Cautions of Using Allele-Based Tests Under Heterosis

Bernard Omolo[1,*], Hongmei Zhang[2] and Wilfried Karmaus[2]

[1]*Division of Mathematics & Computer Science, University of South Carolina - Upstate, Spartanburg, SC 29303, USA*

[2]*Department of Epidemiology and Biostatistics, University of South Carolina - Columbia, Columbia, SC 29208, USA*

Abstract: In genetic studies, heterotic effects are commonly assessed as dominant, additive, or recessive effects for a given genetic marker. However, the distorting effect of heterosis on statistical tests is non-trivial. An inheritance model needs to be carefully chosen to achieve highest testing power. We assess this through simulations *via* allele- and genotype-based tests. Chi-square test statistics for different inheritance models are formulated as a function of relative risks and allele frequencies. The results indicate that testing power from the commonly used allele-based tests can be substantially diminished by heterosis. Assessing the existence of heterosis is thus recommended to avoid false negative findings.

Keywords: Chi-square test, dynamic systems theory, genotype, haplotype, statistical power.

1. INTRODUCTION

The genetic phenomenon whereby the phenotypic levels for the heterozygous genotypes are either greater or less than either of the homozygous genotypes is called heterosis. Positive heterosis presents when there is an increased phenotypic value for the heterozygous genotype, whereas negative heterosis presents when there is a reduced value. At the molecular level, heterosis appears to be counterintuitive to the expectation that one homozygous genotype should be associated with increased adverse outcomes, the other homozygous genotype with decreased adverse outcomes, and the heterozygous genotype should be intermediate [1]. The underlying mechanism of heterosis and its influence on phenotypes coincides with the dynamic systems theory in mathematics and physics and its relation with development, in the sense that dynamic systems provides theoretical principles for the formulation of these complex outcomes [2]. Recently, heterosis received increased attention. For the period between 1995 and 1999, a total of 147 publications on heterosis are listed in PubMed. The number increased to 751 between 2000 and 2004, and to 1635 articles between 2005 and 2009.

Heterosis has a long and an unsolved history. In 1917, Jones stated "that a stimulation resulting from hybridization in both plants and animals has long been recognized. The increased growth as the result of crossing is so common an occurrence that it is probably familiar to everyone who has made any hybridization experiments. This stimulation, variously spoken of as "hybrid vigor", stimulus due to heterozygosis, heterosis, etc., was clearly established as an organic phenomenon by the abundant cases cited by early investigators such as Kölreuter (1766), Gärtner (1849), Darwin (1877) and Focke (1881), as well as a large number of other investigators at that time and an increasingly large number since then" [3]. Jones continued "Concrete explanations as to the cause of these results have not accompanied the accumulation of facts. Various hypotheses have attempted to account for the results, but they have been little more than outlines of the problem". Still, more than ninety years later, there is no conclusive explanation for this phenomenon [4]. An instructive description of these controversies explaining heterosis was provided by Crow in 2008 [5]. Explanations include, for instance, overdominance, dominance, pseudo-overdominance, and gene repression and activation (epigenetic regulation) [6].

Heterosis occurs in plant, animal, and human studies [7-9]. Hybrid vigor (superior growth and fertility over their parents) is frequently used in agriculture. It is also estimated that 65% of the maize production world-wide is hybrid-based [9, 10]. Beef production has increased due to hybrid vigor [11]. Heterozygous advantage (heterosis) has also been thought to confer resistance to certain strains of malaria in patients heterozygous for the sickle-cell gene *HbS* [12]. Whilst the phenomenon is well described in the plant and animal kingdom, it has largely been ignored in human genetics [1]. Typically in human genetics, heterosis is detected in studies that focused on specific genes [1,

*Address correspondence to this author at the Division of Mathematics & Computer Science, University of South Carolina - Upstate, Spartanburg, SC 29303, USA; Tel: 864-503-5362; Fax: 864-503-5930;
E-mail: bomolo@uscupstate.edu

E-ISSN: 1929-6029/13

12-17] but not in genome-wide association studies. The numerical explanation for this bias is simple. Figure **1** shows an example of presence of positive heterosis. The standard genotype-based analysis based on a single nucleotide polymorphism (SNP) will detect an increased disease frequency in the heterozygous "*Gg*" genotype (20/100 vs. 15/100 in the homozygous groups), but the allele-based analysis (*g* vs. *G*;50/300 each) does not show increased occurrence of disease in the two alleles. These allele-based tests, however, are the basis of haplotype analyses and genome-wide association studies [18].

Comings *et al*. noted that heterosis may occur in up to 50% of all genetic associations [1]. Yet most studies select inheritance models without taking into account the possible existence of heterosis. No study in the genetic literature has determined if heterosis (positive or negative) will cause substantial power loss and lead to erroneous conclusions. The high prevalence of heterosis and the uncertain impact of heterosis in association studies motivated the work presented in this article. Through simulations, we examined the influence of heterosis on testing power in allele- or genotype-based association tests. Generic test statistics were derived for different inheritance models (dominant, recessive, and additive), the standard genotype-based test (based on the standard 2x3 contingency table), as well as the allele-based test.

