

Trend Tests for Case-Control Studies of Genetic Markers: Power, Sample Size and Robustness

Boris Freidlin^a Gang Zheng^c Zhaohai Li^{b,d,e} Joseph L. Gastwirth^{b,e}

^aBiometric Research Branch and ^bBiostatistics Branch, National Cancer Institute, Rockville, Md.,

^cOffice of Biostatistics Research, National Heart, Lung and Blood Institute, Bethesda, Md.,

^dBiostatistics Center, George Washington University, Rockville, Md., and ^eDepartment of Statistics, George Washington University, Washington, D.C., USA

Key Words

Candidate gene associations · Cochran-Armitage trend test · MAX test · Maximin efficiency robust test · Power approximation · Sample size

Abstract

The Cochran-Armitage trend test is commonly used as a genotype-based test for candidate gene association. Corresponding to each underlying genetic model there is a particular set of scores assigned to the genotypes that maximizes its power. When the variance of the test statistic is known, the formulas for approximate power and associated sample size are readily obtained. In practice, however, the variance of the test statistic needs to be estimated. We present formulas for the required sample size to achieve a prespecified power that account for the need to estimate the variance of the test statistic. When the underlying genetic model is unknown one can incur a substantial loss of power when a test suitable for one mode of inheritance is used where another mode is the true one. Thus, tests having good power properties relative to the optimal tests for each model are useful. These tests are called efficiency robust and we study two of

them: the maximin efficiency robust test is a linear combination of the standardized optimal tests that has high efficiency and the MAX test, the maximum of the standardized optimal tests. Simulation results of the robustness of these two tests indicate that the more computationally involved MAX test is preferable.

Copyright © 2002 S. Karger AG, Basel

Introduction

In case-control studies evaluating association between a candidate allele and a disease, allele-based tests and genotype-based tests are equivalent when Hardy-Weinberg equilibrium (HWE) holds [1]. However, when HWE is not satisfied, Sasieni [1] showed that the allele-based test is invalid and that the genotype-based Cochran-Armitage (CA) trend test [2, 3] can be used. The CA trend test utilizes a set of scores that can be obtained as an efficient score test for a logistic regression [4]. Hence it is a locally optimal test for the given set of scores. The statistical properties of the optimal test for the additive model were investigated by Slager and Schaid [5].

In applications, the theoretical variance (σ_0^2) of the test statistic may not be known and is replaced by a consistent estimate under the null hypothesis. When the alternative is true, however, the estimated standard deviation converges to σ^* , which may differ from σ_0 . Therefore, power and sample size calculations should be based on σ^* . We present formulas to approximate the sample size needed to achieve a desired power that account for the estimation of the variance of the test statistic. We demonstrate that use of the recessive scores for the recessive model instead of the test optimal for additive model typically leads to noticeably smaller required sample sizes. Only when the allele frequency is high is there a substantial difference between the optimal tests for dominant and additive models when the underlying model of inheritance is dominant. Therefore we apply results from efficiency robustness theory [6-10] to obtain tests that have relatively high power for any of the three commonly used genetic models. The maximin efficiency robust test (MERT) is a linear combination of the tests optimal for the recessive and dominant models. The MAX (the maximum of the standardized version of the three optimal tests) is generally more powerful but is also computationally more intensive.

Cochran-Armitage Trend Test

The data available from a case-control study are represented in table 1, where *A* is a high risk candidate allele and *a* is any of the other alleles. In table 1, *R* and *S* are sizes of random samples of case and control, respectively. Assume that (r_0, r_1, r_2) follows a trinomial distribution with probabilities for genotypes *aa*, *aA* and *AA* equal to p_0, p_1 and p_2 , respectively, and that (s_0, s_1, s_2) follows a trinomial distribution with probabilities q_0, q_1 and q_2 . The null hypothesis of no association can be written as $H_0: p_i = q_i$ for $i = 0, 1, 2$.

Define the penetrances of *aa*, *aA* and *AA* as f_0, f_1 and f_2 , respectively. Let *K* denote the disease prevalence and $\gamma_i = f_i/f_0$ and $\delta_i = (1 - f_i)/(1 - f_0)$ for $i = 1, 2$ be genotype relative risks. The population genotype probabilities will be denoted by $g_0 = \text{Pr}(aa)$, $g_1 = \text{Pr}(aA)$, $g_2 = \text{Pr}(AA)$ and $K = \sum_i f_i g_i$. In the above notation, p_i and q_i can be expressed as

$$p_i = \frac{f_i g_i}{K} = \frac{\gamma_i g_i}{\sum_i \gamma_i g_i} \text{ and } q_i = \frac{(1 - f_i) g_i}{1 - K} = \frac{\delta_i g_i}{\sum_i \delta_i g_i}, \quad (1)$$

where $\gamma_0 = \delta_0 = 1$. Hence, the null hypothesis, $p_i = q_i$ for $i = 0, 1, 2$, is equivalent to $H_0: \gamma_1 = \gamma_2 = 1$ (i.e., $f_0 = f_1 = f_2$). The

