HHS Public Access

Author manuscript

*Genet Epidemiol*. Author manuscript; available in PMC 2021 October 07.

Published in final edited form as:

*Genet Epidemiol*. 2021 March ; 45(2): 171–189. doi:10.1002/gepi.22363.

# Innovative approach to identify multigenomic and environmental interactions associated with birth defects in family-based hybrid designs

**Xiang-Yang Lou**[1], **Ting-Ting Hou**[1,2], **Shou-Ye Liu**[2], **Hai-Ming Xu**[2], **Feng Lin**[2], **Xinyu Tang**[3], **Stewart L. MacLeod**[4], **Mario A. Cleves**[5], **Charlotte A. Hobbs**[6]

[1]Department of Biostatistics, College of Public Health and Health Professions and College of Medicine, University of Florida, Gainesville, Florida, USA

[2]Institute of Bioinformatics and Institute of Crop Science, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China

[3]The US Food and Drug Administration, Silver Spring, Maryland, USA

[4]Arkansas Children's Research Institute, Little Rock, Arkansas, USA

[5]Department of Pediatrics, Morsani College of Medicine, Health Informatics Institute, University of South Florida, Tampa, Florida, USA

[6]Rady Children's Institute for Genomic Medicine, San Diego, California, USA

## Abstract

Genes, including those with transgenerational effects, work in concert with behavioral, environmental, and social factors via complex biological networks to determine human health. Understanding complex relationships between causal factors underlying human health is an essential step towards deciphering biological mechanisms. We propose a new analytical framework to investigate the interactions between maternal and offspring genetic variants or their surrogate single nucleotide polymorphisms (SNPs) and environmental factors using family-based hybrid study design. The proposed approach can analyze diverse genetic and environmental factors and accommodate samples from a variety of family units, including case/control–parental triads, and case/control–parental dyads, while minimizing potential bias introduced by population admixture. Comprehensive simulations demonstrated that our innovative approach outperformed the log-linear approach, the best available method for case–control family data. The proposed approach had greater statistical power and was capable to unbiasedly estimate the maternal and child genetic effects and the effects of environmental factors, while controlling the Type I error rate against population stratification. Using our newly developed approach, we analyzed

**Correspondence**: Xiang-Yang Lou, 2004 Mowry Rd, CTRB #5221, PO Box 117450, Gainesville, FL 32611-7450, USA. xlou@ufl.edu.

AUTHOR CONTRIBUTIONS

Xiang-Yang Lou conceived of the analytical idea and developed the theoretical formalism. Ting-Ting Hou, Shou-Ye Liu, Hai-Ming Xu, and Feng Lin performed the numerical simulations and the analytic calculations. Charlotte A. Hobbs supervised the congenital heart defects study and the interpretation of the results. Xiang-Yang Lou and Ting-Ting Hou took the lead in writing the manuscript. All authors discussed the results, provided critical feedback, and contributed to the final version of the manuscript.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

the associations between maternal and fetal SNPs and obstructive and conotruncal heart defects, with adjustment for demographic and lifestyle factors and dietary supplements. Fourteen and 11 fetal SNPs were associated with obstructive and conotruncal heart defects, respectively. Twenty-seven and 17 maternal SNPs were associated with obstructive and conotruncal heart defects, respectively. In addition, maternal body mass index was a significant risk factor for obstructive defects. The proposed approach is a powerful tool for interrogating the etiological mechanism underlying complex traits.

**Keywords**

case-control family design; congenital heart defects; environmental factors; fetal genes; genetic association and linkage; interactions; maternal genes

## 1 | INTRODUCTION

Human health and other complex traits are determined by a complex interaction between genetic, environmental, lifestyle, and other factors (Evans & Stoddart, 1990; Hunter, 2005). On one hand, DNA is the blueprint for life; genes provide a set of instructions for growth and development, governing the particular physiology and anatomy of an individual (Committee on Evaluation of Children's Health, 2004). It is clear that inherited genetic variation contributes to the pathogenesis of disease both directly and indirectly (Hernandez & Blazer, 2006). On the other hand, the influence of genes always exists within an environmental context; genetic makeup alone can seldom determine a phenotype, only the potential for expression of the phenotype in response to a particular lifestyle and environment (Meaney, 2010). Genes, behavior, physical environment, and social environment form a complex system that regulates the occurrence of disorders by acting on a biological network with many interwoven pathways such that each factor does not necessarily influence health in a deterministic way (Newacheck et al., 2006). Furthermore, parental genotypes may act in concert with individual's inherited genotype to confer phenotypic variation via various mediating mechanisms including epigenetic modulation; supplying messenger RNA, hormones, and proteins essential for fetal development; and the maternal environment (Fowden & Moore, 2012; Keverne, 2015; Sferruzzi-Perri et al., 2016), leading to parent-of-origin effects and/or transgenerational genetic effects (Lawson et al., 2013; Nadeau, 2009; Peters, 2014; Reik & Walter, 2001; Youngson & Whitelaw, 2008). Maternal effects, the causal influence of the maternal genotype or phenotype on the offspring phenotype, are one of the most important subcategories of parent-of-origin effects (Wolf & Wade, 2009). The significant role of maternal effects on the phenotype of the offspring has been well documented (Casellas et al., 2009; Jarvis et al., 2005; Mousseau & Fox, 1998; Thomas et al., 1994); this role is particularly true for structural birth defects and other congenital conditions (Hackshaw et al., 2011; Lee et al., 2005; Lupo et al., 2014). In addition, increasing evidence also supports the "developmental origins of health and disease" concept that processes acting in early life may have long-lasting effects on health in adulthood (Gluckman & Hanson, 2004). Thus, mechanisms underlying human health and disease are quite complex. Identifying the interactive causal factors involved in human health and disease is the first step towards deciphering the biological mechanism and will

lay the scientific foundation for precision medicine, helping promote preventive care and decrease morbidity and mortality.

Genetic association studies that aim to detect direct or proxy nonrandom coexistence between one or more genetic polymorphisms and a trait of interest are powerful tools for dissecting the genetic architecture and elucidating the roles of genetic determinants (Cordell & Clayton, 2005). Genetic associations may be tested using either unrelated individuals, family-based samples, or a mixture of both. Family-based designs are a common choice because of unique advantages and the capability for controlling population stratification, investigating parent-of-origin effects, quantifying shared environmental effects, and testing both linkage and association (Laird & Lange, 2006; Ott et al., 2011). Popularly used analytical methods for family-based association studies fall into four categories. The first category includes the transmission/disequilibrium test (TDT; Ewens & Spielman, 1995; Ott, 1989; Spielman & Ewens, 1996; Spielman et al., 1993), an application of McNemar's $\chi^2$ test (McNemar, 1947), and its extensions for missing parental information including sib TDT (Boehnke & Langefeld, 1998; Horvath & Laird, 1998; Spielman & Ewens, 1998), reconstruction-combined TDT (Knapp, 1999), and 1-TDT (Sun et al., 1999); for extended pedigrees such as pedigree disequilibrium test (Martin et al., 2000); for multiallelic markers including generalized TDT (Rice et al., 1995), extended TDT (Bickeboller & Clerget-Darpoux, 1995; Sham & Curtis, 1995), Monte Carlo Tm test (Kaplan et al., 1997), and exact TDT (Cleves et al., 1997); for multiple markers (Lazzeroni & Lange, 1998; Wilson, 1997; Zhao et al., 2000); for differentiating between fetal and maternal genetic effects (Mitchell, 1997); for genomic imprinting and hybrid family designs (Hu & Zhou, 2010; Hu et al., 2007; Hu, Zhou, Sun et al., 2007); and for quantitative traits (Abecasis et al., 2000; Allison, 1997; Monks & Kaplan, 2000; Rabinowitz, 1997; Xiong et al., 1998). The second category is the conditional association tests including family-based association test and pedigree-based association test (Horvath et al., 2001, 2004; Lange & Laird, 2002; Lange et al., 2002, 2004; Rabinowitz & Laird, 2000). The third category is the likelihood-based association tests such as the conditional likelihood method (Schaid & Sommer,1993, 1994), the partial likelihood method for detecting imprinting and maternal effects (Han et al., 2013; Yang & Lin, 2013), the log-linear models (Vermeulen et al., 2009; Weinberg & Umbach, 2005; Weinberg,1999a, 1999b; Wilcox et al., 1998), and the association test in the presence of informative missingness (Allen et al., 2003). The final category is the mixed effect models (Clark et al., 2016) or regression-based unified methods combining triads and unrelated subjects (Cordell & Clayton, 2002; Cordell et al., 2004; Epstein et al., 2005; Nagelkerke et al., 2004). Extensions of family-based methods have also been made for investigating the complex interplay between child and maternal genes, environmental exposures, and lifestyle factors by the inclusion of maternal and fetal genes, environmental factors, and/or their interactions (Kistner et al., 2009; Sinsheimer et al., 2003). However, none of the current methods fully address analytical problems. Specifically, most of these methods such as the TDT cannot model the maternal effects and/or environmental variates. The currently best available method for analysis of case–control family data, the log-linear models (Vermeulen et al., 2009; Weinberg & Umbach, 2005; Weinberg, 1999a; Weinberg et al., 1998; Wilcox et al., 1998), although considering parental effects, has several limitations: (1) only providing a valid test for association, not for association with linkage, potentially leading to false

positives and bias in presence of population structure, (2) not allowing for continuous environmental factors, and (3) not making full use of control phenotypic information. A more powerful analytical strategy is demanded to unveil the multigenomic architecture of complex diseases in which the effects of various genetic and environmental contributors may be highly intertwined.

Congenital heart defects (CHDs) typically arise from a sophisticated interplay among parental lifestyle, environmental, genetic, and epigenetic factors (Vecoli et al., 2014). CHDs are among the most common and serious birth defects, with a worldwide incidence ranging from 8 to 12 per 1000 live births (Hoffman, 2013; S.C. Mitchell et al., 1971). Cardiogenesis is an intricate biological process regulated by a complex genetic network (Kirby, 2002; Srivastava & Olson, 2000); defects in particular regions of the heart may arise from genetic and environmental effects during specific developmental windows of time (Fishman & Olson, 1997; Srivastava, 2001). Maternal genes, as well as maternal metabolic and lifestyle factors, are involved by acting on the intrauterine environment directly or through fetomaternal interactions indirectly (Hobbs et al., 2005; Hobbs et al., 2010, 2011; Lee et al., 2005). The majority of CHDs (>80%) are nonsyndromic (Gelb & Chung, 2014) and thought to have a multifactorial etiology in which both inherited and noninherited causes play key roles (Lage et al., 2012; Nora, 1968; Vecoli et al., 2014). CHDs are a spectrum of disorders that are etiologically heterogeneous and may be classified into various subtypes for which there is evidence of a shared etiology (Botto et al., 2007). The heritability of various forms of CHDs was reported to be 50%–90% (Cripe et al., 2004; Hinton et al., 2007; McBride et al., 2005). Conotruncal heart defects (CTDs) and obstructive heart defects (OHDs) are two common subtypes, accounting for approximately 20%–30% and 25% of all CHDs, respectively (Cleves et al., 2003; Ferencz et al., 1985; van der Linde et al., 2011). CTDs are a category of malformations of cardiac outflow tracts and great arteries that share a common structural origin, derived from cardiac neural crest cells and secondary heart field (Hutson & Kirby, 2007). OHDs are a subgroup of CHDs in which either the left-sided or the right-sided heart valves, arteries, or veins are abnormally narrow or blocked (Botto et al., 2007). Deciphering the risk factors for CTDs and OHDs and better understanding the complex etiological mechanisms are fundamental to improve diagnosis and genetic counseling and to translate our knowledge into tangible benefits for families with CHDs.

In this study, we present a new analytical approach for detecting the associations of fetal and maternal genetic and environmental factors with congenital heart defects. Comprehensive simulations were performed to assess the validity and power of the proposed method. The new method was then applied to investigate the complex interplays between genetic variants and maternal lifestyle factors underlying the subtypes of CHDs, CTDs, and OHDs, using data from two of our previous studies (Hobbs et al., 2014; Tang et al., 2015).

## 2 | STUDY DESIGN AND STATISTICAL METHOD

### 2.1 | Family-based hybrid design

Among several family-based designs advocated to differentiate offspring and maternal genetic contributions to early-onset disorders, the hybrid design based on augmenting a set of case–parent triads with a set of unrelated control–parent triads/dyads or just parents

of controls is well accepted by the research community (Healy et al., 2010; Vermeulen et al., 2009; Weinberg & Umbach, 2005). A hybrid design can bring the strengths of both population-based case–control and family-based case–parent designs together into a single study framework (Weinberg & Umbach, 2005). While the extension of the proposed method is straightforward to other family-based designs (e.g., discordant siblings, concordant siblings, and multiplex families) and/or phenotypes (e.g., quantitative and ordinal), the hybrid design is used here to illustrate the statistical method.