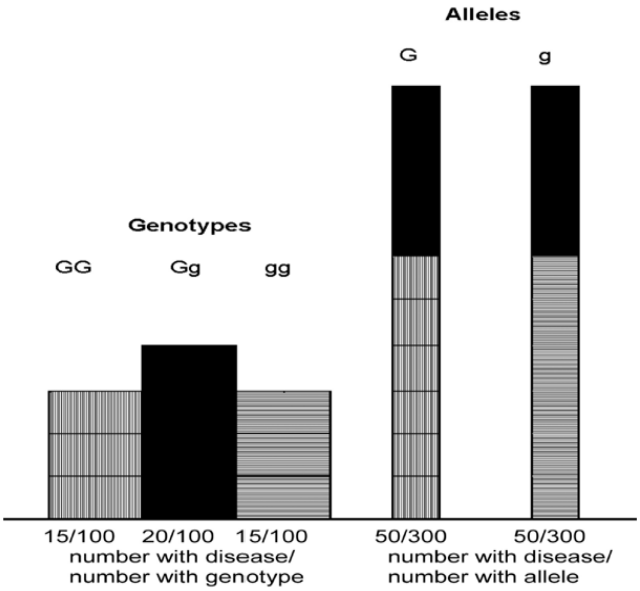

Figure 1: Distortion of the genotype-disease association (left) and the allele analysis (right) in the case of positive heterosis. Each bar represents the proportion of diseased subjects for each genotype or allele. The genotype *Gg* shows an increased proportion of diseased subjects whereas the *G* and *g* alleles indicate no differences in the proportion of diseased subjects.

2. METHODS

In the following, we describe test statistics for the standard genotype-based test (comparing frequencies of *gg*, *Gg*, and *GG*), genotype-based tests under various inheritance models (dominant, recessive, additive) and allele-based test (comparing frequencies of the alleles *g* vs. *G*). Pearson chi-square test statistics are calculated based on cohort studies using a dichotomous outcome variable. Comparable statistics can be determined for case-control studies [19] and for continuous traits. We use the test statistics to compare statistical significance between different tests in the absence and presence of negative and positive heterosis. Statistical power for each test are evaluated based on non-centrality parameters, which are calculated in a similar way but with population parametersincluded in the calculation [20, 21].

2.1. Test Statistics

For all simulations, the Hardy-Weinberg equilibrium $(q = 1 - p; P(GG) = p^2; P(Gg) = 2pq; P(gg) = q^2)$ was maintained for the allele and genotype frequencies. The penetrances for *gg*, *Gg* and *GG* are denoted by $f_0 = P(D|gg), f_1 = P(D|Gg)$ and $f_2 = |P(GG)$ respectively, where *D* represents disease status. Further, let *N* denote the overall sample size, N_0 the observed counts with genotype *gg*, and N_0^D the observed counts for diseased individuals with genotype *gg*. Let N_1, N_1^D, N_2 and N_2^D carry the same notation for *Gg* and *GG*, respectively. Hence, the penetrance estimates are defined as $\hat{f}_0 = \frac{N_0^D}{N_0}$, $\hat{f}_{1=} \frac{N_1^D}{N_1}$, and $\hat{f}_{2=} \frac{N_2^D}{N_2}$.

Denote the estimated frequency of the allele (*G*) as $\hat{p}$; consequently $\hat{q} = 1 - \hat{p}$. The expected frequencies for each disease status corresponding to each genotype can then be described as a function of $N, \hat{p}, \hat{f}_0, \hat{f}_1,$ and $\hat{f}_2$ as follows. For genotype *gg*, let E_0^D denote the expected counts of diseased under the null hypothesis of no association between genotype and disease status and $E_0^{\bar{D}}$ denote the expected counts of non-diseased. Through some algebra, we have $E_0^D = N\hat{q}^2(\hat{q}^2\hat{f}_0 + 2\hat{p}\hat{q}\hat{f}_1 + \hat{p}^2\hat{f}_2)$ and $E_0^{\overline{D}} = N\hat{q}^2(1 - \hat{q}^2\hat{f}_0 - 2\hat{p}\hat{q}\hat{f}_1 - \hat{p}^2\hat{f}_2)$. Let $E_1^D, E_2^D, E_1^{\bar{D}},$ and $E_2^{\bar{D}}$ carry the same notation for genotypes *Gg* and *GG*, respectively. Calculations of these quantities can be done in a similar way. The chi-square test statistic for the standard genotype-based test can be derived as

$$\chi_g^2 = \frac{A + B + C}{(\hat{q}^2\hat{f}_0 + 2\hat{p}\hat{q}\hat{f}_1 + \hat{p}^2\hat{f}_2)(1 - \hat{q}^2\hat{f}_0 - 2\hat{p}\hat{q}\hat{f}_1 - \hat{p}^2\hat{f}_2)}$$

where

$$A = N\hat{q}^2\left[(1-\hat{q}^2)\hat{f}_0 - 2\hat{p}\hat{q}\hat{f}_1 - \hat{p}^2\hat{f}_2\right]^2,$$

$$B = 2N\hat{p}\hat{q}\left[(1-2\hat{p}\hat{q})\hat{f}_1 - \hat{q}^2\hat{f}_0 - \hat{p}^2\hat{f}_2\right]^2, \text{and}$$

$$C = N\hat{p}^2\left[(1-\hat{p}^2)\hat{f}_2 - \hat{q}^2\hat{f}_0 - 2\hat{p}\hat{q}\hat{f}_1\right]^2.$$

