Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 1984 – 1991

doi:10.3906/elk-1401-283

Application of kappa statistics in sequential tests for family-based design

Farid RAJABLI*
Department of Electrical and Electronic Engineering, Faculty of Engineering, Turgut Özal University, Ankara, Turkey

Received: 30.01.2014 • **Accepted/Published Online:** 01.07.2014 • **Final Version:** 23.03.2016

Abstract: Family-based designs are commonly used in genetic association studies to locate markers associated with diseases. It is a challenging task to collect a large enough sample size, perform a statistical test, and obtain the desired statistical power. The sequential probability ratio test (SPRT) was introduced to overcome the limited sample size problem. However, the drawback of SPRT is that, for the sake of accuracy, the test leaves many markers in a gray zone meaning "no decision". In this article, we propose a novel approach: a sequential probability ratio test 'plus' (SPRT+) to reduce the number of these gray zone markers. Using simulated data, the results of SPRT+ are compared with the results of SPRT. SPRT+ shows a promising overall performance in identifying highly and moderately associated markers in the correct association region without a loss of accuracy.

Key words: Markers, sequential probability ratio test, simulation study, family-based association study, kappa statistics

1. Introduction

Genome-wide association studies (GWASs) have initiated a new era in the study of human genetics [1]. Rapid advances in genotyping array technology and the completion of the HapMap Project make it a popular and feasible approach in genetic analysis of complex traits [2]. The goal of a GWAS is to understand the genetic basis of common multifactorial diseases and predict unknown disease-associated variants [3]. Association studies are now being routinely performed in the genetic analysis of common diseases.

The major problem in association studies is the presence of systematic differences in allele frequencies due to the population substructure [4]. This problem is essentially existent in unrelated case-control studies and is susceptible to potential false findings. To overcome this serious danger, family-based design has been proposed [5–7]. Family-based association studies are robust for the discovery of spurious associations and avoid problems of population heterogeneity. It demonstrates the cosegregation of a disease with its genetic markers by using related individuals [8]. The haplotype relative risk approach is the simplest and generally most powerful family-based design consisting of trios: an affected individual and his or her two parents [9].

Several different family-based association tests (FBATs) for this 'trio design' have been developed [10]. They compare the distributions of the transmitted allele to the nontransmitted allele from heterozygous parents and then detect a genetic linkage in the presence of an association [11]. To get reasonable results from FBATs, we need a large enough sample size of a completely known nuclear family genotype [10]. However, for late-onset diseases it is difficult and even sometimes impossible to get the genotype information of the parents [12].

*Correspondence: frecebli@turgutozal.edu.tr

Therefore, in family-based design, the most challenging task is how to avoid a type I error with limited sample size.

Recently, Ilk et al. proposed a sequential test that detects genetic associations even with small sample sizes [13]. Sequential tests are different from classical hypothesis tests that were used in FBATs. In classical hypothesis testing the number of samples is fixed at the beginning of the test, whereas in sequential testing it is not fixed and every sample is analyzed after being collected. A sequential test compares the samples collected up to that moment with certain threshold values and puts the markers into one of the following regions: 'associated zone', 'not-associated zone', and 'gray zone'. A sequential test terminates the sample collection when the marker is in the associated or not-associated zone and continues sampling as long as the marker is in the gray zone [14]. As a result, the primary advantage of this test is high accuracy in defining associated zone and not-associated zone markers, whereas the drawback of this test is undefined gray zone markers. The purpose of this article is to introduce a method for the reduction of the number of gray zone markers in sequential testing.

2. Methods

In this section, first, a general description of the sequential probability ratio test (SPRT) is given and then, to overcome this challenge, the newly proposed algorithm is discussed. The methods are applied to simulated data for performance comparison.

3. SPRT

The SPRT is an application of hypothesis testing that tests a simple null hypothesis H_0 against a single alternative hypothesis H_1 [14]. In classical hypothesis testing, the number of required observations is fixed at the beginning of the experiment and two decision regions are considered: the region of acceptance of the null hypothesis and the region of acceptance of the alternative hypothesis. The SPRT differs from classical hypothesis testing as the required sample size in the SPRT is not predetermined at the beginning of the experiment and the test procedure is carried out sequentially. It splits the sample space into three regions: acceptance of the null hypothesis, acceptance of the alternative hypothesis, and a gray region where the number of observations is not sufficient to make a decision. The advantage of the sequential test is that it requires a smaller number of observations than classical hypothesis testing [15]. Therefore, it is meaningful to apply the SPRT to small samples.