## 2.2 | Statistical model

A logistic regression model for the probability that individual $i$ with child genotype $c$ in a family of maternal genotype $m$ and paternal genotype $f$ is a case can be expressed as follows,

$$\log it(p_{mfci}) = \log\left(\frac{p_{mfci}}{1-p_{mfci}}\right) = \alpha_{mf} + \mathbf{w}_m^T\boldsymbol{\beta}_M + \mathbf{x}_{mfc}^T\boldsymbol{\beta}_C + \mathbf{z}_{mfci}^T\boldsymbol{\gamma} + \left(\mathbf{w}_m^T \otimes \mathbf{z}_{mfci}^T\right)\boldsymbol{\psi}_M + \left(\mathbf{x}_{mfc}^T \otimes \mathbf{z}_{mfci}^T\right)\boldsymbol{\psi}_C, \quad (1)$$

Where $p_{mfci}$ is the probability (in a biallelic case, $m$, $f$, $c = 0, 1, 2, 3$, denote 0, 1, and 2 copies of a given allele and the missing genotype, respectively); $\alpha_{mf}$ is the parameter related to the genotype frequencies and the parental mating type and $\alpha_{mf} = \alpha_{fm}$ under the assumption of mating symmetry; $\boldsymbol{\beta}_M$ is the vector of maternal effects, consisting of additive, and dominance effects; $\mathbf{w}_m$ is the observation vector coding the elements of the incidence (design) matrix related to $\boldsymbol{\beta}_M$ (superscript "T" of a vector denotes its transpose); $\boldsymbol{\beta}_C$ is the vector of child effects, consisting of additive and dominance effects under various parental configurations; $\mathbf{x}_{mfc}$ is the vector coding the elements of the incidence matrix related to $\boldsymbol{\beta}_C$; $\boldsymbol{\gamma}$ is the vector of covariates such as sex, age, and folic acid intake; $\mathbf{z}_{ijkl}$ is the vector of covariate observations; $\boldsymbol{\psi}_M$ is the vector of interactions between a maternal gene and environmental factors; $\mathbf{w}_m^T \otimes \mathbf{z}_{mfci}^T$ is the vector of coding the maternal interactions; $\boldsymbol{\psi}_C$ is the vector of interactions between a child gene and environmental factors; $\mathbf{x}_{mfc}^T \otimes \mathbf{z}_{mfci}^T$ is the vector coding the child interactions; and $\otimes$ represents a Kronecker product. Tables 1 and S1 summarize how to code the design matrix for child effects in terms of the parent–child genotype configurations. The statistical additive and dominance effects in $\boldsymbol{\beta}_C$ (and their interactions $\boldsymbol{\psi}_C$) are the functions of inheritance (and gene–environment interaction) parameters including genotypic frequencies and genotypic disequilibria (i.e., allelic frequencies and disequilibrium coefficient at Hardy–Weinburg equilibrium), associated with factors of $(1 - 2\theta)$ and $(1 - 2\theta)^2$, respectively, when parents are heterozygous, of 1 and 1, respectively, when parents are homozygous (Table 1), and of $(1-\theta)$ and $(1-\theta)^2$, respectively, when the parental genotypic information is missing (Table S1), where $\theta$ is the recombination rate. For an illustration, a child of $MM$ in mating type $MM \times Mm$ receives an $M$ from homozygous parent $MM$ (associated with a factor of 1) and the other $M$ from heterozygous parent $Mm$ [associated with a factor of $(1-2\theta)$], then the coefficient of additive effect is $[1 + (1 - 2\theta)]$ and the coefficient of dominance is $-1 \times (1 - 2\theta) = -(1 - 2\theta)$, while a child of $Mm$ in mating type $MM \times Mm$ receives the $M$ from homozygous parent $MM$ (associated with a factor of 1) and the $m$ from heterozygous parent $Mm$ [associated with a factor of $(1 - 2\theta)$], then the coefficient of additive effect is $[1 - (1 - 2\theta)]$ and the coefficient of dominance is $-1 \times [-(1 - 2\theta)] = (1 - 2\theta)$.

Two choices can be adopted to characterize the model. First, similar to the traditional TDT, we can incorporate the recombination rate $\theta$ into the effect parameters; for example, $a$, $(1-\theta)a$, and $(1-2\theta)a$ are treated as three distinct parameters (e.g., $a'$, $a''$, and $a'''$) that can be estimated from the corresponding components of families. Second, we can also explicitly model the recombination rate $\theta$, offering a parsimonious and potentially more powerful model. With this model, there may be a nonregular problem due to the presence of nonidentifiable parameter(s) when there is no association and/or no genetic effect, requiring a more complicated significance test such as permutation testing (Drton, 2009; Feng & McCulloch, 1996).

Since the grandparent information is unavailable, no maternal population structure is assumed in the above model. If the assumption of no population structure does not hold true, the inclusion of maternal grandparents is needed to avoid potential bias in estimating maternally mediated genetic effects (Mitchell & Weinberg, 2005; Weinberg, 2003). As an alternative, structure assessment and/or principal components analysis (PCA) can also be used to correct for maternal population structure (Price et al., 2006; Pritchard et al., 2000).

The statistical model may also be adapted to analyze the genomic imprinting in which the expression of a gene is dependent on its parent-of-origin. In such a case, an additional parameter, the difference in genotypic value between the reciprocal heterozygotes, may be included in the model to measure the imprinting effect (Strauch et al., 2000).

### 2.3 | Parameter estimation and hypothesis test

Define an indicator variable as follows,

$$d_{mfci} = \begin{cases} 1 & \text{if individual } mfci \text{ is affected} \\ 0 & \text{if individual } mfci \text{ is unaffected} \end{cases}.$$

We have the following likelihood,

$$L(D \mid \boldsymbol{\alpha}, \boldsymbol{\beta}_M, \boldsymbol{\beta}_C, \boldsymbol{\gamma}, \boldsymbol{\psi}_M, \boldsymbol{\psi}_C) \propto \prod_{m,f,c,i} (p_{mfci})^{d_{mfci}} (1-p_{mfci})^{1-d_{mfci}}. \quad (2)$$

The maximum likelihood (ML) method can be used to estimate the statistical parameters including maternal additive and dominance effects, child additive and dominance effects, covariate effects and interaction effects (if applicable), and the recombination rate (if applicable) in Model (1). As in the routine logistic regression, the numerical iterative algorithms such as Fisher's scoring and Newton–Raphson procedure can be used to find the ML estimates. The likelihood ratio (LR) statistic may be computed to test the overall significance between a null model and the corresponding alternative model while the Wald statistics are used to test the significance of individual parameters. Statistical significance can be determined by either the conventional $\chi^2$ or empirical permutation test where the case–control labels are randomly assigned across families within mating-type strata.

We can build diverse LR statistics to test the corresponding hypotheses, such as $H_0$: $(1-2\theta)\delta = 0$ for no linkage or association where $\delta$ is a disequilibrium coefficient $H_0$: $(1-2\theta)\delta =$

$(1-\theta)\delta = \delta$ for tight linkage, and $H_0$: $\theta = 0.5$ for no linkage, based on LR. We can also test $H_0$: $(1-\theta)\delta = \delta = 0$ for a filtering step and $H_0$: $(1-2\theta)\delta = 0$ for a verification step by using two independent and complementary components in the model, of which the former is ignored in the traditional TDT. Benefitting from the filtering step (the markers without association will be filtered out), this approach can reduce multiple testing penalty.

### 2.4 | Simulation scenarios

Simulations were performed based on a set of 2000 families consisting of 1000 nuclear families with an affected child and 1000 nuclear families with an unaffected child under various scenarios to evaluate the validity and statistical power of the new approach. Biallelic markers with a minor allele frequency (MAF) of 0.40 and disease gene with a MAF of 0.30, between which there is a disequilibrium coefficient $\delta$ of either 0.06 or 0.00, were used to generate a base population from which the parents were drawn. Since the disequilibrium coefficient is the characteristic parameter of population stratification (i.e., various admixture scenarios with the same disequilibrium coefficient will give the same result; Slatkin, 2008; VanLiere & Rosenberg, 2008), no specific stratification scenario is considered here. The genotypes of children were simulated according to the parental genotypes assuming a recombination fraction of either 0.50 or 0.00. The phenotype of a child, case or control, was determined based on the logistic regression model,

$$P(y = 1) = \frac{1}{1 + e^{-\eta}},$$

Where $\eta$ is the linear predictor consisting of maternal additive and dominance, child additive and dominance, and covariate effects as defined in Model (1). As the statistical effects in Model (1) are the functions of genetic parameters including gene effects, allele frequencies, and disequilibrium coefficient $\delta$, we used the genetic parameters to specify the data-generating model.

Several scenarios were considered in simulations for various illustrative purposes. Without loss of generality, we assumed a disease gene with mid-parent value $m = 0.0$ and ratio of dominance to additive $d_C/a_C = d_M/a_M = 0.5$ (i.e., a case of semi-dominance where the size of dominance is half that of additive effect), and a continuous environmental factor the observation of which follows the standard normal distribution in all scenarios. To verify the validity of the proposed approach for the null models, in addition to recombination rate $\theta = 0.5$ and child additive effect $a_C = 1.0$, we considered disequilibrium coefficient $\delta = 0.00$ and neither maternal nor covariate effects in Scenario I (i.e., no linkage and no association); $\delta = 0.06$ and neither maternal nor covariate effects in Scenario II (i.e., not linked but associated, the case of nonsyntenic disequilibrium which is equivalent to the spurious association due to population structure); $\delta = 0.06$, a continuous environmental factor with its effect $\gamma = 0.5$, and no maternal effects in Scenario III; and $\delta = 0.06$, maternal additive effect $a_M = 0.5$, and $\gamma = 0.5$, with the inclusion of grandparent genotypes in Scenario IV. To assess statistical power, in addition to $\theta = 0$, $\delta = 0.06$, and $a_C = 1$, we included only child effects in Scenario V (i.e., $a_M = \gamma = 0$), child effects plus a covariate with effect $\gamma = 0.5$ in Scenario VI, and

child effects plus maternal effect with $a_M = 0.5a_C$, $\gamma = 0.5$, and the inclusion of grandparent genotypes in Scenario VII. One thousand simulations were performed for each scenario.

The parsimonious model was used to explicitly estimate the recombination rate. The ML method was used to fit the null and alternative models for the simulated data, and the LR and Wald statistics were computed for hypothesis tests. To circumvent the nonregular problem when there is no association or no genetic effect, permutation testing of 1000 replicates was also performed to construct an empirical distribution of LR statistic under the null hypothesis to avoid an inflated Type I error rate. Type I error rate and power were calculated at a significance level of 0.05. The Type I error rates and power were also assessed at varying significance levels to construct the probability–probability plots of significance level against Type I error rate or power. For comparison, the log-linear method (Vermeulen et al., 2009; Weinberg & Umbach, 2005) was used to analyze the same simulated data implemented with the LEM software (van Den Oord & Vermunt, 2000).

### 2.5 | Candidate gene-based association studies on conotruncal heart defects and obstructive heart defects

To interrogate the risk factors for two subgroups of CHDs (CTDs and OHDs), family-based case–control samples were collected from the participants enrolled in the National Birth Defects Prevention Study (NBDPS) between 1997 and 2008. The NBDPS is a multisite population-based case–control study including 10 states (Arkansas, California, Iowa, Massachusetts, New Jersey, New York, Texas, Georgia, North Carolina, and Utah; Yoon et al., 2001). A total of 616 case families with singleton live-born infants with CTDs and 1645 control families with singleton live-born infants without any major structural birth defects were included in the CTDs study (Hobbs et al., 2014). A total of 586 case families with singleton live-born infants with OHDs and 1702 control families were used for the OHDs study (Tang et al., 2015). All study participants for the analyses gave informed written consent and submitted buccal cells collected using cytobrushes from which DNA was isolated. All the study protocol and forms/procedures were approved by the Institutional Review Boards of the University of Arkansas for Medical Sciences and the NBDPS with protocol oversight by the Centers for Disease Control and Prevention (CDC) Center for Birth Defects and Developmental Disabilities.

