	71	23.5	2	0.000008
	Q1	quant	bi	n	72	23.1	1	0.000002
	Q2		bi	y	72	23.2	1	0.000001
	Q3		multi	n	72	30.6	2	0.0000002
	C1	cens	bi	n	72	18.2	1	0.000020
	C2		bi	y	72	8.6	1	0.003450
	C3		multi	n	72	26.8	2	0.000001
A2M2	D1,D3	dichot	bi	n	47	7.4	1	0.006455
	D2		bi	y	47	7.9	1	0.004885
	Q1,Q3	quant	bi	n	47	7.8	1	0.005223
	Q2		bi	y	47	7.8	1	0.005143
	C1,C3	cens	bi	n	47	7.8	1	0.005233
	C2		bi	y	47	8.6	1	0.538295

[a] Number of informative families.[b] Since A2M is a bi-allelic marker, the multi-allelic test equals the bi-allelic test.

may proceed. We look at three different phenotypes: dichotomous (affection status), quantitative and censored (eg, the age of onset). Within each phenotype category, we present three strategies: the first strategy corresponds to a bi-allelic, additive marker coding; the second strategy has the same marker coding as the first strategy but its phenotype coding involves an offset which is determined by minimising the null variance; the third strategy has the same trait coding as the first strategy but its marker is multi-allelic. So far our method cannot fit models involving an offset and a multi-allelic marker coding. The three strategies involve an additive marker coding but the FBAT software allows one to fit dominant and recessive marker coding as well (see the section on the RC-TDT).

Dealing with dichotomous traits

Strategy D1 (D for dichotomous) is the standard, TDT-type way of dealing with dichotomous traits: it assigns T_{ij}=1 to affecteds and T_{ij}=0 to unaffecteds or unknowns. Further, it uses a bi-allelic (additive) marker coding. We use this phenotype and marker coding since it forms the basis for the TDT and many other tests, eg, the S-TDT, RC-TDT, etc. Strategy D2 uses the same marker coding as strategy D1 but assigns trait values $T_{ij}=1-\mu$ to affecteds, $T_{ij}=-\mu$ to unaffecteds, and T_{ij}=0 to unknowns. Setting μ=0 leads to strategy D1 and is appropriate for rare diseases. With more common diseases, including the unaffecteds by taking $0<\mu<1$ can increase the power of the test. One way would be to define μ as the proportion of affecteds in the sample (analogous to strategies Q1 and C1) but we define strategy D2 to mean that μ is determined by minimising the null variance, which is a standard option in the FBAT software. The usefulness of strategy 2 depends on the structure (and ascertainment) of the data. Note that if only affected offspring have been genotyped, strategy 2 leads to a degenerate test statistic: setting μ=1 leads to a minimum variance of 0. Strategy D3 is the extension of strategy D1 to a multi-allelic, additive marker coding, ie, it effectively extends the RC-TDT to a multi-allelic marker.

Dealing with a quantitative trait

Strategy Q1 (Q for quantitative) codes the traits as $T_{ij}=Y_{ij}-\bar{Y}$ where $\bar{Y}$ denotes the sample mean. Further, it uses a bi-allelic (additive) marker coding. Thus, strategy Q1 generalizes a quantitative trait TDT[9] to the case of missing parental genotypes. Strategy Q2 uses the same marker coding as strategy Q1 but assigns as trait values $T_{ij}=Y_{ij}-\mu$ where the offset μ is determined by minimising the null variance. In general, the minimising μ differs from $\bar{Y}$ so that strategies Q1 and Q2 are different. Strategy Q3 extends strategy Q1 to a multi-allelic marker coding, much in the same way as strategy D3 extends strategy D1. For our analysis of the Alzheimer data set, we treat the age of onset as a quantitative trait by using the age of onset of affecteds while unaffecteds are assigned a maximum age of onset (for the Alzheimer data we chose Y_{ij}=100). The assignment of a maximum age of onset is somewhat arbitrary and strategies C1 – C3 improve on it by making proper use of censoring (affection status) information.

Dealing with a censored trait

Strategy C1 (C for censored) uses age of onset information and takes proper account of the censoring information. Let Y_{ij} be the age of onset if the *ij*th individual is affected; else if the *ij*tj offspring is (was) affected, Y_{ij} is the age at ascertainment or the age at death. Strategy C1 uses the following 'time to onset' coding: $T_{ij}=Y_{ij}-\mu$ for affecteds and $T_{ij}=Y_{ij}$ for unaffecteds. In the following we motivate this coding by showing that in this case, equation (1) takes on the same form as the score equation of a proportional hazards model with an exponential baseline hazard function (this approach can also be generalised to the Cox proportional hazards model.[11] Score equations should merely be considered a (valid) device for coming up with test statistics, see the Discussion.