For a brief discussion of the SPRT method, consider τ as the probability of transmitting the M_1 allele from the M_1M_2 heterozygous parent to the offspring, as in Ilk et al. [13]. The aim is to test the hypothesis H_0: $\tau = \tau_0$ against H_1: $\tau = \tau_1$ where τ_0 is different from τ_1. The decision is then made by using the likelihood ratio statistic λ_n where n is the number of first available observations $x_1, x_2, ..., x_n$ [16]. The likelihood ratio statistic λ_n is then computed for the first n observations using Eq. (1):

$$\lambda_n = \frac{\tau_0^b(1-\tau_0)^c}{\tau_1^b(1-\tau_1)^c} \text{ for } n = 1, 2, \ldots \tag{1}$$

The parameters b and c come from a 2×2 contingency table, where b is the sum of heterozygous parents who transmit the M_1 but not the M_2 allele to the offspring, and c is the sum of heterozygous parents who transmit the M_2 but not the M_1 allele to the offspring.

For a desired statistical power of a test, thresholds k_0 and k_1 with $k_0 < k_1$ were set as boundaries by

Wald's approximations:

$$k_0 = \frac{\alpha}{1-\beta}, \quad k_1 = \frac{1-\alpha}{\beta}, \tag{2}$$

where thresholds k_0 and k_1 depend on type I (α) and type II (β) errors and determine the boundaries between the decision regions. The SPRT then concludes that:

$$accept \quad H_0 \ if \quad \lambda_n \leq k_0, \tag{3}$$

$$reject \quad H_0 \quad if \quad \lambda_n \geq k_1, \tag{4}$$

$$continue \quad sampling \quad if \quad k_0 < \lambda_n < k_1, \tag{5}$$

which means that if the likelihood ratio statistic λ_n is less than or equal to the k_0 threshold, then there is an association between the genetic marker and a susceptible disease gene. Similarly, if the likelihood ratio statistic λ_n is greater than or equal to the k_1 threshold, then there is no association between the genetic marker and a susceptible disease gene. On the other hand, if the likelihood ratio statistic λ_n is within the critical interval (k_0, k_1), a decision has to be postponed until new observations are made [13].

Ilk et al. applied the SPRT to simulated data and showed improved performance of the sequential test over classical hypothesis testing. The SPRT accurately classified markers even with small sample sizes. However, many markers fell within a gray zone and no explanation was given about the gray zone markers.

4. SPRT+

We propose a novel method SPRT+ to reduce gray zone markers by using the marker-term annotation matrix with kappa statistics. The marker-term annotation matrix is one in which rows correspond to the markers and columns correspond to the independent terms. The relationships between the markers and the terms are in binary form. Let $A = [a_{ij}]$ denote an $m \times n$ binary data matrix, where m is the number of markers and n is the number of terms. In this case, if the ith marker is related to the jth term then the element a_{ij} will be equal to 1; if the ith marker is not related to the jth term then the element a_{ij} will be equal to 0. Hence, in A, all associations between markers and terms are represented by a combination of ones and zeros.

Our hypothesis is that similar annotation terms will share similar markers and by kappa statistics we will get the agreement metric between the markers. The kappa statistics determine the distance metric by which the marker-marker and marker set-marker relationship is judged. The metric defines the observed marker-marker relationship compared to its probability according to annotation cooccurrence [17].

We calculate the kappa statistics for the given markers a and b using Eq. (6).

$$K_{ab} = \frac{O_{ab} - A_{ab}}{1 - A_{ab}} \tag{6}$$

Here, O_{ab} is the observed relationship between markers a and b, A_{ab} is the chance cooccurrence of the common markers, and K_{ab} is the distance metric of markers a and b [18].

Notice that in an $m \times m$ kappa score matrix $\mathbf{K}=[k_{ij}]$ with m markers, $k_{ij} = k_{ji}$, i.e. $\mathbf{K}$ is symmetric, and k_{ij} is defined by the kappa statistics between the markers i and j.

According to the defined agreement range of the kappa value, the gray zone markers can then be reclassified into the same three regions once again. Thus, with no need of new trios, the gray zone markers can be refined. As a result, SPRT+ partially overcomes the drawback of the sequential test and reduces the number of gray zone markers.

5. Simulation study

A simulation study was performed to compare the performance of the proposed statistical method, SPRT+, with SPRT. The simulation data were generated with MATLAB. For the simulations, we generated 262,264 markers, 120 trios, and 100 annotation terms.

