Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) 2 (2): 273-277

jeteas.scholarlinkresearch.org

On Statistical Tests for Transmission of Ailments

[1]Dawodu, G A, [1]Asiribo, O E, [1]Adelakun, A A,
[3]Ozoje, M O and [2]Akinwale, T A

[1]Department of Statistics,
[2]Department of Computer Science,
[3]Department of Animal Breeding and Genetics,
University of Agriculture, Ogun State, Nigeria.

Corresponding Author: Dawodu, G A

Abstract
This work is an attempt to deriving a statistic to convincingly and parametrically attest to the fact that an ailment is transmitted when it really is. Researchers into the transmission of ailments are used to presenting their results in tabular $(2 \times f)$ *format and leave their audience to comment as they wish on their results. We feel that it is important that we present all comments by first showing parametrically or otherwise that the identified ailment is transmitted. Towards this end, we derived a statistic that is chi-squared-typed (*$\chi^2_{(h_i-1)}$*) and illustrated its usefulness with respect to checking whether an ailment can be said to be 'statistically' transmitted. Our test was carried out on two secondary data to demonstrate its effectiveness..*

Keywords: transmission, ailment, tests, chi-square distribution

INTRODUCTION

A transmission test is also known as the transmission/disequilibrium test (TDT) (Spielman et al, 1993; Yang, 2010) is a test of both linkage and association based on the frequency of transmission of alleles from heterozygous parents to affected children. It is also a family-based linkage-disequilibrium test that offers a powerful way to test for linkage between alleles and phenotypes that is either causal (i.e., the marker locus is the disease/trait allele) or due to linkage disequilibrium (Allison, 1997). Allison (1997), developed five TDT-type tests for use with quantitative traits whilst Zhao et al (2000) studied six different ones for the multiple tightly linked markers methodology. Those test statistics and their short descriptions are contained in the following Table 1. Zhao et al (2000) gave the summaries of their results concerning those test statistics in Tables 2, 3, 4 and 5.

Table 1: Showing the six test statistics that were studied in Zhao et al (2000)

Test Statistic	Description
T_s	Studies each marker separately
T_d	Discards ambiguous families
T_h	Assumes that haplotype information is known
T_u	Estimates haplotype frequencies only by use of unambiguous families
T_c	Estimates haplotype frequencies by use of both unambiguous families and ambiguous families, by assigning each compatible haplotype group equal probability for each ambiguous family
T_{ml}	Estimates haplotype frequencies by assuming that parents are a random sample of individuals from a population with Hardy-Weinberg equilibrium

Table 2: Observed Type 1 error for different sample sizes and different sample sizes and different population structures

Observed Type I Error Rates for Different Sample Sizes and Different Population Structures

	Type I Error Rate, for (%)											
	N = 100						N = 200					
r	T_s	T_d	T_h	T_u	T_c	T_{ml}	T_s	T_d	T_h	T_u	T_c	T_{ml}
q = .1:												
2	.5	.35	.35	.35	.3	.35	.35	.35	.3	.3	.3	.35
3	.65	.6	.65	.65	.55	.55	.35	.45	.5	.45	.45	.45
4	.4	.4	.5	.5	.4	.45	.45	.6	.4	.4	.4	.4
q = .2:												
2	.6	.6	.4	.6	.55	.55	.6	.55	.55	.3	.3	.35
3	.6	.65	.5	.65	.65	.6	.7	.4	.55	.7	.65	.65
4	.65	.45	.4	.35	.35	.35	.6	.4	.5	.6	.65	.6
q = .3:												
2	.55	.55	.55	.6	.5	.55	.8	.3	.35	.35	.3	.35
3	.65	.35	.65	.5	.5	.5	.65	.5	.55	.6	.6	.6
4	.35	.4	.55	.35	.3	.35	.45	.35	.35	.35	.4	.35
q = .4:												
2	.6	.5	.5	.4	.45	.4	.7	.55	.6	.65	.6	.6
3	.45	.65	.5	.45	.5	.5	.45	.35	.35	.35	.3	.35
4	.7	.4	.55	.6	.6	.6	.65	.5	.4	.45	.4	.45

Source: (Zhao et al, 2000)

Table 3: Statistical test of genotic linkage between the DRD2 locus and alcoholism, in 77 combined german and Hungarian families

