

RESEARCH ARTICLE

Open Access

Estimating the mode of inheritance in genetic association studies of qualitative traits based on the degree of dominance index

Elias Zintzaras^{1,2*} and Mauro Santos³

Abstract

Background: The biological justification for the choice of the genetic mode in genetic association studies (GAS) is seldom available. Then, the mode of inheritance is approximated by investigating a number of non-orthogonal genetic contrasts making the interpretation of results difficult.

Methods: We propose to define the mode of inheritance by the significance of the deviance of the co-dominant contrast and the degree of dominance (h), which is a function of two orthogonal contrasts (the co-dominant and additive). Non-dominance exists when the co-dominant contrast is non-significant and, hence, the risk effect of heterozygotes lies in the middle of the risk of the two homozygotes. Otherwise, dominance (including over- and under-dominance) is present and the direction of dominance depends on the value of h .

Results: Simulations show that h may capture the real mode of inheritance and it is affected by deviations from Hardy-Weinberg equilibrium (HWE). In addition, power for detecting significance of h when the study conforms to HWE rule increases with the degree of dominance and to some extent is related to the mutant allele frequency.

Conclusion: The introduction of the degree of dominance provides useful insights into the mode of inheritance in GAS.

Background

Genetic association studies (GAS), candidate-gene and genome-wide association studies assess the association between a disease and genetic variants (gene polymorphisms) in a population [1]. For a bi-allelic candidate gene with alleles wild-type wt and mutant-type mt in a case-control study, where mt is thought to be associated with a disease, the association is usually assessed using a chi-squared test for the 3×2 contingency table with genotype entries $n_{11}(wtwt)$, $n_{21}(wtmt)$ and $n_{31}(mtmt)$ for the control subjects, and $n_{12}(wtwt)$, $n_{22}(wtmt)$ and $n_{32}(mtmt)$ for the diseased subjects [2]. However, besides evaluating the overall statistical significance, the clinical relevance of a genetic association depends on the magnitude of risk conferred to the carriers of allele mt [3]. Thus, in view of a significant association, the following contrasts of genotypes are defined

by merging information of the genotype distribution [4], and estimated with the odds ratio (OR) and its 95% confidence interval (CI): (i) the additive contrast, which is defined as the comparison between the extreme genotypes $-mtmt$ vs. $wtwt$ (in this contrast the heterozygotes are ignored. Note that this contrast does not correspond to the conventionally examined “additive model” which is tested using the Armitage’s test for trend [5]); (ii) the recessive contrast, which compares genotype $mtmt$ with the merged genotypes $wtwt + wtmt$; (iii) the dominant contrast, which merges genotypes $mtmt + wtmt$ and compares them with genotype $wtwt$; and (iv) the co-dominant contrast, which compares genotype $wtmt$ against the merged genotypes $mtmt + wtwt$ [1].

The biological justification for the choice of the genetic contrasts (which may not necessarily present the genetic model of inheritance) to be tested is, however, seldom available and lack of *a priori* assumption for the specific genetic model is customary practice [6,7]. This actually translates to investigating all four models above in GAS, and the interpretation of results may be confusing; that

* Correspondence: zintza@med.uth.gr

¹Department of Biomathematics, University of Thessaly School of Medicine, 2 Panepistimiou Str, Biopolis, Larissa 41110, Greece

Full list of author information is available at the end of the article

is, it might be the case that more than one contrast or even all of them are statistically significant [1]. The obvious reason for this is because in a case-control 3×2 contingency table there are only 2 degrees of freedom (i.e., there are two independent contrasts at most) and, therefore, the previous contrasts are not statistically independent. Furthermore, when a large number of comparisons are made following a significant genotype effect, some of the contrasts might be significant due to a type I error.

In this paper, we propose an index (h) that measures the degree of dominance and allows inferring the mode of inheritance in GAS in a continuous scale. Then, we numerically analyze how the h -index performs by using of a population genetics model where the real mode of inheritance can be defined *a priori*, and provide estimates of power. We also investigate how deviations from Hardy-Weinberg equilibrium (HWE) can affect inferences of the mode of inheritance using the h -index. Finally, we illustrate the method with an empirical study of published GAS.

Methods

Contrast definition and model analysis

Consider a GAS of a bi-allelic polymorphism which evaluates the risk associated with allele mt . The genotype frequencies are given in a 3×2 contingency table with counts

Genotype	Controls	Cases	
wt/wt	n_{11}	n_{12}	(1)
wt/mt	n_{21}	n_{22}	
mt/mt	n_{31}	n_{32}	

To analyze this case-control study, the following logit model can be fitted to the dataset

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \times g, \quad (2)$$

where π is the probability of being diseased, and g is the genotype effect with 3 levels. Then, the deviance due to the genotype effect (D_g) in the model determines significance of g , and the significance test is based on the χ^2 -distribution with 2 degrees of freedom (df_g) [8].

Let us now define the additive and the co-dominant contrasts [1]. The additive model (L_a) presents an individual contrast (comparison) between the two extreme homozygotes (with a single degree of freedom: $df_{L_a} = 1$)

$$L_a = \sum_i c_i g_i; c = [1 \ 0 \ -1]; \sum_i c_i = 0, \quad (3)$$

and g_i is the effect due to the i th genotype ($i = 1, 2, 3$) [9]. The magnitude of the association corresponding to this contrast is estimated by the natural logarithm (\ln) of the odds ratio $\theta_a = (n_{32} \times n_{11}) / (n_{12} \times n_{31})$ (i.e., we refer

to an additive contrast on the \ln -scale). The estimator $\ln(\theta_a)$ is approximately normally-distributed with asymptotic variance $\text{Var}[\ln(\theta_a)] = \left(\frac{1}{n_{32}} + \frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{31}}\right)$ [10]. Statistical significance of the additive contrast can be tested using a z-test $z = \ln(\theta_a) / \sqrt{\text{Var}[\ln(\theta_a)]}$.

The co-dominant model (L_{co}) is the individual contrast between the heterozygotes and the average of the two homozygotes (with $df_{L_{co}} = 1$)

$$L_{co} = \sum_i c_i g_i, \quad (4)$$

$$c = [-0.5 \ 1 \ -0.5]; \sum_i c_i = 0.$$

The magnitude of the association for this contrast is estimated as the natural logarithm of the odds ratio

$$\theta_{co} = \frac{n_{22} \times (n_{11} + n_{31})}{n_{21} \times (n_{12} + n_{32})}, \quad \text{with asymptotic variance}$$

$\text{Var}[\ln(\theta_{co})] = \left(\frac{1}{n_{22}} + \frac{1}{n_{11} + n_{31}} + \frac{1}{n_{21}} + \frac{1}{n_{12} + n_{32}}\right)$. Statistical significance can also be tested using a z-test as above. The additive and co-dominant contrasts are clearly orthogonal because the dot product of the c_i coefficients defining each contrast is zero (i.e., $[1 \ 0 \ -1] \cdot [-0.5 \ 1 \ -0.5] = 0$) [9].