The marker set was generated with the three different genotypic risk ratio (GRR) subgroups: not significantly associated (GRR $\leq$ 1.5), moderately associated (1.5 <GRR <3.5), and highly associated (GRR $\geq$ 3.5) markers [13]. For a realistic scenario, 98% of the markers were generated under the null hypothesis and only 2% of the markers were generated under the alternative hypothesis. As a consequence, the subgroup with "no significant association" had 257,019 markers, the subgroup with "moderate association" had 3934 markers, and the subgroup with "high association" had 1311 markers. The minimum and maximum sample sizes of simulation data were assumed to be 20 and 120, respectively. Distinct sample sizes in increments of 10 were considered.

The relationships between the marker and the terms were simulated in the same probabilistic range as in subgroups. Namely, there is no relationship among the not-significant associated markers; there is a moderate relationship among the moderately associated markers and a high relationship among the highly associated markers.

Furthermore, the threshold values k_0 and k_1 were calculated with the nominal values $\alpha(0.1\%)$ and $\beta(20\%)$. Different kappa statistic scores ranging from 0.21 (fair agreement) to 0.81 (almost perfect agreement), in increments of 0.1, were considered [17].

6. Results

Consider a scenario in which a family-based haplotype relative risk approach is designed. We have a 'trio design' and want to identify the association between the markers and the disease locus. The sequential test is applied to this limited sample. However, there is a challenge: undefined gray zone markers. The simulation data results and the accuracy versus gray zone percentages for 262,264 markers generated under high association, moderate association, and no association are presented in Table 1 for different kappa scores.

Overall accuracies of SPRT and SPRT+ are calculated by the following equation.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

The accuracy of SPRT exceeds the accuracy of SPRT+ when the kappa score is less than 0.3. When the kappa score value is equal to or greater than 0.3, the accuracy of SPRT+ exceeds the accuracy of SPRT for sample sizes with 40 or less trios.

For samples with a small number of trios, the percentage of gray zone markers in SPRT+ is much less than the percentage of gray zone markers in SPRT. With 30 trios, SPRT classifies about 22% of total markers and leaves about 78% of markers with no decision. With 50 trios, SPRT barely reaches a conclusion for about half of the given markers. Ninety percent of the markers are classified only when the sample size reaches 150

trios. To be fair, one should keep in mind that the classified percentage of markers is less than 90% until reaching 150 trios; the accuracy is always over 90%, meaning that SPRT classifies a marker in a very safe mode.

According to Table 1, when compared to SPRT, SPRT+ always has a higher or equal percentage of decided markers with a decreasing gain as the number of trios in the samples increases.

Table 1. Percentages of true positives and false positives versus trio size and kappa score for markers with moderate association, high association, and no association.

Kappa Statistics	Trio Size		30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200
0.14	Moderate	TP%	93	95	93	90	91	91	92	90	92	92	93	92	92	92	95	95	95	95
	High	TP%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	No	FP%	9.87	6.43	4.73	2.90	1.63	1.50	0.90	0.87	0.70	0.47	0.27	0.23	0.10	0.23	0.17	0.13	0.10	0.10
0.18	Moderate	TP%	88	86	86	84	83	81	82	82	82	86	87	87	86	87	90	91	89	91
	High	TP%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	No	FP%	4.57	2.47	1.47	0.80	0.43	0.43	0.17	0.07	0.17	0.13	0.07	0.07	0.03	0.03	0	0.03	0.03	0
0.22	Moderate	TP%	82	78	75	72	70	72	72	73	73	77	80	80	81	81	83	86	86	86
	High	TP%	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	No	FP%	1.93	0.90	0.47	0.07	0.10	0	0.03	0	0	0.03	0	0.03	0	0	0	0	0	0
0.26	Moderate	TP%	72	67	63	59	57	62	62	65	67	71	76	75	77	77	80	82	83	84
	High	TP%	100	100	100	99	99	100	99	100	99	100	100	100	100	100	100	100	100	100
	No	FP%	0.63	0.17	0.03	0.00	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0
0.3	Moderate	TP%	59	53	49	44	48	51	54	60	62	68	72	72	75	76	78	81	81	83
	High	TP%	100	100	99	99	99	98	99	99	99	98	100	99	99	100	100	100	100	100
	No	FP%	0.10	0.07	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.34	Moderate	TP%	43	40	38	34	39	44	50	56	61	66	70	72	74	76	78	80	81	83
	High	TP%	99	97	99	96	96	96	97	98	96	95	98	99	98	100	100	100	99	100
	No	FP%	0.03	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.38	Moderate	TP%	33	29	29	29	35	41	48	55	60	66	69	71	74	76	78	80	81	83
	High	TP%	98	95	94	94	91	94	96	94	94	93	96	98	97	100	99	100	99	100
	No	FP%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.42	Moderate	TP%	21	20	23	27	33	41	48	55	60	66	69	71	74	76	78	80	81	83
	High	TP%	94	92	86	88	89	90	93	90	91	91	94	95	95	98	98	99	99	99
	No	FP%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.46	Moderate	TP%	14	16	21	27	33	40	48	55	60	66	69	71	74	76	78	80	81	83
	High	TP%	90	87	82	84	84	86	89	88	89	90	92	94	95	98	98	99	99	99
	No	FP%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0.5	Moderate	TP%	11	15	20	27	33	40	48	55	60	66	69	71	74	76	78	80	81	83
	High	TP%	85	80	76	79	81	83	87	86	88	88	92	93	95	98	98	99	99	99
	No	FP%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SPRT	Moderate	TP%	7.6	14	20	27	33	40	48	55	60	66	69	71	74	76	78	80	81	83
	High	TP%	18	42	59	67	74	79	83	85	87	88	91	92	95	98	98	99	99	99