Now, under the dominant model, genotypes *GG* and *Gg* have the same effect on the trait; hence $f_1 = f_2$. Under the recessive model, genotype *Gg* and *gg* have the same effect on the trait and so $f_1 = f_0$. Finally, under the additive model, genotype *Gg* has an intermediate effect on the trait (between *gg* and *GG*) and so $f_1 = \frac{f_0 = f_2}{2}$. However, the relationships between these penetrances will not be true in the presence of heterosis as discussed later in this section. Based on these penetrance relations with heterosis absent, the chi-square tests statistics for different genetic models can be derived. Let $\chi_d^2, \chi_r^2, \chi_c^2, \text{and } \chi_a^2$ denote the chi-square statistics for the dominant, recessive, additive genotype-based tests and the allele-based test, respectively.

Then

$$\chi_d^2 = \frac{N\hat{q}^2(1-\hat{q}^2)(\hat{f}_0-\hat{f}_1)^2}{[\hat{q}^2(\hat{f}_0-\hat{f}_1)+\hat{f}_1][1-\hat{q}^2(\hat{f}_0-\hat{f}_1)-\hat{f}_1]},$$

$$\chi_r^2 = \frac{N\hat{p}^2(1-\hat{p}^2)(\hat{f}_1-\hat{f}_2)^2}{[\hat{f}_1-\hat{p}^2(\hat{f}_1-\hat{f}_2)][1-\hat{f}_1+\hat{p}^2(\hat{f}_1-\hat{f}_2)]}, \text{ and}$$

$$\chi_c^2 = \frac{N^2(N-1)\left[2\hat{p}\left(N(\hat{q}\hat{f}_1+\hat{p}\hat{f}_2)-(\hat{q}^2\hat{f}_0+2\hat{p}\hat{q}\hat{f}_1+\hat{p}^2\hat{f}_2)\right)\right]^2}{[A_1+B_1]},$$

where

$$A_1 = N^2\left(\hat{q}^2\hat{f}_0 + 2\hat{p}\hat{q}\hat{f}_1 + \hat{p}^2\hat{f}_2\right)\left(\hat{q}^2(1-\hat{f}_0) + 2\hat{p}\hat{q}(1-\hat{f}_1) + \hat{p}^2(1-\hat{f}_2)\right)$$

and

$$B_1 = [N(2\hat{p}\hat{q} + 4N\hat{p}^2) - (2N\hat{p}\hat{q} + 2N\hat{p}^2)^2].$$

Lastly,

$$\chi_a^2 = \frac{2N[A_2 - B_2]^2}{C_2 D_2}$$

where

$$A_2 = (\hat{q}^2\hat{f}_0 + \hat{p}\hat{q}\hat{f}_1)(\hat{p}\hat{q}(1-\hat{f}_1) + \hat{p}^2(1-\hat{f}^2)),$$
$$B_2 = (\hat{p}\hat{q}\hat{f}_1 + \hat{p}^2\hat{f}_2)(\hat{q}^2\hat{f}_1 + \hat{p}^2\hat{f}_2)(\hat{q}^2(1-\hat{f}_0) + \hat{p}\hat{q}(1-\hat{f}_1)),$$
$$C_2 = (\hat{q}^2 + \hat{p}\hat{q})(\hat{p}^2 + \hat{p}\hat{q})(\hat{q}^2\hat{f}_0 + 2\hat{p}\hat{q}\hat{f}_1 + \hat{p}^2\hat{f}_2), \text{ and}$$
$$D_2 = (1 - \hat{q}^2\hat{f}_0 - 2\hat{p}\hat{q}\hat{f}_1 - \hat{p}^2\hat{f}_2)$$

In the presence of heterosis, the relationship between $f_0, f_1, \text{and } f_2$ are different from those described above. Positive heterosis is characterized by f_1 being the highest and negative heterosis as f_1 being the smallest. Minelli *et al.* defined heterosis based on relative risks [22]. The genotypic relative risk of *Gg* compared to *gg* was denoted as $\gamma_1 = \frac{f_1}{f_0}$ and the relative risk for *GG* compared to *gg* was denoted as $\gamma_2 = \frac{f_2}{f_0}$. Further, the ratio of log relative risks was denoted as $\gamma = \frac{\log(\gamma_1)}{\log(\gamma_2)}$, so that $\gamma >$ indicates positive heterosis and $\gamma < 0$ denotes negative heterosis [22]. Different values of $|\gamma|$ indicate different strengths of heterosis. Table **1** below summarizes these definitions for the dominant, recessive, additive, and heterotic modes of inheritance in ideal situations.