Table 1. Genotype distribution

	<i>aa</i>	<i>aA</i>	<i>AA</i>	Total
Cases	r_0	r_1	r_2	<i>R</i>
Controls	s_0	s_1	s_2	<i>S</i>
Total	n_0	n_1	n_2	<i>N</i>

alternative hypothesis H_a can be either $\gamma_2 > \gamma_1 \geq 1$ or $\gamma_2 \geq \gamma_1 > 1$.

To test H_0 against H_a using CA trend test, a set of scores $x = (x_0, x_1, x_2)$ must be assigned to genotypes (*aa*, *aA*, *AA*). From Sasieni [1], $x = (0, 1, 2)$ is assigned for the additive model ($\gamma_2 = 2\gamma_1 - 1$), $x = (0, 1, 1)$ for the dominant model ($\gamma_1 = \gamma_2$), and $x = (0, 0, 1)$ for the recessive model ($\gamma_1 = 1$). Since the multiplicative model ($\gamma_2 = \gamma_1^2$) and the additive model are asymptotically equivalent as (γ_1, γ_2) approaches the null value (1, 1), this model is not studied in detail so one can also use $x = (0, 1, 2)$ for it. Given the score x , the CA trend test can be written as

$$Z_T^* = \frac{U}{[\text{var}_{H_0}(U)]^{1/2}}$$

where

$$U = \frac{1}{N} \sum_{i=0}^2 x_i (S r_i - R s_i)$$

and

$$\text{var}_{H_0}(U) = N \sigma_0^2 = \frac{RS}{N} \left[\sum_{i=0}^2 x_i^2 q_i - \left(\sum_{i=0}^2 x_i q_i \right)^2 \right]. \quad (2)$$

In practice, the q_i in (2) may not be known and σ_0 is estimated from the data by representing the q_i by n_i/N . This yields

$$\hat{\text{var}}_{H_0}(U) = N \hat{\sigma}_0^2 = \frac{RS}{N^3} \left[N \sum_{i=0}^2 x_i^2 n_i - \left(\sum_{i=0}^2 x_i q_i \right)^2 \right]. \quad (3)$$

It can be shown that under the null hypothesis $\hat{\sigma}_0$ is a consistent estimator for σ_0 , i.e., $\hat{\sigma}_0$ converges to σ_0 in probability. Denote $U/[\hat{\text{var}}_{H_0}(U)]^{1/2}$ by Z_T . For a two-sided test, the null hypothesis is rejected if $|Z_T| > z_{1-\alpha/2}$ where z_p is the p th percentile of a standard normal distribution and α is the level of significance.

Power and Sample Size

Denote $E_{H_0}(U) = N\mu_a$ and $\text{var}_{H_0}(U) = N\sigma_a^2$. From Slager and Schaid [5],

$$\mu_a = \frac{RS}{N^2} \sum_{i=0}^2 x_i(p_i - q_i),$$

$$\sigma_a^2 = \frac{RS^2}{N^3} \left[\sum_{i=0}^2 x_i^2 p_i - \left(\sum_{i=0}^2 x_i p_i \right)^2 \right] + \frac{R^2 S}{N^3} \left[\sum_{i=0}^2 x_i^2 q_i - \left(\sum_{i=0}^2 x_i q_i \right)^2 \right]. \quad (4)$$

Under H_0 : $p_i = q_i$, σ_a^2 reduces to σ_0^2 given by (2).

Usually, the sample size needed to achieve power $100(1 - \beta)\%$ for a level α test is calculated when σ_0 is known. The corresponding power and sample size for the trend tests are [5]:

$$\begin{aligned} \text{Power}^* &= \Pr_{H_0}(|Z_T^*| > z_{1-\alpha/2}) \\ &= \Phi\left(\frac{-z_{1-\alpha/2}\sigma_0 - N^{1/2}\mu_a}{\sigma_a}\right) + 1 - \Phi\left(\frac{z_{1-\alpha/2}\sigma_0 - N^{1/2}\mu_a}{\sigma_a}\right), \end{aligned} \quad (5)$$

$$N^* = \left(\frac{z_{1-\alpha/2}\sigma_0 + z_{1-\beta}\sigma_a}{\mu_a}\right)^2, \quad (6)$$

where Φ is the cumulative distribution function of the standard normal. In (5) and (6), the values of p_i and q_i can be calculated using (1).