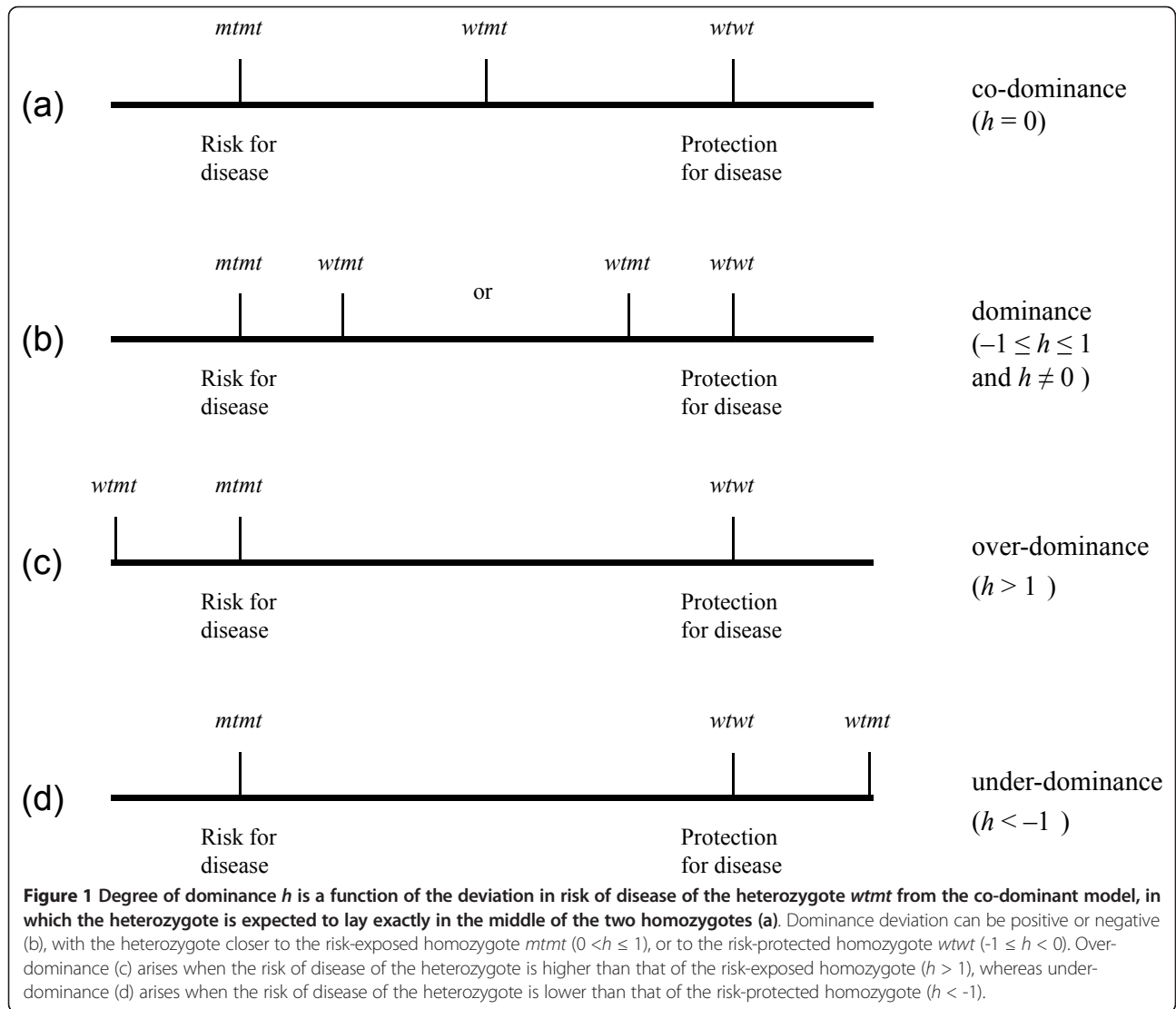
These statistically orthogonal contrasts can be interpreted separately since the deviance of the genotype effect can be split into two independent deviances: one due to the additive contrast (D_{L_a}) and the other due to the co-dominant contrast ($D_{L_{co}}$) [8,9]. In other words, $D_g = D_{L_a} + D_{L_{co}}$, and $df_g = df_{L_a} + df_{L_{co}}$. Now the logit model (2) is equivalent to the model

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 \times L_a + \beta_1 \times L_{co}. \quad (5)$$

Testing the significance of the genotype effect is thus equivalent to simultaneously testing the significance of the additive and co-dominant contrasts based on the respective deviances or z-tests.

Modelling the disease inheritance

We define here for a GAS showing a significant statistical association (i.e., showing significance for at least one of the two orthogonal contrasts) the model of disease inheritance according to the degree of dominance h of the mutant allele mt . In the extreme case where there is non-dominance (i.e., co-dominance or perfect additivity), the heterozygote $wtmt$ "lies" exactly in the middle of the two homozygotes, with $mtmt$ having the maximum susceptibility of being diseased and $wtwt$ having the least (Figure 1). The deviance for the co-dominant contrast ($D_{L_{co}}$) is then non-significant, whereas the deviance (D_{L_a}) for the additive contrast is expected to



be highly significant since we consider GAS with significant association. On the other hand, significance of $D_{L_{co}}$ indicates dominance (i.e., the heterozygote $wtmt$ lies towards $mtmt$ or $wtwt$) irrespectively of the significant level of D_{L_a} .

We can use now the odds ratios of the co-dominant and additive contrasts to define the magnitude and the direction of the degree of dominance. Deviation from perfect additivity can result in dominance (the heterozygote deviates from the middle of the two homozygotes), over-dominance (the heterozygote has a greater risk of disease than the homozygote $mtmt$), or under-dominance (the heterozygote has least risk for the disease than the homozygote $wtwt$) (Figure 1). Therefore, the degree of dominance could be derived from the ratio of the logarithms of the two previous odds ratios as

$$h = \frac{\ln(\theta_{co})}{|\ln(\theta_a)|}, \quad (6)$$

where “|” denotes absolute value. In expression (6) the sign of $\ln(\theta_{co})$ determines the direction of dominance, and the value of $\ln(\theta_{co})$ relative to the absolute value of $\ln(\theta_a)$ the magnitude of dominance deviation h , which can take any value from negative infinity to positive infinity (Figure 2).

Once significance in dominance is detected and h is obtained, the degree of dominance is inferred as follows (assuming, of course, that homozygote $mtmt$ has higher disease risk than homozygote $wtwt$) (Figures 1 and 2). If $-1 \leq h < 0$ there is dominance of the wild-type allele wt , and the heterozygote $wtwt$ is expected to have a risk of being diseased somewhere in between the middle of the