A customized panel of 1536 single nucleotide polymorphisms (SNPs) in 62 candidate genes from the folate, homocysteine, and glutathione/transsulfuration pathways was selected for genotyping. The detailed genotyping procedure has been previously described (Hobbs et al., 2014; Tang et al., 2015). Postgenotyping quality control analysis was performed. Individuals with either high no-call rate and/or high rates of Mendelian inconsistency, and the SNPs with no-call rates more than 10%, Mendelian error rates more than 5%, minor allele frequency more than 5%, and/or significant deviation from Hardy–Weinberg Equilibrium in at least one racial group ($p < 10^{-4}$) were excluded from the subsequent analysis. After quality control, 4648 individuals from 616 case and 1645 control families each with 921 SNPs and 4551 individuals from 569 case, and 1644 control families each with 877 SNPs were included in the final analysis for the CTDs study and the OHDs study, respectively. The minor allele in each SNP was used as the reference allele in the genetic analysis.

Child and maternal genetic effects were estimated with the inclusion of maternal characteristics such as maternal age at delivery, folic acid supplementation, body mass index (BMI) of mother, maternal race, alcohol use, and smoking status, as the covariates. Considering the limited sample size and the potential number of parameters, we did not include the interactions between genes and maternal characteristics in the analysis although it is feasible in principle. Because only parent–child samples were available, the population structure in mothers was examined by PCA and, when necessary, the first five principal components were used to correct for population stratification in maternal effects. Bayesian false-discovery probability (BFDP) that can balance the costs of false discovery and false non-discovery in multiple testing was computed to evaluate the noteworthiness of a tested association, and as suggested, a BFDP of 0.80 was used as the threshold (Wakefield, 2007). All covariates were tested using the Wald statistics.

## 3 | RESULTS

### 3.1 | Simulation study

The simulation results suggested that the proposed approach could control Type I error rate well for diverse scenarios of null models: (1) either the absence of both linkage and association (Figure 1a for Scenario I) or the presence of association but no linkage (Figure 1b–d for Scenarios II–IV) and (2) either with no covariate and no maternal effects (Scenario II), with a covariate but no maternal effects (Scenario III), or with both covariate and maternal effects (Scenario IV). All the Type I error rates were close to the nominal levels. For example, the Type I error rates at the significance level of 0.05 were 0.050 when there were no association and no linkage, 0.052 when only the child gene effects were involved, 0.046 when both the child gene effects and the effect of an environmental factor were engaged, and 0.055 when all child genes and maternal genes and an environmental factor controlled the phenotype. On the other hand, the log-linear method yielded inflated Type I error rates in the presence of nonsyntenic disequilibrium regardless of whether the maternal effects or environment effects were present or absent (Figure 2a–c). Thus, the log-linear method was subject to potential bias stemming from population structure. Furthermore, although a test was proposed for the log-linear method to detect population stratification in a hybrid design via either an LR statistic with 5 *df* or an LR statistic with 1 *df* (Vermeulen et al., 2009; Weinberg & Umbach, 2005), such a test might be invalid and misleading as demonstrated by the inflated Type I error rate in Scenario V, a case where the marker and the causal gene coincide at a certain chromosome positon (a recombination rate of 0.0) but there is no population stratification (Figure 2d). It seems that the real null hypothesis to be tested is whether the mating-type parameters of cases are the same as those of controls or not, and thus the test cannot address the issue of population structure; a rejection or acceptance of the null hypothesis does not necessarily correspond to rejection or acceptance of the absence of population structure.

The comparison of power showed that the proposed approach could substantially improve the statistical power over the log-linear method in Scenarios V and VI (Figure 3a,b). In Scenario VI, the power of the proposed approach and the log-linear method at the significance level of 0.05 were 88.0% and 81.4%, respectively; whereas the counterparts

were 92.8% and 86.9%, respectively, in Scenario V. The magnitude of improvement in Scenario VI seemed to be higher than that in Scenario V, indicating the benefit of accounting for an informative covariate. Although the log-linear method had higher power than the proposed approach in Scenario VII (Figure 3c), the increase might be caused by an increased false-positive rate of 24.6% (Figure 2c). The simulations also suggested that all the estimates of parameters were considered to be unbiased with a small mean squared error, suggesting that the proposed approach also provide valid estimation of parameters (Table 2).

In summary, the simulation studies clearly demonstrated that the proposed approach had higher power than the log-linear model and could unbiasedly estimate both genetic and environmental effects while maintaining the correct Type I error rates in the presence of population structure.

### 3.2 | Obstructive heart defects study

PCA of the OHD data showed that the maternal samples seemed to fall into several groups in the space spanned by the first three principal components, roughly corresponding to the ethnic/racial groups (Figure S1). The top five principal components of mother were therefore included as covariates to account for the potential structure in the maternal population.

Manhattan plots show the distribution of $p$ values for the fetal SNPs (Figure 4) and the maternal SNPs (Figure 5). There were a total of 14 fetal SNPs linked to and associated with the OHDs risk at the noteworthiness level (Table S2). They are located in the MTHFS gene of the folate pathway; the BHMT2, BHMT, and DNMT3L genes of the homocysteine pathway; and the SOD2, GSR, and MGMT genes of the transsulfuration pathway, respectively. SNPs rs526264, rs625879, and rs557302 in the BHMT2 gene were also detected at the noteworthiness level by the previous log-linear analysis (Tang et al., 2015). SNP rs659044, which was not previously reported, was highly associated with rs526264 ($r^2$ = .97) and rs625879 ($r^2$ = .96), respectively (Figure S2). The associations of rs732498 and rs5746105 in the SOD2 gene with OHDs also appeared in the previous report. The newly identified SNP rs645112 in the BHMT gene was highly associated with the previously detected SNP rs490268. Twenty-seven SNPs in 10 genes had a noteworthy maternal influence on the risk for OHDs via either maternal additive, maternal dominance, or both (Table S3). Significant maternal SNPs rs6482747 and rs4751110 in the MGMT gene overlapped with the log-linear analysis. Sixteen significant SNPs in the MGMT gene were located in three LD blocks (Figure S3), implying that the multiplicity of detected associations might be degraded into fewer functional variants. The identified SNPs, rs1556893 and rs4933327, in the MAT1A gene were previously found to be associated with OHDs in obese mothers, but not normal-weight mothers (Tang et al., 2015). Furthermore, the Wald tests revealed that BMI of mother was a significant risk factor for OHDs; the estimated effect ranged from 0.398 to 0.505 in the different SNP models while the $p$-value ranged from .002 to .015. The effects of the other covariates did not reach the significance level.

### 3.3 | Conotruncal heart defects study

According to the PCA results, the maternal ethnic groups appeared to be well separated by the first three principal components (Figure S4). The top five principal components of mothers were included as covariates to account for the potential maternal population structure.

Eleven SNPs were linked to and associated with the risk of developing a CTD at the noteworthiness level (Figure S5). Of these SNPs, five SNPs are located in the MTHFD2 and TYMS genes in the folate pathway and six SNPs are located in the GPX5, GPX6, and GCLC genes in the transsulfuration pathway (Table S4). SNPs rs2612101, rs2847607, rs2847326, and rs2847324 in the TYMS gene were also reported to be associated with CTDs in the previous log-linear analysis (Hobbs et al., 2014). The three SNPs in the GCLC gene that were in high linkage disequilibrium with each other had nearly the same estimated additive and dominance effects (Figure S6), indicating that they potentially share a common causal variant. There were 17 maternal SNPs in 12 genes associated with the risk for CTDs (Figure S7 and Table S5). The associations of maternal SNPs rs572494, rs13212365, rs546726, rs634657, and rs648595 in the GCLC gene and rs6912979 in the SOD2 gene were previously reported (Hobbs et al., 2014), among which rs546726, rs634657, and rs648595 were in the same LD block (Figure S8). Except for age at delivery, which had a marginal effect on the risk for CTDs, the effects of the other covariates did not reach the significance level.

## 4 | DISCUSSION

Population stratification is a well-known confounding factor that can lead to spurious findings in genetic association studies consisting of unrelated individuals (Price et al., 2010). Although the PCA and mixed-effects model are popularly used to correct for population structure (Price et al., 2006; Yu et al., 2006; Zhang et al., 2010), the validity seems to be not guaranteed (Wang et al., 2013; Wu et al., 2011). Family-based studies offer an effective way to bypass the problem of population stratification in association studies via the use of robust tests such as TDT in which neither linkage alone nor disequilibrium alone will generate a positive association (Mackay & Powell, 2007). Furthermore, family data are also potentially informative for interrogating some etiological factors, such as transgenerational genetic effects, underlying complex diseases (Benyamin et al., 2009). Family-based studies offer a unique solution to deciphering the genetic basis of complex traits.

Although log-linear model-based methods (Weinberg & Umbach, 2005; Weinberg, 1999b; Wilcox et al., 1998) may account for parental effects, we experienced several limitations of the models when attempting to unravel complex relationships. First, these models fail to distinguish the effects of population subdivision from linkage disequilibrium (i.e., syntenic association caused by linkage). They are only a valid test for association, not for association with linkage, potentially leading to false positives and estimation bias in presence of population structure. Second, although categorical environmental factors may be included in log-linear models, continuous environmental factors cannot. Lastly, they do not make full use of phenotypic information in controls that may inform the estimation of penetrance parameters. On the other hand, the polytomous logistic method which allows for

continuous environmental cofactor(s) cannot accommodate the parental effects (Kistner et al., 2009). To address these analytical limitations, for a binary phenotype in a family-based hybrid design, we formulated a new approach that offers a unified and coherent statistical framework for the analysis of genetic and nongenetic factors, and their interactions. Specifically, the proposed approach can incorporate information from continuous and/or discrete environmental factors, account for both maternal and child genetic effects, correct for potential population stratification for fetal genetic effects (and for maternal genetic effects when grandparental genotypes are available), and accommodate various family structures and missing data scenarios, such as case–parent triads and control–mother dyads, in a unified analysis with inclusion of the control phenotype. Thus, the proposed approach can make better use of the genetic information available and be potentially more powerful than the existing methods. The simulations corroborated these expectations: the proposed approach adequately controlled Type I error rate, enhanced statistical power, and computed unbiased estimates of the parameters under various simulation scenarios.

In the application of the proposed approach to two candidate gene-based association studies, the major findings revealed in the previous analyses using the log-linear method were replicated (Hobbs et al., 2014; Tang et al., 2015). For example, the fetal SNPs in the BHMT2 and SOD2 genes were associated with OHDs and those in the TYMS gene were associated with CTDs at the significance level of noteworthiness. The associations of maternal SNPs in the MGMT gene with OHDs and the associations of maternal SNPs in the GCLC gene and in the SOD2 gene with CTDs were detected in both the new and the previous analyses. In addition, several new associations might be considered as "being replicated" because they could be accounted for by the high linkage disequilibria between the relevant SNPs. Furthermore, the new analysis also revealed that a high BMI of mother might significantly increase the risk for OHDs. However, there are a few important differences between the two kinds of analyses: (1) we tested both linkage and association for the fetal genome; (2) we used PCA to correct for potential population structure in the maternal genome; (3) we considered both additive and dominance effects in the fetal and maternal genomes; and (4) several demographic and lifestyle factors and folate supplement were included as covariates in the new analysis. Consequently, there were some differences in analytical results.

The potential roles of the identified variants and/or genes in birth defects are extensively supported by the literature. In the OHDs study, the most significant fetal SNPs were rs526264 and rs625879 in the BHMT2 gene that were reported to be associated with nonsyndromic congenital anomaly (Hozyasz, 2010). The significant association of rs645112 in the BHMT gene was involved in nonsyndromic cleft lip and palate (Blanton et al., 2011; Chiquet et al., 2011). BHMT and BHMT2 convert homocysteine to methionine using betaine and S-methylmethionine, respectively, as methyl donor substrates. Anomalies in homocysteine metabolism have been implicated in disorders ranging from vascular disease to neural tube birth defects such as spina bifida (Blom & Smulders, 2011). SNP rs5746105 in the SOD2 gene was found to be associated with nonsydromic myelomeningocele (Kase et al., 2012). Variant of rs12329764 in the CBS gene that encodes cystathionine-beta-synthase to catalyze the conversion of homocysteine to cystathionine was associated with myelomeningocele (Tilley et al., 2012).

