HARDY-WEINBERG EQUILIBRIUM ASSUMPTIONS IN CASE-CONTROL TESTS OF GENETIC ASSOCIATION

by

Myoungkeun Lee

BS, Korea University, Korea, 1999

MS, Catholic University of Korea, Korea, 2001

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Myoungkeun Lee

It was defended on

July 17, 2009

and approved by

Gong Tang, Ph.D.
Assistant Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Candace M. Kammerer, Ph.D.
Associate Professor
Department of Human Genetics
Graduate School of Public Health
University of Pittsburgh

Thesis Director: Eleanor Feingold, Ph.D.
Associate Professor
Departments of Human Genetics and Biostatistics
Graduate School of Public Health
University of Pittsburgh

HARDY-WEINBERG EQUILIBRIUM ASSUMPTIONS IN CASE-CONTROL TESTS OF GENETIC ASSOCIATION

Myoungkeun Lee, M.S.

University of Pittsburgh, 2009

The case-control study design is commonly used in genetic association study with a binary trait using unrelated individuals from a population. To test association with a binary trait in a case-control or cohort study, the standard analysis is a chi-square test or logistic regression model that test to detect a difference in frequencies of alleles or genotypes.

In this thesis, we derive the maximum likelihood estimator, using Chen and Chatterjee's methods, for standard 1 df genetic tests (dominant, recessive, and multiplicative). We then compare these methods that assume HWE with standard Wald tests and chi-squared tests that do not make the HWE assumption. We consider four different HWE scenarios: 1) when HWE holds in both cases and controls, 2) when HWE does not hold in cases and controls follow HWE, 3) when HWE does not hold in controls, and cases follow HWE and 4) when HWE does not hold in either cases or controls.

Our results show that the performances of the three statistics (chi-squared, Wald, and Chen and Chatterjee Wald) are equivalent for multiplicative test under all four HWE scenarios. When HWE holds in both cases and controls, the performances of the three statistics are also equivalent, except for variations attributable to type I error issues. When HWE fails to hold in either cases or controls or both, the 2 df version of the Chan and Chatterjee Wald test (and to a lesser extent the dominant and recessive versions) detects this HWE departure and can therefore "find" a case-control difference even if there is not an allele frequency difference or even a genotype frequency difference. Our results will improve the design and analysis of genetic association studies. Such association studies are a crucial step in understanding the genetic components of many diseases that have a large impact on public health. Better understanding of the etiology of these diseases will lead in the long term to better prevention and treatment strategies.

TABLE OF CONTENTS

LIST OF TABLES

PREFACE

I would like to thank my advisor, Dr. Eleanor Feingold and the committees for their help and support in producing this thesis. I would also like to thank the Department of Biostatistics for their support both financial and otherwise. Finally, I would like to thank my family for all their love and support.

1.0 INTRODUCTION

Genetic association studies are used to test whether phenotypic traits are associated with allele frequencies or genotype frequencies. Genetic association tests can be conducted using unrelated individuals from a population, or using families. Association studies can also use a binary trait. To test association with a binary trait using unrelated individuals (a case-control or cohort study), the standard analysis is a chi-square test or logistic regression model that tests to detect the difference of the frequencies of alleles, or those of genotypes between case and control at the maker locus using unrelated subjects. The cases have been diagnosed with the disease (or non-disease trait) under study, and the controls are randomly selected from the population ("population controls"), or are chosen as "true controls" who are known not to have the disease or trait**.**

In case-control testing, it is often assumed that Hardy-Weinberg equilibrium (HWE) should hold in the overall population or in the controls. This assumption is justified either by assuming that controls are from population controls, or by assuming that controls are from true controls when a disease is rare in the population. In practice, most deviations from HWE are interpreted as evidence of genotyping error, but it is also possible for them to represent real genetic effects due to selection, non-random mating, inbreeding, small population size, population stratification, etc.

The standard chi-squared test and Wald test do not require HWE assumptions in case-control studies. Chen and Chatterjee (2007) proposed a likelihood-based test that assumes HWE in the controls. (Satten and Epstein (2004) also described likelihood-based tests for haplotypes that assume HWE holds in the controls.) However, there are many unresolved issues in Chen and Chatterjee's work. First, Chen and Chatterjee (2007) only considered tests with 2 degrees of freedom. Second, they did not comprehensively examine the behaviors of the statistics when HWE holds and when it does not. An important issue here is whether we want to detect genotype differences and/or departures from HWE, or just allele differences.

To evaluate which methods have better performance, we conducted simulation studies to compare with Chi-squared tests, Wald test, and Chen and Chatterjee’s Wald test when HWE holds and when it does not. In order to do this, we first derived 1 df versions (dominant, recessive, and multiplicative, as described below) of the Chen and Chatterjee test. We considered four different HWE scenarios, as follows: 1) when HWE holds in both cases and controls, 2) when HWE does not hold in both cases and controls, 3) when HWE does not hold in cases and controls follow HWE, 4) when HWE does not hold in controls, and cases follow HWE. We also considered in each HWE scenario the situation in which the allele frequency is the same in cases and controls (the null hypothesis, in some sense), and the situation in which the allele frequency is different in cases and controls. These simulations allow us to answer the following questions. First, how much power is gained by the Chen and Chatterjee approach when the HWE assumption is correct, and second, what can go wrong if the assumption is incorrect.

2.0 METHODS

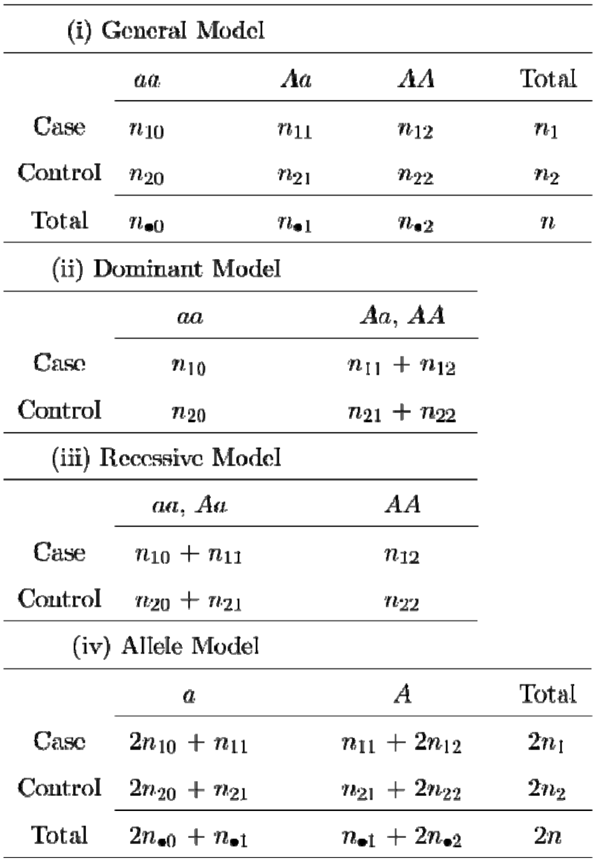
2.1 NON-LIKELIHOOD METHODS

The standard analysis method for a case-control study is a chi-squared test on a contingency table as shown in Table 1. There are several different chi-squared tests that are commonly used. The classical case-control association studies may be analyzed by chi-square test with 2 degrees of freedom (df) using 2×3 contingency table under the null hypothesis of no association. The chi-square test statistic is

$$\chi^2_{df=2} = \sum_{i=1}^{2}\sum_{j=0}^{2}\left(n_{ij} - \frac{n_i n_{\bullet j}}{n}\right)^2 \Bigg/ \frac{n_i n_{\bullet j}}{n}$$

