American Journal of Epidemiology Advance Access published June 1, 2011

American Journal of Epidemiology

DOI: 10.1093/aje/kwr056

## Practice of Epidemiology

# Detection of Parent-of-Origin Effects for Quantitative Traits in Complete and Incomplete Nuclear Families With Multiple Children

**Feng He, Ji-Yuan Zhou*, Yue-Qing Hu, Fengzhu Sun, Jingyuan Yang, Shili Lin, and Wing Kam Fung**

* Correspondence to Dr. Ji-Yuan Zhou, Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, No. 1838, North Guangzhou Avenue, Guangzhou 510515, China (e-mail: zhoujiyuan@gmail.com).

*Initially submitted March 31, 2010; accepted for publication February 14, 2011.*

For a diallelic genetic marker locus, tests like the parental-asymmetry test (PAT) are simple and powerful for detecting parent-of-origin effects. However, these approaches are applicable only to qualitative traits and thus are currently not suitable for quantitative traits. In this paper, the authors propose a novel class of PAT-type parent-of-origin effects tests for quantitative traits in families with both parents and an arbitrary number of children, which is denoted by Q-PAT(*c*) for some constant *c*. The authors further develop Q-1-PAT(*c*) for detection of parent-of-origin effects when information is available on only 1 parent in each family. The authors suggest the Q-C-PAT(*c*) test for combining families with data on both parental genotypes and families with data on only 1 parental genotype. Simulation studies show that the proposed tests control the empirical type I error rates well under the null hypothesis of no parent-of-origin effects. Power comparison also demonstrates that the proposed methods are more powerful than the existing likelihood ratio test. Although normality is commonly assumed in methods for studying quantitative traits, the tests proposed in this paper do not make any assumption about the distribution of the quantitative trait.

genomic imprinting; quantitative trait loci

Abbreviations: AMM, assortative mating model; DBP, diastolic blood pressure; FS, family sample; LRT, likelihood ratio test; PAT, parental-asymmetry test; PSM, population stratification model; SBP, systolic blood pressure; SNP, single nucleotide polymorphism.

*Genomic imprinting* refers to a genetic phenomenon in which a certain gene is differentially expressed between paternal and maternal alleles, which is an important epigenetic factor in the study of complex traits. Morison et al. (1) constructed an imprinted-gene database (http://igc.otago.ac.nz), which contained 61 records for human genes at the time of this writing. For complex diseases, imprinting effects have been demonstrated in Beckwith-Wiedemann, Prader-Willi, and Angleman syndromes (2, 3). For some other complex diseases, such as autism, diabetes, hereditary paagangliomas, intrauterine growth retardation, neural tube detects, obesity, and schizophrenia, imprinting effects are suspected or hypothesized to play an important role (4–8).

Genomic imprinting has generally been examined in analyses of qualitative and quantitative traits by testing for parent-of-origin effects of alleles prior to fine mapping (9). For qualitative trait loci, tests like the parental-asymmetry test (PAT) are simple and powerful for detecting parent-of-origin effects when there is no maternal effect (10, 11). For quantitative trait loci, there are several methods for detecting parent-of-origin effects. van den Oord (12) suggested a finite mixture model, where class membership is known for complete case-parents trios and unknown for incomplete trios, to test for offspring effects, maternal effects, and parent-of-origin effects. However, it is computationally intensive for the situation where there are multiple children in a family and no standard software is available. Allele-sharing methods, such as variance-components approaches and Haseman-Elston regression, have been extended to test for parent-of-origin effects (13–15). However, these approaches require sampling siblings or extended pedigrees in the analysis. Whittaker et al. (16) introduced some simple linear models to allow for the estimation and testing of parent-of-origin effects, which do not take into account missing data on parental genotypes. Further, all the above-mentioned methods assume that the quantitative traits are

normally distributed, which may be violated in practice. On the other hand, making no assumption about the distribution of quantitative traits, Kistner et al. (17) proposed a likelihood ratio test (LRT), based on quantitative polytomous logistic regression (18), to test for maternal effects and parent-of-origin effects, which considers missing data through an expectation-maximization algorithm. However, it can only accommodate families with 1 child.