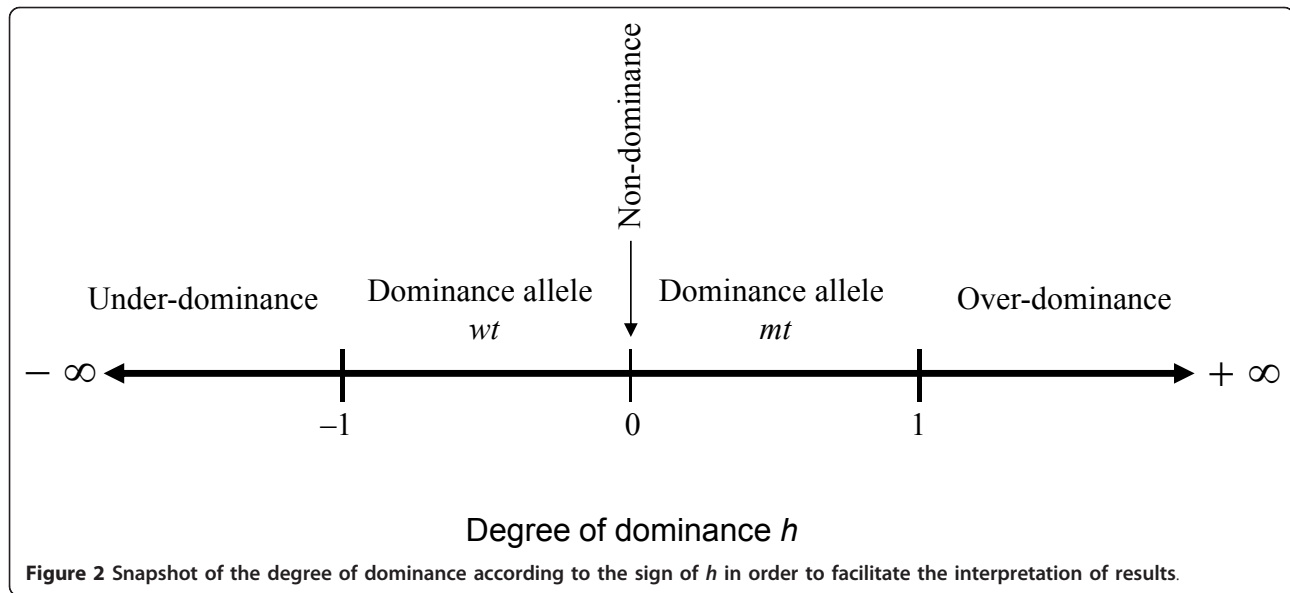


Figure 2 Snapshot of the degree of dominance according to the sign of h in order to facilitate the interpretation of results.

two homozygotes and towards to the homozygote $wtwt$. If $0 < h \leq 1$ there is dominance of the mutant allele mt and the heterozygote $wtmt$ is expected to have a risk of being diseased somewhere in between the middle of the two homozygotes and towards to the homozygote $mtmt$. When $h > 1$ there is over-dominance and the heterozygote has a higher risk of being diseased than the homozygote $mtmt$, whereas if $h < -1$ there is under-dominance and the heterozygous has least chance of being diseased than the homozygote $wtwt$. However, over- and under- dominance is a rare phenomenon in GAS [6].

The statistical test to asses the significance of over- or under-dominance -that is, to test the null hypothesis $H_0: h = 1$ vs. $H_a: h > 1$ (over-dominance), or vs. $H_a: h < -1$ (under-dominance) is equivalent to test:

$$i) \quad H_0: k = \ln\left(\frac{\theta_{co}}{\theta_\alpha}\right) = 0 \quad \text{vs.} \quad H_1: k = \ln\left(\frac{\theta_{co}}{\theta_\alpha}\right) > 0$$

(over-dominance) or $H_1: k = \ln\left(\frac{\theta_{co}}{\theta_\alpha}\right) < 0$ (under-dominance) when $\ln(\theta_\alpha) > 0$,

ii) $H_0: k = \ln(\theta_{co} \times \theta_\alpha) = 0$ vs. $H_1: k = \ln(\theta_{co} \times \theta_\alpha) > 0$ (over-dominance) or $H_1: k = \ln(\theta_{co} \times \theta_\alpha) < 0$ (under-dominance) when $\ln(\theta_\alpha) < 0$.

This can be done using a one-sided z - test where the variance of k is approximately

$$\text{Var}(k) = \text{Var}[\ln(\theta_{co})] + \text{Var}[\ln(\theta_\alpha)] \times \left(1 - \frac{2}{\pi}\right),$$

where $(1-2/\pi)$ is arising from the variance of a half-normal distribution [11]. When the test is non-significant we conclude that there is just dominance or that over-/under-dominance is not beyond chance."

To summarize, inferences regarding any degree of dominance h are obtained from the following order of hypothesis. The first hypothesis states that there is non-dominance, and is tested using the co-dominant contrast. If this hypothesis is true, the risk of disease for the heterozygote $wtmt$ is in the middle of the two homozygotes. If the co-dominant contrast is significant, we then test for the direction of dominance by stating the second hypothesis; that is, $0 < |h| \leq 1$. If this hypothesis is true, the heterozygote $wtmt$ has a risk of disease closer to the homozygote $mtmt$ or the homozygote $wtwt$ according to the sign (+ or -, respectively) of h . Finally, if this hypothesis is false we assess whether or not there is significant over- or under-dominance by testing the hypothesis $|h| > 1$ using k .

Results

We suggest the use of h as defined in (6) to infer the mode of inheritance in GAS, but an obvious question is: how does this index perform? Here, we check (i) the performance of h when the real genetic model is known in advance from a standard population genetics model, then we analyze (ii) power from computer simulations by randomly exploring the parameter space, $0.2 \leq h \leq 1.0$ ($-0.2 \geq h \geq -1.0$) in case-control studies, and (iii) the impact that deviations from HWE in the control subjects have on the h index.

Performance of the h index

We make use of a standard one-locus, two-allele model in population genetics that generalizes the role played by the real degree of dominance (H) on the evolution of a deleterious mutant allele (mt) in a population [12] to

analyze the performance of the proposed index, h . The population genetics model goes as follows

Genotype	Frequencies t_1	Fitness	Frequencies t_2
wtw	p_1^2	1	$\frac{p_1^2}{\bar{w}}$
wmt	$2p_1q_1$	$1 - Hs$	$\frac{2p_1q_1(1 - Hs)}{\bar{w}}$
mtm	q_1^2	$1 - s$	$\frac{q_1^2(1 - s)}{\bar{w}}$

(8)

where $\bar{w} = p_1^2 + 2p_1q_1(1 - Hs) + q_1^2(1 - s)$, and s is the coefficient of selection ($s > 0$, i.e. s is the decrease in fitness of homozygous carriers for the mutant allele relative to the wild-type genotype). In words, we assume that there is an initial cohort of individuals in HWE and that in the time interval $t_1 \rightarrow t_2$ there is a genotype-dependent per capita probability of survival (relative to the wild-type homozygote wtw) given by the values in the column labeled "fitness", with $s > 0$ and $0 \leq H \leq 1$; that is, we know in advance that allele mt is associated with an increased probability of death as time goes on. The parameter H captures the degree of dominance of the mutant allele, which is fully recessive when $H = 0$, co-dominant when $H = 1/2$, and fully dominant when $H = 1$ [13]. To convert this simple model into a hypothetical case-control study we proceeded as follows.

Assume 10,000 individuals at t_1 with allele frequencies in the range $0.05 \leq q_1 \leq 0.5$, and obtain the resulting genotype frequencies at time t_2 in (8) for the whole range of the parameter space $0 \leq H \leq 1$ (changes in the parameter s do not affect the qualitative results). Because the "disease trait" we are studying is the genotype-dependent probability of death in the time interval $t_1 \rightarrow t_2$, the appropriate genotype distribution of cases to be compared with the controls (i.e., the initial cohort) is simply that arising from the individuals that have died. To avoid having entries with zero in the case-control study, we assumed a constant genotype-independent probability of death equal to 5% (this does not obviously affect the qualitative results). This procedure allows us to define parametric controls and cases for any type of inheritance mode (H) and selection coefficient (s). In order to simulate a "sampling" case-control study, $n = 400/400$ cases/controls subjects are sampled randomly from the parametric cohort based on the space defined by the cumulative genotype frequencies [e.g. if a control subject is randomly sampled in the space from $n_1/(n_1 + n_2 + n_3)$ to $(n_1 + n_2)/(n_1 + n_2 + n_3)$, then the subject is assigned as heterozygous].