With the exponential distribution, which is shifted by a constant denoted by γ_0, the probability that the disease onsets at age Y_{ij} is given by $F_{ij}(Y_{ij}-\gamma_0)=1-e^{-\lambda_{ij}(Y_{ij}-\gamma_0)}$. With the proportional hazard assumption, the parameter λ_{ij} is a function of the genetic exposure X_{ij}:$\lambda_{ij}=e^{\beta X_{ij}}/\mu$. The log-likelihood contribution of the *i*th family is given by:

$$ln(L_i)=\sum_j\{\delta_{ij}ln(\lambda_{ij}e^{\lambda_{ij}(Y_{ij}-\gamma_0)}) \quad (1-\delta_{ij})\lambda_{ij}(Y_{ij}-\gamma_0)\}. \qquad (3)$$

where δ_{ij}=1 if Y_{ij} is age-of-onset and δ_{ij}=0 if Y_{ij} is censored (unaffected). Since we are testing whether the genetic exposure X_{ij} has an effect, it is natural to derive the score test for testing the null hypothesis β=0: differentiating equation (3) with respect to β, setting β=0, leads to a statistic

of the general form defined in equation (1) with the 'time to onset' trait coding:

$$T_{ij} = (Y_{ij} \quad y_0) \quad \mu\delta_{ij}. \tag{4}$$

For the exponential model, μ is the average age of onset. Strategies C1 – C3 differ by how they determine the nuisance parameter μ. Strategy C1 uses the maximum likelihood estimate from the exponential model: solving S_i=0 leads the maximum likelihood estimate $\mu=\sum_i\sum_j Y_{ij}/\alpha$ where α is the total number of affected offspring. Strategy C2 determines μ by minimising the null variance, which is a standard option in the FBAT software. Strategy C3 generalises strategy C1 to a multi-allelic marker coding.

Applying the simple strategies to NIMH Alzheimer disease data

As an application, we analysed data by the NIMH Alzheimer Disease Genetics Initiative (see the Acknowledgements). For Alzheimer's disease we focused on two genes: the apolipoprotein E (APOE) gene with three alleles, and the alpha-2 macroglobulin gene (A2M) with two alleles. The 4 (E4) allele of APOE has been consistently found to be over-represented in cases of AD.[12] There is also some evidence for an association between the A2M mutation and Alzheimer disease.[13] We restricted our analysis to those data analysed in[13] for which age-of-onset information was available: 110 sibships (408 individuals), no parental genotypes were available. The quantitative trait is defined as the age of onset for affecteds and (arbitrarily) as 100 for unaffecteds.

Table 1 presents the results of our analysis. We observe that for the APOE locus, the multi-allelic analyses are much better than the bi-allelic (univariate) tests. Probably this is due to the fact that other analyses (not shown) indicate that the 2 allele is protective beyond the effect of merely counterbalancing the predisposing effect of the 4 allele. Further, we observe that for bi-allelic test statistics, it can be worthwhile to use an offset parameter, which is estimated by minimising the null variance, eg, for the APOE4 gene this is true when using a dichotomous or a quantitative trait but not for the censored (time to onset) trait. For the A2M2 allele, using an offset parameter leads to (slightly) better results for the dichotomous and quantitative trait, but strategy C2 leads to an insignificant *P* value. For the APOE locus the quantitative trait leads to the most significant finding and the censored trait is more significant than the dichotomous trait. For both loci, strategy C2 performs poorly.

The RC-TDT is a special case of FBAT

We will argue here that the RC-TDT[7] is essentially equivalent to the test described in strategy D1.

Let us briefly review the RC-TDT: the test is based on reconstructing missing parental genotypes and correcting for the biases involved in this reconstruction by conditioning the test statistic on parental genotype reconstructability. When parental information is available, the RC-TDT equals the TDT. When it is not possible to reconstruct missing information, the RC-TDT equals the S-TDT.[6] The original RC-TDT is a bi-allelic test, which focuses on one allele at a time, and uses an additive marker coding. As part of this paper, we have generalised the RC-TDT to different marker codings.

For each family where the offspring genotypes allow the unique reconstruction of the missing parental genotypes, the RC-TDT is identical to the statistic in strategy D1 (contact the first author for a mathematical proof). When the parental genotypes are not reconstructable the two tests can differ, eg, if there is at least one parent missing and the offspring genotype configuration is {*AB*,*BC*}, the RC-TDT (which equals the S-TDT in this case) differs from the test in strategy D1. For practical purposes, the differences between the RC-TDT and the FBAT are negligible, as can be seen from the simulation studies described in the next section.