In this paper, we propose a novel class of PAT-type parent-of-origin effects tests for quantitative traits in families with both parents and an arbitrary number of children without making any assumption about the distribution of the quantitative trait. This class can accommodate families that have data available on only 1 parental genotype. Extensive simulation studies show that the proposed tests control the empirical type I error rates well under the null hypothesis of no parent-of-origin effects and have good performance in terms of statistical power under the alternative hypothesis compared with existing methods such as LRT (18).

## MATERIALS AND METHODS

### Background and notation

Consider a candidate marker with alleles $M_1$ and $M_2$, where $M_1$ is taken to be the variant allele. Let $F$, $M$, and $C_j$ be the number of copies of allele $M_1$ in the father, mother, and child $j$ of a nuclear family, respectively. Then $F$, $M$, and $C_j$ take possible values of 0, 1, and 2, which represent the genotypes $M_2M_2$, $M_1M_2$, and $M_1M_1$, respectively. To distinguish the parental origin of $M_1$, the following 4 ordered genotypes of the child are used: $M_2/M_2$, $M_2/M_1$, $M_1/M_2$, and $M_1/M_1$, where the left allele is paternal and the right one is maternal. Let $Q$ denote the quantitative trait value of an individual, with the mean value being $\mu_{22}$, $\mu_{21}$, $\mu_{12}$, and $\mu_{11}$ for $M_2/M_2$, $M_2/M_1$, $M_1/M_2$, and $M_1/M_1$, respectively. According to Mendel's law, $\mu_{21} = \mu_{12}$; otherwise, there is a parent-of-origin effect (as in the paper by Weinberg et al. (19)). Specifically, if a larger value of the trait is indicative of a disease, then $\mu_{21} > \mu_{12}$ ($\mu_{21} < \mu_{12}$) represents a paternal (maternal) imprinting effect, or vice versa for an opposite direction of disease and trait value association.

As in earlier work (10, 11), symmetry is assumed across parents within each mating type, that is, $\Pr[F = f, M = m] = \Pr[F = m, M = f]$ for all $f, m = 0, 1, 2$. We further assume that there is no maternal effect. When there are missing data on parental genotypes, we assume that the missingness of a parental genotype is independent of the parent's underlying genotype.

### Methods for use when data on both parents are available

Consider $n$ independent nuclear families, each with known marker genotypes for the father, mother, and children and known quantitative traits for the children. For a child $C$ (with trait value $Q$) and his/her parents $FM$, 15 types of child-parents trio $FMC$ are genetically possible; these are listed in Table 1, together with the notations for the corresponding joint probabilities. For example,

**Table 1.** Classification of All 15 Family Types for Child-Parents Trios, Together With the Notation for the Corresponding Joint Probabilities

| *FMC* | Probability | *FMC* | Probability | *FMC* | Probability |
|---|---|---|---|---|---|
| 212 | $s_1$ | 111 | $s_6$ | 010 | $s_{11}$ |
| 122 | $s_2$ | 110 | $s_7$ | 222 | $s_{12}$ |
| 211 | $s_3$ | 101 | $s_8$ | 201 | $s_{13}$ |
| 121 | $s_4$ | 011 | $s_9$ | 021 | $s_{14}$ |
| 112 | $s_5$ | 100 | $s_{10}$ | 000 | $s_{15}$ |

Abbreviation: *FMC*, father-mother-child.