2.2. Simulations

In the following simulations we compare the statistical power of the allele-based, standard genotype-based, and model-based tests of association considering the status of heterosis. Here model-based tests refer to genotype-based tests under dominant, recessive, and additive inheritance modes. We considered three scenarios for the simulations: (1) no heterosis, (3) possible negative heterosis, and (4) possible positive heterosis. For fixed values of the

Table 1: Ratios of Log Genotype Relative Risks (GRR) for Various Genetic Models. λ* Implies that $\gamma_2 > 1$

Genotype	GRR	Recessive	Dominant	Additive	Negative heterosis	Positive heterosis
Gg	γ_1	1	γ_1	γ_1	$\gamma_1 = \gamma_2^{\lambda}$	$\gamma_1 = \gamma_2^{\lambda}$
GG	γ_2	γ_2	γ_1	$2\gamma_1 - 1$	$\gamma_2 (\gamma_1 < \gamma_2)$	$\gamma_2 (\gamma_1 > \gamma_2)$
λ		0	1	0.5	$\lambda^* < 0$	$\lambda^* > 1$

sample size (*N*) and λ, the penetrance estimates $(\hat{f}_0, \hat{f}_1,$ and $\hat{f}_2)$ and the chi-square test statistics were obtained for each association test. A sample size of *N* = 600 was used. Four values of $\hat{\lambda}$ were considered in the simulations, -0.1, -2.0, 1.1, and 2.0, corresponding to possible negative and positive heterosis, respectively. The values of $(\hat{f}_0, \hat{f}_1,$ and $\hat{f}_2)$ were determined from the given $\hat{\lambda}$ values with $\hat{f}_0$ fixed at 0.1. To account for Monte Carlo error, for each condition of heterosis, 1000 data sets were generated. Each test statistic was calculated for 200 allele frequencies (p) ranging from 0 to 1 and the values were plotted against p. The significance level, α, was fixed at 0.05. The corresponding critical chi-square values were drawn for each scenario (*critical value* = 3.841, *df* = 1 for the test statistic in allele- and model-based tests; and *critical value* = 5.991, *df* = 2 for the test statistic in standard genotype-based tests).

Testing power for each test was calculated based on chi-square non-centrality parameters under the alternative hypothesis. The power at each allele frequency is an average of testing powers over 1000 data sets. All the computations and graphs were produced using SAS 9.2 (SAS Institute, Inc., Cary, North Carolina).

2.3. Examples

To further illustrate the impact of heterosis in allele-based tests and the need to be cautious when using allele-based tests, we describe the results of two asthma studies: the first result showing possible (mild) heterosis and the second showing no pattern of heterosis. The first result was obtained from a prospective study of the natural history of allergic disorders from birth to the age of 18 years, on a cohort from the Isle of Wight (IOW) in the United Kingdom [23]. The genotype and allele frequencies are included in the first part of Table **2**. The percentage of asthma cases for the AA, AG and GG genotypes were used to estimate the penetrances. A (slightly) higher penetrance estimate for the heterozygous genotype (AG) than the homozygous genotypes (AA and GG) suggests the possibility of mild heterosis.

The second result was obtained from the study by Howard *et al*., which examined the association between asthma and atopy phenotypes and IL-13 polymorphisms in a Dutch asthma population [24]. The data is included in the second part of Table **2**. Bronchial hyperesponsiveness (BHR) was used as a biomarker for the asthma phenotype. The percentages of BHR cases for the CC, CT and TT genotypes were used to estimate the penetrances. There was no pattern of heterosis suggested. When there is no pattern of heterosis [24], the allele-based test is more powerful than the genotype-based test, as suggested by the smaller p-value (0.003). However, when mild heterosis appears possible, the allele-based test completely loses its power in identifying the significance of association (p-value = 0.42).

2.4. Results

For the purposes of illustration, we focus on three situations in terms of heterotic status: no heterosis, mild and moderate negative heterosis, and mild and moderate positive heterosis. The results are summarized and discussed as follows.

In the *absence of heterosis* $(\hat{f}_0 = 0.10, \hat{f}_1 = 0.18,$ and $\hat{f}_2 = 0.20)$, the allele-based test has greater statistical power than the recessive model-based and the standard genotype-based tests (data not shown). In a scenario with *mild negative heterosis* $(\hat{\lambda} = -0.1)$ and genotype risks that are still close to the pattern of a recessive model $(\hat{f}_0 = 0.10, \hat{f}_1 = 0.0943,$ and $\hat{f}_2 = 0.18)$, the recessive inheritance model has the highest power while the dominant model has almost no statistical power. It is worth noting that the allele-based test is less powerful than the standard genotype-based test for $p < 0.7$ (Figure **2A**). In this case, even for negative values of $\hat{\lambda}$ close to zero, the allele-based test is inappropriate. In a setting with *moderate negative heterosis* $(\hat{\lambda} = -2.0, \hat{f}_2 = 0.10, \hat{f}_1 = 0.0444,$