Simulations involving strategy D1, the RC-TDT and the S-TDT for different marker codings

We used simulations to compare the power of the RC-TDT with the power of the FBAT test of strategy D1 for an additive, recessive and dominant marker coding. The original RC-TDT was only devised for an additive marker coding but we generalised it to dominant and recessive marker codings.

These simulations closely followed the approach taken by Knapp.[7,14] In brief, we studied a dichotomous trait (affecteds and unaffecteds) which was affected by a disease locus that possessed two alleles. The penetrance f_{DD} for the disease-predisposing genotype was varied: f_{DD}=0.2, 0.5, or 0.8. Dominant, additive, and recessive models were simulated; for each model, a disease prevalence K_P of 5% and an attributable fraction of 50% were assumed. The marker with six concomitant alleles (with frequencies 0.4, 0.2, 0.1, 0.1, 0.1, 0.1) was completely linked to the disease locus. The haplotype frequencies were set to yield the relatively large frequency difference of C=0.25 for the first marker allele between randomly selected affected and unaffected individuals. For each genetic model, we simulated 75 families with four children. R=500 replicate samples were generated and we chose a false positive level of α=0.001.

Table 2 lists the simulated power of the test of strategy D1, the RC-TDT and the S-TDT when both parental genotype are missing. Note that the different marker codings can have a striking effect on the power. For example, if the underlying disease model is additive or recessive it is usually best to use the corresponding marker coding. Surprisingly, for the dominant genetic models that we studied, it is also best to use an additive marker coding. We have found a similar effect in additional simulation studies (unpublished data): the mode of inheritance at the disease locus does not necessarily correspond to the optimal marker coding when the strength of the LD (C) is low; for low LD, the additive marker coding often performs best.

Table 2 Simulation study involving strategy D1, the RC-TDT and the S-TDT when both parents are missing. We use bi-allelic (additive, recessive, dominant) marker codings. The power of strategy D1 is reported first while the power of the RC-TDT and S-TDT are in round and square brackets, respectively. Models have a dominant (additive, recessive) mode of inheritance. False positive rate α=0.001

		Marker coding		
Model	f_{DD}	*Additive*	*Recessive*	*Dominant*
dom	0.2	0.30 (.30) [.25]	0.02 (.02) [.03]	0.29 (.27) [.23]
	0.5	0.52 (.50) [.42]	0.06 (.06) [.07]	0.45 (.43) [.35]
	0.8	0.78 (.75) [.69]	0.17 (.17) [.17]	0.65 (.64) [.58]
add	0.2	0.33 (.32) [.27]	0.04 (.04) [.04]	0.23 (.22) [.20]
	0.5	0.43 (.41) [.33]	0.07 (.07) [.05]	0.27 (.25) [.22]
	0.8	0.54 (.52) [.46]	0.07 (.07) [.06]	0.38 (.36) [.31]
rec	0.2	0.71 (.72) [.72]	0.79 (.79) [.77]	0.03 (.03) [.04]
	0.5	0.86 (.86) [.84}	0.89 (.89) [.87]	0.04 (.04) [.03]
	0.8	0.91 (.92) [.90]	0.95 (.95) [.95]	0.08 (.07) [.06]

Let us now compare the test of strategy D1 to the RC-TDT for different marker codings. One can show that the test of strategy D1 equals the RC-TDT (and also the S-TDT) on sib*pair* data when both parental genotypes are missing. When there are four sibs, the difference in power is less than 3%, ie, for practical applications the test are equivalent.

For missing parental data, the S-TDT has been generalised to different marker codings.[15] For the genetic models described above, our simulation studies show that the test of strategy D1 (and the RC-TDT) can be significantly more powerful than this generalised S-TDT.

Discussion

Statistics of the form (1) have been studied by several authors[8,10] in the context of complete parental data. The theoretical reason why statistics of this form are attractive is that they correspond to score statistics of a generalised linear model that uses a canonical link function to relate the mean phenotype to the marker alleles.[16] We have described elsewhere how to derive different FBAT statistics as score statistics;[17] this derivation should be considered merely a device for obtaining a test statistic because it requires making assumptions which may not be true; for example the score tests assume that the phenotypes of the siblings are independent conditionally on genotype. Fortunately, the false positive rate of the test depends only on how we construct the genotype distribution for the offspring and not how we motivate the test statistic since one conditions on the trait values. The rationale behind the score test approach is that it offers the possibility of constructing powerful tests statistics when the assumed distribution for phenotype given genotype is at least approximately true.

