



Published in final edited form as:

*Genet Epidemiol.* 2009 April ; 33(3): 275–280. doi:10.1002/gepi.20381.

## A Likelihood Ratio Test of Population Hardy Weinberg Equilibrium for Case-Control Studies

Chang Yu<sup>\*1</sup>, Sanguo Zhang<sup>1</sup>, Chuan Zhou<sup>1</sup>, and Saba Sile<sup>2</sup>

<sup>1</sup>Department of Biostatistics, Vanderbilt University School of Medicine Nashville, TN 37232–2158

<sup>2</sup>Department of Medicine, Vanderbilt University School of Medicine Nashville, TN 37232–2158

### Abstract

Testing Hardy-Weinberg Equilibrium (HWE) in the control group is commonly used to detect genotyping errors in genetic association studies. We propose a likelihood ratio test for testing HWE in the study population using both case and control samples. This test incorporates underlying association models. Another feature is that, when we infer the disease-genotype association, we explicitly incorporate HWE or a possible departure from Hardy-Weinberg Equilibrium (DHWE) into the model. Our unified framework enables us to infer the disease-genotype association when a detected DHWE needs to be part of the model after causes for the DHWE are explored. Real datasets are used to illustrate the application of the methodology and its implication in genetic association studies. Our analysis and interpretation touch on genotyping errors, population selection, population stratification, or the study sampling plan, all delicate issues that could be the cause of DHWE.

### Keywords

Likelihood ratio test; Hardy Weinberg equilibrium; case-control study; genotype-disease association

### 1 Introduction

Hardy-Weinberg Equilibrium is used to describe the genotype distribution of a population when it is large, self-contained, and randomly mating. The equilibrium can be summarized as, if  $p$  is the frequency of one allele ( $A$ ) and  $q$  is the frequency of the alternative allele ( $a$ ) for a biallelic locus, then the HWE-expected frequency will be  $p^2$  for the  $AA$  genotype,  $2pq$  for the  $Aa$  genotype, and  $q^2$  for the  $aa$  genotype. The three genotypic proportions should sum to 1, as should the allele frequencies (Hardy, 1908; Weinberg, 1908).

Many methods have been developed to test HWE. Weir (1996) and Emigh (1980) provide summaries of these methods. A  $\chi^2$  test is commonly used to assess a departure from HWE. Exact tests of a departure from HWE have been developed for studies with small sample sizes (Haldane 1954; Wigginton, Cutler, and Abecasis, 2005).

Testing HWE is commonly conducted for genotyping quality control (Gomes, Collins, *et al.*, 1999; Xu, Turner, *et al.*, 2002). Some view the testing as an essential step in genetic association studies (Xu, Turner, *et al.*, 2002; Thakkinstian, McElduff, *et al.*, 2005); however, others caution such use (Nielsen, Ehm, and Weir, 1999; Wittke-Thompson, Pluzhnikov, Cox, 2005; Zou and Donner, 2006). Nielson, Ehm, *et al.* (1999) point out that HWE is

\*Contact for corresponding author: Email: chang.yu@vanderbilt.edu, Phone: (615)–322–8422, Fax: (615)343–4924..

generally expected to be distorted in the case sample in the region of association. Zou and Donner (2006) suggest testing for HWE should not be used as a tool for identifying genotyping errors when it is tested in a single sample. Wittke-Thompson, Pluzhnikov, Cox (2005) provide a framework to guide the interpretation of a DHWE for case-control studies. They suggest that if a DHWE in cases or in both cases and controls is detected, it does not necessarily imply genotyping errors. Rather than discarding the data, the underlying disease-genotype association should be investigated. The association may explain the observed DHWE. If not, other possible explanations such as “genotyping error, chance, failure of assumptions underlying Hardy-Weinberg expectations” should be explored. In their framework, they explicitly assume the genotype is in HWE in the population.

The current work develops a likelihood ratio test for testing HWE in the study population. Our test uses data from both case and control samples, and the procedure accounts for the underlying disease models. We estimate the parameters in the model by minimizing the deviance, thus, the estimates are maximum likelihood estimates. The difference between the deviances of two nested models follows a  $\chi^2$  distribution. This forms a likelihood ratio test for the population HWE.