Statistical Tests of Genetic Linkage between the DRD2 Locus and Alcoholism, in 77 Combined German and Hungarian Families

	No. of Transmissions[a]							
	T_d (P = .053)		T_u (P = .018)		T_c (P = .032)		T_{mi} (P = .025)	
Haplotype	Transmitted	Not Transmitted	Transmitted	Not Transmitted	Transmitted	Not Transmitted	Transmitted	Not Transmitted
111	2	4	9.3	6.8	9.5	7.9	9.0	7.4
112	8	17	11.8	26.6	13.2	25.8	13.0	26.0
121	5	5	17.7	11.4	17.4	10.6	18.0	11.0
122	3	6	8.3	10.3	9.0	10.8	9.1	10.6
211	8	2	15.1	8.9	14.8	9.0	15.0	9.2
212	54	39	68.0	55.8	70.6	57.3	71.1	57.3
221	0	0	0	0	2.5	3.5	2.0	3.3
222	10	17	18	28.3	17.2	29.2	17.0	29.1

Source: (Zhao et al, 2000)

[a]Details of four multilocus TDTs (see Zhao et al, 2000) are discussed in the text and also are summarized in table 1. Data are the number of families in which some eight haplotypes are transmitted and not transmitted from the parents to the affected offspring.

Table 4: Statistical test of genotic linkage between the DRD2 locus and alcoholism, in 55 german families

Statistical Tests of Genetic Linkage between the DRD2 Locus and Alcoholism, in 55 German Families

	No. of Transmissions[a]							
	T_d (P = .556)		T_u (P = .270)		T_c (P = .169)		T_{mi} (P = .201)	
Haplotype	Transmitted	Not Transmitted	Transmitted	Not Transmitted	Transmitted	Not Transmitted	Transmitted	Not Transmitted
111	1	4	8.1	6.1	8.1	7.2	7.9	6.9
112	4	6	5.8	11.8	6.0	10.9	5.9	11.0
121	5	4	15.4	8.9	15.4	8.4	15.7	8.6
122	2	4	4.7	7.3	4.4	7.6	4.4	7.4
211	6	2	12.5	8.0	12.3	7.2	12.6	7.3
212	35	29	44.7	41.1	47.5	42.8	47.6	42.8
221	0	0	0	0	2.2	2.3	1.8	2.2
222	9	13	14.9	22.8	14.0	23.8	14.1	23.8

Source: (Zhao et al, 2000)
[a]Data are as described in the footnote to table 3

Table 3: Statistical test of genetic linkage between the DRD2 locus and alcoholism, in 22 Hungarian families

Statistical Tests of Genetic Linkage between the DRD2 Locus and Alcoholism, in 22 Hungarian Families

	No. of Transmissions[a]							
	T_d (P = .038)		T_u (P = .052)		T_c (P = .150)		T_{mi} (P = .063)	
Haplotype	Transmitted	Not Transmitted	Transmitted	Not Transmitted	Transmitted	Not Transmitted	Transmitted	Not Transmitted
111	1	0	1.2	.5	1.5	.8	1.1	.2
112	4	11	5.8	15.0	7.1	14.9	6.1	15.0
121	0	1	2.3	2.4	1.7	2.2	2.3	2.5
122	1	2	3.7	3.1	4.7	3.2	5.6	3.4
211	2	0	2.5	1.0	2.5	2.3	2.5	2.3
212	19	10	23.5	14.5	23.0	14.5	24.4	14.5
221	0	0	0	0	.4	1.1	.1	1.0
222	1	4	3.0	5.4	3.2	5.5	2.0	5.2

Source: (Zhao et al, 2000)

[a]Data are as described in the footnote to table 3.

