INT. J. BIOL. BIOTECH., 3 (2): 381-386, 2006.

TEST 'S' OF HAPLOTYPE CONCORDANCE AND DISCORDANCE

Mahnaz Khattak[1] , Shuhrat Shah[2] and Salahuddin[2]

[1]*Jinnah College for Women, University of Peshawar, NWFP, Pakistan*
[2]*Department of Statistics, University of Peshawar, NWFP, Pakistan*

ABSTRACT

The test presented here is based upon the proband and his/her affected as well as non-affected siblings. Here, the siblings are analyzed in terms of similarities of haplotypes. The proposed 'S' test is used in testing hypothesis that a particular disease has random pattern of inheritance against the alternative hypothesis that it has non-random pattern of inheritance. Probability distribution, mean and variance of the test is derived under the null hypothesis of random inheritance of the disease. It is then applied to data set of varying size sibships having at least one affected and one unaffected sibs to investigate the existence of linkage disequilibrium.

Key-words: 'S' Test, halophytes, sibling,

INTRODUCTION

To demonstrate the heritability of a trait, one way is to provide evidence for its linkage with a known genetic marker that is the two traits tend to be inherited together more often than would be expected by chance alone. This implies that the loci with alleles determining the two traits are located at the same chromosome with recombination frequency less than 1/2 , that is they are linked.

Penrose (1938, 1953) first discussed the problem of detecting linkage between a quantitative trait and a marker locus. His test uses data on independent pairs. Many workers have contributed a lot for the development and generalization of sib-pair methods in different situations.

The affected sib-pairs (AS) methods assume the presence of a tightly linked disease susceptibility locus (DS) in the vicinity of HLA region. Sib pairs from different families are categorized according to whether they share 0, 1 or two haplotypes identical by descent (IBD).

Sib-pair method can be extended to include any relative that share at least one haploytpe in common and this could be generalized to apply to extended pedigrees (Cantor and Roter, 1987, Cantor, 1989)

De Veries et al. (1976) used the criterion:

F = (maximum – minimum number of haplotypes from one parent) + (maximum – minimum number of haplotypes from the other parent).

Green *et al*. (1983) improved the test '$\boldsymbol{F}$' of De Veries et al. by using criterion '$\boldsymbol{N}$' which takes account of the family size distribution as well. It is given by:

N = maximum haplotype frequency from one parent + maximum haplotype frequency from the other parent.

Criterion '$\boldsymbol{N}$' is essentially equivalent to '$\boldsymbol{F}$'. That is '$\boldsymbol{F}$' is: $\boldsymbol{F = 2(N - s).}$

All these tests use information from affected sibs only, but to utilize fully the data available on genotypes of unaffected as well, tests involving unaffected have been suggested for example, Spielman *et al*. (1980), Clerget-Darpoux *et al*. (1980), Rubenstein *et al*. (1981).

Green and Montassar (1988) proposed another test '$\boldsymbol{T}$' based on haplotype discordance. This is defined as: $\boldsymbol{T = n_1 + n_2 - n_3}$, where n_1 is the $\boldsymbol{N}$ -measure for the affected sibs in a sibship, n_2 is the $\boldsymbol{N}$-value for the unaffected sibs in the same sibship, and n_3 is the $\boldsymbol{N}$ -measure of the whole sibship, with $\boldsymbol{N}$ as defined earlier**.** One advantage of the $\boldsymbol{T}$-test and others using unaffecteds, is that sibships with only one affected and only one unaffected can contribute. A recent survey of these tests, their powers, relative merits and demerits, various extensions and generalizations is given by Green and Shah (1993) and Shah and Green (1993, 1994).

A new method for differentiating between groups of patients according to severity of disease proposed by Shah *et al*. (1995) is found to be an effective tool in analyzing data sets with respect to disease severity. Khattak *et al*. (2005) suggested a simpler and easier method to distinguish between recessive and dominant mode of inheritance of HLA associated diseases and also provides measures to estimate probability of the disease under consideration in the population.

The Proposed 'S' Test:

On the same pattern, we will present a new test, which is based upon the proband and his/her affected as well as non-affected siblings. Here the siblings are analyzed in terms of similarities of haplotypes. The hypothesis is to test

that the disease has random pattern of inheritance against the alternative hypothesis that it has non-random pattern of inheritance.