SPRT+ reduces gray zone markers by classifying moderately and highly associated markers as no further interpretations are expected for not-associated markers. The true-positive (TP) percentages for 5435 markers with high or moderate association and the false-positive (FP) percentages for 257,019 markers with no association are displayed in Table 2. For samples with 30 trios, with a kappa statistics score of 0.26, the FP percentage of markers with no association is less than 1% and the TP percentage of markers with moderate and high association is 72% and 100%, respectively. As sampling continues, the FP percentage reduces to zero and the TP percentage of moderately associated markers reaches 84%. The Figure illustrates the results of the TP percentage of associated markers versus sample size for SPRT and SPRT+ with a kappa score of 0.3.

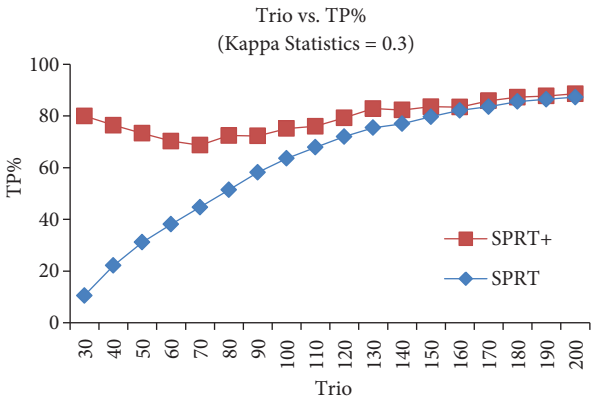

Figure. True-positive percentage results of SPRT+ and SPRT.

Table 2. Percentages of accuracy and gray zone versus trio size and kappa score for SPRT and SPRT+.

Kappa Statistics	Trio Size	30	40	50	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200
0.14	Accuracy %	0.80	0.88	0.91	0.93	0.95	0.95	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.54	0.44	0.36	0.30	0.25	0.21	0.18	0.16	0.13	0.12	0.10	0.09	0.09	0.07	0.06	0.05	0.05	0.05
0.18	Accuracy %	0.88	0.93	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.59	0.49	0.39	0.33	0.27	0.23	0.20	0.17	0.15	0.13	0.11	0.10	0.09	0.08	0.07	0.06	0.06	0.05
0.22	Accuracy %	0.93	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.62	0.51	0.42	0.35	0.30	0.25	0.22	0.19	0.16	0.14	0.12	0.11	0.10	0.09	0.08	0.07	0.06	0.06
0.26	Accuracy %	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.64	0.53	0.43	0.37	0.31	0.26	0.23	0.20	0.17	0.15	0.13	0.12	0.11	0.10	0.08	0.07	0.07	0.06
0.3	Accuracy %	0.97	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.66	0.55	0.45	0.39	0.33	0.28	0.24	0.21	0.18	0.15	0.13	0.12	0.11	0.10	0.09	0.07	0.07	0.06
0.34	Accuracy %	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.69	0.57	0.47	0.40	0.34	0.29	0.25	0.21	0.18	0.16	0.14	0.12	0.11	0.10	0.09	0.08	0.07	0.06
0.38	Accuracy %	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.70	0.59	0.48	0.41	0.35	0.29	0.25	0.21	0.18	0.16	0.14	0.12	0.11	0.10	0.09	0.08	0.07	0.06
0.42	Accuracy %	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.72	0.60	0.50	0.41	0.35	0.30	0.25	0.22	0.19	0.16	0.14	0.13	0.11	0.10	0.09	0.08	0.07	0.06
0.46	Accuracy %	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.73	0.61	0.50	0.42	0.35	0.30	0.26	0.22	0.19	0.16	0.14	0.13	0.11	0.10	0.09	0.08	0.07	0.06
0.5	Accuracy %	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.74	0.62	0.51	0.42	0.36	0.30	0.26	0.22	0.19	0.16	0.14	0.13	0.11	0.10	0.09	0.08	0.07	0.06
SPRT	Accuracy %	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96
	Gray Zone %	0.78	0.64	0.51	0.43	0.36	0.30	0.26	0.22	0.19	0.16	0.14	0.13	0.11	0.10	0.09	0.08	0.07	0.06

Hence, results confirm that SPRT+ reduces the percentage of gray zone markers and accurately classifies the markers with moderate or high association.