Among the fetal SNPs and/or genes associated with CTDs, the TYMS gene codes the enzyme catalyzing the methylation of deoxyuridylate to deoxythymidylate using 5,10-methylenetetrahydrofolate as a cofactor. Several studies pointed to the associations between variants in the TYMS gene and risks of neural tube defects and CTDs (Lupo et al., 2011; Shaw et al., 2009; Volcik et al., 2003; Zhu et al., 2012) and Hispanic infants homozygous for the minor alleles of rs2847326 showed a significant increase in CTD risk (Zhu et al., 2012). GPX6 and GPX5 belong to the glutathione peroxidase family whose major biological role is to protect the organism from oxidative damage. The glutathione peroxidases played a significant role during embryonic development, indicating their biological plausibility for CTDs (Ufer & Wang, 2011). The GCLC gene encodes the catalytic subunit of glutamate–cysteine ligase to catalyze the initial and rate-limiting step of biosynthesis of glutathione that plays a crucial role in the intracellular antioxidant defense systems (Lu, 2009). The associations between variants in the GCLC gene and several human diseases and the prenatal deaths of homozygous knockout mice suggested its importance in fetal development and pathogenesis (Dalton et al., 2000; Gysin et al., 2007; Koide et al., 2003; McKone et al., 2006; Ristoff et al., 2000). The MTHFD2 gene encodes a nuclear-encoded mitochondrial bifunctional enzyme with methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase activities and its expression was shown to be essential for embryonic development (Di Pietro et al., 2002). The polymorphisms within the MTHFD2 gene were associated with risk for numerous pathological states including congenital anomalies (Copp et al., 2013; Ducker & Rabinowitz, 2017; Shaw et al., 2009). Maternal SNP rs1801131 was reported to interact with dietary folate intake in modifying the risk for CTDs (Zhu et al., 2012). Polymorphisms in the MTHFR gene, which provides instructions for making an enzyme called methylenetetrahydrofolate reductase, were also associated with the risk of neural tube defects and spina bifida (de Franchis et al., 2002; Kirke et al., 2004; Tsang et al., 2015; Yan et al., 2012). These convergent pieces of evidence suggest the biological implication of the revealed variants and further follow-up is thus warranted.

The proposed analytical framework is flexible and can be generalized to other types of the phenotype by using appropriate statistical models in place of the logistic model, for example, generalized linear model (Nelder & Wedderburn, 1972) and quasi-likelihood model (McCullagh, 1983) for count and continuous phenotypes, generalized estimating equations model for correlated observations and multiple complex traits (Liang & Zeger, 1986), multinomial logistic model for polytomous data (Begg & Gray, 1984), proportional odds model for ordinal data (Agresti, 1999), and proportional hazards model for survival data (Cox, 1972). It can also be extended to general family-based designs other than child–parent trios.

It is a limitation that the gene–environment interaction effects have not been considered in the real data analyses for conotruncal heart defects and obstructive heart defects because of the problems of estimation stability and convergence in finding the numerical solution for the parameters in the parsimonious model using the iterative Newton–Raphson or Fisher's scoring methods. Such problems are not uncommon for the iteration-based algorithms especially in the case of a large number of parameters to be estimated with limited sample size and/or insufficient data points (Shen & Gao, 2008). The common causes

for nonconvergence and/or instability (e.g., inappropriate convergence to local optimum) include poor initial guess, the existence of stationary points, flat or ridged likelihood surfaces, and multicollinearity among regressors. There may be a few potential ways to mitigate nonconvergence and stabilize estimation in analyzing interactions. To accommodate the interaction effects in our congenital heart defects studies, the full model will include additional tens of interaction parameters between four genetic effects and six maternal characteristics. One feasible solution is to prioritize the interaction effects to be tested in the statistical model on the basis of their biological plausibility and relevance and thereby decrease the sophistication of the statistical model; for example, the BMI–gene interactions may be considered at a high priority within our context. Another is the use of more robust methods such as penalized maximum likelihood (Firth, 1993), homotopy analysis method (Abbasbandy et al., 2007; Watson & Haftka, 1989), and a hybrid method that combines with a grid search method. Furthermore, in practical implementation, it will also be helpful to try multiple nice initial guesses of parameters. It will be crucial in analyzing a large number of interactions to circumvent the problems of convergence and estimation stability.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This study was supported in part by National Institute of Child Health and Human Development Grant R01HD039054, the National Center on Birth Defects and Developmental Disabilities Grant U01DD001039, National Science Foundation Grants DMS2002865 and DMS1701433, the UAMS Research Scholar Pilot Grant Awards in Child Health G1-51898, and the National Natural Science Foundation of China 31470083 and 31671570. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have declared that no competing interests exist.

## DATA AVAILABILITY STATEMENT

The phenotype and genotype data used in this study are available in the database of CDC, which is subject to approval by the NBDPS Data Sharing Committee. The computer program to implement the proposed analysis is freely available from the website: http://ibi.zju.edu.cn/software. The authors affirm that all data necessary for confirming the conclusions of the article are present within the article, figures, and tables.

## REFERENCES

Abbasbandy S, Tan Y, & Liao SJ (2007). Newton-homotopy analysis method for nonlinear equations. Applied Mathematics and Computation, 188(2), 1794–1800. 10.1016/j.amc.2006.11.136

Abecasis GR, Cardon LR, & Cookson WO (2000). A general test of association for quantitative traits in nuclear families. American Journal of Human Genetics, 66(1), 279–292. [PubMed: 10631157]

Agresti A (1999). Modelling ordered categorical data: Recent advances and future challenges. Statistics in Medicine, 18(17–18), 2191–2207. [PubMed: 10474133]

Allen AS, Rathouz PJ, & Satten GA (2003). Informative missingness in genetic association studies: Case-parent designs. American Journal of Human Genetics, 72(3), 671–680. 10.1086/368276 [PubMed: 12592606]

Allison DB (1997). Transmission-disequilibrium tests for quantitative traits. American Journal of Human Genetics, 60(3), 676–690. [PubMed: 9042929]

Begg CB, & Gray R (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika, 71(1), 11–18. 10.2307/2336391

Benyamin B, Visscher PM, & McRae AF (2009). Family-based genome-wide association studies. Pharmacogenomics, 10(2), 181–190. [PubMed: 19207019]

Bickeboller H, & Clerget-Darpoux F (1995). Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. Genetic Epidemiology, 12(6), 865–870. 10.1002/gepi.1370120656 [PubMed: 8788023]

Blanton SH, Henry RR, Yuan Q, Mulliken JB, Stal S, Finnell RH, & Hecht JT (2011). Folate pathway and nonsyndromic cleft lip and palate. Birth Defects Research. Part A, Clinical and Molecular Teratology, 91(1), 50–60. 10.1002/bdra.20740 [PubMed: 21254359]

Blom HJ, & Smulders Y (2011). Overview of homocysteine and folate metabolism. With special references to cardiovascular disease and neural tube defects. Journal of Inherited Metabolic Disease, 34(1), 75–81. 10.1007/s10545-010-9177-4 [PubMed: 20814827]

Boehnke M, & Langefeld CD (1998). Genetic association mapping based on discordant sib pairs: The discordant-alleles test. American Journal of Human Genetics, 62(4), 950–961. [PubMed: 9529345]

Botto LD, Lin AE, Riehle-Colarusso T, Malik S, & Correa A, National Birth Defects Prevention Study. (2007). Seeking causes: Classifying and evaluating congenital heart defects in etiologic studies. Birth Defects Research. Part A, Clinical and Molecular Teratology, 79(10), 714–727. 10.1002/bdra.20403 [PubMed: 17729292]

Casellas J, Farber CR, Gularte RJ, Haus KA, Warden CH, & Medrano JF (2009). Evidence of maternal QTL affecting growth and obesity in adult mice. Mammalian Genome, 20(5), 269–280. 10.1007/s00335-009-9182-9 [PubMed: 19399551]

Chiquet BT, Henry R, Burt A, Mulliken JB, Stal S, Blanton SH, & Hecht JT (2011). Nonsyndromic cleft lip and palate: Crispld genes and the folate gene pathway connection. Birth Defects Research. Part A, Clinical and Molecular Teratology, 91(1), 44–49. 10.1002/bdra.20737 [PubMed: 21254358]

Clark MM, Blangero J, Dyer TD, Sobel EM, & Sinsheimer JS (2016). The quantitative-mfg test: A linear mixed effect model to detect maternal-offspring gene interactions. Annals of Human Genetics, 80(1), 63–80. 10.1111/ahg.12137 [PubMed: 26567478]

Cleves MA, Ghaffar S, Zhao W, Mosley BS, & Hobbs CA (2003). First-year survival of infants born with congenital heart defects in arkansas (1993–1998): A survival analysis using registry data. Birth Defects Research. Part A, Clinical and Molecular Teratology, 67(9), 662–668. 10.1002/bdra.10119 [PubMed: 14703791]

Cleves MA, Olson JM, & Jacobs KB (1997). Exact transmission-disequilibrium tests with multiallelic markers. Genetic Epidemiology, 14(4), 337–347. 10.1002/(SICI)1098-2272(1997)14:4<337::AID-GEPI1>3.0.CO;2-0 [PubMed: 9271708]

National Research Council (US); Institute of Medicine (US). (2004). Children's health, the nation's wealth: Assessing and improving child health. The National Academies Press.

Copp AJ, Stanier P, & Greene ND (2013). Neural tube defects: Recent advances, unsolved questions, and controversies. Lancet Neurology, 12(8), 799–810. 10.1016/S1474-4422(13)70110-8 [PubMed: 23790957]

Cordell HJ, Barratt BJ, & Clayton DG (2004). Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. Genetic Epidemiology, 26(3), 167–185. [PubMed: 15022205]

Cordell HJ, & Clayton DG (2002). A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: Application to HLA in type 1 diabetes. American Journal of Human Genetics, 70(1), 124–141. [PubMed: 11719900]

Cordell HJ, & Clayton DG (2005). Genetic association studies. Lancet, 366(9491), 1121–1131. [PubMed: 16182901]

Cox DR (1972). Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological), 34(2), 187–220.

Cripe L, Andelfinger G, Martin LJ, Shooner K, & Benson DW (2004). Bicuspid aortic valve is heritable. Journal of the American College of Cardiology, 44(1), 138–143. 10.1016/j.jacc.2004.03.050 [PubMed: 15234422]

Dalton TP, Dieter MZ, Yang Y, Shertzer HG, & Nebert DW (2000). Knockout of the mouse glutamate cysteine ligase catalytic subunit (GCLC) gene: Embryonic lethal when homozygous, and proposed model for moderate glutathione deficiency when heterozygous. Biochemical and Biophysical Research Communications, 279(2), 324–329. 10.1006/bbrc.2000.3930 [PubMed: 11118286]

van Den Oord EJ, & Vermunt JK (2000). Testing for linkage disequilibrium, maternal effects, and imprinting with (in) complete case-parent triads, by use of the computer program lem. American Journal of Human Genetics, 66(1), 335–338. 10.1086/302708 [PubMed: 10631165]

Di Pietro E, Sirois J, Tremblay ML, & MacKenzie RE (2002). Mitochondrial nad-dependent methylenetetrahydrofolate dehydrogenase-methenyltetrahydrofolate cyclohydrolase is essential for embryonic development. Molecular and Cellular Biology, 22(12), 4158–4166. [PubMed: 12024029]

Drton M (2009). Likelihood ratio tests and singularities. The Annals of Statistics, 37(2), 979–1012. 10.1214/07-AOS571

Ducker GS, & Rabinowitz JD (2017). One-carbon metabolism in health and disease. Cell Metabolism, 25(1), 27–42. 10.1016/j.cmet.2016.08.009 [PubMed: 27641100]

Epstein MP, Veal CD, Trembath RC, Barker JN, Li C, & Satten GA (2005). Genetic association analysis using data from triads and unrelated subjects. American Journal of Human Genetics, 76(4), 592–608. [PubMed: 15712104]

Evans RG, & Stoddart GL (1990). Producing health, consuming health care. Social Science and Medicine, 31(12), 1347–1463. [PubMed: 2126895]

Ewens WJ, & Spielman RS (1995). The transmission/disequilibrium test: History, subdivision, and admixture. American Journal of Human Genetics, 57(2), 455–464. [PubMed: 7668272]

Feng ZD, & McCulloch CE (1996). Using bootstrap likelihood ratios in finite mixture models. Journal of the Royal Statistical Society. Series B (Methodological), 58(3), 609–617.