Figure 3 plots index h against parameter H assuming $s = 0.2$ and 0.5 . For very low allele frequencies ($0.05 \leq q_1 \leq 0.40$) there is little chance of detecting recessiveness of allele mt (i.e., $H < 0.5$) because, in general, $h \geq -0.2$;

whereas for relatively high allele frequencies ($q_1 = 0.40$), h -index can estimate more efficiently H . In any case, h -index tends to perform better when the selection coefficient increases (i.e. the risk of disease is higher in homozygous mutants).

Power

To estimate power simulations were performed by randomly exploring the parameter space $0.1 \leq h \leq 1.0$ ($0.1 \geq h \geq -1.0$) in case-control studies. Genotype frequencies $n_{i2}(i = 1, 2, 3)$ for the cases in (1) were randomly generated with the restriction $\sum_i n_{i2} = 400$ and resulting frequencies for the disease-associated allele mt bounded by $0.09 \leq q_{cases} \leq 0.11$, $0.09 \leq q_{cases} \leq 0.21$, or $0.39 \leq q_{cases} \leq 0.41$. Genotype frequencies in the controls were generated assuming HWE ($P \geq 0.05$) with the restriction $\sum_i n_{i1} = 400$ and $q_{controls} \leq q_{cases}$. Power to detect dominance was assessed as [14]

$$\text{Power} = \frac{\#(P < 0.05) \text{ co-dominant contrast}}{\# \text{ simulations}} \quad (9)$$

Figure 4 shows that power increases with increasing frequencies of allele mt , and also when h increases (decreases) as expected. There is, however, some asymmetry depending on allele mt being inferred to be dominant ($h > 0$; Figure 2) or recessive ($h < 0$). When mt is recessive power to detect moderate levels of dominance ($h \leq -0.4$) is substantial even at relative low allele frequencies, which might be explained due to the restrictions imposed to genotype frequencies.

Impact of deviations from HWE on h index

We now analyze how estimates of the degree of dominance h in GAS are affected by deviations from Hardy-Weinberg equilibrium (HWE) in the control subjects [15]. We first present a general analytical treatment on the topic, and then illustrate the analysis with simulations.

Following Weir [16], genotype frequencies in the controls can be modeled from (1) as $n_{11} = n_{\bullet 1} \times (p^2 + D)$, $n_{21} = n_{\bullet 1} \times (2pq - 2D)$ and $n_{31} = n_{\bullet 1} \times (q^2 + D)$; where p ($q = 1 - p$) is the frequency of allele wmt , $n_{\bullet 1} = n_{11} + n_{21} + n_{31}$ is the total number of control subjects in the study, and $D \left(D = \frac{n_{11}}{n_{\bullet 1}} - p^2 \right)$ is the Hardy-Weinberg disequilibrium parameter which is expected to be zero when a population has Hardy-Weinberg proportions (i.e., testing for HWE is equivalent to a test of the hypothesis $H_0: D = 0$). Bounds on D are [17]

$$\max[-p^2, -q^2] \leq D \leq pq \quad (10)$$

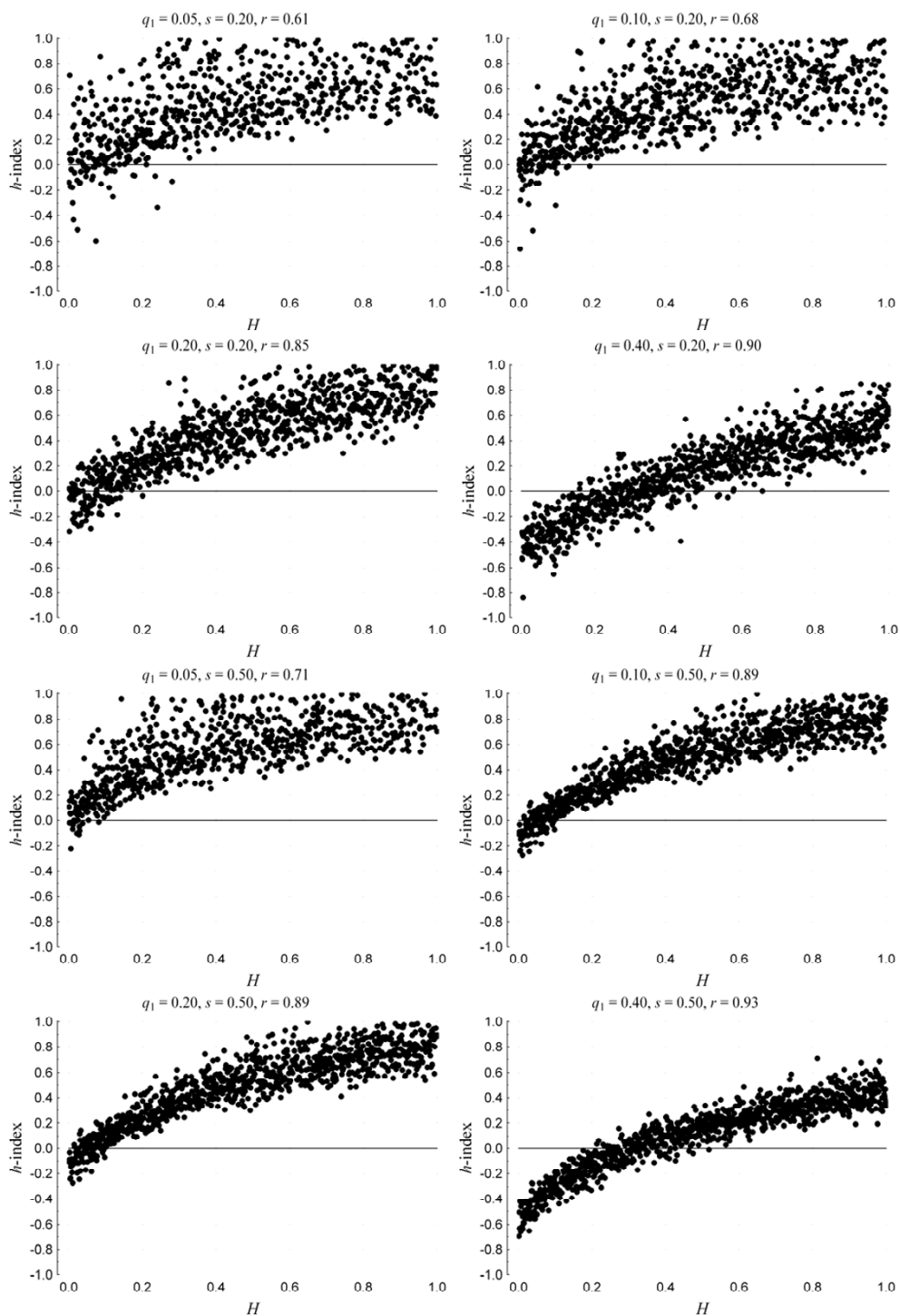
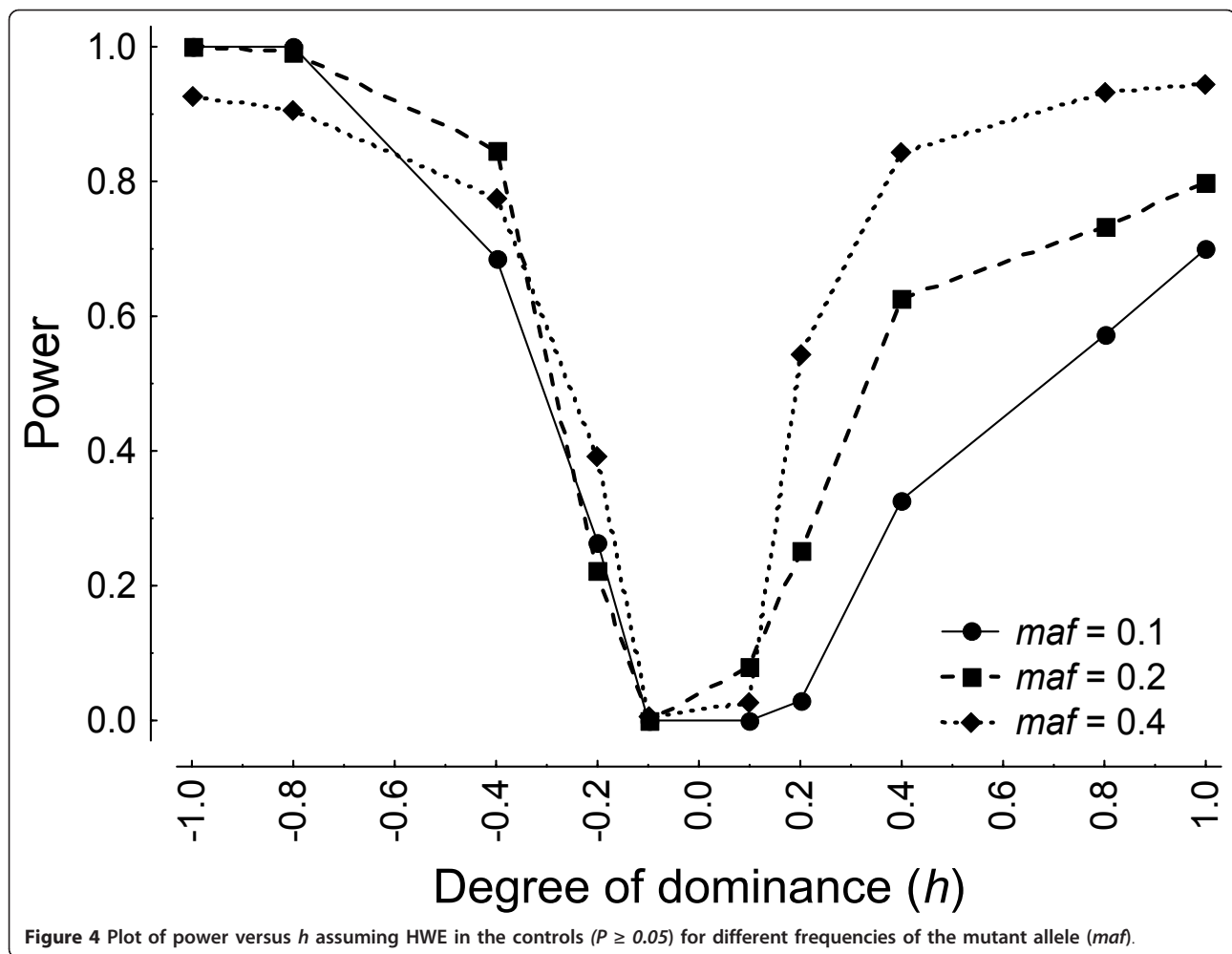