In practice σ_0 often is unknown so Z_T , which utilizes the estimator $\hat{\sigma}_0$ (3), is the statistic applied to the data. Analogs for formulas (5) and (6) will be given. First, note that, under the alternative hypothesis, $\hat{\sigma}_0$ converges to $(\tilde{\sigma}_a^2 + \mu_a^2)^{1/2}$ in probability (see Appendix, A), where

$$\tilde{\sigma}_a^2 = \frac{RS^2}{N^3} \left[\sum_{i=0}^2 x_i^2 p_i - \left(\sum_{i=0}^2 x_i p_i \right)^2 \right] + \frac{R^2 S}{N^3} \left[\sum_{i=0}^2 x_i^2 q_i - \left(\sum_{i=0}^2 x_i q_i \right)^2 \right].$$

Hence

$$\begin{aligned} \text{Power} &= \Pr_{H_0}(|Z_T| > z_{1-\alpha/2}) \\ &= \Pr_{H_0}\left(\frac{U}{N^{1/2}\hat{\sigma}_0} < -z_{1-\alpha/2}\right) + \Pr_{H_0}\left(\frac{U}{N^{1/2}\hat{\sigma}_0} > z_{1-\alpha/2}\right) \\ &\approx \Pr_{H_0}\left(\frac{U}{N^{1/2}(\tilde{\sigma}_a^2 + \mu_a^2)^{1/2}} < -z_{1-\alpha/2}\right) \\ &\quad + \Pr_{H_0}\left(\frac{U}{N^{1/2}(\tilde{\sigma}_a^2 + \mu_a^2)^{1/2}} > z_{1-\alpha/2}\right) \\ &= \Phi\left(\frac{-z_{1-\alpha/2}(\tilde{\sigma}_a^2 + \mu_a^2)^{1/2} - N^{1/2}\mu_a}{\sigma_a}\right) \\ &\quad + 1 - \Phi\left(\frac{z_{1-\alpha/2}(\tilde{\sigma}_a^2 + \mu_a^2)^{1/2} - N^{1/2}\mu_a}{\sigma_a}\right). \end{aligned}$$

Thus the required sample size is

$$N = \left(\frac{z_{1-\alpha/2}(\tilde{\sigma}_a^2 + \mu_a^2)^{1/2} + z_{1-\beta}\sigma_a}{\mu_a}\right)^2. \quad (7)$$

When $R = S$, $\tilde{\sigma}_a^2 = \sigma_a^2$ and using Taylor series expansion, (7) can be approximated by

$$N \approx \frac{[(z_{1-\alpha/2} + z_{1-\beta})\sigma_a]^2}{\mu_a^2} + z_{1-\alpha/2}(z_{1-\alpha/2} + z_{1-\beta}). \quad (8)$$

To illustrate the difference in the sample sizes obtained from formulas (6) and (7), we consider the same example given in Slager and Schaid [5]. We assume that HWE holds, $R = S$, and that the prevalence of a trait is $K = 0.01$ with $\gamma_1 = 2$, $\gamma_2 = 4$ and $p = 0.05$. Using (2) and (4), we obtain $\mu_a = 0.02285$, $\sigma_0^2 = 0.02375$ and $\sigma_a^2 = 0.03331$ ($\sigma_a^2 > \sigma_0^2$). If we assume $\sigma_0^2 = 0.02375$ is known and use formula (6), the sample size for two-sided test with $\alpha = 0.05$ and power 80% is $N^* = 398$. However, when an estimator $\hat{\sigma}_0$ is used, (7) or (8) yield $N = 506$. The sample sizes in table 2 of Slager and Schaid [5] and their simulations are consistent with formula (7) or (8), rather than (6), and are reliable for planning a study. Further discussion of the accuracy of the approximation (8) is given in the Appendix, B.

Slager and Schaid [5] also calculated the sample sizes of the CA trend test using the additive scores $x = (0, 1, 2)$ for each of the additive, recessive, multiplicative, and dominant models. Since the optimal scores for the recessive and dominant models are (0, 0, 1) and (0, 1, 1), respectively, we recalculate the sample sizes N using the sets of scores x assigned to recessive and dominant models. The results are presented in tables 2 and 3 along the corresponding sample sizes for the additive scores given in Slager and Schaid [5].

For low-prevalence alleles ($p < 0.1$) the loss of power when the additive test is used but the disease is recessive is substantial (table 2). For the dominant model, however, the increased power of the optimal test is small in this situation. This is not surprising as the dominant model essentially reduces to the additive one when p is small [8].