$s_1 = \Pr[FMC = 212]$ is the probability that a trio falls into the category $F = 2, M = 1, C = 2$. Under the assumption of mating symmetry, we have $s_3 = s_4$, $s_8 = s_9$, and $s_{13} = s_{14}$. Note that under the null hypothesis of no parent-of-origin effects, we have $\mu \triangleq \mu_{21} = \mu_{12}$. Then, for any constant $c$,

$$\begin{aligned} &E\left[(Q-c)\left(I_{F>M,C=1} - I_{F<M,C=1}\right)\right] \\ &\quad = E\left\{E\left[(Q-c)\left(I_{F>M,C=1} - I_{F<M,C=1}\right)\middle| F, M, C=1\right]\right\} \\ &\quad = E\left[(\mu-c)\left(I_{F>M,C=1} - I_{F<M,C=1}\right)\right] \\ &\quad = (\mu-c)[(s_3+s_8+s_{13}) - (s_4+s_9+s_{14})] = 0, \end{aligned}$$

where $I_{\{\text{comparison statement}\}}$ is 1 if the comparison statement holds and 0 otherwise. Specifically, $I_{F>M,C=1} = 1$ if the father carries more copies of allele $M_1$ than the mother and the child is heterozygous, which indicates that the allele $M_1$ in the child came from the father. Similarly, $I_{F<M,C=1} = 1$ signifies that $M_1$ in the child was inherited from the mother. As such, we consider the difference between $I_{F>M,C=1}$ and $I_{F<M,C=1}$ as a contribution to evidence of parent-of-origin effects for trio $FMC$. Therefore, we construct the following test statistic for detecting parent-of-origin effects:

$$s(c) = \sum_{i=1}^{n} \sum_{j=1}^{l_i} (Q_{ij} - c)\left(I_{F_i>M_i,C_{ij}=1} - I_{F_i<M_i,C_{ij}=1}\right),$$

where $F_i$, $M_i$, $C_{ij}$, and $Q_{ij}$ are the genotypes for the father, mother, and child $j$ and the quantitative trait of child $j$ in family $i$, respectively ($i = 1, \ldots, n$; $j = 1, \ldots, l_i$). Under the null hypothesis, we show in Web Appendix 1 (http://aje.oxfordjournals.org/) that the expectation of $s(c)$ is zero and the variance of $s(c)$ can be unbiasedly estimated by

$$\hat{\sigma}^2(c) = \sum_{i=1}^{n} \left[ \sum_{j=1}^{l_i} (Q_{ij} - c)^2 I_{F_i \neq M_i, C_{ij}=1} + 2 \sum_{j<k} (Q_{ij} - c)(Q_{ik} - c) I_{F_i \neq M_i, C_{ij}=1, C_{ik}=1} \right].$$

Then, the class of statistics is given by $\text{Q-PAT}(c) = s(c)/\sqrt{\hat{\sigma}^2(c)}$. Note that only informative families contribute to the Q-PAT($c$). For a trio to be informative, we require

that the child is heterozygous and the parental source of each allele inherited by the child can be unambiguously determined. The Q-PAT($c$) has an approximately standard normal distribution when the number of informative trios is large. Although this null distribution does not depend on the value $c$, the power of Q-PAT($c$) may be different with different $c$ values under the alternative. As in the paper by Sun et al. (20), we choose $c$ to be the mean trait value of all children in the sample, or the mean trait value of interest among the general population if it is known. Based on our simulation studies, we find that the powers of Q-PAT($c$) for these 2 $c$ values are very similar and are higher than those for other $c$ values (results omitted for brevity).