The chi-square test with 2 degrees of freedom makes no assumptions based on genetic models. It is also common to form tests that have higher power under particular genetic models. For a "dominant" test, we merge the **Aa** and **AA** genotype frequencies, so that a chi-square test with 1 df ($\chi^2_{dom,df=1}$) can use a 2×2 contingency table. A "recessive" test also can be formed as a chi-square test with 1 df ($\chi^2_{rec,df=1}$) by pooling the **aa** and **Aa** genotype frequencies. An "additive" or "multiplicative" test is the Cochran-Armitage (CA) trend test with weights equal to the number of **A** alleles, $x_i =$ (0,1,2) (Cochran, 1954, Armitage, 1955).

Table 1. Contingency table for Case-Control



(i) General Model				
	aa	Aa	AA	Total
Case	n_{10}	n_{11}	n_{12}	n_1
Control	n_{20}	n_{21}	n_{22}	n_2
Total	$n_{\bullet 0}$	$n_{\bullet 1}$	$n_{\bullet 2}$	n
(ii) Dominant Model				
	aa	Aa, AA		
Case	n_{10}	$n_{11} + n_{12}$		
Control	n_{20}	$n_{21} + n_{22}$		
(iii) Recessive Model				
	aa, Aa	AA		
Case	$n_{10} + n_{11}$	n_{12}		
Control	$n_{20} + n_{21}$	n_{22}		
(iv) Allele Model				
	a	A	Total	
Case	$2n_{10} + n_{11}$	$n_{11} + 2n_{12}$	$2n_1$	
Control	$2n_{20} + n_{21}$	$n_{21} + 2n_{22}$	$2n_2$	
Total	$2n_{\bullet 0} + n_{\bullet 1}$	$n_{\bullet 1} + 2n_{\bullet 2}$	$2n$	

The formula for the additive/multiplicative test is as follows.

$$\chi^2_{CA,df=1} = \frac{n\left(\sum_{j=0}^{2} x_j\left(n_1 n_{1j} - n_2 n_{2j}\right)\right)^2}{n_1 n_2\left(n\sum_{j=0}^{2} x_j^2 n_{\bullet j} - \left(\sum_{j=0}^{2} x_j n_{\bullet j}\right)^2\right)}.$$

Another common test is the allele test, which is a chi-square test with 1 df ($\chi^2_{allele,df=1}$) based on the 2×2 contingency table that counts the numbers of **A** and **a** alleles but doubles the total sample size as compared to other tests. The allele test is asymptotically equivalent to the multiplicative test, under certain HWE assumptions (Sasieni, 1997). In general, the allele test is less appropriate, and will not be discussed further in this thesis.

2.2 LIKELIHOOD METHODS (WALD TEST)

For each of the chi-squared tests discussed above, an asymptotically equivalent likelihood-based (Wald) test can be constructed based on the maximum likelihood estimates of the odds ratios for each genotype using the retrospective likelihood function. The estimated odds ratios, based on the notation used in Table 1, are shown as

$$\hat{\psi}_{Aa} = \frac{n_{11}n_{20}}{n_{21}n_{10}} \text{ and } \hat{\psi}_{AA} = \frac{n_{12}n_{20}}{n_{22}n_{10}}.$$

The asymptotic variance-covariance matrix for the logarithm of odds ratios are given by

$$V_{\log\hat{\psi}_{Aa},\log\hat{\psi}_{AA}} = \begin{bmatrix} \frac{1}{n_{10}} + \frac{1}{n_{11}} + \frac{1}{n_{20}} + \frac{1}{n_{21}} & \frac{1}{n_{10}} + \frac{1}{n_{20}} \\ \frac{1}{n_{10}} + \frac{1}{n_{20}} & \frac{1}{n_{10}} + \frac{1}{n_{12}} + \frac{1}{n_{20}} + \frac{1}{n_{22}} \end{bmatrix},$$

and the standard Wald statistic has a chi-square test with 2 df under the null hypothesis of no association, $\psi_{Aa} = \psi_{AA} = 1$,

$$W_{\log\hat{\psi}_{Aa},\log\hat{\psi}_{AA}} = (\log\hat{\psi}_{Aa}, \log\hat{\psi}_{AA})(V_{\log\hat{\psi}_{Aa},\log\hat{\psi}_{AA}})^{-1}(\log\hat{\psi}_{Aa}, \log\hat{\psi}_{AA})^T.$$

The standard Wald tests can be modified as 1-df tests for specific models, similarly to the chi-squared tests. The estimated odds ratios and the asymptotic variance can be estimated by using logistic regression. Here the Wald test for testing the hypothesis of no association between the marker and disease with 1 df can be a dominant test ($\psi_{Aa} = \psi_{AA} = \psi$), a recessive test ($\psi_{Aa} = 1$, $\psi_{AA} = \psi$), or a multiplicative test ($\psi_{Aa} = \psi$, $\psi_{AA} = \psi^2$).

2.3 LIKELIHOOD METHODS (CC WALD TEST)

Chen and Chatterjee (2007) proposed a modified Wald test with 2df that assumes the controls are in HWE. The assumption of HWE for the controls can be justified when the population is under HWE and the disease is rare. The minor allele frequency (MAF, p) of controls is estimated as $\hat{p} = (2n_{22} + n_{21})/2n_2$, using the observed genotype frequencies in the control group, and then the expected genotype frequencies of controls under HWE are computed as follows: $n_{20}^E = n_2(1-\hat{p})^2$, $n_{21}^E = 2n_2\hat{p}(1-\hat{p})$, and $n_{22}^E = n_2\hat{p}^2$. The estimated odds ratios ($\log\hat{\psi}_{Aa}^E$ and $\log\hat{\psi}_{AA}^E$) and the asymptotic variance-covariance matrix ($V_{\log\hat{\psi}_{Aa}^E, \log\hat{\psi}_{AA}^E}$) are then calculated by using the expected genotype frequencies of controls instead of the observed genotype frequencies as follows.