Robust Tests for Candidate Gene Associations

To analyze case-control data (table 1), one can use one of the three trend tests that are optimal for the dominant, additive or recessive models, respectively. If the mode of inheritance is known, the choice of test is clear. In most circumstances, however, the mode is not known. The results in tables 2 and 3 demonstrated that there may be a substantial loss of power when an optimal test for one model is used when the data follow a different model. This occurs primarily when the recessive model is correct. Only when the allele frequency is high (e.g., $p > 0.25$) does this situation arise when the dominant model applies. We consider the optimal tests with their variances estimated from the data using (3) and construct two different robust procedures.

Table 2. Sample sizes N required by the additive (ADD) and recessive (REC) score tests to achieve 80% power ($\alpha = 0.05$) when the recessive model holds

γ_2	K	p	Sample size	
			ADD score	REC score
2	0.01	0.01	16,000,000	460,089
		0.10	18,881	4,684
		0.50	413	298
	0.10	0.01	13,000,000	367,396
		0.10	15,483	3,743
		0.50	339	243
3	0.01	0.01	3,960,000	153,113
		0.10	5,170	1,568
		0.50	151	113
	0.10	0.01	3,270,000	120,118
		0.10	4,212	1,232
		0.50	124	92

ADD scores $x = (0, 1, 2)$; REC scores $x = (0, 0, 1)$.

Table 3. Sample sizes N required by the additive (ADD) and dominant (DOM) score tests to achieve 80% power ($\alpha = 0.05$) when the dominant model holds

γ_1	K	p	Sample size	
			ADD score	DOM score
2	0.01	0.01	2,408	2,396
		0.10	367	347
		0.50	691	422
	0.10	0.01	1,926	1,916
		0.10	298	281
		0.50	576	356
3	0.01	0.01	811	807
		0.10	136	128
		0.50	337	196
	0.10	0.01	638	635
		0.10	109	103
		0.50	282	167

ADD scores $x = (0, 1, 2)$; DOM scores $x = (0, 1, 1)$.

Consider a general case, a set of K alternative models and the corresponding optimal test statistics Z_i , $i = 1, \dots, K$. Given a family (C) of consistent tests with asymptotically normal distributions that includes Z_i for all the models, the one achieving maximin efficiency relative to the other tests in the family is called efficiency robust. Originally, linear combinations of the various optimal test statistics were considered and we call the robust linear combination the MERT. With the advent of the computer era, one can also consider non-linear tests. The maximum of several standardized optimum tests often has higher maximin efficiency than the simpler MERT. Efficiency robust procedures [6, 10] depend on the null correlation of the optimal test statistics as the asymptotic relative efficiency (ARE) of Z_i relative to Z_j when Z_j is optimal is $\text{eff}(Z_i, Z_j) = \rho_{ij}^2$, where $\rho_{ij} = \text{cor}_{H_0}(Z_i, Z_j)$. Thus MERT, Z_{MERT} , a linear combination of several of the optimum test statistics, achieves the

$$\sup_{Z \in C} \inf_{1 \leq i \leq K} \text{eff}(Z, Z_i),$$

where Z_i is the optimal test for the true model, and Z ranges over the family (C) of tests that contains the optimal ones for the three models. The second robust test is the maximum of the optimal statistics, $\max_{1 \leq i \leq K} (Z_i)$, or MAX. Unfortunately, the asymptotic distribution of MAX is not available. Its power and sample size are based on simulation. Freidlin et al. [9] showed that relative per-

formance of MERT and MAX depends on the minimum correlation coefficient, $\rho^* = \min_{1 \leq i, j \leq K} \rho_{ij}$. When $\rho^* \leq 0.6$, MAX is more powerful than Z_{MERT} , but when $\rho^* \geq 0.7$, then two robust tests are virtually equivalent in power. For comparison, we include Z_{MERT} and MAX in the simulation.

Henceforth we focus on the optimal tests for dominant, additive and recessive models denoted by Z_{DOM} , Z_{ADD} and Z_{REC} , respectively (note that the corresponding scores are $x_{\text{DOM}} = (0, 1, 1)$, $x_{\text{ADD}} = (0, 1, 2)$, $x_{\text{REC}} = (0, 0, 1)$). In the Appendix, C, we give the null correlation of the tests and apply the general MERT theory to show that $\rho^* = \text{cor}_{H_0}(Z_{\text{DOM}}, Z_{\text{REC}})$ and

$$Z_{\text{MERT}} = \frac{Z_{\text{DOM}} + Z_{\text{REC}}}{\{2 [1 + \text{cor}_{H_0}(Z_{\text{DOM}}, Z_{\text{REC}})]\}^{1/2}}. \quad (9)$$

In large samples Z_{MERT} has a standard normal distribution and its minimum efficiency for the set of genetic models considered is

$$[1 + \text{cor}_{H_0}(Z_{\text{DOM}}, Z_{\text{REC}})]/2. \quad (10)$$

To assess the robustness of the various tests under the different genetic models, we simulated case-control association studies where each group had 200 (500) members. We chose a low (moderate) prevalence, p , of candidate allele 0.1 (0.3). The variance of each optimal statistic was estimated from (3). The correlations of optimal tests are presented in table 4.