When we infer the disease-genotype association, HWE or a possible DHWE are explicitly incorporated into the model. A DHWE could be due to a variety of reasons such as genotyping errors, population selection, population stratification, or the sampling plan of the study. If they come into play, then it is likely these problems will impact both cases and controls. The purpose of the likelihood ratio test of the population HWE is to prompt the investigators to think about these issues in addition to possible genotyping errors. If a DHWE is detected and genotyping errors are ruled out, our unified framework enables us to model the association under DHWE in contrast to the current practice that testing HWE is a middle step in the analysis of association studies.

The rest of the manuscript is organized as follows. Section 2.1 summarizes common genetic disease models. In Section 2.2, we develop the likelihood ratio test for population HWE. In Sections 3.1 and 3.2, we demonstrate our methods in detail using data from two genetic association studies conducted at Vanderbilt University Medical Center. Then we revisit some examples discussed by Wittke-Thompson, Pluzhnikov and Cox (2005). We developed the analysis software in R and it can be obtained from the corresponding author.

## 2 Statistical Methods

We first summarize the common disease models presented in Wittke-Thompson, Pluzhnikov, Cox (2005). Then we develop a likelihood ratio test for testing HWE in the study population.

### 2.1 Common Disease Models

Wittke-Thompson, Pluzhnikov, Cox (2005) explicitly assume the susceptibility locus is in HWE in the study population. This assumption implies the genotype distribution follows

$$Pr(AA) = p^2, \quad Pr(Aa) = 2pq, \quad Pr(aa) = q^2, \quad (1)$$

where  $p$  is the population frequency of the wild-type allele ( $A$ ), and  $q$  is the population frequency of the disease-susceptibility allele ( $a$ ).

Let  $\alpha$  be the baseline disease penetrance in homozygotes ( $AA$ ), and  $Y = 1, 0$  denote the study outcome diseased or not. They present the following general disease model:

$$Pr(Y=1|AA)=\alpha, \quad Pr(Y=1|Aa)=\alpha\beta, \quad Pr(Y=1|aa)=\alpha\gamma, \quad (2)$$

where  $\beta$  is the relative risk of disease for the heterozygotes  $Aa$  in reference to homozygotes  $AA$  and  $\gamma$  is the relative risk of disease for the homozygotes  $aa$ .

The prevalence of disease in the population is restricted as

$$K_p = p^2\alpha + 2pq\alpha\beta + q^2\alpha\gamma \quad (3)$$

It is recognized that the disease prevalence  $K_p$  can not be estimated in a case-control study, thus it has to be obtained through external studies.

In this framework, Wittke-Thompson, Pluzhnikov, Cox (2005) propose to estimate the parameters  $\theta = (\beta, \gamma, q)$  by minimizing the general  $\chi^2$  statistic.  $\alpha$  is obtained through constraint (3) after the estimates  $(\hat{\beta}, \hat{\gamma}, \text{ and } \hat{q})$  are obtained.

## 2.2 A Likelihood Ratio Test for Population HWE

In this section, we first expand the common disease models to incorporate whether or not the genotype is in HWE in the population. Then we develop a test for the population HWE expressed as a null hypothesis that the susceptibility locus is in HWE in the population versus the alternative hypothesis that the susceptibility locus is not in HWE in the population. Here we define the study population as the population from which cases and controls are drawn and to which study findings will be generalized.

The null hypothesis  $H_0$  is expressible as (1), and under  $H_0$  the disease models are described in Section 2.1. The alternative hypothesis  $H_a$  can be expressed as the genotype distribution in the population

$$Pr(AA)=p_0, \quad Pr(Aa)=p_1, \quad Pr(aa)=p_2, \quad (4)$$

where  $p_2 = 1 - p_0 - p_1$ . Under  $H_a$ ,  $K_p$  is constrained as

$$K_p = p_0\alpha + p_1\alpha\beta + p_2\alpha\gamma \quad (5)$$

Table 1 summarizes a typical data set from a traditional unmatched case-control study. The objective is to explore the relationship between disease status and a single-locus 2-allele genotype denoted as  $AA$ ,  $Aa$ , and  $aa$ .