Ferencz C, Rubin JD, McCarter RJ, Brenner JI, Neill CA, Perry LW, Hepner SI, & Downing JW (1985). Congenital heart disease: Prevalence at livebirth. The Baltimore-Washington infant study. American Journal of Epidemiology, 121(1), 31–36. [PubMed: 3964990]

Firth D (1993). Bias reduction of maximum-likelihood-estimates. Biometrika, 80(1), 27–38. 10.2307/2336755

Fishman MC, & Olson EN (1997). Parsing the heart: Genetic modules for organ assembly. Cell, 91(2), 153–156. [PubMed: 9346232]

de Franchis R, Botto LD, Sebastio G, Ricci R, Iolascon A, Capra V, Andria G, & Mastroiacovo P (2002). Spina bifida and folate-related genes: A study of gene-gene interactions. Genetics in Medicine, 4(3), 126–130. 10.1097/00125817-200205000-00005 [PubMed: 12180146]

Fowden AL, & Moore T (2012). Maternal-fetal resource allocation: Co-operation and conflict. Placenta, 33(Suppl 2), e11–e15. 10.1016/j.placenta.2012.05.002 [PubMed: 22652046]

Gelb BD, & Chung WK (2014). Complex genetics and the etiology of human congenital heart disease. Cold Spring Harbor Perspectives in Medicine, 4(7), a013953. 10.1101/cshperspect.a013953 [PubMed: 24985128]

Gluckman PD, & Hanson MA (2004). Living with the past: Evolution, development, and patterns of disease. Science, 305(5691), 1733–1736. 10.1126/science.1095292 [PubMed: 15375258]

Gysin R, Kraftsik R, Sandell J, Bovet P, Chappuis C, Conus P, & Do KQ (2007). Impaired glutathione synthesis in schizophrenia: Convergent genetic and functional evidence. Proceedings of the National Academy of Sciences of the United States of America, 104(42), 16621–16626. 10.1073/pnas.0706778104 [PubMed: 17921251]

Hackshaw A, Rodeck C, & Boniface S (2011). Maternal smoking in pregnancy and birth defects: A systematic review based on 173 687 malformed cases and 11.7 million controls. Human Reproduction Update, 17(5), 589–604. 10.1093/humupd/dmr022 [PubMed: 21747128]

Han M, Hu YQ, & Lin S (2013). Joint detection of association, imprinting and maternal effects using all children and their parents. European Journal of Human Genetics, 21(12), 1449–1456. 10.1038/ejhg.2013.49 [PubMed: 23531864]

Healy J, Bourgey M, Richer C, Sinnett D, & Roy-Gagnon MH (2010). Detection of fetomaternal genotype associations in early-onset disorders: Evaluation of different methods and their application to childhood leukemia. Journal of Biomedicine and Biotechnology, 2010, 369534–13. 10.1155/2010/369534 [PubMed: 20617153]

Hernandez LM, & Blazer DG (2006). Genes, behavior, and the social environment: Moving beyond the nature/nurture debate. National Academies Press.

Hinton RB Jr., Martin LJ, Tabangin ME, Mazwi ML, Cripe LH, & Benson DW (2007). Hypoplastic left heart syndrome is heritable. Journal of the American College of Cardiology, 50(16), 1590–1595. 10.1016/j.jacc.2007.07.021 [PubMed: 17936159]

Hobbs CA, Cleves MA, Karim MA, Zhao W, & MacLeod SL (2010). Maternal folate-related gene environment interactions and congenital heart defects. Obstetrics and Gynecology, 116(2 Pt 1), 316–322. 10.1097/AOG.0b013e3181e80979 [PubMed: 20664391]

Hobbs CA, Cleves MA, Macleod SL, Erickson SW, Tang X, Li J, Li M, Nick T, & Malik S, National Birth Defects Prevention, Study. (2014). Conotruncal heart defects and common variants in maternal and fetal genes in folate, homocysteine, and transsulfuration pathways. Birth Defects Research. Part A, Clinical and Molecular Teratology, 100(2), 116–126. 10.1002/bdra.23225 [PubMed: 24535845]

Hobbs CA, Cleves MA, Melnyk S, Zhao W, & James SJ (2005). Congenital heart defects and abnormal maternal biomarkers of methionine and homocysteine metabolism. American Journal of Clinical Nutrition, 81(1), 147–153.

Hobbs CA, MacLeod SL, Jill James S, & Cleves MA (2011). Congenital heart defects and maternal genetic, metabolic, and lifestyle factors. Birth Defects Research. Part A, Clinical and Molecular Teratology, 91(4), 195–203. 10.1002/bdra.20784 [PubMed: 21384532]

Hoffman J (2013). The global burden of congenital heart disease. Cardiovascular Journal of Africa, 24(4), 141–145. 10.5830/CVJA-2013-028 [PubMed: 24217047]

Horvath S, & Laird NM (1998). A discordant-sibship test for disequilibrium and linkage: No need for parental data. American Journal of Human Genetics, 63(6), 1886–1897. [PubMed: 9837840]

Horvath S, Xu X, & Laird NM (2001). The family based association test method: Strategies for studying general genotype–phenotype associations. European Journal of Human Genetics, 9(4), 301–306. [PubMed: 11313775]

Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, & Laird NM (2004). Family-based tests for associating haplotypes with general phenotype data: Application to asthma genetics. Genetic Epidemiology, 26(1), 61–69. [PubMed: 14691957]

Hozyasz KK (2010). The search for risk factors that contribute to the etiology of non-syndromic cleft lip with or without cleft palate (cl/p) in the polish population. Pediatria Polska, 85(6), 609–623. 10.1016/S0031-3939(10)70562-X

Hu YQ, & Zhou JY (2010). Inferring haplotype/disease association by joint use of case-parents trios and case-parent pairs. Annals of Human Genetics, 74(3), 263–274. 10.1111/j.1469-1809.2010.00563.x [PubMed: 20529016]

Hu YQ, Zhou JY, & Fung WK (2007). An extension of the transmission disequilibrium test incorporating imprinting. Genetics, 175(3), 1489–1504. 10.1534/genetics.106.058461 [PubMed: 17194789]

Hu YQ, Zhou JY, Sun F, & Fung WK (2007). The transmission disequilibrium test and imprinting effects test based on case-parent pairs. Genetic Epidemiology, 31(4), 273–287. 10.1002/gepi.20208 [PubMed: 17266118]

Hunter DJ (2005). Gene-environment interactions in human diseases. Nature Reviews Genetics, 6(4), 287–298.

Hutson MR, & Kirby ML (2007). Model systems for the study of heart development and disease. Cardiac neural crest and conotruncal malformations. Seminars in Cell and Developmental Biology, 18(1), 101–110. 10.1016/j.semcdb.2006.12.004 [PubMed: 17224285]

Jarvis JP, Kenney-Hunt J, Ehrich TH, Pletscher LS, Semenkovich CF, & Cheverud JM (2005). Maternal genotype affects adult offspring lipid, obesity, and diabetes phenotypes in lgxsm recombinant inbred strains. Journal of Lipid Research, 46(8), 1692–1702. 10.1194/jlr.M500073-JLR200 [PubMed: 15897602]

Kaplan NL, Martin ER, & Weir BS (1997). Power studies for the transmission/disequilibrium tests with multiple alleles. American Journal of Human Genetics, 60(3), 691–702. [PubMed: 9042930]

Kase BA, Northrup H, Morrison AC, Davidson CM, Goiffon AM, Fletcher JM, Ostermaier KK, Tyerman GH, & Au KS (2012). Association of copper-zinc superoxide dismutase (sod1) and manganese superoxide dismutase (SOD2) genes with nonsyndromic myelomeningocele. Birth Defects Research. Part A, Clinical and Molecular Teratology, 94(10), 762–769. 10.1002/bdra.23065 [PubMed: 22972774]

Keverne EB (2015). Genomic imprinting, action, and interaction of maternal and fetal genomes. Proceedings of the National Academy of Sciences of the United States of America, 112(22), 6834–6840. 10.1073/pnas.1411253111 [PubMed: 25404322]

Kirby ML (2002). Molecular embryogenesis of the heart. Pediatric and Developmental Pathology, 5(6), 516–543. 10.1007/s10024-002-0004-2 [PubMed: 12297889]

Kirke PN, Mills JL, Molloy AM, Brody LC, O'Leary VB, Daly L, Murray S, Conley M, Mayne PD, Smith O, & Scott JM (2004). Impact of the MTHFR C677T polymorphism on risk of neural tube defects: Case-control study. BMJ, 328(7455), 1535–1536. 10.1136/bmj.38036.646030.EE [PubMed: 15155469]

Kistner EO, Shi M, & Weinberg CR (2009). Using cases and parents to study multiplicative gene-by-environment interaction. American Journal of Epidemiology, 170(3), 393–400. 10.1093/aje/kwp118 [PubMed: 19483188]

Knapp M (1999). The transmission/disequilibrium test and parental-genotype reconstruction: The reconstruction-combined transmission/disequilibrium test. American Journal of Human Genetics, 64(3), 861–870. [PubMed: 10053021]

Koide S, Kugiyama K, Sugiyama S, Nakamura S, Fukushima H, Honda O, Yoshimura M, & Ogawa H (2003). Association of polymorphism in glutamate-cysteine ligase catalytic subunit gene with coronary vasomotor dysfunction and myocardial infarction. Journal of the American College of Cardiology, 41(4), 539–545. [PubMed: 12598062]

Lage K, Greenway SC, Rosenfeld JA, Wakimoto H, Gorham JM, Segre AV, Roberts AE, Smoot LB, Pu WT, Pereira AC, Mesquita SM, Tommerup N, Brunak S, Ballif BC, Shaffer LG, Donahoe PK, Daly MJ, Seidman JG, Seidman CE, & Larsen LA (2012). Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. Proceedings of the National Academy of Sciences of the United States of America, 109(35), 14035–14040. 10.1073/pnas.1210730109 [PubMed: 22904188]

Laird NM, & Lange C (2006). Family-based designs in the age of large-scale gene-association studies. Nature Reviews Genetics, 7(5), 385–394.

Lange C, DeMeo D, Silverman EK, Weiss ST, & Laird NM (2004). Pbat: Tools for family-based association studies. American Journal of Human Genetics, 74(2), 367–369. [PubMed: 14740322]

Lange C, DeMeo DL, & Laird NM (2002). Power and design considerations for a general class of family-based association tests: Quantitative traits. American Journal of Human Genetics, 71(6), 1330–1341. [PubMed: 12454799]

Lange C, & Laird NM (2002). Power calculations for a general class of family-based association tests: Dichotomous traits. American Journal of Human Genetics, 71(3), 575–584. [PubMed: 12181775]

Lawson HA, Cheverud JM, & Wolf JB (2013). Genomic imprinting and parent-of-origin effects on complex traits. Nature Reviews Genetics, 14(9), 609–617. 10.1038/nrg3543

Lazzeroni LC, & Lange K (1998). A conditional inference framework for extending the transmission/disequilibrium test. Human Heredity, 48(2), 67–81. [PubMed: 9526165]

Lee PJ, Ridout D, Walter JH, & Cockburn F (2005). Maternal phenylketonuria: Report from the united kingdom registry 1978–97. Archives of Disease in Childhood, 90(2), 143–146. 10.1136/adc.2003.037762 [PubMed: 15665165]

Liang KY, & Zeger SL (1986). Longitudinal data-analysis using generalized linear-models. Biometrika, 73(1), 13–22.

van der Linde D, Konings EE, Slager MA, Witsenburg M, Helbing WA, Takkenberg JJ, & Roos-Hesselink JW (2011). Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. Journal of the American College of Cardiology, 58(21), 2241–2247. 10.1016/j.jacc.2011.08.025 [PubMed: 22078432]

Lu SC (2009). Regulation of glutathione synthesis. Molecular Aspects of Medicine, 30(1–2), 42–59. 10.1016/j.mam.2008.05.005 [PubMed: 18601945]

Lupo PJ, Mitchell LE, Canfield MA, Shaw GM, Olshan AF, Finnell RH, & Zhu H, National Birth Defects Prevention Study. (2014). Maternal-fetal metabolic gene-gene interactions and risk of neural tube defects. Molecular Genetics and Metabolism, 111(1), 46–51. 10.1016/j.ymgme.2013.11.004 [PubMed: 24332798]

Lupo PJ, Mitchell LE, & Goldmuntz E (2011). Nat1, nos3, and tyms genotypes and the risk of conotruncal cardiac defects. Birth Defects Research. Part A, Clinical and Molecular Teratology, 91(1), 61–65. 10.1002/bdra.20745 [PubMed: 21254360]

Mackay I, & Powell W (2007). Methods for linkage disequilibrium mapping in crops. Trends in Plant Science, 12(2), 57–63. [PubMed: 17224302]

Martin ER, Monks SA, Warren LL, & Kaplan NL (2000). A test for linkage and association in general pedigrees: The pedigree disequilibrium test. American Journal of Human Genetics, 67(1), 146–154. [PubMed: 10825280]

McBride KL, Pignatelli R, Lewin M, Ho T, Fernbach S, Menesses A, Lam W, Leal SM, Kaplan N, Schliekelman P, Towbin JA, & Belmont JW (2005). Inheritance analysis of congenital left ventricular outflow tract obstruction malformations: Segregation, multiplex relative risk, and heritability. American Journal of Medical Genetics. Part A, 134A(2), 180–186. 10.1002/ajmg.a.30602 [PubMed: 15690347]

McCullagh P (1983). Quasi-likelihood functions. Annals of Statistics, 11(1), 59–67.