$$\hat{\psi}_{Aa}^E = \frac{n_{11}n_{20}^E}{n_{21}^E n_{10}} \text{ and } \hat{\psi}_{Aa}^E = \frac{n_{12}n_{20}^E}{n_{22}^E n_{10}},$$

$$V_{\log\hat{\psi}_{Aa}^E, \log\hat{\psi}_{AA}^E} = \begin{bmatrix} \frac{1}{n_{10}} + \frac{1}{n_{11}} + \frac{1}{2n_{20}^E + n_{21}^E} + \frac{1}{2n_{22}^E + n_{21}^E} & \frac{1}{n_{10}} + \frac{2}{n_{21}^E} \\ \frac{1}{n_{10}} + \frac{2}{n_{21}^E} & \frac{1}{n_{10}} + \frac{1}{n_{12}} + \frac{4}{2n_{22}^E + n_{21}^E} + \frac{4}{2n_{20}^E + n_{21}^E} \end{bmatrix},$$

Their proposed Wald statistic has a chi-square test with 2 df assuming the controls under HWE as shown in

$$W_{\log\hat{\psi}_{Aa}^E, \log\hat{\psi}_{AA}^E}^E = (\log\hat{\psi}_{Aa}^E, \log\hat{\psi}_{AA}^E)(V_{\log\hat{\psi}_{Aa}^E, \log\hat{\psi}_{AA}^E})^{-1}(\log\hat{\psi}_{Aa}^E, \log\hat{\psi}_{AA}^E)^T.$$

Chen and Chatterjee (2007) showed that their proposed 2 df Wald test had more power than the standard 2 df Wald test because the difference between an asymptotic variance-covariance matrix of their test and that of the standard Wald test was asymptotically negative definite and the non-centrality parameter of their proposed Wald test was greater than the standard Wald test. Their test gained efficiency under a recessive disease test and made no difference under a multiplicative test. They also noted that the test would be expected to have inflated type I error under the null hypothesis when the HWE assumption does not hold, but they did not comprehensively consider the behavior under various HWE scenarios.

Chen and Chatterjee type Wald tests can be modified as 1-df tests for specific models, analogously to the chi-squared tests. The estimated odds ratios and the asymptotic variance can be estimated by using the expected genotype frequencies of controls instead of the observed genotype frequencies as follows,

$$\hat{\psi}^{E}_{Dom} = \frac{(n_{11}+n_{12})n^{E}_{20}}{n_{10}(n^{E}_{21}+n^{E}_{22})} \text{ , and } V_{\log\hat{\psi}^{E}_{Dom}} = \frac{1}{n_{10}} + \frac{1}{n_{11}+n_{12}} + \frac{2\hat{p}}{n_2(1-\hat{p})(2\hat{p}-\hat{p}^2)^2} \text{ ,}$$

$$\hat{\psi}^{E}_{\text{Rec}} = \frac{n_{12}(n^{E}_{20}+n^{E}_{21})}{(n_{10}+n_{11})n^{E}_{22}} \text{ , and } V_{\log\hat{\psi}^{E}_{\text{Rec}}} = \frac{1}{n_{10}+n_{11}} + \frac{1}{n_{12}} + \frac{1-\hat{p}}{2n_2\hat{p}(1-\hat{p}^2)^2} \text{ ,}$$

$$\hat{\psi}^{E}_{Mul} = \frac{(n_{11}+2n_{12})n^{E}_{21}}{(2n_{10}+n_{11})2n^{E}_{22}} \text{ , and } V_{\log\hat{\psi}^{E}_{Mul}} = \frac{1}{2n_{10}+n_{11}} + \frac{1}{n_{11}+2n_{12}} + \frac{1}{2n_2\hat{p}(1-\hat{p})} \text{ .}$$

Here the Wald test for testing the hypothesis of no association between the marker and disease with 1 df can be a dominant test ($\psi_{Aa} = \psi_{AA} = \psi$), a recessive test ($\psi_{Aa} = 1,\ \psi_{AA} = \psi$), or a multiplicative test ($\psi_{Aa} = \psi,\ \psi_{AA} = \psi^2$).

3.0 SIMULATION

We conducted simulation studies to compare the behavior of the Chi-squared test, Wald test, and Chen and Chatterjee's Wald test. For each test, we considered the 2 df test and the three 1-df tests previously discussed (dominant, recessive, multiplicative). We also considered four different HWE scenarios: 1) HWE for both case and control, 2) Not HWE for both case and control, 3) Not HWE for case and HWE for control, 4) HWE for case and Not HWE for control. For each HWE scenario, we considered the null hypothesis situation of no difference in allele frequencies between cases and controls as well as the alternative hypothesis.

We constructed the four HWE scenarios as follows. When the violation of HWE is excess heterozygotsity, we subtract small amounts (**e**=0.01, 0.02), $(\max[-p^2, -(1-p)^2] \leq \mathrm{e} \leq p(1-p))$, from the homozygous genotype frequencies, **aa** and **AA**, and add 2**e** to the heterozygous genotype frequency **Aa**. When violation of HWE is excess homozygosity, we add **e** in the homozygous genotypes, **aa** and **AA**, and subtract 2**e** in the heterozygous genotype **Aa**.

Table 2. Genotype frequencies when not HWE

Genotype	excess heterozygosity	excess homozygosity
aa	$(1-p)^2 - e$	$(1-p)^2 + e$
Aa	$2p(1-p) + 2e$	$2p(1-p) - 2e$
AA	$p^2 - e$	$p^2 + e$

Our simulated data used equal sample sizes of 200 and 300 in both case and control with 10,000 replicates. For each HWE scenario, we considered the same MAF=0.3 in both case and control and MAF=0.3 in case and MAF=0.2 in control. We show the trinomial distribution probability of the genotypes in Table 3. The frequencies of genotypes are generated from the trinomial distributions in both cases and controls. The odds ratios for the model implied by the allele frequencies above are ψ_{Aa} = 1.71 and ψ_{AA} = 2.94.

For each replicate dataset, we performed all genetic tests: Chi-squared test, Wald test, and Chen and Chatterjee's Wald test, with both 1-df and 2-df versions of each. The estimated rejection probabilities rates are the proportion of replicated data in which the p-value of each statistics is less than the nominal level 0.05 under the null hypothesis or under the alternative hypothesis.