Assuming common disease model (2), under either  $H_0$  or  $H_a$ , the conditional distributions of the genotype in cases and in controls are

$$Pr(AA|Y=1) = \frac{Pr(AA)\alpha}{K_p}, \quad Pr(Aa|Y=1) = \frac{Pr(Aa)\alpha\beta}{K_p}, \quad Pr(aa|Y=1) = \frac{Pr(aa)\alpha\gamma}{K_p} \quad (6)$$

and

$$\begin{aligned}
 Pr(AA|Y=0) &= \frac{Pr(AA)(1-\alpha)}{(1-K_p)}, \\
 Pr(Aa|Y=0) &= \frac{Pr(Aa)(1-\alpha\beta)}{(1-K_p)}, \\
 Pr(aa|Y=0) &= \frac{Pr(aa)(1-\alpha\gamma)}{(1-K_p)},
 \end{aligned}
 \tag{7}$$

respectively.

Conditional on the study design parameters,  $n_1$  cases and  $n_2$  controls, the expected numbers of genotypes in cases are:

$$En_{11}=n_1Pr(AA|Y=1), \quad En_{21}=n_1Pr(Aa|Y=1), \quad En_{31}=n_1Pr(aa|Y=1)$$

The expected number of genotypes in controls can be expressed as replacing  $n_1$  with  $n_2$  and the probability expressions with (7).

We fit models by minimizing the deviance function

$$\Lambda(\theta) = 2 \sum_{i,j} n_{ij} \log \left\{ n_{ij} / E(n_{ij}) \right\} \tag{8}$$

over the parameter space of  $\theta$ , where  $\theta$  depends on the specific model. Due to the limitation that the data has only 4 independent observations, only models more parsimonious than the general models can be estimated under the alternative hypothesis. General models under  $H_a$  are saturated with  $\theta = (p_0, p_1, \beta, \gamma)$ . In Section 3, we will only consider examples for which recessive disease models (i.e.  $\beta = 1$ ) or dominant disease models (i.e.  $\beta = \gamma$ ) are appropriate.

We use the R function `nlm` to obtain parameter estimates in the models. The standard error of the estimates is obtained through the inverse of the Hessian matrix and through the delta method since proper transformations of the parameters are needed to facilitate the algorithm.

Under  $H_0$  and  $H_a$ , we fit the same disease model (2), and the difference of the deviances follows a  $\chi^2$  distribution with one degree of freedom under regularity conditions (Section 4.5.4 of Agresti, 2002, page 141–142). This forms a likelihood ratio test for the population HWE,  $H_0$  versus  $H_a$ .

### 3 Examples

In this section, we illustrate the application of our methodology and its implication in genetic association studies using several real datasets. The example in Section 3.1 explores a set of possible explanations for the observed DHWE. Section 3.2 discusses the potential impact of the sampling plan on the analysis. For both examples, our analysis results always prompted the investigators to reevaluate genotyping quality. After genotyping errors were ruled out, additional issues were explored as possible explanations for the DHWE in the study population. Consequently, the inference of the disease-genotype association would be made under the best-fit disease model.

#### 3.1 Association of BSND-V43I and Essential Hypertension

We applied our methodology to an association study involving genetic variations in an accessory chloride channel subunit and hypertension in the Ghanaian population. BSND-V43I was identified as a common polymorphism in the non-Caucasian population.

Functional examination of this variant demonstrated a partial loss-of-function variant when heterologously expressed with CIC-Kb in cultured cells. The BSND-V43I genotypes (*GG*, *AG*, *AA*) are (155, 27, 4) in the cases and (408, 55, 15) in the controls (Sile, Gillani, *et al.*, (2007)). In current practice, where HWE is separately tested in cases and in controls, both demonstrate a significant departure from HWE ( $p = 0.043$ , and  $p < 0.001$ , respectively). Applying the general disease model, the best fit is a model with  $\alpha = 0.2006$ ,  $\beta = 0.99$ ,  $\gamma = 0.85$ , and  $q = 0.11$  using the essential hypertension prevalence of  $K_p = 0.20$  (based on World Health Organization report (2005)). Here *GG* is the reference group and  $q$  is the *A* allele frequency in the population. The best-fit model has a lack-of-fit statistic of 36.66 with 1 degree of freedom, exhibiting a significant lack of fit.