McKone EF, Shao J, Frangolias DD, Keener CL, Shephard CA, Farin FM, Tonelli MR, Pare PD, Sandford AJ, Aitken ML, & Kavanagh TJ (2006). Variants in the glutamate-cysteine-ligase gene are associated with cystic fibrosis lung disease. American Journal of Respiratory and Critical Care Medicine, 174(4), 415–419. 10.1164/rccm.200508-1281OC [PubMed: 16690975]

McNemar Q (1947). Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika, 12(2), 153–157. 10.1007/bf02295996 [PubMed: 20254758]

Meaney MJ (2010). Epigenetics and the biological definition of gene x environment interactions. Child Development, 81(1), 41–79. 10.1111/j.1467-8624.2009.01381.x [PubMed: 20331654]

Mitchell LE (1997). Differentiating between fetal and maternal genotypic effects, using the transmission test for linkage disequilibrium. American Journal of Human Genetics, 60(4), 1006–1007. [PubMed: 9106551]

Mitchell LE, & Weinberg CR (2005). Evaluation of offspring and maternal genetic effects on disease risk using a family-based approach: The “pent” design. American Journal of Epidemiology, 162(7), 676–685. 10.1093/aje/kwi249 [PubMed: 16093287]

Mitchell SC, Korones SB, & Berendes HW (1971). Congenital heart disease in 56,109 births. Incidence and natural history. Circulation, 43(3), 323–332. [PubMed: 5102136]

Monks SA, & Kaplan NL (2000). Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. American Journal of Human Genetics, 66(2), 576–592. [PubMed: 10677318]

Mousseau TA, & Fox CW (1998). The adaptive significance of maternal effects. Trends in Ecology and Evolution, 13(10), 403–407. [PubMed: 21238360]

Nadeau JH (2009). Transgenerational genetic effects on phenotypic variation and disease risk. Human Molecular Genetics, 18(R2), R202–R210. 10.1093/hmg/ddp366 [PubMed: 19808797]

Nagelkerke NJD, Hoebee B, Teunis P, & Kimman TG (2004). Combining the transmission disequilibrium test and case-control methodology using generalized logistic regression. European Journal of Human Genetics, 12(11), 964–970. [PubMed: 15340361]

Nelder JA, & Wedderburn RWM (1972). Generalized linear models. Journal of the Royal Statistical Society. Series A, 135(3), 370–384.

Newacheck PW, Rising JP, & Kim SE (2006). Children at risk for special health care needs. Pediatrics, 118(1), 334–342. 10.1542/peds.2005-2238 [PubMed: 16818583]

Nora JJ (1968). Multifactorial inheritance hypothesis for the etiology of congenital heart diseases. The genetic-environmental interaction. Circulation, 38(3), 604–617. [PubMed: 4876982]

Ott J (1989). Statistical properties of the haplotype relative risk. Genetic Epidemiology, 6(1), 127–130. [PubMed: 2731704]

Ott J, Kamatani Y, & Lathrop M (2011). Family-based designs for genome-wide association studies. Nature Reviews Genetics, 12(7), 465–474.

Peters J (2014). The role of genomic imprinting in biology and disease: An expanding view. Nature Reviews Genetics, 15(8), 517–530. 10.1038/nrg3766

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, & Reich D (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. 10.1038/ng1847 [PubMed: 16862161]

Price AL, Zaitlen NA, Reich D, & Patterson N (2010). New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics, 11(7), 459–463.

Pritchard J, Stephens M, Rosenberg N, & Donnelly P (2000). Association mapping in structured populations. American Journal of Human Genetics, 67(1), 170–181. 10.1086/302959 [PubMed: 10827107]

Rabinowitz D (1997). A transmission disequilibrium test for quantitative trait loci. Human Heredity, 47(6), 342–350. [PubMed: 9391826]

Rabinowitz D, & Laird N (2000). A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. Human Heredity, 50(4), 211–223. [PubMed: 10782012]

Reik W, & Walter J (2001). Genomic imprinting: Parental influence on the genome. Nature Reviews Genetics, 2(1), 21–32. 10.1038/35047554

Rice JP, Neuman RJ, Hoshaw SL, Daw EW, & Gu C (1995). TDT with covariates and genomic screens with mod scores: Their behavior on simulated data. Genetic Epidemiology, 12(6), 659–664. 10.1002/gepi.1370120623 [PubMed: 8787990]

Ristoff E, Augustson C, Geissler J, Rijk T, Carlsson K, Luo JL, Andersson K, Weening RS, van Zwieten R, Larsson A, & Roos D (2000). A missense mutation in the heavy subunit of gamma-glutamylcysteine synthetase gene causes hemolytic anemia. Blood, 95(7), 2193–2196. de. [PubMed: 10733484]

Schaid DJ, & Sommer SS (1993). Genotype relative risks: Methods for design and analysis of candidate-gene association studies. American Journal of Human Genetics, 53(5), 1114–1126. [PubMed: 8213835]

Schaid DJ, & Sommer SS (1994). Comparison of statistics for candidate-gene association studies using cases and parents. American Journal of Human Genetics, 55(2), 402–409. [PubMed: 8037216]

Sferruzzi-Perri AN, Lopez-Tello J, Fowden AL, & Constancia M (2016). Maternal and fetal genomes interplay through phosphoinositol 3-kinase(pi3k)-p110alpha signaling to modify placental resource allocation. Proceedings of the National Academy of Sciences of the United States of America, 113(40), 11255–11260. 10.1073/pnas.1602012113 [PubMed: 27621448]

Sham PC, & Curtis D (1995). An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. Annals of Human Genetics, 59(Pt 3), 323–336. [PubMed: 7486838]

Shaw GM, Lu W, Zhu H, Yang W, Briggs FB, Carmichael SL, Barcellos LF, Lammer EJ, & Finnell RH (2009). 118 SNPS of folate-related genes and risks of spina bifida and conotruncal heart defects. BMC Medical Genetics, 10, 49. 10.1186/1471-2350-10-49 [PubMed: 19493349]

Shen J, & Gao S (2008). A solution to separation and multicollinearity in multiple logistic regression. Journal of Data Science, 6(4), 515–531. [PubMed: 20376286]

Sinsheimer JS, Palmer CGS, & Woodward JA (2003). Detecting genotype combinations that increase risk for disease: The maternal-fetal genotype incompatibility test. Genetic Epidemiology, 24(1), 1–13. 10.1002/gepi.10211 [PubMed: 12508251]

Slatkin M (2008). Linkage disequilibrium: Understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics, 9(6), 477–485.

Spielman RS, & Ewens WJ (1996). The TDT and other family-based tests for linkage disequilibrium and association. American Journal of Human Genetics, 59(5), 983–989. [PubMed: 8900224]

Spielman RS, & Ewens WJ (1998). A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test. American Journal of Human Genetics, 62(2), 450–458. [PubMed: 9463321]

Spielman RS, McGinnis RE, & Ewens WJ (1993). Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). American Journal of Human Genetics, 52(3), 506–516. [PubMed: 8447318]

Srivastava D (2001). Genetic assembly of the heart: Implications for congenital heart disease. Annual Review of Physiology, 63, 451–469. 10.1146/annurev.physiol.63.1.451

Srivastava D, & Olson EN (2000). A genetic blueprint for cardiac development. Nature, 407(6801), 221–226. 10.1038/35025190 [PubMed: 11001064]

Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, & Baur MP (2000). Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: Application to mite sensitization. American Journal of Human Genetics, 66(6), 1945–1957. [PubMed: 10796874]

Sun FZ, Flanders WD, Yang QH, & Khoury MJ (1999). Transmission disequilibrium test (TDT) when only one parent is available: The 1-TDT. American Journal of Epidemiology, 150(1), 97–104. [PubMed: 10400559]

Tang X, Cleves MA, Nick TG, Li M, MacLeod SL, Erickson SW, Li J, Shaw GM, Mosley BS, & Hobbs CA, National Birth Defects Prevention Study. (2015). Obstructive heart defects associated with candidate genes, maternal obesity, and folic acid supplementation. American Journal of Medical Genetics. Part A, 167(6), 1231–1442. 10.1002/ajmg.a.36867 [PubMed: 25846410]

Thomas F, Balkau B, Vauzelle-Kervroedan F, & Papoz L (1994). Maternal effect and familial aggregation in niddm. The codiab study. Codiab-inserm-zeneca study group. Diabetes, 43(1), 63–67. [PubMed: 8262318]

Tilley MM, Northrup H, & Au KS (2012). Genetic studies of the cystathionine beta-synthase gene and myelomeningocele. Birth Defects Research. Part A, Clinical and Molecular Teratology, 94(1), 52–56. 10.1002/bdra.22855 [PubMed: 21957013]

Tsang BL, Devine OJ, Cordero AM, Marchetta CM, Mulinare J, Mersereau P, Guo J, Qi YP, Berry RJ, Rosenthal J, Crider KS, & Hamner HC (2015). Assessing the association between the methylenetetrahydrofolate reductase (mthfr) 677c>t polymorphism and blood folate concentrations: A systematic review and meta-analysis of trials and observational studies. American Journal of Clinical Nutrition, 101(6), 1286–1294. 10.3945/ajcn.114.099994

Ufer C, & Wang CC (2011). The roles of glutathione peroxidases during embryo development. Frontiers in Molecular Neuroscience, 4, 12. 10.3389/fnmol.2011.00012 [PubMed: 21847368]

VanLiere JM, & Rosenberg NA (2008). Mathematical properties of the r(2) measure of linkage disequilibrium. Theoretical Population Biology, 74(1), 130–137. 10.1016/j.tpb.2008.05.006 [PubMed: 18572214]

Vecoli C, Pulignani S, Foffa I, & Andreassi MG (2014). Congenital heart disease: The crossroads of genetics, epigenetics and environment. Current Genomics, 15(5), 390–399. 10.2174/1389202915666140716175634 [PubMed: 25435801]

Vermeulen SH, Shi M, Weinberg CR, & Umbach DM (2009). A hybrid design: Case-parent triads supplemented by control-mother dyads. Genetic Epidemiology, 33(2), 136–144. 10.1002/gepi.20365 [PubMed: 18759250]

Volcik KA, Shaw GM, Zhu H, Lammer EJ, Laurent C, & Finnell RH (2003). Associations between polymorphisms within the thymidylate synthase gene and spina bifida. Birth Defects Research. Part A, Clinical and Molecular Teratology, 67(11), 924–928. 10.1002/bdra.10029 [PubMed: 14745930]

Wakefield J (2007). A bayesian measure of the probability of false discovery in genetic epidemiology studies. American Journal of Human Genetics, 81(2), 208–227. [PubMed: 17668372]

Wang K, Hu X, & Peng Y (2013). An analytical comparison of the principal component method and the mixed effects model for association studies in the presence of cryptic relatedness and population stratification. Human Heredity, 76(1), 1–9. 10.1159/000353345 [PubMed: 23921716]

Watson LT, & Haftka RT (1989). Modern homotopy methods in optimization. Computer Methods in Applied Mechanics and Engineering, 74(3), 289–305. 10.1016/0045-7825(89)90053-4

Weinberg CR (1999a). Allowing for missing parents in genetic studies of case-parent triads. American Journal of Human Genetics, 64(4), 1186–1193. [PubMed: 10090904]

Weinberg CR (1999b). Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. American Journal of Human Genetics, 65(1), 229–235. 10.1086/302466 [PubMed: 10364536]

Weinberg CR (2003). Studying parents and grandparents to assess genetic contributions to early-onset disease. American Journal of Human Genetics, 72(2), 438–447. 10.1086/346171 [PubMed: 12533786]

Weinberg CR, & Umbach DM (2005). A hybrid design for studying genetic influences on risk of diseases with onset early in life. American Journal of Human Genetics, 77(4), 627–636. 10.1086/496900 [PubMed: 16175508]

Weinberg CR, Wilcox AJ, & Lie RT (1998). A log-linear approach to case-parent-triad data: Assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. American Journal of Human Genetics, 62(4), 969–978. 10.1086/301802 [PubMed: 9529360]

Wilcox AJ, Weinberg CR, & Lie RT (1998). Distinguishing the effects of maternal and offspring genes through studies of “case-parent triads”. American Journal of Epidemiology, 148(9), 893–901. [PubMed: 9801020]

Wilson SR (1997). On extending the transmission/disequilibrium test (TDT). Annals of Human Genetics, 61(Pt 2), 151–161. 10.1046/j.1469-1809.1997.6120151.x [PubMed: 9177122]

Wolf JB, & Wade MJ (2009). What are maternal effects (and what are they not)? Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364(1520), 1107–1115. 10.1098/rstb.2008.0238 [PubMed: 19324615]

Wu CQ, DeWan A, Hoh J, & Wang ZH (2011). A comparison of association methods correcting for population stratification in case-control studies. Annals of Human Genetics, 75, 418–427. [PubMed: 21281271]

Xiong MM, Krushkal J, & Boerwinkle E (1998). TDT statistics for mapping quantitative trait loci. Annals of Human Genetics, 62(Pt 5), 431–452. [PubMed: 10088040]

Yan L, Zhao L, Long Y, Zou P, Ji G, Gu A, & Zhao P (2012). Association of the maternal mthfr c677t polymorphism with susceptibility to neural tube defects in offsprings: Evidence from 25 case-control studies. PLoS One, 7(10), e41689. 10.1371/journal.pone.0041689 [PubMed: 23056169]

Yang J, & Lin S (2013). Robust partial likelihood approach for detecting imprinting and maternal effects using case-control families. The Annals of Applied Statistics, 7(1), 249–268.