Table 2: Examples of Mild Heterosis (Asthma Phenotype) and No Heterosis (BHR Phenotype) from the IOW Cohort and a Dutch Population, Respectively

SNP	Phenotype	Genotype			Allele		P-value (χ^2)	
Exon 4 (3'UTR or C8932052)	Asthma	AA (35)	AG (212)	GG (475)	A (282)	G (1162)	Genotype	Allele
	% cases	14.3	35.4	32.0	30.1	32.6	0.046	0.42
Promoter	BHR	CC (213)	CT (107)	TT (15)	C (533)	T (137)	0.014	0.003
	% cases	59.2	72.0	86.7	61.7	75.2		

and $\hat{f}_2 = 0.15$) the allele-based test has a lower statistical power than the dominant model-based test for $p < 0.5$ (Figure **2B**) and is overall lower than the standard-genotype-base model. However, at extreme allele frequencies, allele-based test tends to have higher statistical power than the standard genotype-based test. The reason is that extreme allele frequencies are related to lower numbers in either one of the homozygous groups. For instance, following the Hardy-Weinberg law, if $p = 0.1$, then the probability of *GG* is 0.01 and if $p = 0.9$ then the probability of *gg* is 0.01 ($(1 - 0.9)^2$). Thus, in both cases, one homozygous group no longer contributes substantially to the test statistic based on genotypes and the allele-based test accumulates most information in the two alleles. We observed similar patterns for smaller p (data not shown).

Under *mild positive heterosis* with $\hat{\lambda} = 1.1 (\hat{f}_0 = 0.10, \hat{f}_1 = 0.214,$ and $\hat{f}_2 = 0.20)$ the allele-based test is less powerful than the dominant model-based test since in this example the genotype risks reflect a dominant mode of inheritance (Figure **3A**). In addition, the allele-based test has lower statistical power than the standard genotype-based test for most values of the allele frequency. Similar situations are obtained for *moderate positive heterosis* $\hat{\lambda} = 2.0 (\hat{f}_0 = 0.10, \hat{f}_1 = 0.20,$ and $\hat{f}_2 = 0.1414)$. For $\hat{\lambda} = 2.0$ the allele-based test is less powerful than the dominant model-based and the standard genotype-based tests. In particular, when the allele frequency is in the mid-range, the allele-based has lowest power (Figure **3B**). For larger $\hat{\lambda}$ we achieved similar results (data not shown).

3. DISCUSSION

In the absence of heterosis, allele-based association tests are, as expected, more powerful than the standard genotype-based and other model-based tests. Under possible negative or positive heterosis, allele-based tests possess lower statistical power than the standard genotype-based and the model-based tests, particularly in the mid-range of the allele frequency distribution, depending on the mode of inheritance. These simulations emphasize the fact that blindly applying allele-based tests without assessing the existence of heterosis can result in misleading inferences due to substantial power loss. To some degree this is avoided by relying on the standard genotype (2 d.f.) test.

It has been noted that, in the human genome, more than 50% of the allele frequencies fall in the middle range. For instance, Kruglyak and Nickerson estimated that a minimal allele frequency of 30% is found in 23–27% of single nucleotide polymorphisms and 40% in 24–28% [25]. Hence, the reduction of statistical power

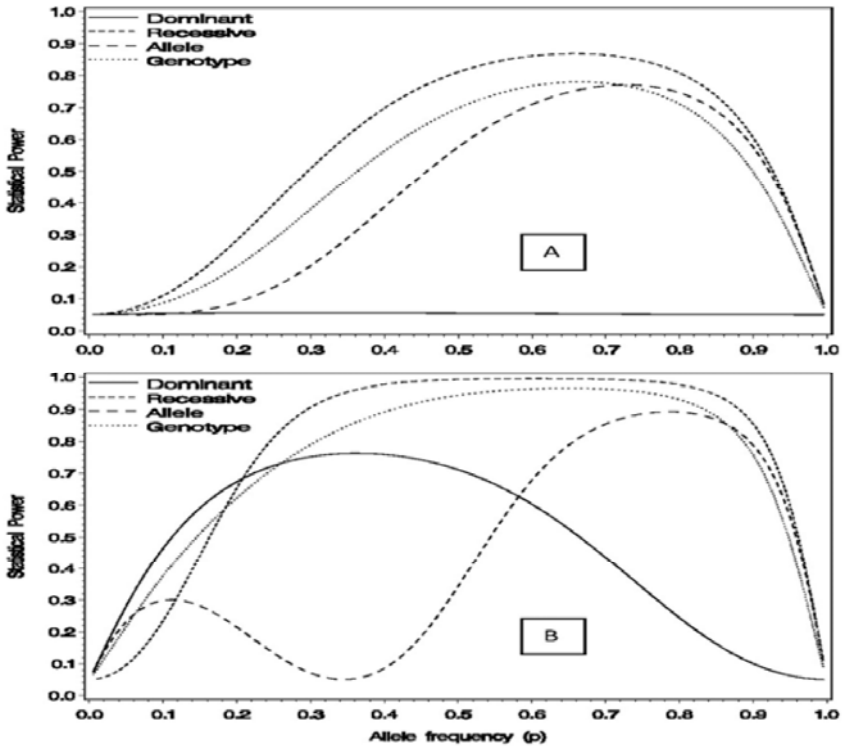

Figure 2: Power plots under negative heterosis. λ = -0.1[A]; λ = -2.0[B].