The estimated  $\beta$  indicates a recessive disease model. Therefore, we separately fit a recessive model under  $H_0$  and  $H_a$ . Results are presented in Table 2. The recessive model under  $H_0$  had a similar fit to the general model. A note should be made that the deviance for the general model under  $H_0$  is actually the general  $\chi^2$  statistic and it is different from the deviance defined in (8) when we fit our models. We did that since we would like to analyze this data set using both the methods of Wittke-Thompson, Pluzhnikov and Cox (2005) and our methods. This explains why the deviance for the recessive model under  $H_0$  is smaller than the general model under  $H_0$ . The deviance indicates the recessive model under  $H_a$  offers significant improvement over the same model under  $H_0$ . The recessive model under  $H_a$  no longer demonstrates a significant lack-of-fit. The likelihood ratio test suggests significant evidence against  $H_0$ . Hence, we conclude that the genotype is not in HWE in the Ghanaian population from which the cases and the controls were drawn.

There are several possible explanations for the departure from HWE in this example with the BSND-V43I allele. Some of the explanations involve genotyping errors, population stratification, sampling methods, selection, non-random mating, and possibly chance.

Historically, deviation from HWE has been attributed to genotyping errors; however the investigators had several measures in place to avoid such problems. These measures included unstructured sample-numbering system regarding cases and controls, blanks and sequence-verified controls in each plate (Sile, Gillani, *et al.*, 2007).

In addition, deviation from HWE could stem from population stratification or admixture. However, previous studies by Adeyemo, Chen, *et al.*, (2005) demonstrated that these issues are negligible in the Ghanaian population that Sile, Gillani, *et al.*, (2007) examined.

Furthermore, sampling methods could not explain this observation of deviation from HWE. The samples were not ascertained with regards to certain clinical or genetic phenotype or prior knowledge of any disease status. The exclusion criterion was the presence of an acute illness. We believe the cases and the controls are random samples.

Finally, selection appears to be a possible explanation. Unfortunately, there is no data regarding other markers in linkage disequilibrium (LD) in this region to further examine the issue of selection. Using electrophysiology patch clamping, Sile, Gillani, *et al.*, (2007) demonstrated that BSND-V43I is a partial loss-of-function polymorphism. They concluded, based on their functional data, that susceptible subjects with this allele might be protected from developing hypertension. Based on their conclusion we think that selection might be an explanation for the observed deviation from HWE. Further studies examining patterns of variations and LD are needed to determine if selection is indeed present in this region. Additional possible explanations include non-random mating and chance, for which we can not evaluate in this study.

### 3.2 Association of TGF $\beta$ 1 Codon 10 Polymorphism and Familial Pulmonary Arterial Hypertension

We also applied our methodology to an association study of TGF $\beta$ 1 Codon 10 Polymorphism and Familial Pulmonary Arterial Hypertension (FPAH). Dr. John Phillips III and his group genotyped TGF $\beta$ 1 Codon 10 on a cohort of 120 FPAH patients (proband) and 51 of their relatives. Every person in the cohort has the BMPR2 mutation. The hypothesis is that codon 10 T to C transition increases expression and circulating levels of TGF $\beta$ 1, thus increasing the chance for FPAH. The TGF $\beta$ 1 codon 10 SNP genotypes (*TT*, *CT*, *CC*) are (29, 78, 13) in the cases and (17, 28, 6) in the controls (Phillips III, Poling, *et al.*, 2008). When tested separately, cases demonstrated a significant departure from HWE ( $p = 0.0004$ ), but controls did not. Fitting the general disease model, the best fit is a model with  $\alpha = 0.0000066$ ,  $\beta = 2.06$ ,  $\gamma = 1.05$ , and  $q = 0.39$  using a FPAH prevalence of  $K_p = 0.00001$  (Online Mendelian Inheritance in Man, 2007). Here *TT* is the reference group,  $\beta$  and  $\gamma$  are the relative risk of *CT* and *CC* groups, respectively, and  $q$  is the C allele frequency in the population. This best-fit model has a lack-of-fit statistic of 1.17 with 1 degree of freedom, indicating no significant lack of fit. We conclude that the underlying disease model explains the observed departure from HWE in the cases, though it is difficult to interpret the model with  $\beta = 2.06$  and  $\gamma = 1.05$ .