We have shown that the test of strategy D1, which counts the number of a specific allele among the affecteds, is practically identical to the RC-TDT.[7] This result has three benefits: First, it shows how to extend the RC-TDT to more general, possibly multi-allelic, marker codings and to different phenotypes. Second, by showing the equivalence of two seemingly different approaches we bring some light into the thicket of the many family-based association methods that deal with missing parental information. Last but not least, we can transfer all the nice intuition surrounding the RC-TDT to interpret the relatively abstract algorithm introduced by Rabinowitz and Laird;[3] actually, their approach can roughly be interpreted as generalising the RC-TDT to the case when parental genotypes are not uniquely reconstructable.

Software availability

We have implemented the FBAT method in a program called FBAT. The program and its documentation can be downloaded from our website at: www.biostat.harvard.edu/fbat/default.html.

Acknowledgements

We thank Daniel Rabinowitz, Michael Knapp and Kathryn Lunetta for valuable comments on the manuscript. The Alzheimer Disease Data were collected in three projects that participated in the National Institute of Mental Health (NIMH) Alzheimer Disease Genetics Initiative. From 1991-98, the Principal Investigators and Co-Investigators were: Massachusetts General Hospital, U01 MH46281, MS Albert, and D Blacker; Johns Hopkins University, U01 MH46290, SS Bassett, GA Chase, and MF Folstein; University of Alabama, U01 MH46373, RCP Go, and LE Harrell. The APOE genotypes were generated from biomaterials in the NIMH Alzheimer Disease Genetics Initiative in the laboratory of RE Tanzi, Massachusetts General Hospital. This manuscript was supported by N.I.H. grant MH59532-02 and a BONFOR scholarship (S Horvath) of the University of Bonn, Germany.

References

1 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993; **52**: 502–516.
2 Terwilliger J, Ott J: A haplotype based 'haplotype relative risk' approach to detecting allelic associations. *Hum Hered* 1992; **42**: 337–346.
3 Rabinowitz D, Laird N: A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 2000; **50**: 227–233.
4 Martin ER, Kaplan NL, Weir BS: Tests for linkage and association in nuclear families. *Am J Hum Genet* 1997; **61**: 439–448.
5 Lazzeroni LC, Lange K: A Conditional Inference Framework for Extending the Transmission/Disequilibrium Test. *Hum Hered* 1998; **48**: 67–81.
6 Spielman RS, Ewens WJ: A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am J Hum Genet* 1998; **62**: 450–458.
7 Knapp M: The transmission/disequilibrium test and parental-genotype reconstruction: The reconstruction-combined transmission/disequilibrium test. *Am J Hum Genet* 1999; **64**: 861–870.

8 Schaid DJ: General score tests for association of genetic markers with disease using cases and their parents. *Genet Epidemiol* 1996; **13**: 423 – 449.
9 Rabinowitz D: A transmission disequilibrium test for quantitative trait loci. *Hum Hered* 1997; **47**: 342 – 350.
10 Lunetta KL, Faraone SV, Biedermann J, Laird NM: Family-based tests of association that use unaffected sibs, covariates, and interactions. *Am J Hum Genet* 2000; **66**: 605 – 614.
11 Mokliatchouk O, Blacker D, Rabinowitz D: Association tests for traits with variable age of onset. *Hum Hered* 2001; **51**: 46 – 53.
12 Blacker D, Haines JL, Rodes L *et al*: APOE-4 and age-at-onset of Alzheimer's disease: the NIMH Genetics Initiative. *Neurology* 1997; **48**: 139 – 147.
13 Blacker D, Wilcox MA, Laird NM *et al*: Alpha-2 macroglobulin is genetically associated with Alzheimer's disease. *Nature Genet* 1998; **19**: 357 – 360.
14 Knapp M: Using exact *P* values to compare the power between RC-TDT and S-TDT. *Am J Hum Genet* 1999; **65**: 1208 – 1210.
15 Schaid DJ, Rowland C: Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. *Am J Hum Genet* 1998; **63**: 1492 – 1506.
16 Clayton D: A generalization of the transmission/disequilibrium test for uncertain haplotype transmission. *Am J Hum Genet* 1999; **65**: 1170 – 1177.
17 Laird NM, Horvath S, Xu X: Implementing a unified approach to family based tests of association. *Genet Epidemiol* 2000; **19 (Suppl 1)**: 36 – 42.