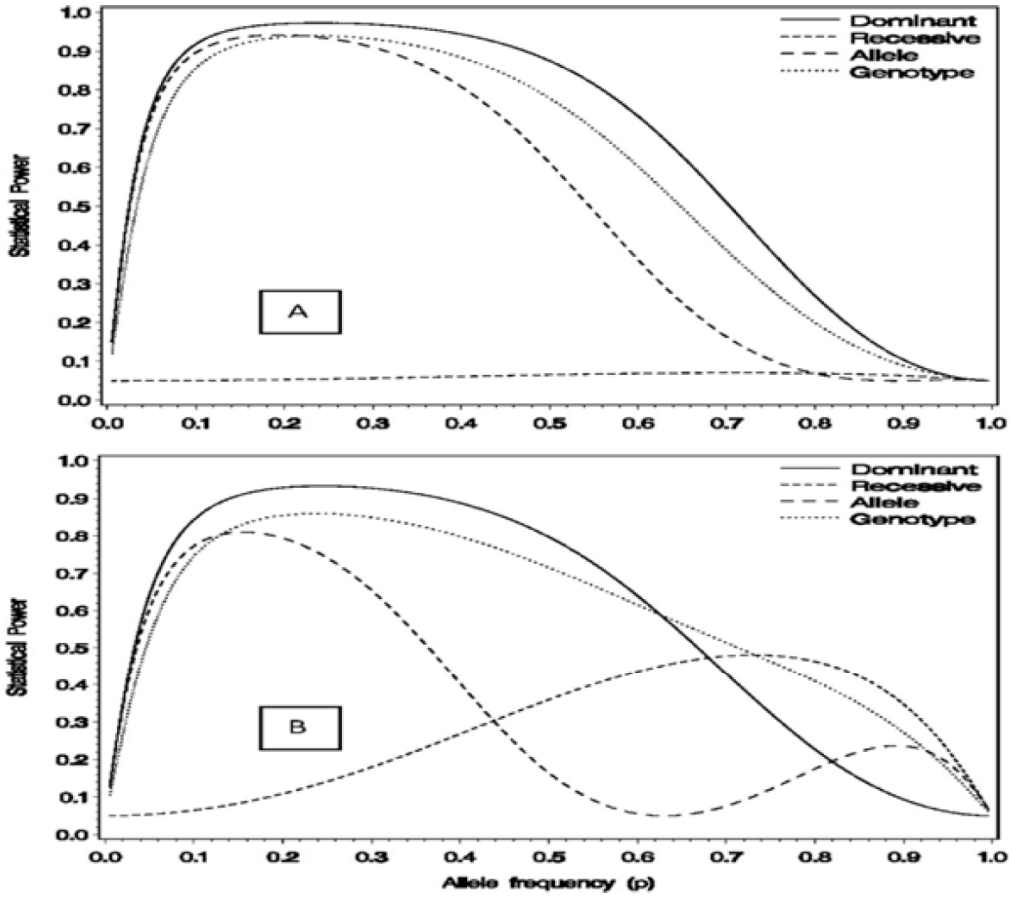

Figure 3: Power plots under positive heterosis. λ = 1.1 [A]; λ = 2.0 [B].

due to heterosis may affect as many as 50% of allele-based tests. Individual alleles from multiple SNPs are used to compute haplotypes [26-28]. Haplotype construction has generated tremendous interest among computational biologists and statistical geneticists. A large number of statistical programs have been developed to estimate haplotypes [27]. Given the fact that allele-based analyses are favored in a large number of studies and also are the backbone of haplotype-based association tests, the results of this study suggests that negative findings of allele-based analyses should be interpreted with caution when heterosis was not assessed. For example, if a haplotype that incorporates three single nucleotide polymorphisms (SNPs) and one of these is heterotic, then the chance of this haplotype to be associated with phenotypic outcomes may be reduced. The extent to which heterotic associations may reduce the statistical power of haplotype association studies is not yet known. In addition, it is not known whether the use of haplotype pairs (diplotypes) will compensate for the reduced power of haplotype association studies. Since genome-wide association studies (GWAS) also utilize allele-based tests (for instance [18, 29]), it is unclear to what degree GWAS are underachieving because heterosis is ignored. In light of the study of Comings *et al.* [1], we speculate that the burden due to this limiting assumption may not be ignorable.