Based on preliminary analyses conducted by Dr. John Phillips III and his group, it appears that the association between this genotype and FPAH follows a dominant disease model. We separately fit a dominant model under  $H_0$  (i.e. in the population the TGF $\beta$ 1 Codon 10 SNP genotypes are in HWE) and  $H_a$  (i.e. in the population the TGF $\beta$ 1 Codon 10 SNP genotypes are not in HWE). The results are summarized in Table 3. The dominant model does not indicate significant lack-of-fit under either  $H_0$  or  $H_a$ . The likelihood ratio test, however, suggests significant evidence against  $H_0$ . Hence, we conclude that the genotype is not in HWE in the population from which the cases and the controls were drawn. We communicated this finding to the investigator. *The investigation team first re-genotyped this SNP*; and genotyping errors were ruled out. One possible explanation for the DHWE in the population might be the plan of the study ascertainment. FPAH is a rare disease, and it is established that people with the BMPR2 mutation are at a higher risk for this disease. In the study sample, every subject has the BMPR2 mutation. Thus, the study sample comprises a stratum from the general population, rather than a representative sample. Secondly, the controls are relatives of the probands. This works against the principles of case-control studies. In this example, however, the primary objective, given the well-established risk factor of the BMPR2 mutation, is to evaluate the additional effect of TGF $\beta$ 1 Codon 10 SNP on this rare disease. Thus the design is appropriate. We recognize that since they are related, the controls might have the tendency to show a similar distribution of TGF $\beta$ 1 Codon 10 genotype to the diseased. However, any bias introduced would be toward the null and makes the investigators' findings conservative.

### 3.3 Reanalysis of the examples analyzed by Wittke-Thompson, Pluzhnikov, and Cox

Among the examples discussed by Wittke-Thompson, Pluzhnikov and Cox (2005), we focus on the studies that showed significant lack-of-fit by their best-fit recessive models. The study by Ozaki, Ohnishi, *et al.* (2002) had the heterozygote relative risk estimate of 1.024, so it is also included. Under the assumption that the genotype is in HWE in the population, the lack-of-fit by the best-fit genetic disease models suggests the genotype disease association is an unlikely explanation for the observed DHWE in patients or in controls.

We first fit recessive models, i.e., with  $\beta$  fixed at 1, under the hypothesis that the genotype is in HWE for these examples. Table 4 summarizes the results. The parameter estimates and the goodness-of-fit statistics are almost identical or very close to those reported in Table 1 of

Wittke-Thompson, Pluzhnikov and Cox (2005). All the best-fit recessive models show significant lack-of-fit.

We then fit the same recessive models, but under the assumption that the population is not in HWE. Results are reported in Table 5. With two parameters,  $p_0$  and  $p_1$ , to represent the genotype frequencies, the allele frequency  $q$  is calculated as  $p_1/2 + (1 - p_0 - p_1)$ . The estimates of  $q$  and  $\alpha$  are similar to those in Table 4, and some estimates of  $\gamma$  have changed significantly. As expected, the deviances are now smaller, and all but one demonstrate significant improvement over the models reported in Table 4. For the study with PubMed ID 11889073 (see Table 5), the recessive model still shows significant lack-of-fit and the investigator needs to look for possible reasons other than the DHWE in the population for an explanation of the lack-of-fit.

Comparing the models under  $H_0$  and  $H_a$  using the likelihood ratio test, we see that much of the lack-of-fit shown in the recessive disease models (Table 4) can be explained by assuming the genotypes are not in HWE in the population. The likelihood ratio tests are reported in Table 6.

These findings raise questions for the investigators about the reason for the departure from HWE in their study population. Sections 3.1 and 3.2 present two detailed examples in which a set of possible explanations were explored to explain the departure. In addition to genotyping errors, we suggest the investigators look into similar issues for possible explanations of the DHWE. On the other hand, whether the genotype is in HWE in the study population plays an important role in making inference about the genotype disease association. Therefore, this assumption should be assessed explicitly in the model.

## 4 Discussion

HWE is commonly tested separately in cases and in controls for genotyping quality control. Several researchers have expressed concern about this practice (Nielsen, Ehm, and Weir, 1999; Wittke-Thompson, Pluzhnikov, Cox, 2005; Zou and Donner, 2006). The current work proposes a likelihood ratio test for testing HWE using both case and control samples. If the problems that HWE testing is intended to detect, such as genotyping errors or population stratification, come into play, these problems are likely to impact both cases and controls. Rather than the current approach of separately testing HWE in cases and in controls as a middle step in the analysis of association studies, our methods test HWE in the study population. We explicitly incorporate HWE or a possible DHWE into the model when we infer the underlying disease-genotype association. If genotyping errors are ruled out and the DHWE is plausible, our methods provide a means to study the association. The observation that some of the estimates of  $\gamma$  in Table 5 changed significantly from the estimates in Table 4 underlines the message that the association estimates depend on the assumption.