Table 3. Trinomial distribution probability of the frequencies of genotypes (aa, Aa, AA)

HWE			e=0.01		e=0.02	
Case	Control	MAF	Case	Control	Case	Control
HWE	HWE	ca:0.3 co:0.3	(0.49,0.42,0.09)	(0.49,0.42,0.09)	(0.49,0.42,0.09)	(0.49,0.42,0.09)
		ca:0.3 co:0.2	(0.49,0.42,0.09)	(0.64,0.32,0.04)	(0.49,0.42,0.09)	(0.64,0.32,0.04)
not HWE	not HWE	ca:0.3 co:0.3	(0.48,0.44,0.08)	(0.48,0.44,0.08)	(0.47,0.46,0.07)	(0.47,0.46,0.07)
		ca:0.3 co:0.2	(0.48,0.44,0.08)	(0.63,0.34,0.03)	(0.47,0.46,0.07)	(0.62,0.36,0.02)
not HWE	HWE	ca:0.3 co:0.3	(0.48,0.44,0.08)	(0.49,0.42,0.09)	(0.47,0.46,0.07)	(0.49,0.42,0.09)
		ca:0.3 co:0.2	(0.48,0.44,0.08)	(0.64,0.32,0.04)	(0.47,0.46,0.07)	(0.64,0.32,0.04)
HWE	not HWE	ca:0.3 co:0.3	(0.49,0.42,0.09)	(0.47,0.46,0.07)	(0.49,0.42,0.09)	(0.47,0.46,0.07)
		ca:0.3 co:0.2	(0.49,0.42,0.09)	(0.63,0.34,0.03)	(0.49,0.42,0.09)	(0.62,0.36,0.02)

4.0 RESULTS

We performed simulations to compare rejection probabilities of three statistics: Chi-squared tests, Wald test, and Chen and Chatterjee's Wald test, with 4 different versions of each test (2 df, dominant, multiplicative, recessive). In Table 4 and Table 5, we present simulation results for the situation of excess heterozygosity (when there is an HWE deviation). The only difference between Table 4 and Table 5 is the sample size in each dataset.

Under HWE for both cases and controls, when the same MAF=0.3 in both case and control, Chen and Chatterjee's Wald test with 2 df and the Wald test with 2 df, and Chi-squared test with 2 df have rejection probabilities under the nominal level (0.05) with no genetic assumption, and those of 1df tests have similar rejection probabilities for the dominant test and for the multiplicative test. But the rejection probability in Chen and Chatterjee's methods is inflated in the recessive test. When the MAF=0.3 for case and MAF=0.2 for control, the 2 df tests have the same rejection probability, and the 1 df tests have similar rejection probabilities for the dominant test and the same rejection probability for the multiplicative test. The Wald test has more rejection probability than the chi-squared test under the recessive test. When HWE holds for both cases and controls, the performances of three statistics are equivalent, except that Chen and Chatterjee's Wald test has the highest rejection probability with MAF=0.3 for case and 0.2 for control for the recessive test, presumably because of the inflated rejection probability with the same MAF=0.3 in both case and control.

Under no HWE for both cases and controls, when the same MAF=0.3 in both case and control, Chen and Chatterjee’s methods are more inflated rejection probabilities for the recessive test, and also more inflated in 2 df tests than the Wald test and Chi-squared test. All multiplicative tests are similar. When the deviation of HWE (e) is large, Chen and Chatterjee’s Wald tests increase but Chi-squared tests and Wald tests remain the same rejection probabilities for the dominant and recessive tests, and no genetic test. When the MAF=0.3 for cases and MAF=0.2 for controls, Wald test and chi-squared test are similar with no genetic assumption, and are similar in dominant and multiplicative tests. The Wald test has more rejection probability than chi-squared test for the recessive test. When HWE does not hold for cases and controls, the performances of three statistics are equivalent for multiplicative test, but Wald test shows better performances than chi-squared test. Note that in this scenario our cases and controls are out of HWE by exactly the same amount. Of course in a realistic scenario the amount of deviation from HWE may not be identical in cases and controls.

Under no HWE for cases and HWE for controls, when the same MAF=0.3 in both case and control, all the multiplicative tests are similar. When the deviation of HWE (e) is large, the rejection probabilities are inflated in the dominant test, the recessive test, and the 2df test, but the multiplicative test remains the same when the deviation of HWE (e) is large. If we consider the deviation from HWE to be a difference that we want to detect, the 2 df tests detect a violation of HWE for all 3 tests. If we consider that the deviation from HWE must be some other kind of error that we don't want to detect, the multiplicative test does not detect the deviation from

HWE, whether we use chi-squared, Wald, or CC Wald. When the MAF=0.3 for case and MAF=0.2 for control, the performances of 3 statistics are equivalent for the multiplicative test.

Under HWE for cases and no HWE for controls, when the same MAF=0.3 in both case and control, the multiplicative tests are similar to each other. When the deviation of HWE (e) is large, the rejection probabilities are inflated in the Wald test and Chi-squared test for the dominant test, for the recessive test, and with no genetic assumption but Chen and Chatterjee's Wald test remains the same. If we consider the deviation from HWE to be a difference that we want to detect, the chi-square and Wald detect a violation of HWE for all 3 tests. When the MAF=0.3 for case and MAF=0.2 for control, 1df tests are similar under the dominant model and for the multiplicative model. The performances of 3 statistics are equivalent for multiplicative test and 2df for no genetic assumption, and Chen and Chatterjee's Wald tests, 1df for the dominant test has the highest rejection probability.

In Table 6, the simulation results are shown when the deviation of HWE (e) is 0.01 where there are more homozygous genotypes than expected under HWE. The results are qualitatively similar to those in Table 4 and Table 5.

In Table 7, we show the mean of estimated OR in the standard and Chen and Chatterjee's methods when sample size is 200, and the deviation of HWE (e) is 0.01 where there are more heterozygous genotypes than expected under HWE. When controls are not in HWE, the mean of the estimated OR in Chen and Chatterjee's method is the same when HWE holds in both cases and controls because the estimated MAF of controls based on the observed genotype frequencies when HWE does not hold are the same as the observed genotype frequencies when HWE holds.

Table 4. Probability of rejecting hypotheses when excess heterozygote (n=200)

HWE				e=0.01			e=0.02		
Case	Control	MAF	Test	Chisq	Wald	ccWald	Chisq	Wald	ccWald
HWE	HWE	ca:0.3 co:0.3	Dom	0.040	0.050	0.048			
			Rec	0.031	0.051	0.069			
			Mul	0.047	0.048	0.047			
			Geno	0.049	0.051	0.045			
		ca:0.3 co:0.2	Dom	0.832	0.856	0.880			
			Rec	0.454	0.542	0.792			
			Mul	0.900	0.901	0.902			
			Geno	0.838	0.841	0.844			
not HWE	not HWE	ca:0.3 co:0.3	Dom	0.039	0.050	0.050	0.043	0.056	0.067
			Rec	0.030	0.052	0.081	0.030	0.051	0.140
			Mul	0.050	0.050	0.044	0.052	0.052	0.041
			Geno	0.046	0.048	0.069	0.053	0.055	0.180
		ca:0.3 co:0.2	Dom	0.833	0.856	0.918	0.836	0.859	0.951
			Rec	0.535	0.631	0.669	0.606	0.717	0.479
			Mul	0.921	0.922	0.912	0.937	0.937	0.921
			Geno	0.870	0.873	0.868	0.901	0.906	0.911
not HWE	HWE	ca:0.3 co:0.3	Dom	0.047	0.060	0.060	0.061	0.073	0.075
			Rec	0.044	0.071	0.081	0.079	0.113	0.145
			Mul	0.053	0.053	0.049	0.052	0.052	0.048
			Geno	0.067	0.069	0.074	0.128	0.136	0.189
		ca:0.3 co:0.2	Dom	0.875	0.893	0.916	0.916	0.931	0.946
			Rec	0.314	0.400	0.653	0.193	0.262	0.475
			Mul	0.909	0.910	0.903	0.920	0.920	0.912
			Geno	0.852	0.854	0.858	0.876	0.878	0.898
HWE	not HWE	ca:0.3 co:0.3	Dom	0.042	0.054	0.042	0.058	0.071	0.046
			Rec	0.042	0.066	0.065	0.082	0.118	0.067
			Mul	0.046	0.046	0.044	0.051	0.051	0.045
			Geno	0.064	0.067	0.041	0.125	0.128	0.045
		ca:0.3 co:0.2	Dom	0.782	0.810	0.886	0.722	0.752	0.888
			Rec	0.668	0.749	0.802	0.852	0.908	0.809
			Mul	0.914	0.915	0.909	0.926	0.926	0.916
			Geno	0.872	0.876	0.855	0.928	0.933	0.856