Table 4. Average correlations of optimal test statistics (based on 10,000 replications)

	Z_{DOM}	Z_{ADD}	Z_{REC}
$p = 0.1$	1.000	0.974	0.200
		1.000	0.419
			1.000
$p = 0.3$	1.000	0.909	0.306
		1.000	0.674
			1.000

Table 5. Empirical power for optimal and robust tests ($p = 0.1$, based on 10,000 replications)

$R = S$	Model	Z_{DOM}	Z_{ADD}	Z_{REC}	Z_{MERT}	MAX
200	Null	0.052	0.049	0.056	0.045	0.046
	DOM	0.805	0.785	0.145	0.619	0.729
	ADD	0.794	0.804	0.256	0.716	0.744
	REC	0.175	0.362	0.800	0.575	0.667
500	Null	0.052	0.053	0.051	0.052	0.049
	DOM	0.806	0.785	0.134	0.612	0.730
	ADD	0.806	0.817	0.261	0.727	0.759
	REC	0.162	0.357	0.794	0.614	0.714

Table 6. Empirical power for optimal and robust tests ($p = 0.3$, based on 10,000 replications)

$R = S$	Model	Z_{DOM}	Z_{ADD}	Z_{REC}	Z_{MERT}	MAX
200	Null	0.052	0.051	0.046	0.049	0.050
	DOM	0.792	0.711	0.184	0.631	0.712
	ADD	0.737	0.790	0.470	0.773	0.744
	REC	0.219	0.550	0.802	0.644	0.729
500	Null	0.051	0.050	0.049	0.050	0.050
	DOM	0.805	0.720	0.180	0.634	0.725
	ADD	0.776	0.827	0.507	0.809	0.781
	REC	0.208	0.540	0.795	0.644	0.720

Table 7. Melanoma data from Shahbazi et al. [11]: in situ vs. control

	AA	AG	GG	Total
In situ	6	8	10	24
Control	32	47	20	99
Total	38	55	30	123

Table 8. p Values for various tests for the data in table 7

Test	p Value
Z_{DOM}	0.240
Z_{ADD}	0.044
Z_{REC}	0.014
Z_{MERT} (DOM, ADD, REC)	0.041
MAX (DOM, ADD, REC)	0.031
Z_{MERT} (ADD, REC)	0.020
MAX (ADD, REC)	0.021

Notice that correlations of the standardized optimum tests depend on the prevalence, as both robust tests are functions of this multivariate distribution their distributions will also depend on p . When the correlations are estimated from the data and used in the formula for the MERT or the simulation of the distribution of the MAX, the effect of the prevalence is accounted for. The simulation results confirm that the minimum correlation, ρ^* , of all three optimal tests is $\text{cor}_{H_0}(Z_{DOM}, Z_{REC})$, which is about 0.200 (0.306) for $p = 0.1$ ($p = 0.3$). This indicates that the MAX should be the more robust test.

The simulated power of the tests are given in table 5 ($p = 0.1$) and table 6 ($p = 0.3$). When $p = 0.3$, the levels of tests were quite close to their nominal levels under H_0 . When $p = 0.1$, however, the robust tests tended to be conservative (slightly below 0.05) when $R = S = 200$. In all cases the penetrances were selected so that the optimal test for a given model had power near 0.80. Even when $R = S = 500$, the power of Z_{DOM} (Z_{REC}) when the recessive (dominant) model was true was quite small, < 0.21 . On the other hand, when $R = S = 500$, Z_{ADD} had a power 0.36 (0.54) for recessive traits when $p = 0.1$ (0.3). In contrast, the MAX had a power > 0.70 , while Z_{MERT} had a power > 0.60 for all models. Thus, both robust methods provide protection against model uncertainty, and we recommend the MAX for general use.

An Example

To illustrate the methods, we reanalyze data from Shahbazi et al. [11] showing an association between a variant of the epidermal growth factor (EGF) gene and malignant melanoma. Those investigators examined a candidate allele (G), which produces more EGF than the alternative allele (A). The data for all controls and in situ cases are presented in table 7.