Testing HWE in the study population also has implications beyond genotyping quality control. Some genetic methods depend on the assumption that the population is in HWE (Cheng and Chen (2005) and Cheng and Lin (2005)). Our methods provide investigators a useful tool to test HWE in the study population before they apply the analysis methods in those settings.

The examples in Section 3 illustrate the application of the methodology and its implication in genetic association studies. As demonstrated in the examples, a DHWE in the study population could be due to reasons other than genotyping errors, such as population stratification, population selection, the study sampling plan, or *failure of the assumptions underlying HWE*. Although our methods appear to carry a simple message, they touch on

these delicate issues. We suggest a detected DHWE in the study population be investigated with this in mind.

One limitation of our methods is that we assume the disease prevalence is fixed. It is recognized the disease prevalence can not be estimated in a case-control study, thus it has to be obtained from external sources. In the example of Section 3.1, we also analyzed the data using essential hypertension prevalence of  $K = 0.15$  and  $K = 0.25$ . Results were similar to the reported and therefore the conclusion is unchanged. In practice, we suggest investigators conduct sensitivity analyses using a range of plausible estimates of disease prevalence.

## Acknowledgments

We acknowledge John Phillips III, *Scott Williams*, Chun Li, and Daniel Zelterman for helpful discussions. We also acknowledge John Phillips III for giving us access to the data set of TGF $\beta$ 1 codon 10 polymorphism and familial pulmonary arterial hypertension association study. This research was supported in part by the U.S. National Institutes of Health with grant RR00095 awarded to the General Clinical Research Center at Vanderbilt University Medical Center.

## References

- Adeyemo AA, Chen G, Chen Y, Rotimi C. Genetic structure in four West African population groups. *BMC Genet.* 2005; 6:38. [PubMed: 15978124]
- Agresti, A. *Categorical data analysis.* John Wiley & Sons; New York; Chichester: 2002.
- Cheng KF, Chen JH. Bayesian models for population-based case-control studies when the population is in Hardy-Weinberg equilibrium. *Genetic Epidemiology.* 2005; 28:183–192. [PubMed: 15593279]
- Cheng KF, Lin WJ. Retrospective analysis of case-control studies when the population is in Hardy-Weinberg equilibrium. *Statistics in Medicine.* 2005; 24:3289–3310. [PubMed: 16206241]
- Emigh TH. A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics.* 1980; 36:627–642.
- Gomes I, Collins A, Lonjou C, Thomas NS, Wilkinson J, Watson M, Morton N. Hardy-Weinberg quality control. *Annals of Human Genetics.* 1999; 63:535–538. [PubMed: 11246455]
- Haldane JBS. An exact test for randomness of mating. *Journal of Genetics.* 1954; 52:631–635.
- Hardy GH. Mendelian proportions in a mixed population. *Science.* 1908; 28:49–50. [PubMed: 17779291]
- Nielsen D, Ehm MG, Weir BS. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *American Journal of Human Genetics.* 1999; 63:1531–1540. [PubMed: 9867708]
- Online Mendelian Inheritance in Man, OMIM (TM). Johns Hopkins University; Baltimore, MD: 2007. MIM Number: 178600URL: <http://www.ncbi.nlm.nih.gov/omim/>
- Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y. Functional SNPs in the lymphotoxin- $\alpha$  gene that are associated with susceptibility to myocardial infarction. *Nature Genetics.* 2002; 32:650–654. [PubMed: 12426569]
- Phillips JA III, Poling JS, Phillips CA, Stanton KC, Austin ED, Cogan JD, Wheeler L, Yu C, Newman JE, Dietz HC, Loyd JE. Synergistic heterozygosity for TGF $\beta$ 1 SNPs and BMPR2 mutations modulates the age at diagnosis and penetrance of familial pulmonary arterial hypertension. *Genet Med.* 2008; 10(5) (in press).
- Sile S, Gillani NB, Velez DR, Vanoye CG, Yu C, Byrne LM, Gainer JV, Brown NJ, Williams SM, George AL Jr. Functional BSND variants in essential hypertension. *Am J Hypertens.* 2007; 20(11):1176–1182. [PubMed: 17954364]
- Thakkinstian A, McElduff P, D'Este C, Duffy D, Attia J. A method for meta-analysis of molecular association studies. *Statistics in Medicine.* 2005; 24:1291–1306. [PubMed: 15568190]
- Weinberg, W. On the demonstration of heredity in man, in *Papers on Human Genetics.* Boyer, SH., editor. Prentice-Hall; Englewood Cliffs, NJ: 1908. 1963.
- Weir, BS. *Genetic Data Analysis II Methods for Discrete Population Genetic Data.* Sinauer Associates, Inc.; Sunderland, Massachusetts: 1996.



- Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*. 2005; 76:887–893. [PubMed: 15789306]
- World Health Organization Regional Office for Africa. Cardiovascular Disease in the African Region: Current Situation and Perspectives. AFR/RC55/12. June.2005 2005.
- Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational Inference about Departures from Hardy-Weinberg Equilibrium. *American Journal of Human Genetics*. 2005; bf 76:967–986. [PubMed: 15834813]
- Xu J, Turner A, Little J, Bleecker ER, Meyers DA. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? *Human Genetics*. 2002; 111:573–574. [PubMed: 12516594]
- Zou G,Y, Donner A. The merits of testing Hardy-Weinberg equilibrium in the analysis of unmatched genetic case-control data: A cautionary note. *Annals of Human Genetics*. 2006; 70:923–933. [PubMed: 17044866]

**Table 1**

Genotype Distribution in a Case-Control Study

Genotype	Disease Status		
	Case	Control	total
AA	$n_{11}$	$n_{12}$	$n_{1.}$
Aa	$n_{21}$	$n_{22}$	$n_{2.}$
aa	$n_{31}$	$n_{32}$	$n_{3.}$
	$n_{.1}$	$n_{.2}$	$N$

**Table 2**

Analysis of BSND Variant V43I and Essential Hypertension

Disease Model	$q^{\dagger}$	$\alpha$	$\beta$	$\gamma$	Deviance	$p$ -Value
General Under $H_0$	0.1073	0.2006	0.99	0.85	36.66*	<0.001
Recessive Under $H_0$	0.0904	0.2001	1.0	0.94	29.71	<0.001
Recessive Under $H_a$	<u>0.0911</u>	<u>0.2016</u>	<u>1.0</u>	<u>0.73</u>	<u>1.00</u>	<u>0.317</u>
Likelihood Ratio Test of $H_0$ vs. $H_a$					28.71	<0.001

\* This is the general  $\chi^2$  statistic instead of the deviance defined as (8).

$q^{\dagger}$  is the frequency of the minor allele  $A$ .

**Table 3**

Association of TGFβ1 Codon 10 Polymorphism and FPAH

Disease Model	$q^{\dagger}$	$\alpha$	$\beta$	$\gamma$	Deviance	p-Value
General Under $H_0$	0.39	$6.6 \times 10^{-6}$	2.06 <sup>**</sup>	1.05	1.17 <sup>*</sup>	0.2795
Dominant Under $H_0$	0.34	$5.5 \times 10^{-6}$	2.45 <sup>**</sup>	2.45 <sup>**</sup>	4.74	0.0937
Dominant Under $H_a$	0.38	$7.3 \times 10^{-6}$	1.57	1.57	0.21	0.6456
Likelihood Ratio Test of $H_0$ vs. $H_a$					4.52	0.0334

\* This is the general  $\chi^2$  statistic instead of the deviance defined as (8).

\*\* Significantly different from 1 at 0.05 level.

$q^{\dagger}$  is the frequency of the minor allele C.

Table 4

Fitting recessive disease models ( $\beta$  is fixed at 1) to ten case-control studies analyzed by Wittke-Thompson, Pluzhnikov and Cox (2005) under the null hypothesis  $H_0$ : the genotype is in HWE in the study population.