Yoon PW, Rasmussen SA, Lynberg MC, Moore CA, Anderka M, Carmichael SL, Costa P, Druschel C, Hobbs CA, Romitti PA, Langlois PH, & Edmonds LD (2001). The national birth defects prevention study. Public Health Reports, 116(Suppl 1), 32–40.

Youngson NA, & Whitelaw E (2008). Transgenerational epigenetic effects. Annual Review of Genomics and Human Genetics, 9, 233–257. 10.1146/annurev.genom.9.081307.164445

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, & Buckler ES (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 38(2), 203–208. [PubMed: 16380716]

Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, & Buckler ES (2010). Mixed linear model approach adapted for genome-wide association studies. Nature Genetics, 42(4), 355–U118. [PubMed: 20208535]

Zhao H, Zhang S, Merikangas KR, Trixler M, Wildenauer DB, Sun F, & Kidd KK (2000). Transmission/disequilibrium tests using multiple tightly linked markers. American Journal of Human Genetics, 67(4), 936–946. [PubMed: 10968775]

Zhu H, Yang W, Lu W, Etheredge AJ, Lammer EJ, Finnell RH, Carmichael SL, & Shaw GM (2012). Gene variants in the folate-mediated one-carbon metabolism (focm) pathway as risk factors for conotruncal heart defects. American Journal of Medical Genetics. Part A, 158A(5), 1124–1134. 10.1002/ajmg.a.35313 [PubMed: 22495907]

Zhu H, Yang W, Shaw N, Perloff S, Carmichael SL, Finnell RH, Shaw GM, & Lammer EJ (2012). Thymidylate synthase polymorphisms and risk of conotruncal heart defects. American Journal of Medical Genetics. Part A, 158A(9), 2194–2203. 10.1002/ajmg.a.35310 [PubMed: 22887475]

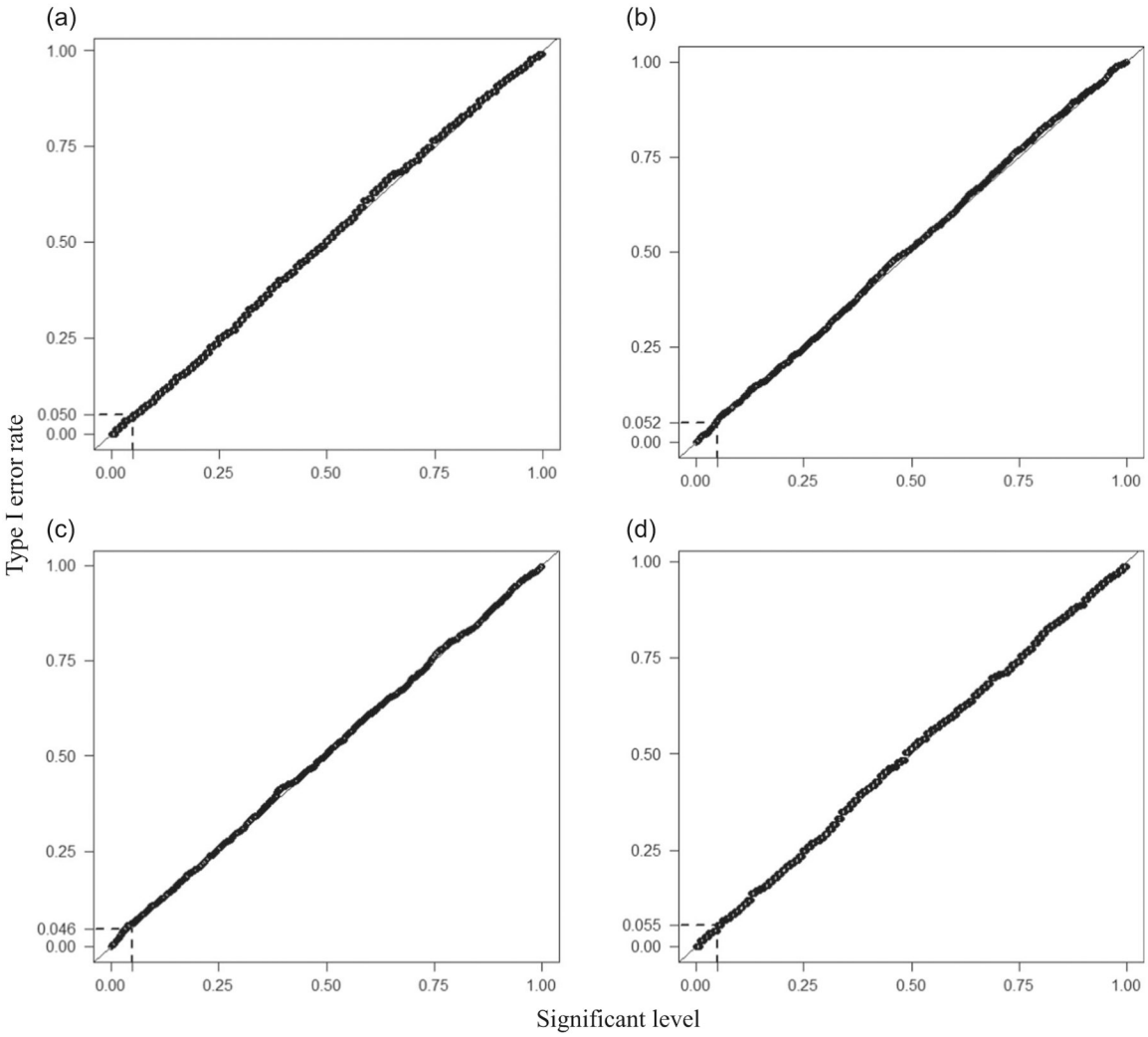


**FIGURE 1.**

Probability–probability plot of significance level versus Type I error rate for the logistic approach. The horizontal axis represents significance level while the vertical axis is Type I error rate. The reference line is the diagonal line with unit slope through the origin. To evaluate the Type I error rate, the scenarios of no linkage were considered: (a) the absence of association, (b) the presence of association due to child effects, (c) the presence of association due to child effects as well as the effect of an environmental factor, and (d) the presence of association due to child and maternal effects as well as the effect of an environmental factor

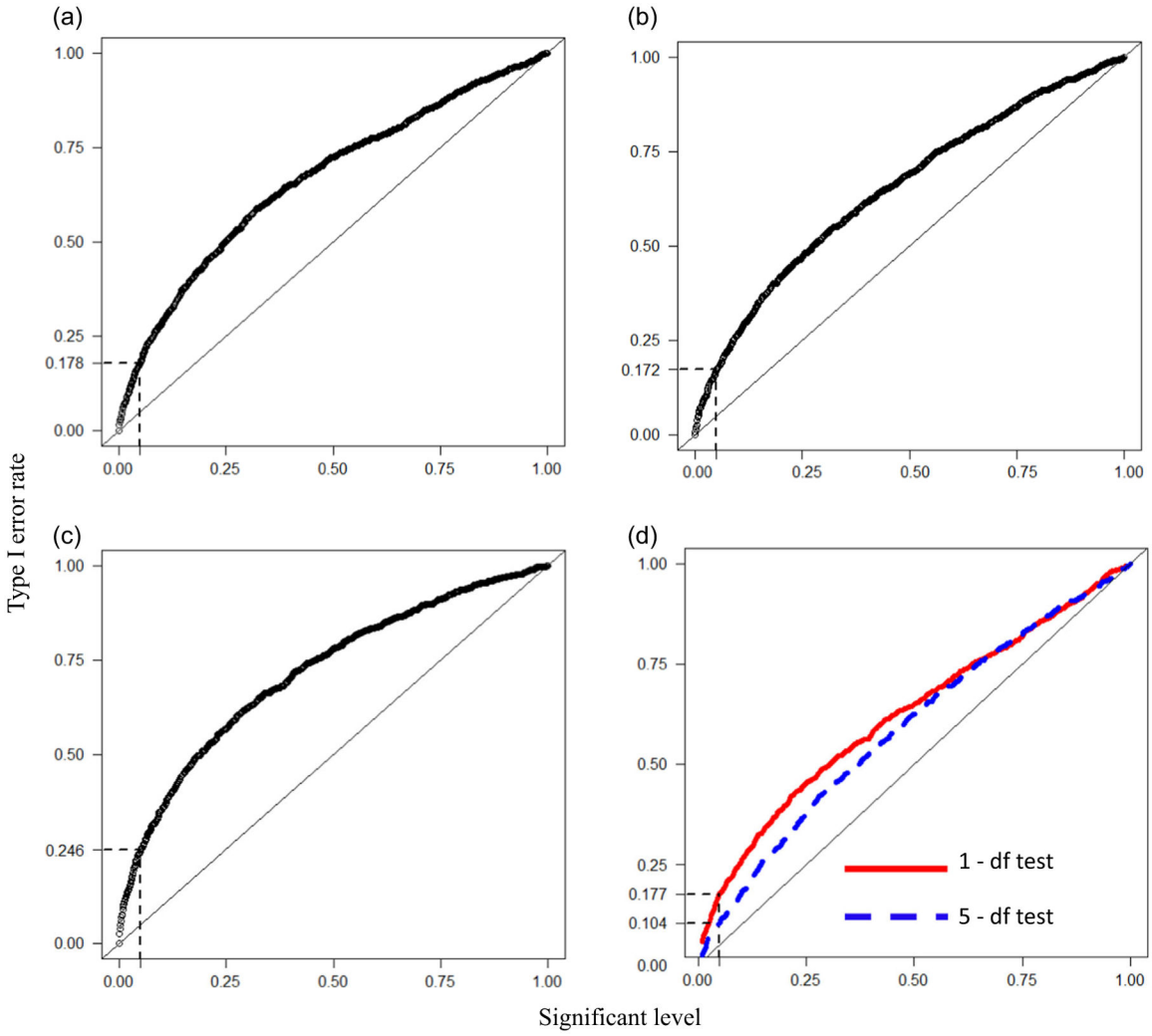


**FIGURE 2.**

Probability–probability plot of significance level versus Type I error rate for the log-linear method. The horizontal axis represents the significance level while the vertical axis is Type I error rate. The reference line is the diagonal line with unit slope through the origin. To evaluate the Type I error rate, the scenarios of no linkage were considered: (a) the presence of association due to child effects, (b) the presence of association due to child effects as well as the effect of an environmental factor, (c) the presence of association due to child and maternal effects as well as the effect of an environmental factor, and (d) is the Type I error rates for the likelihood ratio tests with 5 *df* and 1 *df* for population structure in the log-linear method, in the scenario of no population structure but the recombination rate of 0.0 between the causal locus and marker. The red solid line represents the likelihood ratio (LR) test with 1 *df* and the blue dash line represents the LR test with 5 *df*