Table 5. Probability of rejecting hypotheses when excess heterozygote (n=300)

HWE				e=0.01			e=0.02		
Case	Control	MAF	Test	Chisq	Wald	ccWald	Chisq	Wald	ccWald
HWE	HWE	ca:0.3 co:0.3	Dom	0.039	0.051	0.046			
			Rec	0.035	0.052	0.072			
			Mul	0.047	0.047	0.046			
			Geno	0.046	0.047	0.049			
		ca:0.3 co:0.2	Dom	0.953	0.961	0.967			
			Rec	0.651	0.724	0.914			
			Mul	0.980	0.980	0.979			
			Geno	0.953	0.955	0.956			
not HWE	not HWE	ca:0.3 co:0.3	Dom	0.041	0.054	0.054	0.044	0.055	0.076
			Rec	0.032	0.047	0.092	0.033	0.052	0.204
			Mul	0.047	0.047	0.041	0.050	0.050	0.039
			Geno	0.050	0.051	0.092	0.050	0.051	0.278
		ca:0.3 co:0.2	Dom	0.956	0.965	0.986	0.955	0.964	0.995
			Rec	0.727	0.791	0.811	0.806	0.856	0.606
			Mul	0.988	0.988	0.985	0.991	0.991	0.987
			Geno	0.970	0.971	0.970	0.980	0.981	0.985
not HWE	HWE	ca:0.3 co:0.3	Dom	0.046	0.057	0.057	0.067	0.082	0.078
			Rec	0.053	0.074	0.098	0.115	0.154	0.214
			Mul	0.050	0.050	0.046	0.045	0.046	0.041
			Geno	0.075	0.076	0.097	0.172	0.174	0.286
		ca:0.3 co:0.2	Dom	0.977	0.981	0.985	0.986	0.988	0.992
			Rec	0.482	0.558	0.806	0.292	0.365	0.614
			Mul	0.985	0.985	0.983	0.984	0.984	0.981
			Geno	0.967	0.968	0.968	0.975	0.975	0.979
HWE	not HWE	ca:0.3 co:0.3	Dom	0.054	0.068	0.052	0.069	0.086	0.043
			Rec	0.054	0.074	0.075	0.111	0.149	0.070
			Mul	0.057	0.057	0.053	0.050	0.050	0.045
			Geno	0.081	0.083	0.047	0.170	0.172	0.045
		ca:0.3 co:0.2	Dom	0.925	0.936	0.972	0.882	0.897	0.973
			Rec	0.849	0.892	0.912	0.964	0.978	0.916
			Mul	0.983	0.983	0.982	0.986	0.986	0.983
			Geno	0.970	0.972	0.964	0.989	0.990	0.962

Table 6. Probability of rejecting hypotheses when excess homozygote (e=0.01)

HWE				n=200			n=300		
Case	Control	MAF	Test	Chisq	Wald	ccWald	Chisq	Wald	ccWald
HWE	HWE	ca:0.3 co:0.3	Dom	0.039	0.050	0.049	0.044	0.056	0.049
			Rec	0.031	0051	0.072	0.036	0.050	0.074
			Mul	0.047	0.048	0.047	0.048	0.048	0.048
			Geno	0.050	0.052	0.047	0.052	0.052	0.049
		ca:0.3 co:0.2	Dom	0.835	0.859	0.882	0.069	0.966	0.975
			Rec	0.463	0.548	0.795	0.651	0.718	0.915
			Mul	0.904	0.905	0.902	0.984	0.984	0.984
			Geno	0.847	0.850	0.848	0.960	0.961	0.964
not HWE	not HWE	ca:0.3 co:0.3	Dom	0.042	0.052	0.057	0.044	0.057	0.057
			Rec	0.035	0.054	0.110	0.035	0.051	0.123
			Mul	0.052	0.053	0.057	0.051	0.052	0.056
			Geno	0.051	0.053	0.092	0.050	0.051	0.112
		ca:0.3 co:0.2	Dom	0.839	0.862	0.837	0.957	0.964	0.945
			Rec	0.407	0.497	0.891	0.583	0.653	0.966
			Mul	0.896	0.897	0.904	0.977	0.977	0.980
			Geno	0.829	0.832	0.863	0.949	0.949	0.963
not HWE	HWE	ca:0.3 co:0.3	Dom	0.044	0.055	0.056	0.049	0.062	0.058
			Rec	0045	0.068	0.106	0.048	0.070	0.116
			Mul	0.053	0.03	0.055	0.050	0.051	0.053
			Geno	0.068	0.069	0.089	0.072	0.074	0.107
		ca:0.3 co:0.2	Dom	0.782	0.811	0.838	0.927	0.939	0.948
			Rec	0.593	0.677	0.897	0.787	0.839	0.969
			Mul	0.903	0.903	0.906	0.976	0.976	0.977
			Geno	0.841	0.846	0.863	0.955	0.955	0.964
HWE	not HWE	ca:0.3 co:0.3	Dom	0.045	0.055	0.053	0.044	0.058	0.050
			Rec	0.041	0.066	0.072	0.052	0.072	0.078
			Mul	0.051	0.051	0.053	0.048	0.049	0.050
			Geno	0.064	0.065	0.047	0.072	0.074	0.050
		ca:0.3 co:0.2	Dom	0.883	0.902	0.881	0.976	0.980	0.969
			Rec	0.280	0.362	0.803	0.424	0.500	0.913
			Mul	0.901	0.902	0.906	0.977	0.977	0.979
			Geno	0.853	0.856	0.846	0.961	0.961	0.959

Table 7. The mean of estimated OR when excess heterozygote (n=200)