Figure 3 Performance of the h -index against the true underlying mode of inheritance given by parameter H for allele frequencies in the range $(0.05 \leq q_1 \leq 0.4$ and selection coefficient $s = 0.2$ and 0.50 , when $400/400$ controls/cases are sampled from the parametric cohort. Each plot also gives the Spearman correlation between H and h -index.



Additive and dominant contrasts can now be generally written as

$$\begin{aligned} \theta_a &= \left(\frac{n_{32}}{n_{12}} \right) \times \left(\frac{p^2 + D}{q^2 + D} \right) \\ \theta_{co} &= \left(\frac{n_{22}}{n_{12} + n_{32}} \right) \times \left(\frac{p^2 + q^2 + 2D}{2pq - 2D} \right). \end{aligned} \quad (11)$$

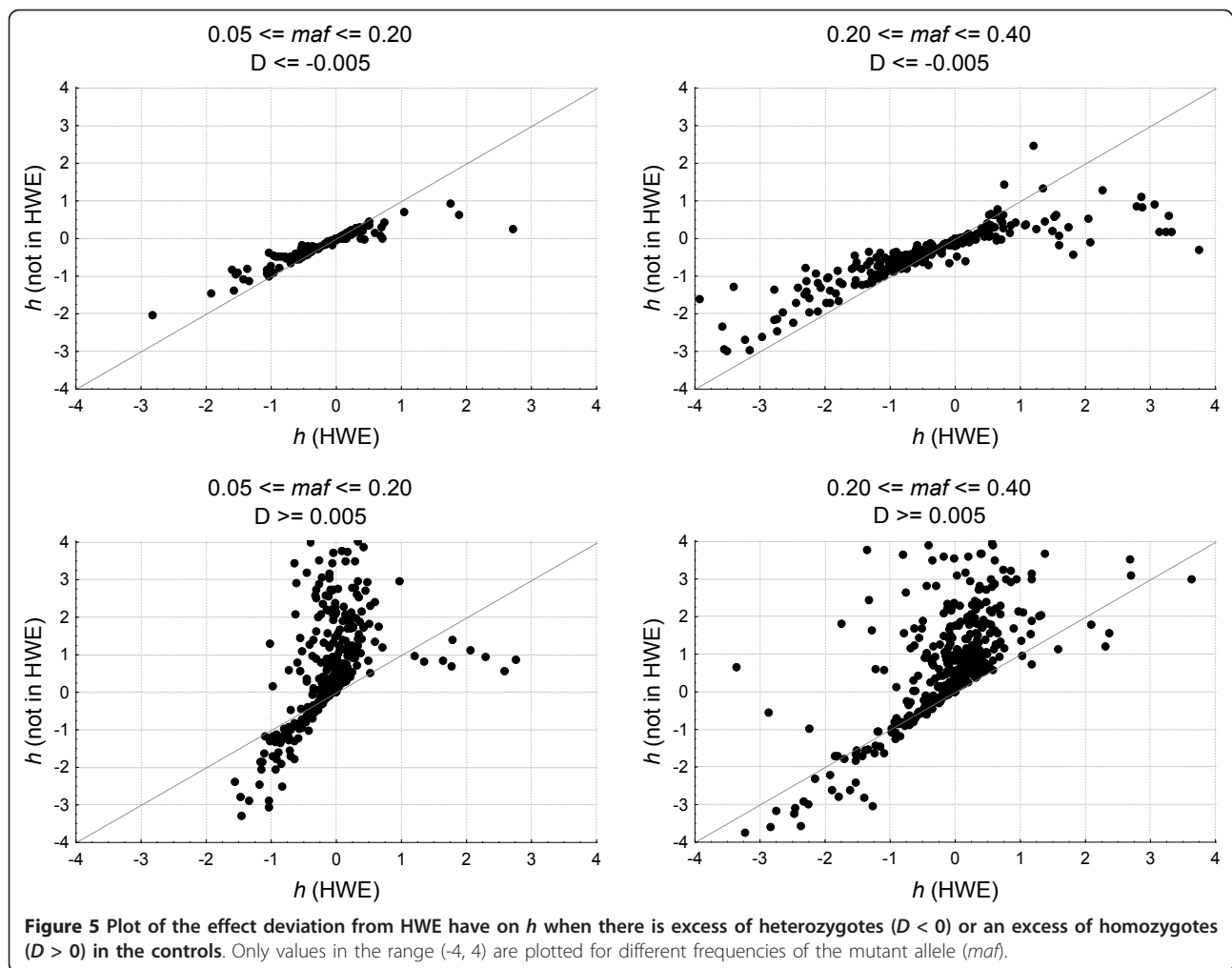
From (10) it is straightforward to estimate the impact deviations from HWE will have in the two contrasts (reasonably assuming $p \geq q$)