If we have no prior knowledge of the underlying genetic mechanism we would use a robust test based on the three optimal tests Z_{DOM} , Z_{ADD} and Z_{REC} . If prior studies indicated that the mode of inheritance was either additive or recessive, then we would just use a robust test based on Z_{ADD} and Z_{REC} . The estimated correlation matrix obtained from the correlations in the Appendix, C, using the genotype frequencies of the cases and controls combined, were $\text{cor}_{H_0}(Z_{DOM}, Z_{ADD}) = 0.843$, $\text{cor}_{H_0}(Z_{DOM}, Z_{REC}) = 0.380$ and $\text{cor}_{H_0}(Z_{ADD}, Z_{REC}) = 0.817$. The results of the individual optimal tests and the robust procedures are given in table 8. Notice that with the exception of the optimal test for the dominant model all the tests are significant at the 0.05 level. When all three models are plausible, the MAX test yields a somewhat lower p value than the MERT. This is consistent with the guidelines indicating that the MAX test is more powerful than the MERT when the minimum correlation is < 0.7 . When one can eliminate the dominant model based on prior genetic knowledge, the p values of both robust tests are much closer as the correlation of the tests for additive and recessive models is about 0.8.

Discussion

When calculating sample sizes based on formulas similar to (6), which assumes that σ_0 is known, one needs implicitly to account for the fact that σ_0 is estimated by $\hat{\sigma}_0$. Under the alternative, $\hat{\sigma}_0^2$ converges to $\sigma_a^2 + \mu_a^2$, which may exceed σ_0^2 . Thus, formulas (7) or (8) should be used to determine the sample size.

If one knows the mode of inheritance, one should use the optimum test for that model. If the underlying genetic model is not known, then one should use a robust test. Usually, the MAX will be superior to the simpler MERT. However, in situations when the plausible set of models is smaller, e.g., a recessive model is not plausible, one can use the simpler method. Both robust procedures appear to provide more protection against an incorrect choice of model than using the additive scores.

As the sample sizes to achieve prespecified type I and type II errors are calculated under an assumed alternative and use of the corresponding optimal test, when the investigator does not know the specific model underlying the data, a robust test should be used. The sample sizes determined by using the optimal test for a particular model are too small if a different model holds. Hence, they should be increased to allow for the loss of power incurred by model uncertainty. The use of an efficiency robust method helps

to minimize the worst loss of power or needed increase in sample size over the range of plausible models.

Acknowledgments

We would like to thank two referees for their helpful suggestions which greatly improved the manuscript. The research of Z.L. was supported in part by EY14478 of the National Eye Institute and the work of J.L.G. was supported in part by SBR-9807731 of the National Science Foundation.

Appendix

A. Prove that $\hat{\sigma}_0^2$ converges to $\sigma_a^2 + \mu_a^2$ in probability under H_a

Assume $R = \varphi N$ and $S = (1 - \varphi)N$ where $\varphi \in (0, 1)$. Hence, for $i = 0, 1, 2$, as $N \rightarrow \infty$, $n_i/N \rightarrow \varphi p_i + (1 - \varphi)q_i$ in probability. Therefore, under H_a , we obtain

$$\begin{aligned} \hat{\sigma}_0^2 &\rightarrow \varphi(1 - \varphi) \left\{ \sum_i x_i^2 (\varphi p_i + (1 - \varphi)q_i) - \left[\sum_i x_i (\varphi p_i + (1 - \varphi)q_i) \right]^2 \right\} \\ &= \varphi^2(1 - \varphi) \sum_i x_i^2 p_i + \varphi(1 - \varphi)^2 \sum_i x_i^2 q_i - \varphi^3(1 - \varphi) \left(\sum_i x_i p_i \right)^2 \\ &\quad - \varphi(1 - \varphi)^3 \left(\sum_i x_i q_i \right)^2 - 2\varphi^2(1 - \varphi)^2 \left(\sum_i x_i p_i \right) \left(\sum_i x_i q_i \right) \\ &= \sigma_a^2 + \varphi^2(1 - \varphi)^2 \left[\sum_i x_i (p_i - q_i) \right]^2 = \sigma_a^2 + \mu_a^2. \end{aligned}$$

B. Accuracy of expression (8)

Let $R = S$. Under H_a , if $\mu_a^2 \ll \sigma_a^2$, then, by Taylor series expansion, (7) can be approximated by (8). The error of this approximation can be written as

$$-\frac{1}{4} z_{1-\alpha/2} (z_{1-\beta}) \left(\frac{\mu_a}{\sigma_a} \right)^2 + O((\mu_a/\sigma_a)^4), \quad (11)$$

where $O((\mu_a/\sigma_a)^4)$ is the order $(\mu_a/\sigma_a)^4$. Note that if we replace $z_{1-\beta}$ and $z_{1-\alpha/2}$ by 0.842 (power = 80%) and 1.96 ($\alpha = 0.05$), respectively, then the absolute value of the error term in (11) is < 1 if $\mu_a^2/\sigma_a^2 < 2.424$. When the inbreeding coefficient [12, 13] $F = 0$ or when there is inbreeding ($F = 0.10$), we calculate the values of μ_a^2/σ_a^2 for three different genetic models with $K = 0.01, 0.10, p = 0.01, 0.10, 0.50, R = S$, and the γ 's, which were considered in [5]. The largest value of μ_a^2/σ_a^2 is 0.0912, corresponding to the recessive model with $\gamma_2 = 3, K = 0.10$ and $p = 0.50$ and $F = 0$. Thus (8) is a very good approximation for calculating the required sample size when σ_0 is not known.