HWE				e=0.01	
Case	Control	MAF	Test	Chisq, Wald	ccWald
HWE	HWE	ca:0.3 co:0.3	Dom	1.02(0.205)	1.02(0.196)
			Rec	1.07(0.410)	1.02(0.310)
			Mul	1.01(0.158)	1.01(0.157)
			Geno	1.02,1.07(0.217,0.425)	1.02,1.02(0.189,0.357)
		ca:0.3 co:0.2	Dom	1.89(0.391)	1.89(0.379)
			Rec	2.80(1.768)	2.46(0.825)
			Mul	1.74(0.299)	1.74(0.296)
			Geno	1.76, 3.48(0.381,2.21)	1.75,3.09(0.343,1.154)
not HWE	not HWE	ca:0.3 co:0.3	Dom	1.02(0.206)	1.05(0.203)
			Rec	1.07(0.445)	0.90(0.278)
			Mul	1.00(0.162)	1.01(0.157)
			Geno	1.02,1.08(0.215,0.461)	1.09,0.94(0.200,0.328)
		ca:0.3 co:0.2	Dom	1.88(0.390)	1.97(0.398)
			Rec	3.57(2.585)	2.16(0.719)
			Mul	1.80(0.313)	1.77(0.301)
			Geno	1.74,4.45(0.373,3.248)	1.87,2.78(0.362,1.033)
not HWE	HWE	ca:0.3 co:0.3	Dom	1.06(0.219)	1.06(0.209)
			Rec	0.94(0.373)	0.90(0.286)
			Mul	1.01(0.163)	1.01(0.162)
			Geno	1.09,0.98(0.236,0.401)	1.09,0.94(0.205,0.340)
		ca:0.3 co:0.2	Dom	1.97(0.410)	1.97(0.398)
			Rec	2.46(1.561)	2.16(0.750)
			Mul	1.77(0.313)	1.77(0.311)
			Geno	1.88,3.16(0.408,2.012)	1.88,2.80(0.368,1.079)
HWE	not HWE	ca:0.3 co:0.3	Dom	0.98(0.198)	1.02(0.194)
			Rec	1.23(0.491)	1.03(0.305)
			Mul	1.01(0.158)	1.01(0.154)
			Geno	0.95,1.19(0.203,0.489)	1.01,1.04(0.188,0.350)
		ca:0.3 co:0.2	Dom	1.81(0.372)	1.89(0.372)
			Rec	4.04(2.910)	2.46(0.805)
			Mul	1.77(0.301)	1.74(0.290)
			Geno	1.63,4.89(0.347,3.573)	1.75,3.09(0.338,1.121)

5.0 DISCUSSION

Our simulation results show that the multiplicative test is essentially the same under all 4 HWE scenarios. This is to be expected, as the multiplicative test is basically a test of allele frequency difference. By contrast, the 2 df test is a very sensitive test of genotype frequency difference, so we expect it to behave very differently in different HWE scenarios. The recessive and dominant tests are less-sensitive tests of genotype frequency difference.

When HWE holds for both cases and controls, the three statistics (chi-squared, Wald, and Chen and Chatterjee Wald) are more or less equivalent for the 2 df and 1 df tests except the recessive test. The recessive test has the highest rejection probability under the alternative hypothesis in the Chen and Chatterjee version, but it also has an inflated rejection probability under the null hypothesis above the nominal level. Chen and Chatterjee (2007) considered only 2 df test with underlying genetic models and indicated that their 2 df test had gained more power with larger ORs and smaller MAFs when the underlying genetic model was the recessive model. We did not consider an underlying genetic model but our underlying model was more likely additive/multiplicative model. We considered more tests; 1 df (dominant, recessive, and multiplicative) tests and 2 df test but fewer genetic models than Chen and Chatterjee did. Our results did not have a discrepancy because we did not test underlying the recessive model.

When HWE fails to hold in either cases or controls or both, the 2 df version of the Chan and Chatterjee Wald test (and to a lesser extent the dominant and recessive versions) detects this

HWE departure and can therefore "find" a case-control difference even if there is not an allele frequency difference or even a genotype frequency difference. In principle, it could be considered desirable to "detect" HWE departures in cases only and consider this evidence of association even in the absence of allele frequency differences, but we believe that the "detection" of association when there are no allele or genotype frequency differences is problematic. We note that Chen and Chatterjee did point out this problem, describing their method as having incorrect type I error when the HWE assumption is violated.

We compared the performances of the three statistics based on a single SNP. A critical question is how these statistics would be compared if used in a genome scan. In a genome scan context, we expect that some loci will be in HWE and others will not. Loci with large departures from HWE are typically excluded in genome scanning, but the cutoff p-value is typically quite extreme, so many retained loci would show moderate departures from HWE. If one uses the CC Wald statistic with 2df for a genome scan, it should detect all true positives, but it might also detect many loci in which there is no association but just a HWE departure. Thus we do not recommend the use of the Chen and Chatterjee statistic for genome scanning. Simulation studies with realistic genome scan data could tell us more about the size of this effect, however.

On the other hand, when one of HWE assumptions such as population stratification is violated, these methods can be used after it adjusts for PCAs, or genetic association tests can be conducted using family-based methods: case trio, case trio + unrelated control, case trio + control trio, and case duo + control duo instead of using unrelated individuals from a population.

There are limitations of this study. We only consider that the same MAF is 0.3 in both case and control, and that the different MAFs are 0.3 in case and 0.2 in control for each HWE scenario. We also considered 200 and 300 sample sizes only. Nor did we evaluate performances of 3 statistics in other MAFs and sample sizes.

APPENDIX A

DERIVATION OF ODDS RATIO AND ITS VARIANCE

The genotype probability in cases is as follows

$$P(G = g \mid D = 1) = \frac{\psi_g \times P(G = g)}{\sum_g \psi_g \times P(G = g)},$$

where $P(G = g)$ is the probability of genotype in controls, ψ_g is the odds ratio for genotype, and genotype g=aa, Aa, AA.