A perusal of Tables 3 through 5 will acquaint a reader of the fact that there are two 'opposing' columns; the 'transmitted' column and the 'Not transmitted' column, concerning each Haplotype/Tests. The intuitive question, that this paper proposes to answer, with respect to each and all of the highlighted Haplotype/Test cases is "Can we, at any given instance, statistically conclude that with respect to a particular test, there is transmission or non-transmission within a haplotype?" Our answer is "We can". That is, we can derive a statistic that can generally convince us that, concerning a given ailment, there is a transmission from founders to offspring, through a specified haplotype, when indeed there is one. The answer to this question will enable us to be statistically convinced that there is a transmission or mutation, of a specified ailment, from the founder to an offspring in a given pedigree. Zhao et al (2000) is not the only group that will usually present their results in the form contained in tables 3 through 5, other examples can be found in the works of (Clarke et al, 1959; Crow, 1986; Paterson et al, 2009). Our motivation for this work is borne out of the fact we wish to cast out of the minds of readers, of those works, all elements of guessing and doubts that are usually conceived, with respect to transmission, whenever results of works are presented in the discussed form.

MATERIALS AND METHODS

Let us reconsider Tables 3 through 5 from Zhao et al (2000), we can conveniently represent the pairs of 'transmitted' and 'Non-transmitted' in the form of $(2 \times H)$ matrices and assume we have contingency tables. By so doing the column for haplotypes will be

considered as that of "labels" and we proceed with the methodology for analyzing any contingency table[1]. The test statistic for contingency table is χ^2 distributed (Rao, 2007).

RESULTS

Now let c_{ij} be the count appearing in the i[th] row, j[th] column of our table. If h_j heterozygous parents carry allele j, then under the null hypothesis of Mendelian transmission, c_{ij} is binomially distributed with h_j trials and success probability 0.5. That is;

$$c_{ij} \to Bin(h_j, 0.5), \quad E(c_{ij}) = 0.5 * h_j = \frac{h_j}{2} \quad (1), \quad V(c_{ij}) = 0.5 * 0.5 * h_j = \frac{h_j}{4} \quad (2)$$

However this can be standardized to obtain the standardized residual corresponding to c_{ij}, that is;

$$Z_{ij} = \frac{c_{ij} - \frac{h_j}{2}}{\sqrt{\frac{h_j}{4}}} \quad (3)$$

The chi-square statistic; $\chi^2 = \sum_{i}^{2} \sum_{j=1}^{f} Z_{ij}^2$ (4)

is an all-encompassing test for departure from the null hypothesis of Mendelian segregation to sufferers. The maximum standardized residual:

$$Z_{\max} = \max_{i,j} Z_{ij} \quad (5)$$

ought to be sensitive to preferential transmission of a single allele to sufferers since its similitude is done in case-control association problems.

Under the null hypothesis, each relevant transmission event is independent and equally likely to involve either gene of the transmitting parent. Another statistic can be obtained through a constructed table of number of alleles of each type been transmitted to sufferers. If we let T_i be the value of the statistic T for the i[th] randomly generated table from a sample of n such independent tables, then the p-value of the observed statistic T_{obs} can be approximated by the sample proportion

$$\bar{P}r(T \geq T_{obs}) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\{T_i \geq T_{obs}\}} \quad (6)$$

where

$$\delta_{\{T_i \geq T_{obs}\}} = \begin{cases} 1, & T_i \geq T_{obs} \\ 0, & otherwise \end{cases} \quad (7)$$

EXAMPLE 1: Transmission/Disequilibrium Test for Costa Rican *Ataxia-telangiectasia* (AT) Families (Lange, 2004): The following table 6 summarizes marker data on 16 Costa Rican children afflicted with the recessive disease *Ataxia-telangiectasia* (AT). At the chromosome-11 marker D11S1817, 28 of their 32 fully-typed parents are heterozygous

Table 6: transmission pattern/allele to summarize marker data on 16 Costa Rican children afflicted with the recessive disease *Ataxia-telangiectasia* (AT)

Transmission Pattern /Allele	1	3	4	5	7	8	10	11	20	21
Transmitted	3	0	22	0	1	0	0	0	0	2
Not Transmitted	0	4	0	4	3	4	1	1	2	9

Source: (Lange, 2004).

Table 7: On the transmission pattern/allele to summarize marker data on 16 Costa Rican children afflicted with the recessive disease *Ataxia-telangiectasia* (AT). This contains the "expected" situation if our null hypothesis is true.

