

A Note on Robust TDT-Type Test under Informative Parental

Missingness

J.H. Chen^a and K.F. Cheng^{b,c}

^aBiostatistics Center and Graduate Institute of Biostatistics, China Medical University,
Taichung, Taiwan (ROC)

^bBiostatistics Center and College of Public Health, China Medical University, Taichung,
Taiwan (ROC)

^cGraduate Institute of Statistics, National Central University, Chungli, Taiwan (ROC)

Short Title: Robust TDT-type Association Test

Correspondence to Professor: K.F. Cheng, Biostatistics Center, China Medical University,
Taichung, Taiwan (ROC). E-mail: kfcheng@mail.cmu.edu.tw.

Phone number: 886-4-2207-8539. Fax: 886-4-22078539.

Many family-based association tests rely on the random transmission of alleles from parents to offspring. Among them, the transmission/disequilibrium test (TDT) may be considered to be the most popular statistical test. The TDT statistic was proposed to evaluate nonrandom transmission of alleles from parents to the diseased children. However, in family studies, parental genotypes are not always available. Quite often, the offspring genotype affects the severity of offspring phenotype or/and the age at onset and in turn affects the parental missingness. In such case, the nonrandom transmission of alleles may also occur even when the gene and disease are not associated. As a consequence, the usual TDT or its variations would produce excessive false positive conclusions in association studies. In this note, we propose a TDT-type association test which is not only simple in computation but also robust to the joint effect of population stratification and informative parental missingness. The test statistic does not rely on any model and also allows for having different mechanisms of parental missingness across subpopulations. We use a simulation study to compare the performance of new test and the TDT and point out the advantage of the new method.

Keywords: Association test; Case-parents study; Informative missingness; Robustness;

Transmission /disequilibrium test

1. Introduction

Testing association between genetic markers and disease usually consists of a comparison of genotypes from a sample of diseased individuals with those from a certain sample of nondiseased individuals. The usual case-parents study suggests using genotype data of the diseased children and their parents for making inference about gene-disease association. Well known tests based on parental controls include the transmission/disequilibrium test (TDT) proposed by Spielman *et al.* [1], and the conditional-on-parental-genotypes (CPG) tests proposed by Schaid and Sommer [2] (see also [3]-[7]) for related approaches. The TDT and the CPG tests are identical under additive genetic model. However, the CPG approach is generally more powerful than the TDT approach under other genetic models.

In case-parents study, the cases and controls are matched in genetic ancestry. Thus, the analysis based on the TDT or CPG tests is free of bias arising from population stratification. This is an important property for valid association tests. However, these tests may still produce biased results if informative parental missingness exists in the study. The effect of missing parental genotype and its correction were studied by Clayton [8], Sun *et al.* [9], Weinberg [10], Cervino and Hill [11], Allen *et al.* [12], and Chen [13] (see also Robinwiz and Laird [14]; Robinwiz [15] for tests based on general families.) However, many of these methods often require assumptions such as missing-at-random (MAR, conditional on offspring and available parent, the genotype frequencies among missing parents and among

observed parents are the same) or missing-independent-of-offspring-genotype (MIOG, conditional on parental genotypes, the parental missingness is independent of offspring's genotype). Since there is no genotype information available on the missing parents, thus these important assumptions are usually difficult to justify in real applications. Another assumption also often required in some association tests is that the response probabilities of parents can be modeled by the same parametric function across all families in the study sample. For example, Allen *et al.* [12] required that the response-odds parameters satisfy relatively simple models across all studied families. This assumption may not be credible either, if the overall population consists of several subpopulations and response rates have different forms across subpopulations.