From (6) and (7), if $R = S$ and $\sigma_a^2 + \mu_a^2 = \sigma_0^2$, then $N^* = N$. However, it can be shown (results not presented here) that, under HWE and a recessive model ($\gamma_1 = 1$ and $\delta_1 = 1$) with $p \leq 0.50, \sigma_a > \sigma_0$ when $\gamma_2 < 9$ and $f_0 + f_2 < 1$. That is, $N > N^*$. For the dominant model ($\gamma_1 = \gamma_2 = \gamma$ and $\delta_1 = \delta_2 = \delta$), $\Delta = \gamma\delta > 1$ when $f_0 + f_2 < 1$. Under HWE, it can be shown (results not presented here) that, if $\Delta = 2$ (and $\Delta = 3$), $\sigma_a > \sigma_0$ if and only if $p < 0.235$ (and $p < 0.204$).

C. The extreme pair and the MERT

For a family of three optimal tests, often the MERT is a linear combination of the two optimal tests for the extreme members of the family, i.e., the tests with the minimum correlation. Two conditions need to be satisfied [6, 8]: (i) $\rho^* \geq 0$, and (ii) $\rho^* + 1$ must be less than the sum of the other two correlations. We now show that for the family of optimal test statistics Z_{DOM} , Z_{ADD} and Z_{REC} and the corresponding U_{DOM} , U_{ADD} and U_{REC} [4], Z_{MERT} can be expressed as (9) with ARE given by (10), where Z_{MERT} asymptotically follows a standard normal distribution.

Since, under H_0 , (r_0, r_1, r_2) and (s_0, s_1, s_2) follow the same trinomial distribution, we obtain:

$$\begin{aligned} \text{cor}_{H_0}(Z_{\text{DOM}}, Z_{\text{ADD}}) &= \text{cor}_{H_0}(U_{\text{DOM}}, U_{\text{ADD}}) \\ &= \frac{p_2(p_1 + 2p_0)}{[p_2(1 - p_2)]^{1/2}[(p_1 + 2p_2)p_0 + (p_1 + 2p_0)p_2]^{1/2}} \\ \text{cor}_{H_0}(Z_{\text{ADD}}, Z_{\text{REC}}) &= \text{cor}_{H_0}(U_{\text{ADD}}, U_{\text{REC}}) \\ &= \frac{p_0(p_1 + 2p_2)}{[p_0(1 - p_0)]^{1/2}[(p_1 + 2p_2)p_0 + (p_1 + 2p_0)p_2]^{1/2}}, \\ \text{cor}_{H_0}(Z_{\text{DOM}}, Z_{\text{REC}}) &= \text{cor}_{H_0}(U_{\text{DOM}}, U_{\text{REC}}) \\ &= \frac{p_0p_2}{[p_0(1 - p_0)]^{1/2}[p_2(1 - p_2)]^{1/2}}. \end{aligned}$$

Using the above expressions and routine algebra, one can show that

$$\begin{aligned} \text{cor}_{H_0}(U_{\text{DOM}}, U_{\text{REC}}) &< \text{cor}_{H_0}(U_{\text{DOM}}, U_{\text{ADD}}), \\ \text{cor}_{H_0}(U_{\text{DOM}}, U_{\text{REC}}) &< \text{cor}_{H_0}(U_{\text{ADD}}, U_{\text{REC}}), \end{aligned}$$

which implies (i).