The new test 'S' is modification of test M of haplotype concordance in which affected as well as unaffected sibs in a sibship are considered. The test is applicable only when there is at least one unaffected sib in a sibship. Moreover in case of all affected sibs in a sibship, test S is essentially the same as test M.

The test S is stated as:

S = [Sum of haplotypes from both parents in the affected siblings - the number of distinct haplotypes in the affected siblings] – [Sum of haplotypes from both parents in the unaffected siblings - the number of distinct haplotypes in the unaffected siblings]

That is, $S = (2m - k_1) - (2r - k_2)$, where m is the number of affected siblings and r is the number of unaffected siblings. k_1 and k_2 takes values 2, 3, or 4 as the siblings may share 2, 3 or 4 different genes from two parents having a, b, c and d genes in affected and non-affected sibs. We assume here that the two parents are heterozygous.

If we proceed with m= 2 and r = 1, then the possible values of *S* will be calculated as in the following Table(1).

Table 1. Possible combinations of haplotypes in affected and unaffected sibs with their S scores.

m = 2	r=1	S_{21}	m = 3	r = 2	S_{31}	S_{32}
ac ac	Ac	2	ac ac ac	ac ac	4	2
ac ad	Ac	1	ac ac ad	ac ac	3	1
ac bc	Ac	1	ac ac bc	ac ac	3	1
ac bd	Ac	0	ac ac bd	ac ac	2	0
ac ac	Ad	2	ac ac ac	ac ad	4	3
ac ad	Ad	1	ac ac ad	ac ad	3	2
ac bc	Ad	1	ac ac bc	ac ad	3	2
ac bd	Ad	0	ac ac bd	ac ad	2	1
ac ac	Bc	2	ac ac ac	ac bc	4	3
ac ad	Bc	1	ac ac ad	ac bc	3	2
ac bc	Bc	1	ac ac bc	ac bc	3	2
ac bd	Bc	0	ac ac bd	ac bc	2	1
ac ac	Bd	2	ac ac ac	ac bd	4	3
ac ad	Bd	1	ac ac ad	ac bd	3	2
ac bc	Bd	1	ac ac bc	ac bd	3	2
ac bd	Bd	0	ac ac bd	ac bd	2	1

m = 4				r=1	r = 2		r = 3			S_{41}	S_{42}	S_{43}
ac	Ac	ac	ac	ac	ac	ac	ac	ac	ac	6	4	2
ac	Ac	ac	ad	ac	ac	ac	ac	ac	ac	5	3	1
ac	ac	ac	bc	ac	ac	ac	ac	ac	ac	5	3	1
ac	ac	ac	bd	ac	ac	ac	ac	ac	ac	4	2	0
ac	ac	ac	ac	ad	ac	ad	ac	ac	ad	6	5	3
ac	ac	ac	ad	ad	ac	ad	ac	ac	ad	5	4	2
ac	ac	ac	bc	ad	ac	ad	ac	ac	ad	5	4	2
ac	ac	ac	bd	ad	ac	ad	ac	ac	ad	4	3	1
ac	ac	ac	ac	bc	ac	bc	ac	ac	bc	6	5	3
ac	ac	ac	ad	bc	ac	bc	ac	ac	bc	5	4	2

Table continued……

ac	ac	ac	bc	bc	ac	bc	ac	ac	bc	5	4	2
ac	ac	ac	bd	bc	ac	bc	ac	ac	bc	4	3	1
ac	ac	ac	ac	bd	ac	bd	ac	ac	bd	6	6	4
ac	ac	ac	ad	bd	ac	bd	ac	ac	bd	5	5	3
ac	ac	ac	bc	bd	ac	bd	ac	ac	bd	5	5	3
ac	ac	ac	bd	bd	ac	bd	ac	ac	bd	4	4	2

S_{mr} indicates the S- score for 'm' number of affected and 'r' number of unaffected sibs in a sibship.
These are some possible combinations for variable values of m and r (m >r) and their respective S values where $S = (2m - k_1) - (2_r - k_2) = 2(m - r) - (k_1 - k_2)$.
Now, for k_1 and k_2 taking values only 2,3 or 4, the variable 'S' gets only the following five possible values.