In this note, we first point out that when there is no disease-gene association, and both parents are observed, the probability of offspring's genotype conditional on the parental genotypes (general CPG probabilities) are no longer the same as the usual Mendelian proportions, if the parental missingness also depends on the offspring genotype. In this case, many tests such as TDT or its variations, depending on using the properties of Schaid-Sommer's CPG probability, would produce biased association results. **This particular case may occur when, for example, the offspring genotype affects the severity of offspring phenotype or/and the age at onset and in turn affects the parental missingness.**

According to the previous discussions, we find that in the literature, there exists no

association test which is simultaneously robust to the effects of population stratification and general informative parental missingness. In this note, we intend to propose a truly robust association test based on case-parents data. The proposed test is novel, simple, and derived from using the conditional probability of the offspring's genotype given parental genotypes when they are both observed. Thus, it is robust to the effect of population stratification. We emphasize that the new test does not require any assumption or model for parental missingness. That is, we let the probability of parental missingness simultaneously depend not only on the parental genotypes but also the offspring's genotype, and be model free. In the case of population stratification, we also allow this probability depend on the ethnicity. Thus the mechanism of the missingness considered in this note is the most general form of informative parental missingness (GIPM), under which many important association tests may become invalid. In this note, we also present simulation results to compare the performance of the usual TDT test and the new test using only the complete case-parents data. Under some scenarios where the MIOG condition fails, we show the TDT test tends to have excessive false positive association results. This indicates that many approaches based on the Schaid-Sommer's CPG probability when both parental genotypes are observed [12, 13] may be invalid too. In contrast, the new test has satisfactory performance in the sense that its type I error can be approximately controlled at the desired significance level and its power is in general sufficiently large so that at least moderate genetic effect can be detected using

reasonable number of family data. In the simulation study, we consider scenarios where conditions such as MAR, MIOG or GIPM are satisfied. We also consider the situation where the general population consists of two subpopulations with different allele frequencies at the candidate marker and different mechanisms for the parental missingness. Under all conditions studied in the simulation, we find out that the new test is insensitive to the joint effect of population stratification and GIPM.

2. Method

We assume that the candidate gene has two alleles, coded as a (normal allele) and A (candidate disease allele), or can be divided into two groups of alleles. The genotype of the diseased offspring is denoted by G_0 . The set of parental genotypes is denoted by (G_m, G_f) , where G_m is the maternal genotype and G_f is the paternal genotype. G_0 represents the number of copies of the A allele in the offspring genotype (taking the values 0, 1, and 2) with the same convention for G_m and G_f . The missing pattern is denoted as (R_m, R_f) , where R_m (R_f) equals one if the maternal (paternal) genotype is available in the study and zero, otherwise.

In the following discussion, we focus on using complete family trios, where both parental and maternal genotypes are observed. The probability of an offspring genotype G_0 conditional on his/her parental genotypes (G_m, G_f) , parental missingness pattern $(R_m, R_f) = (1, 1)$, and offspring's phenotype D_0 is given by

$$\frac{\gamma_{G_0}(G_m, G_f) \varphi_{G_0} P[G_0 | G_m, G_f]}{\sum_{g=0}^2 \sum_{(G_m, G_f)} \gamma_g(G_m, G_f) \varphi_g P[G_0 = g | G_m, G_f]}, \quad (2.1)$$

where $\varphi_{G_0} = P[D_0 | G_0] / P[D_0 | G_0 = 0]$ are the genotype relative risk parameters, and $\gamma_{G_0}(G_m, G_f) = P[R_m = 1, R_f = 1 | G_0, G_m, G_f, D_0] / P[R_m = 1, R_f = 1 | G_0 = 0, G_m, G_f, D_0]$ is a ratio of missingness probabilities under offspring genotype G_0 versus that under baseline. Note that the general CPG probability (2.1) is derived under the usual assumption that the offspring's phenotype and parental genotypes are independent conditional on the offspring's genotype. If the overall population consists of several subpopulations, we require this assumption to be held within each subpopulation too. We point out that the general CPG probability can be reduced to the Schaid-Sommer's CPG probability [2], if $\gamma_{G_0}(G_m, G_f)$ is a constant with respect to G_0 . The latter condition holds when, for example, the MAR or MIOG condition holds. On the other hand, if $\gamma_{G_0}(G_m, G_f)$ is not a constant, then any test based on the Schaid-Sommer's CPG probability may be invalid.