$$\begin{aligned} \frac{n_{32}}{n_{12}} \times \frac{p}{q} \leq \theta_a (\text{HWE}) \leq \infty \\ \infty \geq \theta_{co} (\text{HWE}) \geq \left(\frac{n_{22}}{n_{12} + n_{32}} \right), \end{aligned} \quad (12)$$

where the first terms in the inequalities are for a positive value of D ($D = pq$, which obviously implies that there is an excess of homozygotes), and the last terms are for negative values ($D = -q^2$, with an excess of

heterozygotes). From (12) it is now clear that population stratification (i.e., when a sample is composed of subsamples that differ in allele frequencies, thus generating and excess of homozygotes due to the well-known Wahlund's principle in population genetics [12]) will generally upward bias the absolute value of h , where the opposite is true when there is an excess of heterozygotes in the controls.

Simulations were performed by randomly generating subject cases with genotype frequencies $n_{i2} (i = 1, 2, 3)$ under the restriction $\sum_i n_{i2} = 400$ and $0.05 \leq q_{\text{cases}} \leq 0.40$ for allele mt . Genotype frequencies in the controls were modeled following Weir [16] as indicated above, with the restriction $q_{\text{controls}} \leq q_{\text{cases}}$ and $D \geq 0.005$ ($D \leq -0.005$). Note that when modeled in this way we are assuming parametric genotype frequencies in the controls; that is, we assume perfect HWE in the controls with allele frequencies equal to the simulated values. Figure 5 illustrates the bias incurred when estimating h , and also points to an asymmetry in the sense that within



the range $-1 \leq h \leq 1$ studies where $D < 0$ could perhaps be included without the bias being too serious (note, however, that in some cases h changes sign), but when $D > 0$ occurs we may not capture the true direction of dominance.

We now illustrate the proposed methodology by applying it to three working examples. Thereafter, an empirical study using a database of 831 GAS was carried out.

GAS with non-dominance

A GAS investigating the association between the *ACE D/I* polymorphism and coronary artery disease (CAD) produced the following genotype distributions [18]:

Genotype	Controls	Cases
II	64 (30.6%)	34 (16.6%)
ID	94 (45.0%)	89 (43.4%)
DD	51 (24.4%)	82 (40%)

After fitting the logit model (2) the change in deviance due to the genotype effect was $D_g = 16.72$ with $df_g = 2$ ($P < 0.01$); thus, there is significant association between *ACE D/I* polymorphism and CAD. Now, we expect at least one of the orthogonal contrasts to be significant. When the genotype effect was split into its two independent contrasts and the logit model was re-fitted following equation (5), the changes in deviance were $D_{L_a} = 16.62$ with $df_{L_a} = 1$ ($P < 0.01$), and $D_{L_{co}} = 0.10$ with $df_{L_{co}} = 1$ ($P = 0.75$), for the additive and co-dominant contrasts; respectively. $D_{L_a} + D_{L_{co}} = 16.72$ as expected because the two degrees of freedom of the genotype effect were orthogonally decomposed into its two genetic components. Given that $D_{L_{co}}$ was not significant the data suggest that the mode of inheritance can be non-dominance. The interpretation would then be that the heterozygote *ID* has a risk of being diseased that lies in the middle of the risk-protected *II* and risk-exposed *DD* homozygous genotypes.

GAS with dominance

A GAS investigating the association between the alleles *ADH2*1* and *ADH2*2* with alcoholism produced the following genotype distributions [19]:

Genotype	Controls	Cases
*2/*2	448 (82.2%)	238 (70.0%)
*2/*1	93 (17.1%)	85 (25.0%)
*1/*1	4 (0.7%)	17 (5.0%)

The logit model with the genotype effect was fitted and the deviances were significant for both contrasts ($P < 0.01$ for $D_{L_{co}}$ and D_{L_a}). Because the co-dominant model is significant, we proceed to inquiry about the degree of dominance, which here is $h = \ln(\theta_{co})/|\ln(\theta_a)| = 0.48/|2.08| = 0.23$, suggesting dominance for the risk-associated allele *1. Therefore, we may conclude that the homozygous *1/*1 has a greater risk of being alcoholic than the homozygous *2/*2, and that the heterozygote *2/*1 has a risk of alcoholism closer to the *1/*1 homozygote than to the midpoint between the two homozygotes.

GAS with under-dominance

A GAS investigating the association between the *BDNF G196A* polymorphism and schizophrenia produced the following genotype distributions [20]

Genotype	Controls	Cases
GG	208 (59.4%)	229 (71.3%)
GA	131 (37.4%)	83 (25.9%)
AA	11 (3.1%)	9 (2.8%)

In this last example, the co-dominant contrast was significant ($P < 0.01$) whereas the deviance for the additive contrast was not ($P < 0.52$). The degree of dominance is $h = \ln(\theta_{co})/|\ln(\theta_a)| = -0.54/|0.30| = -1.82$, therefore there is indication for under-dominance since $h < -1$. The one-sided z -test for under-dominance ($h < -1$ or $k < 0$ or) is statistically significant ($P = 0.02$). Thus, it seems that the heterozygote *GA* has the least risk of disease (or higher protection) than both homozygotes, which do not statistically differ between them.

Empirical study

A database of 831 GAS archived in the Department of Biomathematics, University of Thessaly ([http://biomath.](http://biomath.med.uth.gr/)

med.uth.gr/), was utilized for the empirical study [15]. A GAS was considered eligible when (i) it examined bi-allelic polymorphisms; (ii) it provided the complete genotype distribution for diseased subjects and controls of individual studies included in the meta-analysis; (iii) controls were non-diseased; (iv) it was written in English; and (v) considered binary outcomes.

In 208 GAS, the genotype distribution showed a significant association after fitting the logit model (2) with the genotype effect. These studies involved 58 variants from 20 genes in association with 12 phenotypes. Table 1 summarizes the inferred degree of dominance (h) in the GAS, and groups studies according to the mode of inheritance (non-dominance, dominance, over-dominance, and under-dominance), and also according to the significant associations found for the contrasts that were of customary used in previous analyses.