2 (m-r)-2 when $k_1 = 4$, $k_2 = 2$
2 (m-r) -1 when $k_1 = 4$, $k_2 = 3$ or $k_1 = 3$, $k_2 = 2$
2 (m-r) when $k_1 = k_2 = 2$ or 3 or 4
2 (m-r) +1 when $k_1 = 2$, $k_2 = 3$ or $k_1 = 3$, $k_2 = 4$
2 (m-r) +2 when $k_1 = 2$, $k_2 = 4$

Now we can present this new variable *S* along with their probabilities in the following Table (2).

Table 2. Probability distribution of 'S' for m+r size shibships.

S	P(S = s)
2 (m-r)-2	$(1-2^{-m+1})^2 (2^{-r+1})^2$
2 (m-r)-1	$(1-2^{-m+1})^2 \ 2 \ (2^{-r+1}) (1-2^{-r+1}) + 2 (2^{-m+1}) (1-2^{-m+1}) (2^{-r+1})$
2 (m-r)	$(2^{-m+1})^2 (2^{-r+1})^2 + 4 (4^{-m+1}) (2^{-r+1}) \ (1-2^{-m+1}) (1-2^{-r+1}) + (1-2^{-m+1})^2 (1-2^{-r+1})^2$
2 (m-r)+1	$2(2^{-m+1})^2 (2^{-r+1}) (1-2^{-r+1}) + 2 (2^{-m+1}) (1- 2^{-m+1}) (1-2^{-r-+1})^2$
2 (m-r)+2	$(2^{-m+1)})^2 (1-2^{-r+1})^2$

The sum of probabilities is equal to one, hence it shows that it is a complete probability distribution, and we can derive its mean and variance easily.

Derivation of mean and variance of 'S' under H_0

E (S) $= [2(m-r)-2] ((1-2^{-m+1})^2(2^{-r+1})^2 + [2(m-r)-1] [2(1-2^{-m+1})^2 (2^{-r+1}) (1-2^{r+1})+ 2(2^{-m+1}) (1-2^{-m+1}) (2^{r+1})^2]+ [2(m-r)] [(2^{-m+1})^2 \ [(2^{-r+1})^2 + 4 (2^{-m+1}) (2^{-r+1}) (1-2^{-m+1}) (1-2^{-r+1}) + (1-2^{-m+1})^2 (1-2^{-r+1})^2]+ [2(m-r)+1] [2(2^{-m+1})^2 (2^{-r+1}) (1-2^{-r+1}) + 2 (2^{-m+1}) (1- 2^{-m+1}) (1-2^{-r-+1})^2]+ [2 (m-r)+2] [(2^{-m+1)})^2 (1-2^{-r+1})^2]$

$= 2(m-r^{)} +2^{-m+1}-2^{-r+1}$

$= 2 [(m-r) + 2^{-m}-2^{-r}]$

Similarly for variance of S

Var (S) = E $(S^2) - [E(S)]^2$ where

E $(S^2) = [(2(m-r)-2]^2 (1-2^{-m+1})^2 (2^{-r+1})^2 + [(2(m-r) -1]^2 [(1-2^{-m+1})^2 2(2^{-r+1}) (1-2^{-r+1})+ 2 (2^{-m+1}) (1-2^{-m+1}) (2^{-r+1})] +[2(m-r)]^2 [(2^{-m+1})^2 (2^{-r+1})^2 + 4 (4^{-m+1}) (2^{-r+1}) \ (1-2^{-m+1}) (1-2^{-r+1})+ (1-2^{-m+1})^2 (1-2^{-r+1})^2 + [2(m-r)+1]^2 2(2^{-m+1})^2 (2^{-r+1}) (1-2^{-r+1}) + 2 (2^{-m+1}) (1- 2^{-m+1}) (1-2^{-r-+1})^2+ [2(m-r)+2]^2 [(2^{-m+1)})^2 (1-2^{-r+1})^2]$

Var (S) = E $(S^2) - [E(S)]^2$

$= 2[2^{-m+1} (1-2^{-m+1}) + 2^{-r+1} (1-2^{-r+1})]$

Thus mean and variance of test S are $2[(m-r) + 2^{-m}-2^{-r}]$ and $2[2^{-m+1}(1-2^{-m+1}) + 2^{-r+1}(1-2^{-r+1})]$, respectively.
Now for any number of affected (m) and non-affected sibs (r), in a sibship of size m + r, we can find the expected mean and variance of test S.
We can present different number of affected and non-affected siblings using the test 'S' along with their probabilities and expected means and variances in a tabular form in table 6.4. We will assume that the parents are heterozygous i.e. they have no common haplotypes and the number of affected sibs is greater than the number of unaffected sibs i.e. $m \geq r$.