The general CPG probability depends on the relative risk parameters, ratios of missingness probabilities, and Mendelian proportions. If we define $\varepsilon_g(G_m, G_f) = \gamma_0(G_m, G_f) - \gamma_g(G_m, G_f) = 1 - \gamma_g(G_m, G_f)$ and assume that with respect to g , $\varepsilon_g(G_m, G_f)$ are small and approximately equal (denoted as $\varepsilon(G_m, G_f)$) for each fixed (G_m, G_f) , then the general CPG probabilities can be greatly simplified after applying Taylor's expansion. Note that this assumption essentially requires that the probability of parental missingness do not deviate too much under different offspring's genotypes. Simulation results

presented in this paper confirm that even the differences are moderate ($|\varepsilon_g(G_m, G_f) - \varepsilon(G_m, G_f)| < 0.1$), the test proposed in this paper still has satisfactory performance. In contrast, the usual TDT has type I errors seriously inflated under this scenario. In our formulation of the testing procedure we consider approximations of the general CPG probabilities by ignoring all terms involving $\varepsilon_g(G_m, G_f)^a$, $a \geq 2$ in their Taylor's expansions. Under the null hypothesis of no gene-disease association, the first-order approximations of the general CPG probabilities are given in Table I.

In view of the approximation results of Table I, we consider association analysis using only the informative family data. Let $\hat{P}_{2+3}(i)$ denote the sample proportion of an offspring carrying i risk alleles under parental mating types 2 or 3. $\hat{P}_{7+8}(i)$ and $\hat{P}_6(i)$ represent similar sample proportions under parental mating types 7 or 8 and mating type 6, respectively. The results in Table I imply that

$$S = N_{2+3} \left[P_{2+3}(2) - P_{2+3}(1) \right] + N_6 \left[2P_6(2) - P_6(1) \right] + N_{7+8} \left[P_{7+8}(0) - P_{7+8}(1) \right] \approx 0, \quad (2.2)$$

under null association. The variance estimate of S is given by

$$\begin{aligned} Var = & \left[4P_{2+3}(1)(1 - P_{2+3}(1)) \right] N_{2+3} + \left[4P_6(2)(1 - P_6(2)) + P_6(1)(1 - P_6(1)) + 4P_6(2)P_6(1) \right] N_6 \\ & + \left[4P_{7+8}(1)(1 - P_{7+8}(1)) \right] N_{7+8} \end{aligned}$$

where $N_k(N_{k+j})$ is the number of complete families with mating type k (mating types k or j). Thus, a simple TDT-type association test can be defined as $T = S^2 / Var$. The P-value of

the test is given by $\Pr[\chi_1^2 \geq T]$, where χ_1^2 is a chi-square random variable with one degree of freedom. We point out that the test is still valid under population stratification, where the probability function of parental missingness differs in subpopulation.

3. Simulation Results

We have conducted a simulation study to investigate the performance of the new association test T and compared the results with those for the traditional TDT based on the **complete trios**.

According to Chen [13], the methods of Allen *et al.* [12] and Chen [13] had the best overall performance under various missingness models satisfying MIOG condition. However, under **complete trios**, the methods of Allen *et al.* and Chen are the same as or variations of the traditional TDT, thus we excluded their methods in our simulation study. To study the performance of type I error, we assumed the relative risks satisfied $\varphi_1 = \varphi_2 = 1$ in the simulations. To study the power performance, we considered three genetic models: dominant model with $\varphi_1 = \varphi_2 = 5$, recessive model with $\varphi_1 = 1, \varphi_2 = 5$ and additive model with $\varphi_1 = 5, \varphi_2 = 9$.

In the simulation study, we considered three missingness models satisfying MAR, MIOG, or GIPM condition, respectively. We assumed that the joint missingness probability was the product of maternal and paternal missingness probabilities:

$$\begin{aligned} &P(R_m = 1, R_f = 1 | G_m = g_m, G_f = g_f, G_0 = g_0, D = 1) \\ &= P(R_m = 1 | G_m = g_m, G_0 = g_0, D = 1) \times P(R_f = 1 | G_f = g_f, G_0 = g_0, D = 1). \end{aligned}$$