### Methods for use when information on only 1 parent is available

Suppose we have $n_M$ single-mother families (i.e., nuclear families in which information on the father's genotype is not available) and $n_F$ single-father families. For single-mother families, let $M_i$ and $C_{ij}$ be the genotypes of the mother and child $j$ and let $Q_{ij}$ be the trait value of the child ($i = 1, \ldots, n_M; j = 1, \ldots, l_i$). Define $F_i$, $C_{ij}$, and $Q_{ij}$ similarly for single-father families ($i = n_M + 1, \ldots, n_I = n_M + n_F; j = 1, \ldots, l_i$). Then, for any constant $c$, we design the following test statistic for single-parent families:

$$s_1(c) = w\sum_{i=1}^{n_M}\sum_{j=1}^{l_i}(Q_{ij}-c)\left(I_{M_i<C_{ij},\,C_{ij}=1} - I_{M_i>C_{ij},\,C_{ij}=1}\right) + (1-w)\sum_{i=n_M+1}^{n_I}\sum_{j=1}^{l_i}(Q_{ij}-c)\left(I_{F_i>C_{ij},\,C_{ij}=1} - I_{F_i<C_{ij},\,C_{ij}=1}\right),$$

where $w = n_{CF}/(n_{CF}+n_{CM})$, $n_{CM} = \sum_{i=1}^{n_M} l_i$, and $n_{CF} = \sum_{i=n_M+1}^{n_I} l_i$. It is shown in Web Appendix 2 that the expectation of $s_1(c)$ is zero under no parent-of-origin effects. Further, $\hat{\sigma}_1^2(c)$ is an unbiased estimator of the variance of $s_1(c)$ under the null, where

$$\begin{aligned}\hat{\sigma}_1^2(c) = {} & w^2\sum_{i=1}^{n_M}\left[\sum_{j=1}^{l_i}(Q_{ij}-c)^2 I_{M_i\neq C_{ij},\,C_{ij}=1} + 2\sum_{j<k}(Q_{ij}-c)(Q_{ik}-c)I_{M_i\neq 1,\,C_{ij}=1,\,C_{ik}=1}\right]\\ & + (1-w)^2\sum_{i=n_M+1}^{n_I}\left[\sum_{j=1}^{l_i}(Q_{ij}-c)^2 I_{F_i\neq C_{ij},\,C_{ij}=1} + 2\sum_{j<k}(Q_{ij}-c)(Q_{ik}-c)I_{F_i\neq 1,\,C_{ij}=1,\,C_{ik}=1}\right]\\ & + \frac{n_{CF}^2\sum_{i=1}^{n_M} l_i^2 + n_{CM}^2\sum_{i=n_M+1}^{n_I} l_i^2}{n_{CM}n_{CF}(n_{CM}+n_{CF})^2}\sum_{i=1}^{n_M}\sum_{j=1}^{l_i}(Q_{ij}-c)\left(I_{M_i<C_{ij},\,C_{ij}=1} - I_{M_i>C_{ij},\,C_{ij}=1}\right)\\ & \times\sum_{i=n_M+1}^{n_I}\sum_{j=1}^{l_i}(Q_{ij}-c)\left(I_{F_i>C_{ij},\,C_{ij}=1} - I_{F_i<C_{ij},\,C_{ij}=1}\right).\end{aligned}$$

Thus, we propose a new class of statistics Q-1-PAT($c$) $= s_1(c)/\sqrt{\hat{\sigma}_1^2(c)}$ to test for parent-of-origin effects in situations where information on only 1 parent is available in each family. The Q-1-PAT($c$) is asymptotically normally distributed, and $c$ can be chosen similarly as in Q-PAT($c$) for power consideration.

### Method for combining data on complete and incomplete families

Now suppose we have a mixture of $n$ complete families and $n_I$ incomplete families. We propose the following combined statistics to test for parent-of-origin effects:

$$\text{Q-C-PAT}(c) = \frac{s(c) + s_1(c)}{\sqrt{\hat{\sigma}^2(c) + \hat{\sigma}_1^2(c)}},$$

which has an approximately standard normal distribution under the null.