More than ninety years after Jones' description of heterosis there is no conclusive molecular explanation for this phenomenon [4]. There are numerous attempts in the plant kingdom to understand the molecular basis of heterosis [4, 30]. Recent investigations have suggested non-additive gene expression, small RNAs, and epigenetic regulation as an explanation for heterosis [6]. However, in human genetics, heterosis is often not accepted and authors have had to defend findings of heterosis [13, 17, 31, 32] in particular when allele-based associations do not agree with the findings from genotype-based tests [1]. Other authors circumvent the problem and contrast alleles from the heterozygous with only one homozygous group [33]. Hence, there is a need to explore the molecular basis of heterosis and to improve scientific acceptance of heterotic findings in human epidemiology. To better understand heterosis and its impact in different populations, future studies can include additional

factors such as gender, age, and race [1, 16, 34, 35]. In addition, heterosis may also vary with phenotype [1]. The same SNP, for example, may show heterosis for allergic asthma but not for allergic sensitization.

It is true that researchers should attempt to use unbiased methods, but this is not always the case, as we tend be largely unaware of our misconceptions. Therefore, incorporating background knowledge to some extent has the potential to diminish the occurrence of misconception and further increase statistical power. On the other hand, when the model of inheritance is not known, we propose applying a two-step analytic approach in the spirit of the method proposed earlier [36]. The first step is to select the appropriate inheritance mode by statistically comparing dominant, recessive, additive and heterotic models. In the second step, association studies are conducted based on the model selected. In the first step, since all four models have the same degrees of freedom, a likelihood ratio test cannot be applied. Instead, in order to select the most likely inheritance mode, we propose conducting a comparison of the likelihoods of each model; this is a special case of the Akaike Information Criterion (AIC) that is commonly used in model selection. The model with the highest maximum likelihood is selected and utilized in the subsequent association studies.

The concepts discussed in this work can be applied to scientific research in general. The presence of heterosis can possibly generate different results. This is coherent with the dynamic systems framework, that is, different intercommunications may produce very distinct systems and research should take this into account. Furthermore, novel research could benefit from applying multiple methods in combination, thereby providing more powerful and more easily interpretable results. The possible contradictions emerging from such studies may also afford the necessary insights to further understand the phenomena under study.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. D.C. Rao at Washington University in St. Louis and the anonymous referees for their insightful comments that helped to enhance the quality of the manuscript. The work of Bernard Omolo was partially supported by a grant from the National Heart, Lung, and Blood Institute (R25-HL085040). Hongmei Zhang was partially supported by a grant from the National Heart, Lung, and Blood Institute (1R03HL095429-01). The work of Wilfried Karmaus was in part supported by grants of the National Institute of Allergy and Infectious Diseases (1R01AI061471-01A1) and the National Heart, Lung, and Blood Institute (1R01HL082925-01A2).

REFERENCES

[1] Comings DE, MacMurray JP. Molecular heterosis: a review. Mol Genet Metab 2000; 71: 19-31. http://dx.doi.org/10.1006/mgme.2000.3015

[2] Thelen E, Smith LB. Dynamic Systems Theories. In: Damon W, Lerner RM, editors. Handbook of Child Psychology Volume 1, Theoretical Models of Human Development. 6th ed. 2006; pp. 258-312.

[3] Jones DF. Dominance of linked factors as a means of accounting for heterosis. Genetics 1917; 2: 466-79.

[4] Lipman ZB, Samir D. Heterosis: revisiting the magic. Trends Genet 2007; 23: 60-66. http://dx.doi.org/10.1016/j.tig.2006.12.006

[5] Crow JF. Mid-century controversies in population genetics. Annu Rev Genet 2008; 42: 1-16. http://dx.doi.org/10.1146/annurev.genet.42.110807.091612

[6] Chen ZJ. Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci 2010; 15: 57-71. http://dx.doi.org/10.1016/j.tplants.2009.12.003

[7] Hochholdinger F, Hoecker N. Towards the molecular basis of heterosis. Trends Plant Sci 2007; 12: 427-32. http://dx.doi.org/10.1016/j.tplants.2007.08.005

[8] Penrose LS. Evidence of heterosis in man. In: Pros R Sock Lund B Biol Sci 1955; 144: 203-13.

[9] Duvick DN. Heterosis: feeding people and protecting natural resources. In: Coors JG, Pander S, editors, The genetics and Exploitation of Heterosis in Crops. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America 1999; pp. 19-29.

[10] Duvick DN. Biotechnology in the 1930s: the development of hybrid maize. Nat Rev Genet 2001; 2: 69-74. http://dx.doi.org/10.1038/35047587

[11] Marshall DM. Breed differences and genetic parameters for body composition traits in beef cattle. J Amin Sci 1994; 72: 2745-55.

[12] Hedrick PW. Population genetics of malaria resistance in humans. Heredity 2011; 107: 283-304. http://dx.doi.org/10.1038/hdy.2011.16

[13] Comings DE. Molecular heterosis as the explanation for the controversy about the effect of the drd2 gene on dopamine d2 receptor density. Mol Psychiatry 1999; 4: 213-15. http://dx.doi.org/10.1038/sj.mp.4000500

[14] Reuter M, Hannig J, Ambling M, *et al*. The role of the tph1 and tph2 genes for nicotine dependence: a genetic association study in two different age cohorts. Neuropsychobiology 2007; 56: 47-54. http://dx.doi.org/10.1159/000110728

[15] Gosso MF, de Geus EJ, Polderman TJ, Boomsma DI, Heutink P, Posthuma D. Catechol o-methyl transferase and dopamine d2 receptor gene polymorphisms: evidence of positive heterosis and gene-gene interaction on working memory functioning. Eur J Hum Genet 2008; 16: 1075-82. http://dx.doi.org/10.1038/ejhg.2008.57

[16] Eny KM, Corey PN, El-Sohemy A. Dopamine d2 receptor genotype (c957t) and habitual consumption of sugars in a free-living population of men and women. J Nutrigenet Nutrigenomics 2: 235-42.