We also assumed that each marginal missingness probability satisfied a logistic regression

model:

$$P(R_m = 1 | G_m = g_m, G_0 = g_0, D = 1) = \frac{1}{1 + \exp(-\alpha_m - g_m\beta_m - g_0\gamma_m)}$$

and

$$P(R_f = 1 | G_f = g_f, G_0 = g_0, D = 1) = \frac{1}{1 + \exp(-\alpha_f - g_f\beta_f - g_0\gamma_f)}.$$

Under MAR condition, we assumed $\alpha_m = 1.7346$, $\alpha_f = 1.0986$, and the remaining parameter values were zeroes. This is equivalent to assuming maternal response rate equal to a constant 0.85 and paternal response rate equal to 0.75. Under MIOG condition, we assumed $\alpha_m = 1.3863$, $\beta_m = -0.5390$, $\alpha_f = 0.8473$, $\beta_f = -0.4418$, and $\gamma_m = \gamma_f = 0$. This is equivalent to having maternal response rate ranging from 0.5765 to 0.8000 and paternal response rate ranging from 0.4909 to 0.7000. Two models satisfying GIPM condition were assumed in the study. GIPM (1) model assumed $\alpha_m = 1.7346$, $\beta_m = -0.2183$, $\gamma_m = -0.3445$, $\alpha_f = 1.3863$, $\beta_f = -0.1206$, and $\gamma_f = -0.2559$. This is equivalent to assuming maternal response rate ranging from 0.6523 to 0.8500 and paternal response rate ranging from 0.6532 to 0.8000. Note that this is a weak GIPM model. GIPM (2) model assumed $\alpha_m = 0.8473$, $\beta_m = 0.2513$, $\gamma_m = -0.3466$, $\alpha_f = 0.4055$, $\beta_f = -0.0827$, and $\gamma_f = -0.1614$. and $\gamma_f = 0.1614$. In this case, the range of maternal response rate is (0.5400, 0.7000) and that for the paternal response rate is (0.4800, 0.6000). This is a moderate GIPM model.

We also studied the effect of population stratification. We assumed that the studied population consisted of two subpopulations with high risk allele frequencies $p_1 = 0.4$, and

$p_2 = 0.2$, respectively, and each subpopulation satisfied Hardy-Weinberg equilibrium condition. We assumed the total complete trios for study is 300 and the proportion p of the family trios is from subpopulation 1. If $p = 1$ ($p = 0$) then the studied population was subpopulation 1 (2) with allele frequency 0.4 (0.2). The simulation results reported in the tables are based on 10,000 replications. Each size (or power) is the proportion of times that 10,000 simulated p-values < 0.05 .

In Tables II and III, we report the simulated sizes and powers of the association tests T and TDT under different combinations of missingness model and population structure. The results in Table II were based on one population and therefore there was no effect of population stratification. Under this case, the range of the size of the T test was (0.0506, 0.0565) and that of the TDT was (0.0519, 0.2685), when the risk allele frequency was 0.4. On the other hand, when the risk allele frequency became 0.2, the corresponding ranges changed to (0.0534, 0.0762) and (0.0491, 0.1807), respectively. These results showed that the size of the new test was basically consistent with the nominal value of 5% under most simulation conditions. The exceptional case occurred when the allele frequency was small and GIPM level was moderate. In contrast, the size of the TDT tended to be inflated under GIPM models. The amount of increase in size also depends on the GIPM level. Under the same case, the powers of the T test were in general greater than 0.9800. The exceptional cases occurred when the allele frequency was high and the genetic model was additive or allele frequency

was low and genetic model was recessive. However, we pointed out that the power of the new test was at least 0.70 under combinations of any genetic model and GIPM model. This indicates that the new test is rather efficient. The results in Table III were derived under two subpopulations with identical or different missingness models. Under these cases, the effects of population stratification were present. Therefore, from Table III one can study the joint effects of population stratification and GIPM when the new test T or TDT were used. According to Table III, we first found out that the size of the new test ranged from 0.0530 to 0.0586 and that of TDT ranged from 0.0528 to 0.2075 under all study conditions. This means that using the new test, we were able to control its type I error at the predetermined significance level, while the TDT cannot. It is also of interest to point out that the new test seems to have better power performance when there is population stratification, comparing with that under no population stratification. Table III showed that the power of the new test were in general greater than 0.900. The exceptional case happened under MAR and additive genetic model where the smallest power was 0.7782. These results concluded that the new test was efficient in detecting true associations under population stratification and any missingness