To show (ii), let F be inbreeding coefficient. Under the null hypothesis, $p_i = q_i = g_i$, $i = 0, 1, 2$. Hence, we can write $p_0 = qF +$

$(1 - F)q^2$, $p_1 = 2pq(1 - F)$ and $p_2 = pF + (1 - F)p^2$. Then $p_1 + 2p_2 = 2p$, $p_1 + 2p_0 = 2q$ and $(p_1 + 2p_2)p_0 + (p_1 + 2p_0)p_2 = 2pq(1 + F)$. Denote $I(p) = p + pq(1 - F)$. Hence (ii) is equivalent to

$$\begin{aligned} \frac{2p[p_0]^{1/2}}{[I(p)]^{1/2}} + \frac{2q[p_2]^{1/2}}{[I(q)]^{1/2}} &\geq [2pq(1 + F)]^{1/2} \left\{ 1 + \frac{[p_0p_2]^{1/2}}{[I(p)I(q)]^{1/2}} \right\} \\ \Leftrightarrow 2p[p_0I(q)]^{1/2} + 2q[p_2I(p)]^{1/2} &\geq [2pq(1 + F)]^{1/2} \{ [I(p)I(q)]^{1/2} + \\ &[p_0p_2]^{1/2} \} \\ \Leftrightarrow 4p^2p_0I(q) + 4q^2p_2I(p) + 8pq[p_0p_2I(p)I(q)]^{1/2} \\ &\geq 2pq(1 + F) \{ I(p)I(q) + p_0p_2 + 2[p_0p_2I(p)I(q)]^{1/2} \}. \end{aligned} \quad (12)$$

Denote $[p_0p_2I(p)I(q)]^{1/2}$ by h . Then, from (12), (ii) is equivalent to

$$\begin{aligned} 2pI(q) [2pp_0 - q(1 + F)I(p)] + 2qp_2 [2qI(p) - p(1 + F)p_0] \\ + 4pq(1 - F)h \geq 0. \end{aligned} \quad (13)$$

where, after some algebra, $2pp_0 - q(1 + F)I(p) = pq[(1 - F)^2q - (1 - F)]$ and $2qI(p) - p(1 + F)p_0 = pq[2 - (1 + F)/F + (1 - F)^2q]$. Substituting the above two terms into (13), we obtain that (13) is equivalent to

$$4pq(1 - F)h + 2p^2q^2 f(F, p) \geq 0,$$

where

$$\begin{aligned} f(F, p) &= [1 + p(1 - F)][(1 - F)^2q - (1 - F)] \\ &+ [F + (1 - F)p][2 - (1 - F)F + (1 - F)^2q] \\ &= F^3(2 - 4p + 2p^2) - F^2(2 - 8p + 6p^2) + F(2 - 6p + 6p^2). \end{aligned}$$

Hence, to show (ii), it remains to show $f(F, p) \geq 0$ for $0 \leq F, p \leq 1$. Note that, for any $0 \leq F \leq 1$, $(\partial^2/\partial p^2)f(F, p) = -4(1 - F)^3 \leq 0$. Thus $f(F, p)$ is concave with respect to p , this is, $f(F, p) \geq \min \{ f(F, 0), f(F, 1) \}$, where $f(F, 0) = 2F(F^2 - F + 1) \geq 0$ and $f(F, 1) = 2F \geq 0$. This shows that (ii) holds.

References

- Sasieni PD: From genotypes to genes: Doubling the sample size. *Biometrics* 1997;53: 1253-1261.
- Cochran WG: Some methods for strengthening the common chi-squared tests. *Biometrics* 1954;10:417-451.
- Armitage P: Tests for linear trends in proportions and frequencies. *Biometrics* 1955;11: 375-386.
- Tarone RE, Gart JJ: On the robustness of combined tests for trends in proportions. *J Am Stat Assoc* 1980;75:110-116.
- Slager SL, Schaid DJ: Case-control studies of genetic markers: Power and sample size approximations for Armitage's test for trend. *Hum Hered* 2001;52:149-153.
- Gastwirth JL: The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *J Am Stat Assoc* 1985; 80:380-384.
- Podgor MJ, Gastwirth JL, Mehta CR: Efficiency robust tests of independence in contingency tables with ordered classifications. *Stat Med* 1996;15:2095-2105.
- Whittemore AS, Tu I-P: Simple, robust linkage tests for affected sibs. *Am J Hum Genet* 1998; 62:1228-1242.
- Freidlin B, Podgor MJ, Gastwirth JL: Efficiency robust tests for survival or ordered categorical data. *Biometrics* 1999;55:883-886.
- Shih M, Whittemore AS: Allele-sharing among affected relatives: Non-parametric methods for identifying genes. *Stat Methods Med Res* 2001; 10:27-55.
- Shahbazi M, Pravica P, Nasreen N, Fakhoury H, Fryer AA, Strange RC, Hutchinson PE, Osborne JE, Lear JT, Smith AG, Hutchinson IV: Association between functional polymorphism in EGF gene and malignant melanoma. *Lancet* 2002;359:397-401.
- Vogel F, Motulsky AG: *Human Genetics: Problems and Approaches*, ed 3. New York, Springer, 1997.
- Ott J: *Analysis of Human Genetic Linkage*, ed 3. Baltimore, JHU Press, 1999.

Material may be protected by copyright law (Title 17, U.S. Code)