Transmission Pattern /Allele	1	3	4	5	7	8	10	11	20	21
Transmitted	3/2	2	11	2	2	2	½	1/2	1	11/2
Not Transmitted	3/2	2	11	2	2	2	½	1/2	1	11/2

A thorough perusal of table 6 will strongly suggest that at least, allele 4 of this marker is preferentially transmitted to the sufferers. This suspicion is confirmed by our two statistics as shown in our computations (I). Even in cases in which there are no obvious signs of transmission, this technique help detect if there is transmission. An example is the case of the Framingham Heart study (Paterson et al, 2009). If we check the transmission of the minor allele we shall obtain the result in computation II.

Table 8: An extract of the data on a Heart Study (Paterson et al (2009))

Transmitted	103	58	107	28	27	59	535	500
Not transmitted	178	124	218	79	76	130	394	663

The above data (Table 8) is our observed whilst our table of expected (table 9) is as follows

Table 9: The table of the expected if our null hypothesis is true.

Transmitted	$\frac{281}{2}$	91	$\frac{325}{2}$	$\frac{107}{2}$	$\frac{103}{2}$	$\frac{189}{2}$	$\frac{929}{2}$	$\frac{1163}{2}$
Not transmitted	$\frac{281}{2}$	91	$\frac{325}{2}$	$\frac{107}{2}$	$\frac{103}{2}$	$\frac{189}{2}$	$\frac{929}{2}$	$\frac{1163}{2}$

[1] A frequency table in which a sample is classified according to two different attributes, is called a contingency table.

COMPUTATION I (USING OUR Z_{ij}): The computations goes thus;

$$Z_{1,1} = \frac{3 - \frac{3}{2}}{\sqrt{\frac{3}{4}}} = \sqrt{3}, \quad \left(Z_{1,1}^2 = 3\right)$$

$$Z_{1,2} = \frac{0 - 2}{\sqrt{\frac{4}{4}}} = -2, \quad \left(Z_{1,2}^2 = 4\right)$$

$$Z_{1,3} = \frac{22 - 11}{\sqrt{\frac{22}{4}}} = \sqrt{22}, \quad \left(Z_{1,3}^2 = 22\right)$$

…

$$Z_{1,10} = \frac{2 - \frac{11}{2}}{\sqrt{\frac{11}{4}}} = \frac{-7}{\sqrt{11}}, \quad \left(Z_{1,10}^2 = \frac{49}{11}\right)$$

$$Z_{2,1} = \frac{0 - \frac{3}{2}}{\sqrt{\frac{3}{4}}} = -\sqrt{3}, \quad \left(Z_{2,1}^2 = 3\right)$$

$$Z_{2,2} = \frac{4 - \frac{4}{2}}{\sqrt{\frac{4}{4}}} = 2, \quad \left(Z_{2,2}^2 = 4\right)$$

…

$$Z_{2,10} = \frac{9 - \frac{11}{2}}{\sqrt{\frac{11}{4}}} = \frac{7}{\sqrt{11}}, \quad \left(Z_{2,10}^2 = \frac{49}{11}\right)$$

$consequently;$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{10} Z_{ij}^2 = 92.91, \quad Z_{\max} = \sqrt{22} = 4.69$$

$\chi^2 = 92.91$ is highly significant. As such, we affirm that the ailment, AT, is transmitted from parent to siblings among Costa Ricans (based on the data).

USING THE p-value: Armed with the value of chi-square, $\chi^2 = 92.91$, we search the tabulated values and out of the 10^6 independent trials none of the simulated statistics was as large as our corresponding observe statistics, $\chi^2 = 92.91$. Hence we say, in this case, that the p-value does not exist.

COMPUTATION II (USING OUR Z_{ij}): The computations goes thus;

$$Z_{11} = \frac{103 - \frac{281}{2}}{\sqrt{\frac{281}{4}}} = -4.4741 \qquad \left(Z_{11}^2 = 20.0178\right)$$

$$Z_{12} = \frac{58 - 91}{\sqrt{\frac{182}{4}}} = -0.7253 \qquad \left(Z_{12}^2 = 0.5260\right)$$

…

$$Z_{18} = \frac{500 - \frac{1163}{2}}{\sqrt{\frac{1163}{4}}} = -4.7797 \qquad \left(Z_{18}^2 = 22.8452\right)$$

…

$$Z_{28} = \frac{663 - \frac{1163}{2}}{\sqrt{\frac{1163}{4}}} = 12.7671 \qquad \left(Z_{28}^2 = 163\right)$$

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{8} Z_{ij}^2 = 517.5454 \qquad Z_{\max} = 12.7671$$

$\chi^2 = 517.5454$ is highly significant. As such, we also say that the minor allele involved with this heart study is transmitted (based on the data).