The important point here is that in most GAS (114 out of 208; i.e., 55%) the proposed method defines a non-dominance inheritance, whereas the conventional contrasts applied to these studies would render a statistical significant dominant contrast (64 out of 114; i.e., 56%) or recessive contrast (86 out of 114; i.e., 75%). However, 19% (i.e., 18 out of 94) of the studies with dominance deviated from HWE with $D > 0$ and 20% with $D < 0$; and with non-dominance the values are 3% (i.e., 3 out of 114) with $D > 0$ and 9% with $D < 0$. The majority of studies with non-dominance (86%) or dominance (84%) were underpowered (power < 0.5); and only 6% and 5%, respectively, have power > 0.8 .

Also important is that in a significant proportion of studies (25%) our method detects under- (12%) or over-dominance (13%). Figure 6 plots $-\log[P]$ as a function of h for the co-dominant and the additive contrasts. Statistical significance of the additive model was (with a few exceptions) only found when under- or over-dominance were absent.

Discussion

Herein, we have proposed to identify the mode of inheritance in a continuous scale using the degree of dominance h , which is based on the ratio of the odds ratio of the co-dominant contrast divided by the absolute value of the odds ratio of the additive contrast. Numerical

Table 1 Results from 208 GAS studies that render a statistically significant association.

Degree of dominance	Total (*)	Number of GAS with significant contrasts			
		Co-dominant	Additive	Dominant	Recessive
Non-dominance ($h = 0$)	114	0	114	64	86
Dominance ($h \neq 0$)	94	94	22	59	29
Under-dominance ($h < -1$)	26(18)	26	1	10	9
Dominance ($-1 \leq h < 0$) or ($0 < h \leq 1$)	41	41	21	34	16
Over-dominance ($h > 1$)	27(15)	27	0	15	4

The degree of dominance is presented. The results produced by applying the dominant and recessive contrasts are also shown.

* Studies with significant ($P < 0.05$) under- or over-dominance

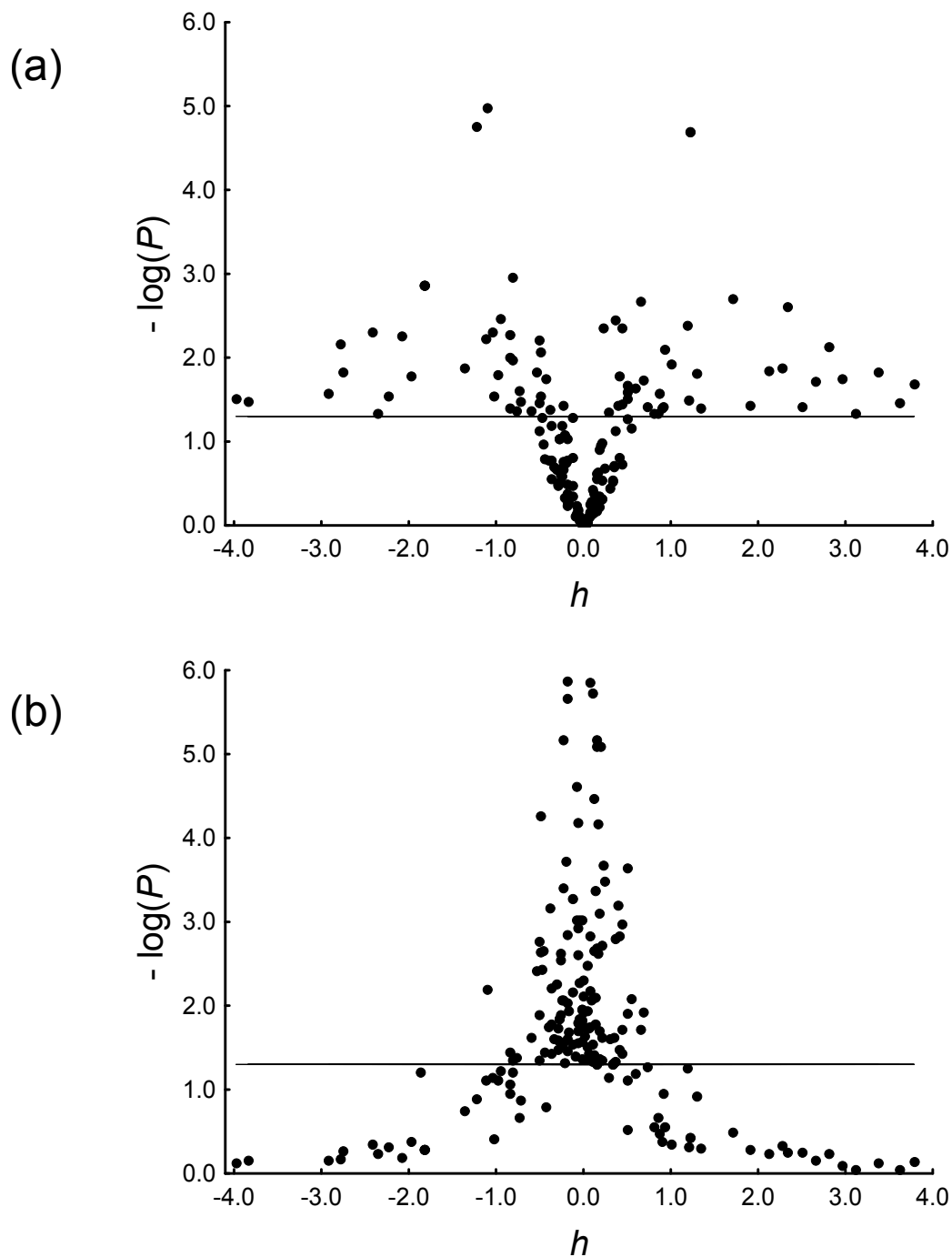


Figure 6 Empirical behavior of the degree of dominance h from 208 GAS showing significant genotype effects. The plots show $-\log(P)$ as a function of h for (a) the co-dominant and (b) the additive contrasts, and the horizontal line at $-\log(P) = 1.3$ represents the critical value above which statistical significance ($P < 0.05$) is attained. There is a clear reverse behaviour in the plots and, as predicted from our numerical study, statistical significance of the additive model was found when under- or over-dominance were absent (i.e., $|h| \leq 1$).

results suggest that the h index captures the essence of what should be understood by genetic model or mode of inheritance. A meticulous analysis has been performed to check performance against an *a priori* model where we already know that a mutant allele is associated to a

disease and also the degree of dominance of this allele. Simulations also show that the degree of dominance h is affected by deviations from HWE, although the bias is more serious when there is population stratification. In these cases the findings should be interpreted carefully,

and adjustments for departures from HWE might be applied [1,10]. Furthermore, power for detecting significance for h when the study conforms HWE rule increases with the degree of dominance and to some extent is related to the mutant allele frequency.

The empirical study we carried out showed the degree of dominance may sufficiently indicate the mode of inheritance. Any degree of dominance exists when the co-dominant contrast is significant irrespectively to the additive contrast. The co-dominant and additive contrasts show a reverse pattern in h and, also important, in the range of over- or under-dominance the additive contrast is non-significant. In general, candidate-gene studies have a tendency to lack power for detecting dominance arising from weak genetic contributions of common variants; though, large genome-wide association studies have been undertaken recently and an effort to create consortia for data sharing is initiated [21,22]. An underpowered GAS will underestimate the significance of the orthogonal contrasts and, therefore, the value of h . Nevertheless, the power to detect the significance of the co-dominant contrast and/or the additive contrast is the same with the power to detect a significance association between the genotype distribution and the disease using a logistic regression with explanatory variable the genotype with three levels. The proposed index may be applied to both types of GAS (candidate-gene studies and GWAS) in the same way (of course the recording of the genotype distribution is a necessary condition). However, in testing the significance of the orthogonal contrasts for an individual variant of a GWAS a multiplicity adjustment should be considered.