Study	Genotype Distributions						$\gamma$	Deviance ( $df=2$ )	$p$ -value
	PubMed ID/Disease/Gene	Patients	Controls	$K_p$	$q$	$\alpha$			
11468325 / Alzheimer disease /CST3	137-34-8	180-40-8	.06	0.121	0.0585	2.750	7.877	.019	
11124296 / Macular degeneration /mEPHX	42-24-32	95-38-33	.01	.299	0.00745	4.827	39.326	<.001	
11889073 / Colorectal cancer /MTHFR	28-64-94	114-560-553	.05	.671	0.0448	1.257	9.570	.008	
12631667 / Crohn disease /CYP1A1	2-20-129	5-22-122	.05	.892	0.0360	1.489	6.386	.041	
9607207 / Hypertension (nephroangiosclerosis) / ACE	6-10-21	8-48-19	.05	.568	0.0314	2.829	7.529	.023	
10712418 / Myocardial infarction /TF-1208	227-361-218	197-349-186	.1	.482	0.0954	1.205	6.410	.041	
10680782 / Multiple sclerosis /GSTM3	276-97-14	221-64-15	.05	.154	0.0494	1.467	10.044	.007	
10430441 / Stroke /NOS3	109-125-31	154-203-36	.15	.351	0.1496	1.022	7.675	.022	
11027931 / Venous thrombosis/ $\beta$ -fibrinogen	2-6-82	0-22-163	.01	.934	0.00693	1.508	5.116	.077	
12426569 /MI/lymphotoxin- $\alpha$ gene	416-504-213	378-512-116	.01	.372	0.00942	1.441	8.566	.014	

Fitting recessive disease models ( $\beta$  is fixed at 1) to ten case-control studies analyzed by Wittke-Thompson, Pluzhnikov and Cox (2005) under the alternative hypothesis  $H_d$ : the genotype is not in HWE in the study population. Table 4 has more study details.

**Table 5**

Study	Genotype Distributions		$K_p$	q	$P_0$	$P_1$	$\alpha$	$\gamma$	Deviance ( $df=1$ )	p-value
	Patients	Controls								
11468325	137-34-8	180-40-8	.06	0.127	.781	.182	0.0594	1.265	0.181	.671
11124296	42-24-32	95-38-33	.01	.324	.551	.249	0.00842	1.939	1.233	.267
11889073	28-64-94	114-560-553	.05	.671	.102	.452	0.0445	1.275	8.760	.003
12631667	2-20-129	5-22-122	.05	.897	.025	.153	0.0406	1.282	0.912	.339
9607207	6-10-21	8-48-19	.05	.563	.142	.589	0.0296	3.566	3.832	.050
10712418	227-361-218	197-349-186	.1	.489	.278	.466	0.0980	1.079	0.771	.380
10680782	276-97-14	221-64-15	.05	.166	.718	.232	0.0507	0.724	1.107	.292
10430441	109-125-31	154-203-36	.15	.346	.403	.502	0.1464	1.256	0.678	.410
11027931	2-6-82	0-22-163	.01	.934	.010	.111	0.00735	1.411	3.504	.061
12426569	416-504-213	378-512-116	.01	.364	.388	.496	0.00919	1.764	1.385	.239

**Table 6**

Likelihood ratio test of population HWE for ten case-control studies analyzed by Wittke-Thompson, Pluzhnikov and Cox (2005).

Study PubMed ID/Disease/Gene	Genotype Distributions		Difference of Deviance( <i>df</i> =1)	<i>p</i> -value
	Patients	Controls		
11468325/Alzheimer disease/CST3	137-34-8	180-40-8	7.696	.006
11124296/Macular degeneration/mEPHX	42-24-32	95-38-33	38.093	<.001
11889073/Colorectal cancer/MTHFR	28-64-94	114-560-553	0.810	.368
12631667/Crohn disease/CYP1A1	2-20-129	5-22-122	5.474	.019
9607207/Hypertension(nephroangiosclerosis)/ACE	6-10-21	8-48-19	3.697	.055
10712418/Myocardial infarction/TF-1208	227-361-218	197-349-186	5.638	.018
10680782/Multiple sclerosis/GSTM3	276-97-14	221-64-15	8.937	.003
10430441/Stroke/NOS3	109-125-31	154-203-36	6.997	.008
11027931/Venous thrombosis/ $\beta$ -fibrogen	2-6-82	0-22-163	1.612	.204
12426569 /MI/lymphotoxin- $\alpha$ gene	416-504-213	378-512-116	7.181	.007