4. Real data analysis

We next considered a real study to investigate the performance of the TDT and new association test under null association. The study was to examine transforming growth factor beta-1 SNPs in relation to asthma risk and degree of atopy among 546 case-parent triads (Li

et al.[16]), consisting of asthmatics aged 4-17 years and their parents in Mexico City. Five SNPs were considered in the study. Here, we focus only on SNP rs8179181. Both TDT and the new test showed that no statistically significant association exists between this SNP and asthma risk (P-value=0.457901, and 0.797963, respectively). We used GIPM model ($\alpha_m = 1.9924$, $\beta_m = -0.2578$, $\gamma_m = -0.3180$ $\alpha_f = 1.7346$, $\beta_f = -0.3483$, and $\gamma_f = -0.2685$.) as described above to randomly generate incomplete family triads. Figure 1 shows the p-value histograms for the TDT and the new test based on 10,000 replications. The original study has 136 informative families (consisting of at least one heterozygous parent) . Under our missingness model, the averaged number of informative and complete families is 92. That is, about 1/3 of the informative families have missing parental genotypes. The figure shows that the TDT has excessive number of small p-values, indicating that the analysis based on the TDT has produced too many false positive results. In contrast, the new association test still maintains satisfactory performance under complicated missingness scenario.

5. Discussion

Several family-based tests of association or linkage of genetic marker and a disease susceptible locus have been proposed in the literature. These tests have gained popularity because of their insensitivity to population stratification. However, these tests may still be biased because of missing parental information, which would be typical for diseases of old age. Some of these tests accommodate missing parental information, but they also require

important assumptions such as MAR or MIOG. Unfortunately, these assumptions are difficult to justify based on the incomplete family data, particularly when the population under study is heterogeneous. Under our simulation settings, we found that if the parental missingness also depended on the genotypic outcome of the diseased offspring, then the largest empirical type I error rate of the usual TDT, based on using 300 complete trios, would be 0.2685, when in fact the predetermined significance level was only 0.0500. Since many recently proposed tests for correcting bias in case-parents studies, by Allen *et al.* [12], or Chen [13] for examples, were the same as or a variation of the TDT under complete trios, therefore, one needs to be cautious in using these tests. Guo *et al.* [16, 17] considered the missing parental haplotype problem based on the EM algorithm approach. However, they also assumed that MAR or MIOG conditions were satisfied.

We note that under general parental missingness, Rabinowitz [15] also developed an analysis based on a regression-adjusted score statistic to adjust for population heterogeneity. The proposed method provided a general framework for developing valid association tests with incomplete family data. However, the test depends on the choice of score vector and specification of the conditional probability of the missing genotype(s). Guidance on the choice of these important functions and the related sensitivity analysis so far remain unsolved.

In this note, we consider a simple TDT-type test based on complete families with at least one heterozygous parent. The test statistic depends on the proportions of the transmission of

the risk allele from parents to their diseased children. Thus it is simple in computation and robust to the effect of population stratification. The test allows the parental missingness depending on all genotype information of the family and the subpopulations involved in the study. It is also nonparametric in the sense that there is no model ever being used in the analysis. We remark that our analysis is based on using those family data where both parents respond to the study. In the development of the new test we have used a Taylor's expansion for the joint response probability conditional on the offspring's genotype, with the requirement that the conditional probability does not deviate too much with respect to the offspring's genotype. Thus, theoretically speaking, if the offspring's genotypic outcome would greatly influence the parental missingness, then the approximation used in the analysis may not be valid and the new test could be biased too. However, according to our simulation results, if the differences of these conditional response probabilities are less than 10%, the performance of our new test is still satisfactory. We consider such differences to be rather reasonable in practical applications, especially when the parental response rates are moderate or high.