Table 3. Probability distribution of S for m+r size sibships along with their means and variances.

S_{21} m =2, r=1	P (S_{21})	S_{31} m=3, r =1	P (S_{31})	S_{32} m=3, r=2	P (S_{32})	S_{41} m=4, r=1	P(S_{41})
0	¼	2	9/16	0	9/64	4	49/64
1	2/4	3	6/16	1	24/64	5	14/64
2	¼	4	1/6	2	22/64	6	1/64
				3	8/64		
				4	1/64		
Mean	1		5/2		3/2		17/4
Variance	1/2		3/8		7/8		7/32

S_{42}	P (S_{42})	S_{43}	P (S_{43})	S_{mr}	P (S_{mr})
2	49/256	0	49/1024	2(m-r)-2	$(1-2^{-m+1})^2 (2^{-r+1})^2$
3	112/256	1	308/1024	2(m-r)-1	$(1-2^{-m+1})^2 2(2^{-r+1}) (1-2^{-r+1})+ + 2 (2^{-m+1})$ $(1-2^{-m+1}) (2^{-r+1})$
4	78/256	2	526/1024	2(m-r)	$(2^{-m+1})^2 (2^{-r+1})^2 + 4 (4^{-m+1}) (2^{-r+1}) (1-2^{-m+1}) (1-2^{-r+1})+$ $(1-2^{-m+1})^2 (1-2^{-r+1})^2$
5	16/256	3	132/1024	2(m-r)+1	$(1-2^{-m+1})^2 (1-2^{-m+1})^2 + 4(2^{m+1}) (1-2^{-m+1}) (2^{-v+1}) (1-2^{-r}) + (2^{-m+1})^2 (2^{-r+1})^2$
6	1/256	4	9/1024	2(m-r)+2	$(2^{-m+1})^2 (1-2^{-r+1})^2$
Mean	13/4		7/4		$2[(m-r)+2^{-m}-2^{-r}]$
Variance	23/32		19/32		$2[2^{-m+1}(1-2^{-m+1})+ 2^{-r+1}(1-2^{-r+1})]$

Hence it reveals some sort of linkage which is perhaps sufficient to suggest that the possibility of association between HLA and disease genes will be worth exploring when further samples are taken and of course the sample size be enlarged and the data be well randomized.

Application to Data Sets:

To illustrate the application of 'S', we analyze data relating to rheumatoid arthritis which consist of 22 families with at least one member affected by the disease and was collected by Cud-worth and Woodrow (1975) found in the literature. The families were typed for HLA-A, B, Cw and DR alleles by the hematology department. The analysis of the rest of the families is shown in Table (4).

CONCLUSION

The test gives expected means and variance for any number of affected and non-affected siblings. If the number of unaffected is greater than the number of affected, that is, if for example we have S_{13} instead of S_{31} then the variance will remain same but mean will have same value with negative sign. Further, for equal number of affected and non-affected siblings, that is, m = r, mean will always be zero and variance will be $4[2^{-m+1}(1-2^{-m+1})]$. The new test 'S' is simpler and easily applicable then the already established test for disease association with genes and their non-random inheritance.
The new test not only takes haplotype concordance among the affected siblings but it also considers haplotype discordance in the whole sibship. Usually the information given by non- affected siblings of diseased person were ignored in past. The new test 'S' provides room for the incorporation of information contained in unaffected siblings. The distribution of this test under H_0 was the only way to calculate its mean and variance, and then apply it to data set for detecting linkage disequilibrium.

Table 4. Cudworth's family data on rheumatoid arthritis and HLA typing.