With small sample sizes, it is difficult to classify markers into associated or not-associated groups. SPRT overcomes this challenge, but some results remain in a huge gray zone with unclassified markers. The results show that SPRT+ effectively refines the gray zone markers and sustains the desired statistical power. With a 0.26 kappa statistics score, the moderately and highly associated markers are identified as significantly associated with the disease marker. When comparing the two methods, for SPRT+, the TP percentages of associated

markers lie between 70% and 80% even with a small sample size, whereas for SPRT, 150 and more trios are needed to achieve the same power.

Note that, although not classifying all moderately associated markers with even 200 trios, 100% of highly associated markers are classified as associated with the disease with only 30 trios.

7. Conclusion

One of the important challenges in family-based association studies is reducing the needed sample size. Traditional statistical methods require more than 200 trios to give powerful results and classify the markers as associated with the disease or not. However, especially for the late-onset diseases like some types of cancer, it is very difficult to gather trios with offspring whose parents are still alive. The sequential test SPRT was proposed to analyze samples with small sizes and became useful for accurate association analysis. The SPRT does not only classify the markers into two groups as associated and not associated markers, but also for some other markers says that there is not enough evidence for correct classification. These markers require more trios and are put into a 'gray zone'. This conserves the method's statistical accuracy. However, now a new problem arises. For small sample sizes, the percentage of the decided markers may be very low, which results in a high percentage of 'gray zone' markers. To overcome this problem, a novel approach (SPRT+) is proposed in this study.

Simulation results have shown that SPRT+ conserves the statistical accuracy of the SPRT while partially solving the 'gray zone' problem by using a marker-term annotation matrix and kappa statistics.

References

[1] Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. BMC Med Genet 2009; 10: 6.

[2] Ku CS, Loy EY, Pawitan Y, Chia KS. The pursuit of genome-wide association studies: where are we now? J Hum Genet 2010; 55: 195-206

[3] Bush WS, Moore JH. Genome-wide association studies. PLoS Comput Biol 2012; 8: 12-18.

[4] Laird NM, Lange C. The role of family-based designs in genome-wide association studies. Stat Sci 2009; 24: 388-397.

[5] Falk CT, Rubinstein P. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 1987; 51: 227-233.

[6] Ott J. Statistical properties of the haplotype relative risk. Genet Epidemiol 1989; 6: 127-130.

[7] 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: 506-516.

[8] Lasky-Su J, Won S, Mick E, Anney RJL, Franke B, Neale B, Biederman J, Smalley SL, Loo SK, Todorov A et al. On genome-wide association studies for family-based designs: an ıntegrative analysis approach combining ascertained family samples with unselected controls. Am J Hum Genet 2010; 86: 573-580.

[9] Ott J, Kamatani Y, Lathrop M. Family-based designs for genome-wide association studies. Nat Rev Genet 2011; 12: 465-474.

[10] Laird NM, Lange C. Family-based methods for linkage and association analysis. Adv Genet 2008; 60: 219-252.

[11] Spielman RS, Ewens WJ. The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet 1996; 59: 983-989.

[12] Wang D, Sun F. Sample sizes for the transmission disequilibrium tests: TDT, S-TDT and 1-TDT. Commun Stat Theory 2000; 29: 1129-1142.

[13] Ilk O, Rajabli F, Dungul-Ciglidag D, Ozdag H, Ilk HG. A novel approach for small sample size family-based association studies: sequential tests. Eur J Hum Genet 2011; 19: 915-920.

[14] Anderson TW. A modification of the sequential probability ratio test to reduce the sample size. Ann Math Stat 1959; 34: 165-197.

[15] Wald A. Sequential Analysis. New York, NY, USA: John Wiley & Sons, 1947.

[16] Spigelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio test: application to Bristol, Shipman and adult cardiac surgery. Int J Qual Health Care 2003; 15: 7-13.

[17] Viera AJ, Garrett JM. Understanding ınterobserver agreement: the kappa statistic. Fam Med 2005; 37: 360-363.

[18] Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol 2007; 8: R183.