Many family-based association tests also include in **complete trios**, such as dyads or monads, in their analysis. However, the trade-off is that they also require strong assumptions such MAR or MIOG be satisfied. To keep full robustness and model-free in our association analysis, we find that the genotype data from incomplete families contribute no additional

information, if the approach for analyzing complete data was modified for incomplete data. This is because that the probability of the offspring's genotype conditional on the (one) observed parent's genotype still has two unknown parameters under the null hypothesis. So far, it is not clear if there exists such a method that includes incomplete trios in the analysis without making any assumption about the probability of missingness. It is of interest to investigate this issue in the future.

Acknowledgements

This research was supported in part by a grand from National Science Council and a joint research grand from China Medical University and Asia University.

References

1. Spielman, R. S., McGinnis, R. E. AND Ewens, W. J.. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 1993; **52**: 506–516.
2. Schaid, D. J. AND Sommer, S. S. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *American Journal of Human Genetics* 1993; **53**: 1114–1126.
3. Ott, J. Statistical properties of the haplotype relative risk. *Genetic Epidemiology* 1989; **6**: 127–130.
4. Terwilliger, J. D. AND Ott, J. A haplotype-based “halotype relative risk” approach to

- detecting allelic associations. *Human Heredity* 1992; **42**: 337–346.
5. Ewens, W. J. AND Spielman, R. S. The transmission /disequilibrium test: history, subdivision and admixture. *American Journal of Human Genetics* 1995; **57**: 455–464.
 6. Thomson, G. Analysis of complex human genetic traits. An ordered-notation method and new tests for model of inheritance. *American Journal of Human Genetics* 1995a; **57**: 474–486.
 7. Thomson, G. Mapping disease genes: Family-based association studies. *American Journal of Human Genetics* 1995b; **57**: 487–498.
 8. Clayton, D. A generalization of the transmission/disequilibrium test for uncertain haplotype transmission. *American Journal of Human Genetics* 1999; **65**: 1170–1177.
 9. Sun, F., Flanders, W. D., Yang, Q. AND Khoury, M. J. Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. *American Journal of Epidemiology* 1999; **150**: 97–104.
 10. Weinberg, C. R. Allowing for missing parents in genetic studies of case-parent triads. *American Journal of Human Genetics* 1999; **64**: 1186–1193.
 11. Cervino, A. C. AND Hill, A. V. Comparison of tests for association and linkage in incomplete families. *American Journal of Human Genetics* 2000; **67**: 120–132.
 12. Allen, A. S., Rathouz, P. J. AND Satten, G. A. Informative missingness in genetic association studies: case-parent designs. *American Journal of Human Genetics* 2003; **72**:

671–680.

13. Chen, Y. H. New Approach to association testing in case-parent designs under informative parental missingness. *Genetic Epidemiology* 2004; **27**: 131–140.
14. Rabinowitz, D. AND Laird, N. A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Human Heredity* 2000; **50**: 211–223.
15. Rabinowitz, D. Adjusting for population heterogeneity and misspecified haplotype frequencies when testing nonparametric null hypotheses in statistical genetics. *Journal of the American Statistical Association* 2002; **97**: 742–758.
16. Li, H., Romieu, I., Wu, H., Sienna-Monge, J.J., Ramírez-Aguilar, M., del Río-Navarro, B.E., del Lara-Sánchez, I.C., Kistner, E.O., Gjessing, H.K., London, S.J. Genetic polymorphisms in transforming growth factor beta-1 (*TGFBI*) and childhood asthma and atopy. *Human Genetics* 2007; **121**: 529–538.
17. Guo, C.Y., DeStefano, A.L., Lunetta, K.L., Dupuis, J., and Cupples, L. A. Expectation maximization Algorithm based haplotype relative risk (EM-HRR) test of linkage disequilibrium using incomplete case-parents trios. *Human Heredity* 2005; **59**: 125-135.
18. Guo, C.Y., Gui, J., and Cupples, L. A. Impact of non-ignorable missingness on genetic tests of linkage and/or association using case-parent trios. *BMC Genetics* 2005; **6**: (Suppl 1):S90.