USING THE p-value: Just like in the AT case, the p-value does not exist.

DISCUSSION

To check statistically whether an ailment is transmitted we first consider the marker alleles potentially contributed by heterozygous parents to sampled sufferers which are usually arranged in a $2 \times f$ contingency table, with one row counting parental alleles passed to sufferers and the other counting alleles not passed to the sufferers. The f columns correspond to the f different alleles seen among the parents. By excluding homozygous parents in our scheme because they do not tell us anything about the transmission distortion. Contingency table data, of this sort, are usually explicitly condition on the parental genotypes in order to eliminate ethnic association. After doing this, there is no harm done to the scheme if we merely count the alleles transmitted to siblings and to related individual sufferers scattered around the pedigree as well as the source of pollution. Two inviolable rules to observe, during the survey, are that both parents of the sufferer should be typed and that the marker typing should be done in one part of a family without regard to the outcomes of marker typing in another part of the family.

Now let c_{ij} be the count appearing in the i[th] row, j[th] column of our table. If h_j heterozygous parents carry allele j, then under the null hypothesis of Mendelian transmission, c_{ij} is binomially distributed with

h_j trials and success probability 0.5. The standardized residual corresponding to c_{ij}, as such, is given by;

$$Z_{ij} = \frac{c_{ij} - \frac{h_j}{2}}{\sqrt{\frac{h_j}{4}}} \quad (3)$$

The chi-square statistic $\chi^2 = \sum_{i}^{2} \sum_{j=1}^{f} Z_{ij}^2$ (4)

is an all-encompassing test for departure from the null hypothesis of Mendelian segregation to sufferers. The maximum standardized residual

$$Z_{\max} = \max_{i,j} Z_{ij} \quad (5)$$

ought to be sensitive to preferential transmission of a single allele to sufferers since its similitude is done in case-control association problems.

Under the null hypothesis, each relevant transmission event is independent and equally likely to involve either gene of the transmitting parent. Another statistic can be obtained through a constructed table of number of alleles of each type been transmitted to sufferers. If we let T_i be the value of the statistic T for the i^{th} randomly generated table from a sample of n such independent tables, then the p-value of the observed statistic T_{obs} can be approximated by the sample proportion

$$\hat{P}r\left(T \geq T_{obs}\right) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\{T_i \geq T_{obs}\}} \quad (6)$$

$where$

$$\delta_{\{T_i \geq T_{obs}\}} = \begin{cases} 1, & T_i \geq T_{obs} \\ 0, & otherwise \end{cases} \quad (7)$$

We illustrated these statistics with examples.

REFERENCES

Allison D B (1997) Transmission-Disequilibrium Tests for Quantitative Traits. Am. J. Hum. Genet. 60:676-690.

Clarke C A, Price-Evans D A, McConnell R B and Sheppard P M (1959) Secretion of Blood Group Antigens and Peptic Ulcers. Brit. Med. J 1: pgs 603-607.

Crow J F (1986) Basic Concepts in Population, Quantitative and Ecological Genetics. Freeman, San Francisco.

Lange K (2004). Mathematical and Statistical Methods for Genetic Analysis (Second Edition). Springer-Verlag. New Delhi-110001, India.

Paterson A D, Waggott D, Schillert A, Infante-Rivard C, Bull S B, Yoo Y J and Pinnaduwage D (2009). Transmission-ratio distortion in the Framingham Heart Study.

Rao, K V (2007). Biostatistics: A Manual of Statistical Methods for Use in Health, Nutrition and Anthropology (Second Edition). Jaypee Brothers, Medical Publishers (P) Ltd. New Delhi

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–516.

Yang J (2010) Likelihood Approaches for Detecting Imprinting and Maternal Effects in Family-Based Association Studies. (Ph.D. Dissertation). The Ohio State University.

Zhao H, Zhang S, Merikangas K R, Trixler M, Wildenauer D B, Sun F and Kidd K K (2000). Transmission/Disequilibrium Tests Using Multiple Tightly Linked Markers. Am. J. Hum. Genet. 67:936–946.