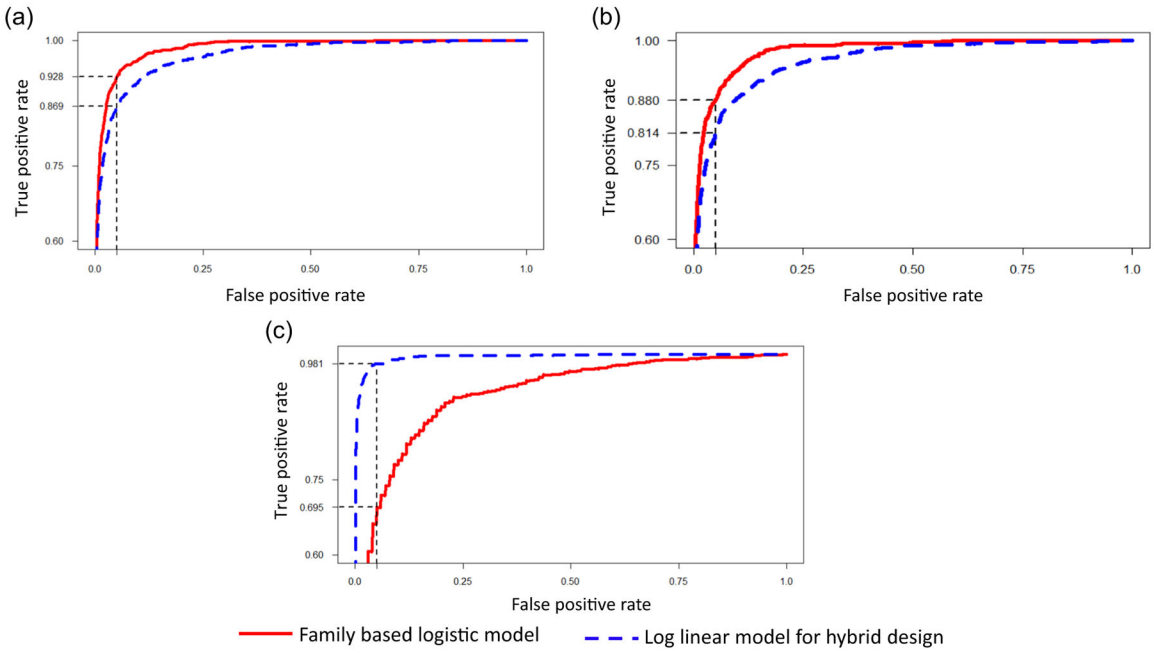


**FIGURE 3.**

Comparison of receiver operating characteristic curves between the logistic model and the log-linear model. The horizontal axis represents the false-positive rate, while the vertical axis is the true positive rate. Red solid line is for the family-based logistic model and blue dash line is for the log-linear model. To assess the statistical power, the scenarios of complete linkage with a linkage disequilibrium were considered, (a) the presence of only child effects, (b) the presence of child effects and an environmental effect, and (c) the presence of child and maternal effects and an environmental effect

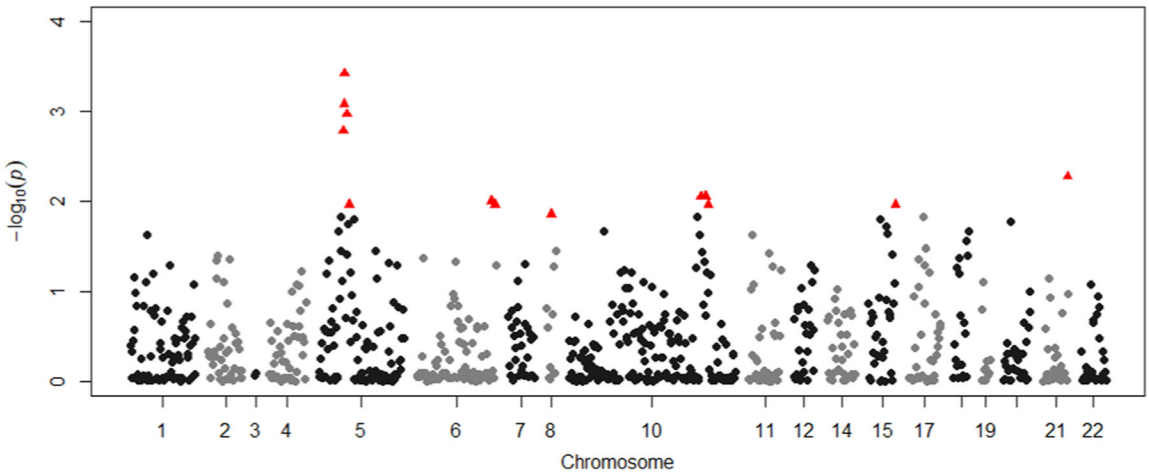


**FIGURE 4.**

Manhattan plot of the fetal SNPs associated with obstructive heart defects. The horizontal axis denotes the SNP position, while the vertical axis represents the minus logarithm of an observed *p*-value. The red triangle points denote the SNPs reaching the noteworthiness level at a BFDP of ≤ .80, while the black circle points represent the SNPs not reaching the noteworthiness level at a BFDP of > .80. BFDP, Bayesian false-discovery probability; SNPs, single nucleotide polymorphisms

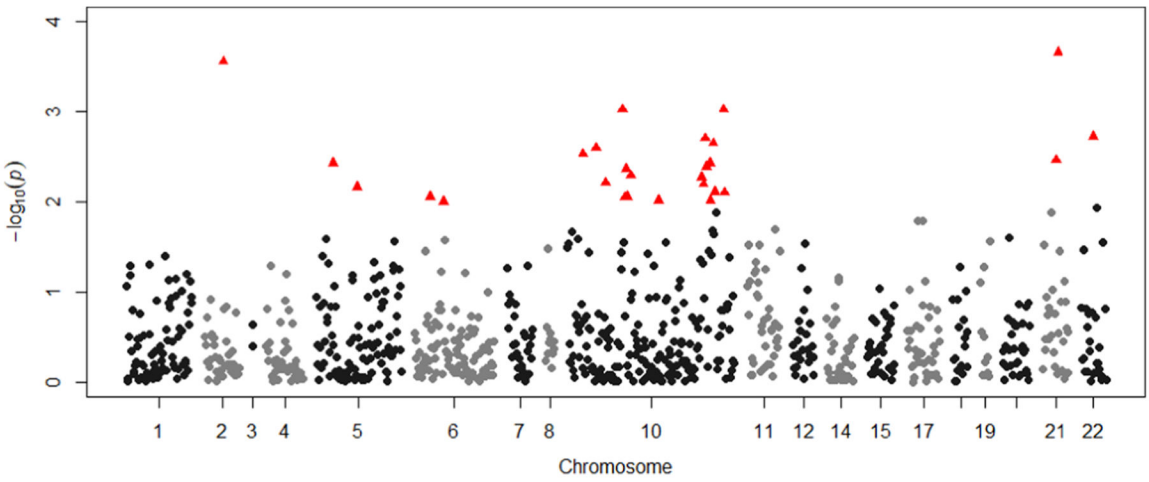


**FIGURE 5.**

Manhattan plot of the maternal SNPs associated with obstructive heart defects. The horizontal axis denotes the SNP position, while the vertical axis represents the minus logarithm of an observed *p*-value. The red triangle points denote the SNPs reaching the noteworthiness level at a BFDP of ≤ .80, while the black circle points represent the SNPs not reaching the noteworthiness level at a BFDP of > .80. BFDP, Bayesian false-discovery probability; SNPs, single nucleotide polymorphisms

**TABLE 1**

Coding for the genetic effects when all the genotypes are available

| Parental genotype | Mating frequency[a] | Child genotype | Frequency given parental mating type | Genetic model[b] |
|---|---|---|---|---|
| MM × MM | $(P_{MM})^2$ | MM | 1.0 | $\mu + (1 + 1)\alpha - \beta$ |
| MM × Mm | $2P_{MM}P_{Mm}$ | MM | 0.5 | $\mu + [1 + (1 - 2\theta)]\alpha - (1 - 2\theta)\beta$ |
| | | Mm | 0.5 | $\mu + [1 - (1 - 2\theta)]\alpha + (1 - 2\theta)\beta$ |
| MM × mm | $2P_{MM}P_{mm}$ | Mm | 1.0 | $\mu + (1 - 1)\alpha + \beta$ |
| Mm × Mm | $(P_{Mm})^2$ | MM | 0.25 | $\mu + [(1 - 2\theta) + (1 - 2\theta)]\alpha - (1 - 2\theta)^2\beta$ |
| | | Mm | 0.5 | $\mu + [(1 - 2\theta) - (1 - 2\theta)]\alpha + (1 - 2\theta)^2\beta$ |
| | | mm | 0.25 | $\mu - [(1 - 2\theta) + (1 - 2\theta)]\alpha - (1 - 2\theta)^2\beta$ |
| Mm × mm | $2P_{Mm}P_{mm}$ | Mm | 0.5 | $\mu - [1 - (1 - 2\theta)]\alpha + (1 - 2\theta)\beta$ |
| | | mm | 0.5 | $\mu - [1 + (1 - 2\theta)]\alpha - (1 - 2\theta)\beta$ |
| mm × mm | $(P_{mm})^2$ | mm | 1.0 | $\mu + (-1 - 1)\alpha - \beta$ |

[a] Mating frequency under the assumption of random mating, where $P_{MM}$, $P_{Mm}$, and $P_{mm}$ are the genotypic frequencies of marker, respectively (this assumption can be relaxed by adding additional parameters reflecting the mating type-specific frequencies).

[b] $\theta$ is the recombination rate between a genetic marker and the causal gene. There is a functional correspondence between the statistical parameters, population mean $\mu$, average effect $\alpha$, and dominance deviation $\beta$, and the inheritance parameters, mid-parent value $m$ additive effect $a$, dominance effect $d$ as well as genotypic frequencies and genotypic disequilibria. When the assumption of random mating holds true, there are $\mu = m + \left[(2p_Q - 1) - \frac{(2p_M - 1)\delta}{p_M(1 - p_M)}\right]a + \frac{1 - \left[(2p_Q - 1) - \frac{(2p_M - 1)\delta}{p_M(1 - p_M)}\right]^2}{2}d$, $\alpha = \frac{\delta}{2p_M(1 - p_M)}\left[a - (2p_Q - 1)d + \frac{(2p_M - 1)\delta}{p_M(1 - p_M)}d\right]$, $\beta = \frac{\delta^2}{2(p_M)^2(1 - p_M)^2}d$, where $\delta$ is the linkage disequilibrium coefficient, and $p_Q$ and $p_M$ are allelic frequencies at the causal gene. In the case of nonrandom mating, additional parameter(s) for genotypic disequilibria can be included in the functional relationship.

**TABLE 2**

Biases and mean squared errors of estimates under different scenarios

| | | Bias (MSE) | | | | | |
|---|---|---|---|---|---|---|---|
| **Scenarios**[a] | $\theta$ | $\alpha_c$ | $\beta_c$ | $\alpha_m$ | $\beta_m$ | **Covariate** | $\theta$ |
| II | 0.50 | 0.009 (0.003) | −0.006 (0.010) | NA | NA | NA | 0.000 (0.044) |
| V | 0.00 | 0.000 (0.001) | 0.003 (0.003) | NA | NA | NA | 0.041 (0.004) |
| III | 0.50 | 0.012 (0.003) | 0.001 (0.009) | NA | NA | −0.064 (0.006) | 0.003 (0.048) |
| VI | 0.00 | 0.002 (0.001) | 0.002 (0.003) | NA | NA | −0.062 (0.006) | 0.040 (0.004) |
| IV | 0.50 | 0.023 (0.004) | −0.001 (0.009) | −0.027 (0.004) | −0.002 (0.009) | −0.103 (0.013) | 0.083 (0.080) |
| VII | 0.00 | −0.006 (0.003) | 0.008 (0.005) | 0.018 (0.003) | 0.001 (0.004) | −0.100 (0.012) | 0.089 (0.023) |

[a]In all scenarios, the association parameter $\delta$ is 0.06, the minor allele frequencies of causal gene and of marker are 0.3 and 0.4, respectively, the ratio of dominance effect to additive effect is 0.5, and child additive effect is 1.0. In Scenarios II and V, there are only child effects (i.e., a maternal additive effect of 0.0 and an environmental effect of 0.0), in Scenarios III and VI, there are an environmental effect of 0.5 and a maternal additive effect of 0.0, and in Scenarios IV and VII, there are maternal effects of 0.5 and an environmental effect of 0.5.