Table I. First-order approximations of the general CPG probabilities for the complete trio under the null hypothesis of no association

Mating type	Parental genotype $G_m \times G_f$	Offspring genotype		
		AA	Aa	aa
1	$AA \times AA$	1	0	0
2	$AA \times Aa$	1/2	1/2	0
3	$Aa \times AA$	1/2	1/2	0
4	$AA \times aa$	0	1	0
5	$aa \times AA$	0	1	0
6	$Aa \times Aa$	$\frac{[4 + \varepsilon(Aa, Aa)]}{16}$	$\frac{[8 + 2\varepsilon(Aa, Aa)]}{16}$	$\frac{[4 - 3\varepsilon(Aa, Aa)]}{16}$
7	$Aa \times aa$	0	1/2	1/2
8	$aa \times Aa$	0	1/2	1/2
9	$aa \times aa$	0	0	1

Table II. Sizes and Powers of the Association Tests Under One Population

Missingness Model	Hypothesis	Sampling proportion p=1		Sampling proportion p=0	
		T	TDT	T	TDT
MAR	Null	0.0538	0.0519	0.0535	0.0491
	Dominant ($\varphi_2 = 5$)	0.9890	1.0000	0.9996	1.0000
	Recessive ($\varphi_2 = 5$)	1.0000	1.0000	0.7489	0.9898
	Additive ($\varphi_2 = 9$)	0.4525	1.0000	0.9982	1.0000
MIOG	Null	0.0506	0.0530	0.0534	0.0537
	Dominant ($\varphi_2 = 5$)	0.9972	1.0000	0.9956	1.0000
	Recessive ($\varphi_2 = 5$)	0.9997	1.0000	0.6033	0.9740
	Additive ($\varphi_2 = 9$)	0.7013	1.0000	0.9942	1.0000
GIPM(1)	Null	0.0565	0.1661	0.0692	0.1126
	Dominant ($\varphi_2 = 5$)	0.9984	1.0000	0.9978	1.0000
	Recessive ($\varphi_2 = 5$)	0.9998	1.0000	0.7172	0.9339
	Additive ($\varphi_2 = 9$)	0.7773	1.0000	0.9962	1.0000
GIPM(2)	Null	0.0559	0.2685	0.0762	0.1807
	Dominant ($\varphi_2 = 5$)	0.9978	1.0000	0.9987	1.0000
	Recessive ($\varphi_2 = 5$)	0.9998	1.0000	0.7857	0.9335
	Additive ($\varphi_2 = 9$)	0.7206	1.0000	0.9975	1.0000

Table III. Sizes and Powers of the Association Tests Under Two Populations

Missingness Model		Hypothesis	Sampling proportion p=2/3		Sampling proportion p=1/3	
Subpopulation 1	Subpopulation 2		T	TDT	T	TDT
MAR	MAR	Null	0.0558	0.5450	0.0586	0.0528
		Dominant ($\varphi_2 = 5$)	0.9973	1.0000	0.9997	1.0000
		Recessive ($\varphi_2 = 5$)	0.9989	1.0000	0.9725	0.9999
		Additive ($\varphi_2 = 9$)	0.7782	1.0000	0.9664	1.0000
MIOG	MIOG	Null	0.0539	0.0545	0.0536	0.0528
		Dominant ($\varphi_2 = 5$)	0.9994	1.0000	0.9999	1.0000
		Recessive ($\varphi_2 = 5$)	0.9933	1.0000	0.9725	0.9999
		Additive ($\varphi_2 = 9$)	0.9042	1.0000	0.9664	1.0000
GIPM(1)	GIPM(1)	Null	0.0554	0.1601	0.0546	0.1373
		Dominant ($\varphi_2 = 5$)	0.9994	1.0000	0.9999	1.0000
		Recessive ($\varphi_2 = 5$)	0.9926	0.9995	0.9455	0.9949
		Additive ($\varphi_2 = 9$)	0.9183	1.0000	0.9869	1.0000
GIPM(2)	GIPM(1)	Null	0.0550	0.2075	0.0530	0.1560
		Dominant ($\varphi_2 = 5$)	0.9990	1.0000	1.0000	1.0000
		Recessive ($\varphi_2 = 5$)	0.9937	0.9998	0.9485	0.9944
		Additive ($\varphi_2 = 9$)	0.9084	1.0000	0.9855	1.0000