In order to avoid the obstacle of the multiple genetic contrasts, Zintzaras [4] proposed the concept of generalized odds ratio (OR_G) for analyzing and meta-analyzing GAS. OR_G is a single statistic that summarizes the magnitude and significance of the association without considering the hash of possible contrasts, and provides a straightforward interpretation of the results in GAS. The OR_G utilizes the complete genotype distribution and provides an estimate of the magnitude of the association, given that the mutational load and the phenotype (bi-allelic or multi-allelic) are treated as a graded exposure (case-control or disease progression). Specifically, OR_G express the probability of a subject being diseased relative to probability of being free of disease, given that the diseased subject has a higher mutational load than the non-diseased. OR_G with values greater than one suggests that disease is proportional to increased genetic exposure and inversely proportional for values less than one. Thus, the application of OR_G may overcome the shortcomings of multiple model testing or erroneous model specification and provides an alternative and robust way for genetic association testing.

Regarding meta-analysis, model-free approaches have been proposed to estimate the genetic model [6,23]. However, we would like to stress that merging studies with potentially heterogeneous modes of inheritance should be avoided since we could entirely miss the true biological meaning underlying disease susceptibility. The application of our proposed method to identify the mode of inheritance warrants further investigation in this context. Although the methodology proposed here is straightforward, a Bayesian approach for implementing the method might be more desirable, especially when there is prior estimation of the magnitude and accuracy of the genetic risk effect and the degree of dominance [24,25].

Conclusion

The introduction of the degree of dominance h , which is the ratio of the two orthogonal contrasts, may provide useful insights into the mode of disease inheritance.

Acknowledgements

E.Z. thanks Chris Schmid, Issa Dahabreh and George Kitsios at Tufts Medical Center for useful discussion on the subject. M.S. is supported by grants 2009SGR 636 from Generalitat de Catalunya to the Grup de Biologia Evolutiva and CGL2010-15395 from the Ministerio de Ciencia e Innovación (Spain), and by the ICREA Acadèmia program.

Author details

¹Department of Biomathematics, University of Thessaly School of Medicine, 2 Panepistimiou Str, Biopolis, Larissa 41110, Greece. ²The Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Tufts University School of Medicine, 800 Washington Str, Boston, MA 02111, USA.

³Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

Authors' contributions

Both authors contributed equally to this paper and they read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 19 July 2011 Accepted: 21 December 2011

Published: 21 December 2011

References

1. Zintzaras E, Lau J: Synthesis of genetic association studies for pertinent gene-disease associations requires appropriate methodological and statistical approaches. *Journal of Clinical Epidemiology* 2008, **61**:634-645.
2. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K: A comprehensive review of genetic association studies. *Genetics in Medicine* 2004, **4**:45-61.
3. Munafò MR, Flint J: Meta-analysis of genetic association studies. *Trends in Genetics* 2004, **20**:439-44.
4. Zintzaras E: The generalized odds ratio as a measure of genetic risk effect in the analysis and meta-analysis of association studies. *Statistical Applications in Genetics and Molecular Biology* 2010, **9**, Article 21.
5. Lewis CM: Genetic association studies: design, analysis and interpretation. *Brief Bioinform* 2002, **3**(2):146-53.
6. Minelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J: The choice of a genetic model in the meta-analysis of molecular association studies. *International Journal of Epidemiology* 2005, **34**:1319-1328.
7. Attia J, Thakkinstian A, D'Este C: Meta-analyses of molecular association studies: methodologic lessons for genetic epidemiology. *Journal of Clinical Epidemiology* 2003, **56**:297-303.

8. McCullagh P, Nelder JA: **Generalized Linear Models**. Chapman and Hall: London; 1989.
9. Mead R: **The Design of Experiments**. Cambridge University Press: Cambridge; 1988.
10. Zintzaras E: **Variance estimation of allele-based odds ratio in the absence of Hardy-Weinberg equilibrium**. *European Journal of Epidemiology* 2008, **23**:323-326.
11. Kokoska S, Nevison C: **Statistical Tables and Formulae**. Springer-Verlag: New York; 1992.
12. Crow JF, Kimura M: **An Introduction to Population Genetics Theory**. Harper & Row: London; 1970.
13. Falconer DS, Mackay TFC: **Introduction to Quantitative Genetics**. Longmans Green: Harlow; 4 1996.
14. Thakkinstian A, Thompson JR, Minelli C, Attia J: **Choosing between per-genotype, per-allele, and trend approaches for initial detection of gene-disease association**. *Journal of Applied Statistics* 2009, **36**:633-646.
15. Zintzaras E: **Impact of Hardy-Weinberg equilibrium deviation on allele-based risk effect of genetic association studies and meta-analysis**. *European Journal of Epidemiology* 2010, **25**:553-560.
16. Weir BS: **Genetic Data Analysis**. Sinauer Associates: Massachusetts; 1990.
17. Hernández JL, Weir BS: **A disequilibrium coefficient approach to Hardy-Weinberg testing**. *Biometrics* 1989, **45**:53-70.
18. Fatini C, Abbate R, Pepe G, Battaglini B, Gensini F, Ruggiano G, Gensini GF, Guazzelli R: **Searching for a better assessment of the individual coronary risk profile. The role of angiotensin-converting enzyme, angiotensin II type 1 receptor and angiotensinogen gene polymorphisms**. *Eur Heart J* 2000, **21**(8):633-8.
19. Zintzaras E, Stefanidis I, Santos M, Vidal F: **Do alcohol-metabolizing enzyme gene polymorphisms increase the risk of alcoholism and alcoholic liver disease?** *Hepatology* 2006, **43**:352-361.
20. Neves-Pereira M, Cheung JK, Pasdar A, Zhang F, Breen G, Yates P, Sinclair M, Crombie C, Walker N, St Clair DM: **BDNF gene is a risk factor for schizophrenia in a Scottish population**. *Mol Psychiatry* 2005, **10**(2):208-12.
21. Zintzaras E, Lau J: **Trends in meta-analysis of genetic association studies**. *Journal of Human Genetics* 2008, **53**:1-9.
22. Wang WY, Barratt BJ, Clayton DG, Todd JA: **Genome-wide association studies: theoretical and practical concerns**. *Nature Review Genetics* 2005, **6**:109-18.
23. 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.
24. Minelli C, Thompson JR, Abrams KR, Lambert PC: **Bayesian implementation of a genetic model-free approach to the meta-analysis of genetic association studies**. *Statistics in Medicine* 2005, **24**:3845-3861.
25. Spiegelhalter DJ, Abrams KR, Myles JP: **Bayesian Approaches to Clinical Trials and Health Care Evaluation**. Wiley: New York; 2004.

Pre-publication history

The pre-publication history for this paper can be accessed here:
<http://www.biomedcentral.com/1471-2288/11/171/prepub>

doi:10.1186/1471-2288-11-171

Cite this article as: Zintzaras and Santos: Estimating the mode of inheritance in genetic association studies of qualitative traits based on the degree of dominance index. *BMC Medical Research Methodology* 2011 **11**:171.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