Assuming n cases and J controls, the whole likelihood for the case-control is as follows

$$L = \prod_{i=1}^{n} P(G_i \mid D_i = 1) \times \prod_{j=1}^{n} P(G_j \mid D_j = 0),$$

Taking log of the likelihood,

$$\begin{aligned}\log L &= \sum_{i=1} P(G_i \mid D_i = 1) + \sum_{j=1} P(G_j \mid D_j = 0) \\ &= (n_{11} + n_{21}) \log P(G = aa) + (n_{12} + n_{22}) \log P(G = Aa) + (n_{13} + n_{23}) \log P(G = AA) \\ &\quad + n_{22} \log \psi_{Aa} + n_{23} \log \psi_{AA} - n_2 \log\left(P(G = aa) + \psi_{Aa} P(G = Aa) + \psi_{Aa} P(G = AA)\right).\end{aligned}$$

Taking the derivatives of log L with respect to a dominant test (ψ_{Aa} = ψ_{AA} = ψ), a recessive test (ψ_{Aa} = 1, ψ_{AA} = ψ), and a multiplicative test (ψ_{Aa} = ψ, ψ_{AA} = ψ^2).

$$\frac{\partial \log L}{\partial \psi_{DOM}} = \frac{n_{11}+n_{12}}{\psi} - \frac{n_1(P(G=Aa)+P(G=AA))}{P(G=aa)+\psi P(G=Aa)+\psi P(G=AA)},$$

$$\frac{\partial \log L}{\partial \psi_{REC}} = \frac{n_{12}}{\psi} - \frac{n_1 P(G=AA)}{P(G=aa)+P(G=Aa)+\psi P(G=AA)},$$

$$\frac{\partial \log L}{\partial \psi_{MUL}} = \frac{n_{11}+2n_{12}}{\psi} - \frac{n_1(P(G=Aa)+2\psi P(G=AA))}{P(G=aa)+\psi P(G=Aa)+\psi^2 P(G=AA)}.$$

By solving the equation, the estimated odds ratios can be estimated by using the expected genotype frequencies of controls instead of the observed genotype frequencies as follows,

$$\hat{\psi}^E_{DOM} = \frac{(n_{11}+n_{12})n^E_{20}}{n_{10}(n^E_{21}+n^E_{22})}, \quad \hat{\psi}^E_{REC} = \frac{n_{12}(n^E_{20}+n^E_{21})}{(n_{10}+n_{11})n^E_{22}}, \ , \text{ and } \hat{\psi}^E_{MUL} = \frac{(n_{11}+2n_{12})n^E_{21}}{(2n_{10}+n_{11})2n^E_{22}}. \ .$$

where $n^E_{20} = n_2(1-\hat{p})^2$, $n^E_{21} = 2n_2\hat{p}(1-\hat{p})$, and $n^E_{22} = n_2\hat{p}^2$ are the expected genotype frequencies of controls under HWE and $\hat{p} = (2n_{22}+n_{21})/2n_2$ is the observed genotype frequencies in the control group.

The asymptotic variance can be estimated by using $\log\hat{\psi}^E_{DOM}$, $\log\hat{\psi}^E_{REC}$, and $\log\hat{\psi}^E_{MUL}$ as follows,

$$V_{\log\hat{\psi}^E_{Dom}} = \frac{1}{n_{10}} + \frac{1}{n_{11}+n_{12}} + \frac{2\hat{p}}{n_2(1-\hat{p})(2\hat{p}-\hat{p}^2)^2},$$

$$V_{\log\hat{\psi}^E_{\mathrm{Rec}}} = \frac{1}{n_{10}+n_{11}} + \frac{1}{n_{12}} + \frac{1-\hat{p}}{2n_2\hat{p}(1-\hat{p}^2)^2},$$

$$V_{\log\hat{\psi}^E_{Mul}} = \frac{1}{2n_{10}+n_{11}} + \frac{1}{n_{11}+2n_{12}} + \frac{1}{2n_2\hat{p}(1-\hat{p})}.$$