Family	Affected sibs with haplotypes	Unaffected sibs with haplotypes	S	E (S)	V (S)
# 1	a/b, c/b	c/b	1	1	½
# 2	a/d, a/b	-	1	1	1/2
# 3	a/b, a/b	-	2	1	1/2
# 4	a/b, a/b	c/d	2	1	1/2
# 5	a/b, a/d	-	1	1	1/2
# 6	a/b, c/b	-	1	1	1/2
# 7	a/b, a/d, a/b	c/d	3	5/2	3/8
# 8	c/b, c/b	-	2	1	1/2
# 9	a/b, c/d	-	0	1	1/2
# 10	a/d, a/d	-	2	1	1/2
# 11	a/b, c/b	-	1	1	1/2
# 12	a/b, c/b, c/b	-	3	5/2	3/8
# 13	a/b, a/b	a/d	2	1	1/2
# 14	a/d, c/d	-	1	1	1/2
# 15	a/d	a/d	0	0	1/2
# 16	a/b	a/b	0	0	1/2
# 17	a/d, c/d, a/b	c/d	2	5/2	0
# 18	c/b, c/d, c/d	a/d, c/d	2	3/2	0
# 19	c/b	-	0	0	3/8
# 20	a/b, c/d	-	0	1	7/8
# 21	a/b, a/b	-	2	1	0
# 22	a/b	c/b	0	0	1/2
Total			28	23	9

The test criterion yielded by S is: $T = [\sum S - \sum E(S)] / (\sum V(S))^{1/2}$, that is
T = (28-.05-23)/3=1.667 and p=. 0475 which is just significant at 5% level.

REFERANCES

Cantor, R.M. (1989). A linkage test with identity-by-descent marker data from pairs of affected relatives. *Prog. Clin. Boil. Res*., 329: 111-116.

Cantor, R.M. and J.I. Rotter (1987). Marker concordance in pairs of distant relatives: a new method of linkage analysis for common diseases. *An.J. Hum. Genet.*, 51: A 252.

Clerget-Darpoux, F., C. Bonaiti-Pellie, J. Hors, I. Deschamps and N. Fein-gold (1980). Application of the Lod score method to detection of linkage between HLA and juvenile insulin-dependent diabetes. *Clinical Genetics,* 18: 51-57.

Cudworth, A.G. and J.C. Woodrow (1975). Evidence for HLA linked genes in 'Juvenile diabetese". *Br. Med. J.,* 3: 133-135.

De Veries, R.R.P., A. Lai and R.F.M. Fat (1976). HLA- linked genetic control of host response to Mycobacterian Leprae *Lancet,* 2: 1328-1330.

Green, J.R., and M. Montasser (1988). HLA haplotype discordance. *Biometrics,* 44: 941-950.

Green, J. R and J.C. Woodrow (1977). Sibling Method for detecting HLA- linked genes and disease. *Tissue antigens*, 9: 31-35.

Green, J. R., H.C. Low and J.C. Woodrow (1983). Inference on the inheritance of disease using the repetitions of HLA haplotypes in affected siblings. *Ann. Hum. Genet*., 47: 73-82.

Green, J.R and S. Shah (1993). Power comparison of various sibships tests of association. *An J. Hum. Genet*. 57: 151-158.

Khattak, M., Shah, S. and Salahuddin. (2005). Mode of inheritance of HLA- associated diseases. *Pak. J. Statist*. 21: 203-208.

Rubinstein, P., F. Ginsberg and C. Falk (1981). Genetics of type I diabetes mellitus: a single gene mapping between HLA-B and GLO. *An. J. Hum. Genet.,* 45: 865-882.

Shah, S. and J.R. Green (1993). Testing for a haplotype concordance with incomplete parental data, modified Q and correction for modified N. *An. J. Hum, Genet*., 57: 239-247.

Shah, S. and J.R. Green (1994). The distribution of Q: A powerful sibship test of association. *Ann. Hum. Genet*., 58: 163-173..

Shah, S., M. Khattak and Qazi Ubaidullah. (1995). A test of inheritance for disease severity. *Pak. J. Hist. & Phil.*, 1: 29-36.

Spielman, R. S., L. Baker and C.M. Zmijewski (1980). Gene dosage and susceptibility to insuline dependent diabetes. *Ann. Hum. Genet*., 44: 135-150.

Penrose, L.S. (1938). Genetic linkage in graded human characters. *Ann. Eugen.*, 8: 233-237.

Penrose, L.S. (1953). The general-purpose sib-pair linkage test. *Ann. Eugen.*, 18: 120-124.

(*Accepted for publication April 2006*)