[17] Morahan G, Huang D, Wu M, *et al*. Association of IL 12b promoter polymorphism with severity of atopic and non-atopic asthma in children. Lancet 2002; 360: 455-59. http://dx.doi.org/10.1016/S0140-6736(02)09676-9

[18] Amos CI. Successful design and conduct of genome-wide association studies. Hum Mol Genet 2007; 16(2): R220-225. http://dx.doi.org/10.1093/hmg/ddm161

[19] Gail MH, Pee D, Benichou J, Carroll R. Designing studies to estimate the penetrance of an identified autosomal dominant mutation: cohort, case-control, and genotyped-proband designs. Genet Epidemiol 1999; 16: 15-39. http://dx.doi.org/10.1002/(SICI)1098-2272(1999)16:1<15::AID-GEPI3>3.0.CO;2-8

[20] Mitra SK. On the limiting power function of the frequency chi-square test. Ann Math Statist 19: 1221-33.

[21] Eyduran E, Ozdemir T, Kazim MK, Keskin S, Cak B. A study of power of chi-square and g statistics in biological sciences. Pak J Biol Sci 2006; 9: 1324-27. http://dx.doi.org/10.3923/pjbs.2006.1324.1327

[22] Minnelli C, Thompson JR, Abrams KR, Thakkinstian A, Attia J. The choice of a genetic model in the meta-analysis of molecular association studies. Int J Epidemiol 2005; 34: 1319-28. http://dx.doi.org/10.1093/ije/dyi169

[23] Arshad SH, Hide DW. Effect of environmental factors on the development of allergic disorders in infancy. J Allergy Clin Immunol 1992; 90: 235-41. http://dx.doi.org/10.1016/0091-6749(92)90077-F

[24] Howard TD, Whittacker PA, Zaiman AL, Koppelman GH, Xu J, Hanley MT, *et al*. Identification of association of polymorphisms in the interleukin-13 gene with asthma and atopy in a Dutch population. Am J Respirat Cell Mol Biol 2001; 25: 377-84.

[25] Kruglyak L, Nickerson DA. Variation is the spice of life. Nat Genet 2001; 27: 234-36. http://dx.doi.org/10.1038/85776

[26] Clark AG. Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol 1990; 7: 111-22.

[27] Xu J. Extracting haplotypes from diploid organisms. Curr Issues Mol Biol 2006; 8: 113-22.

[28] Garner C, Slatkin M. On selecting markers for association studies: patterns of linkage disequilibrium between two and three diallelic loci. Genet Epidemiol 2003; 24: 57-67. http://dx.doi.org/10.1002/gepi.10217

[29] Thorleifsson G, Walters GB, Gudbjartsson DF, *et al*. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet 2009; 41: 18-24. http://dx.doi.org/10.1038/ng.274

[30] Birchler JA, Auger DL, Riddle NC. In search of the molecular basis of heterosis. Plant Cell 2003; 15: 2236-39. http://dx.doi.org/10.1105/tpc.151030

[31] Hart SP. Il12b promoter polymorphism and asthma. Lancet 2002; 360: 2085. http://dx.doi.org/10.1016/S0140-6736(02)11980-5

[32] Birchler JA, Auger DL, Riddle NC. Homing in on the asthma gene. Lancet 2002; 360: 422-23. http://dx.doi.org/10.1016/S0140-6736(02)09693-9

[33] Sobacchi C, Vezzoni P, Reid DM, *et al*. Association between a polymorphism affecting an ap1 binding site in the promoter of the tcirg1 gene and bone mass in women. Calcif Tissue Int 2004; 74: 35-41. http://dx.doi.org/10.1007/s00223-002-0004-2

[34] Lehmann DJ, Cortina-Borja M, Warden DR, *et al*. Large meta-analysis establishes the ace insertion-deletion polymorphism as a marker of Alzheimer's disease. Am J Epidemiol 2005; 162: 305-17. http://dx.doi.org/10.1093/aje/kwi202

[35] Lee HS. Gender-specific molecular heterosis and association studies: dopamine d2 receptor gene and smoking. Am J Med Genet B Neuropsychiatr Genet 2003; 118B: 55-59. http://dx.doi.org/10.1002/ajmg.b.10036

[36] Zheng G, Ng HK. Genetic model selection in two-phase analysis for case control association studies. Biostatistics 2008; 9: 391-99. http://dx.doi.org/10.1093/biostatistics/kxm039

Received on 16-12-2012 Accepted on 05-01-2013 Published on 31-01-2013

http://dx.doi.org/10.6000/1929-6029.2013.02.01.06