APPENDIX B

R PROGRAM FOR SIMULATION

```
################################
##### Generating Case Data #####
################################
##  Define number of familiy and replication  ##
n.rep     <- 10000                # Number of Replication #
ss        <- 200                  # sample size           #
##  Minor Allele Frequency and HWD  ##
case.maf    <- 0.3;     case.e <- 0
control.maf <- 0.3;  control.e <- 0
for (k in 1:n.rep){
    ##  Minor Allele Frequency  ##
    p <- case.maf                   # p = P(A)          #
    q <- 1-p                        # q = P(a) = 1-P(A) #
    ##  Genotype by Mendelian rule: P(G) ##
    p.AA <- p^2   -   case.e        # P(G=AA) #
    p.Aa <- 2*p*q + 2*case.e        # P(G=Aa) #
    p.aa <- q^2   -   case.e        # P(G=aa) #
    # generating random number based on P(G)
    case.r <- t(rmultinom(1, size=ss, prob=c(p.aa, p.Aa, p.AA)))
    colnames(case.r)<-c("aa","Aa","AA"); rownames(case.r)<-c("case")
    ###################################
    ##### Generating Control Data #####
    ###################################
    ##  Minor Allele Frequency  ##
    p <- control.maf                # p = P(A)          #
    q <- 1-p                        # q = P(a) = 1-P(A) #
    ##  Genotype by Mendelian rule: P(G) ##
    p.AA <- p^2   -   control.e   # P(G=AA) #
    p.Aa <- 2*p*q + 2*control.e   # P(G=Aa) #
    p.aa <- q^2   -   control.e   # P(G=aa) #
    # generating random number based on P(G)
    control.r <- t(rmultinom(1, size=ss, prob=c(p.aa, p.Aa, p.AA)))
    colnames(control.r)<-c("aa","Aa","AA");
    rownames(control.r)<-c("control")
    table<-rbind(control.r,case.r)
```

```
###################
##### Testing #####
###################
#### Chi-squared test #####
## No genetic ##
#   aa  Aa  AA    ##
chisq<-chisq.test(table)
## Dominant test   ##
#    aa    Aa+AA    ##
c.c.dom<-matrix(c(table[1,1], table[2,1], table[1,2]+table[1,3],
                  table[2,2]+table[2,3]),ncol=2)
chisq.dom<-chisq.test(c.c.dom)
## Recesscive test   ##
#         aa+Aa   AA ##
c.c.rec<-matrix(c(table[1,1]+table[1,2], table[2,1]+table[2,2],
                 table[1,3], table[2,3]),ncol=2)
chisq.rec<-chisq.test(c.c.rec)
## CA Trend test  ##
#   aa  Aa  AA    ##
trend<-c(0,1,2)
xbar<-(sum(table[,1])*trend[1]+sum(table[,2])*trend[2]+
       sum(table[,3])*trend[3])/sum(table)
b<-( sum(table[,1])*((table[2,1]/sum(table[,1]))-
     sum(table[1,])/sum(table)))* (trend[1]-xbar)+
     sum(table[,2])*((table[2,2]/sum(table[,2]))-
    (sum(table[1,])/sum(table)))*(trend[2]-xbar)+
     sum(table[,3])*((table[2,3]/sum(table[,3]))-
    (sum(table[1,])/sum(table)))*(trend[3]-xbar))/
       (sum(table[,1])*((trend[1]-xbar)^2) +
        sum(table[,2])*((trend[2]-xbar)^2)+
        sum(table[,3])*((trend[3]-xbar)^2))
c.a.trend<-((b^2)/(sum(table[1,])*sum(table[2,])/sum(table)^2))*
             (sum(table[,1])*((trend[1]-xbar)^2)+
              sum(table[,2])*((trend[2]-xbar)^2)+
              sum(table[,3])*((trend[3]-xbar)^2))
trend.p<-1-pchisq(c.a.trend,df=1)

### Standard Wald test ###
## Geno test ##
b_Aa<-log((table[1,1] * table[2,2])/(table[1,2] * table[2,1]))
b_AA<-log((table[1,1] * table[2,3])/(table[1,3] * table[2,1]))
v<-matrix(c(1/table[1,1]+1/table[2,2]+1/table[1,2]+1/table[2,1],
            1/table[1,1]+1/table[2,1],
            1/table[1,1]+1/table[2,1],
            1/table[1,1]+1/table[2,3]+1/table[1,3]+1/table[2,1]),ncol=2)
wald<-t(c(b_Aa,b_AA))%*%solve(v)%*%(c(b_Aa,b_AA))
wald.p<-1-pchisq(wald,df=2)
#Geno
count<-c(table[1,1], table[2,1], table[1,2],
         table[2,2], table[1,3], table[2,3])
gr<-rep(0:1,3)
geno<-rep(0:2,c(2,2,2))
lg<-data.frame(geno, gr, count)
w.g<-anova(glm(gr~factor(geno), data=lg, weight=count,
               family=binomial()),test="Chisq")
#Dom
count<-c(table[1,1],table[2,1],
         table[1,2]+table[1,3],table[2,2]+table[2,3])
gr<-rep(0:1,2)
geno<-rep(0:1,c(2,2))
ld<-data.frame(geno, gr, count)
or.d<-log((table[1,1]*(table[2,2]+table[2,3]))/
          ((table[1,2]+table[1,3]) * table[2,1]))
w.d<-anova(glm(gr~geno, data=ld, weight=count,
              family=binomial()),test="Chisq")
#Rec
count<-c(table[1,1]+table[1,2], table[2,1]+table[2,2],
         table[1,3], table[2,3])
gr<-rep(0:1,2)
geno<-rep(0:1,c(2,2))
lr<-data.frame(geno, gr, count)
or.r<-log(((table[1,1]+table[1,2])*table[2,3])/
           (table[1,3]*(table[2,1]+table[2,2])))
w.r<-anova(glm(gr~geno, data=lr, weight=count,
               family=binomial()),test="Chisq")
#Multi
w.m<-anova(glm(gr~geno,data=lg,weight=count,
               family=binomial()),test="Chisq")
or.m<-summary(glm(gr~geno, data=lg, weight=count, family=binomial()))


### Chen & Chattergee's Wald test ###
f<-(2*table[1,3]+table[1,2])/(2*(table[1,1]+table[1,2]+table[1,3]))
n_0 <-sum(table[1,])
n_00<-n_0*(1-f)^2
n_01<-n_0*2*(1-f)*f
n_02<-n_0*(f)^2

cc.b_Aa<-log((n_00 * table[2,2])/(n_01 * table[2,1]))
cc.b_AA<-log((n_00 * table[2,3])/(n_02 * table[2,1]))
cc.v<-matrix(c(1/(2*n_02+n_01)+1/table[2,2]+1/(2*n_00+n_01)+1/table[2,1],
               1/table[2,1] + 1/(n_0*f*(1-f)),
               1/(n_0*f*(1-f))+1/table[2,1],
               4/(2*n_02+n_01)+1/table[2,3]+1/table[2,1]+4/(2*n_00+n_01)),
               ncol=2)
cc.wald<-t(c(cc.b_Aa,cc.b_AA))%*%solve(cc.v)%*%(c(cc.b_Aa,cc.b_AA))
cc.wald.p<-1-pchisq(cc.wald,df=2)

#Dom
cc.dom<-log(((table[2,2]+table[2,3])/table[2,1])*((1-f)^2/(2*f-f^2)))
cc.dom.v<-1/(table[2,2]+table[2,3])+1/table[2,1]+
          2*f/(n_0*(1-f)*(2*f-f^2)^2)
cc.w.d<-cc.dom^2/cc.dom.v
cc.w.d.p<-1-pchisq(cc.w.d,df=1)

#Rec
cc.rec<-log((table[2,3]/(table[2,1]+table[2,2]))*((1-f^2)/(f^2)))
cc.rec.v<-1/(table[2,3])+1/(table[2,1]+table[2,2])+(1-f)/(2*n_0*f*(1-f)^2)
cc.w.r<-cc.rec^2/cc.rec.v
cc.w.r.p<-1-pchisq(cc.w.r,df=1)

#Multi
cc.mul<-log((table[2,2]+2*table[2,3])/(2*table[2,1]+table[2,2])*((1-f)/f))
cc.mul.v<-1/(table[2,2]+2*table[2,3])+1/(2*table[2,1]+table[2,2])+
             1/(2*n_02+n_01)+1/(2*n_00+n_01)
   cc.w.m<-cc.mul^2/cc.mul.v
   cc.w.m.p<-1-pchisq(cc.w.m,df=1)

   d.p1<-c(chisq.dom$statistic,chisq.dom$p.value,
           w.d[2,2],w.d[2,5],cc.w.d,cc.w.d.p)
   r.p1<-c(chisq.rec$statistic, chisq.rec$p.value,
           w.r[2,2], w.r[2,5], cc.w.r, cc.w.r.p)
   m.p1<-c(c.a.trend, trend.p,
           w.m[2,2], w.m[2,5], cc.w.m, cc.w.m.p)
   g.p2<-c(chisq$statistic,     chisq$p.value,
          w.g[2,2], w.g[2,5], cc.wald, cc.wald.p)
   dom.p1<-rbind(dom.p1,d.p1)
   rec.p1<-rbind(rec.p1,r.p1)
   mul.p1<-rbind(mul.p1,m.p1)
   geno.p2<-rbind(geno.p2,g.p2)
}
write.table(dom.p1[-1,], paste(dataDir, "/dom.df1.txt", sep=""), sep="\t",
            col.names=FALSE,row.names = FALSE)
write.table(rec.p1[-1,], paste(dataDir, "/rec.df1.txt", sep=""), sep="\t",
           col.names=FALSE,row.names = FALSE)
write.table(mul.p1[-1,], paste(dataDir, "/mul.df1.txt", sep=""), sep="\t",
            col.names=FALSE,row.names = FALSE)
write.table(geno.p2[-1,], paste(dataDir, "/geno.df2.txt", sep=""), sep="\t",
           col.names=FALSE,row.names = FALSE)
```

BIBLIOGRAPHY

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* 11, 375-386.

Cochran, W.G. (1954). Some methods for strengthening the common chi-square tests. *Biometric* 10, 417-451.

Chen, J., Chatterjee, N. (2007). Exploiting Hardy-Weinberg Equilibrium for Efficient Screening of single SNP Associations from Case-control Studies. *Human Heredity* 63, 196-204.

Sasieni, P.D.(1997). From Genotypes to Genes: Doubling the Sample size. *Biometrics* 53, 1253-1261

Satten, G.A., Epstein, M.P. (2004). Comparison of prospective and retrospective methods for haplotype inference in case-control studies. Genetic Epidemiology 27,192-201