## SIMULATION STUDY

### Settings

We consider 2 population models in our simulation: the population stratification model (PSM) and the assortative mating model (AMM). In PSM, we assume 2 subpopulations with equal proportions and in Hardy-Weinberg equilibrium within each subpopulation, as in the paper by Kistner et al. (17). For the first (second) subpopulation, the allele frequency of $M_1$ is 0.5 (0.1), and the population mean quantitative trait value is 0 (1.5). We assume that the quantitative trait is normally distributed with variance 1 in both subpopulations for simulating the data, although the assumption is not needed in the analysis. The parent-of-origin effects of the marker are simulated by imposing a shift, $\lambda$, on the trait value for the person inheriting a maternal copy of $M_1$. The mean trait values for genotypes $M_2/M_2$, $M_2/M_1$, $M_1/M_2$, and $M_1/M_1$ in the first (second) subpopulation are $-0.5\lambda$, $0.5\lambda$, $-0.5\lambda$, and $-0.5\lambda$ ($1.5 - 0.1\lambda$, $1.5 + 0.9\lambda$, $1.5 - 0.1\lambda$, and $1.5 + 0.9\lambda$), respectively. Note that $\lambda = 0$ means no parent-of-origin effects for the quantitative trait,

which is used to study the size of the proposed tests; other nonzero λ values (ranging from 0.3 to 1.2) are for studying the power.

In AMM, 80% of the families are generated through random mating and the remaining 20% through assortative mating, where the difference between 2 parental trait values is between −0.5 and 0.5. Note that the allele frequency of $M_1$ is taken as 0.5 or 0.1 in the PSM and AMM models. To further investigate how allele frequencies affect the performance of the proposed tests, we also consider some other allele frequencies. Based on 300 families from a homogeneous population, the corresponding sizes of the tests are listed in Web Table 1, and power is shown in Web Figures 1 and 2 (see Web Appendices 3–5).

In the first simulation for Q-PAT($c$), the constant $c$ is taken as the sample mean and the population mean of the quantitative trait. The remaining simulations are all performed with $c$ being the sample mean. When there are missing data on parental genotypes, we use incomplete-family rate τ and father-missing rate β to determine the probability that a family is incomplete (i.e., information is available on only 1 parent) and the probability that the father's information is missing given that a parent's information is missing, respectively (11). To assess the size and power of Q-1-PAT($c$), we consider β in the range of 0.2–0.8 with an increment of 0.1. In the study of the power of Q-C-PAT($c$), we fix β to be 0.5. The incomplete-family rate τ varies within the range 0–1 and an increment of 0.1, unless noted otherwise.

To investigate the effect of family structure on the proposed methods, we utilize 3 types of family samples (FS), each with 300 children: FS1, representing 300 families with 1 child; FS2, representing 150 families with 2 children; and FS3, representing 150 families with 1 child and 75 families with 2 children. For power comparison with other methods, we also consider the family sample type FS4: 500 families with 1 child. For each set of parameter values, we evaluate the empirical size and power by simulation with 10,000 replicates at the significance level $\alpha = 5\%$. Note that not all families in the sample are informative when testing for parent-of-origin effects. On average, there are only 81 and 89 families out of the 300 informing with the PSM and AMM models for family sample FS1, respectively; only 40 (44) families out of 150 are informing with the PSM (AMM) model for FS2; and only 61 (68) families out of 225 for FS3. For FS4 and PSM, approximately 135 families out of 500 are informative.

### Size and power of Q-C-PAT(*c*) with both complete and incomplete families

The size and power results for Q-PAT($c$) and Q-1-PAT($c$) are given in Web Appendices 6 and 7, respectively. Table 2 shows the actual size of Q-C-PAT($c$) for family samples FS1 and FS3 under population models PSM and AMM. The empirical sizes all stay close to the nominal 5% level, signifying the validity of Q-C-PAT($c$) as a test for parent-of-origin effects.

Figure 1 shows the power of Q-C-PAT($c$) for different incomplete-family rates under family samples FS1 and FS3 and population models PSM and AMM, when λ ranges from 0.3 to 1.2 in increments of 0.3, where $\beta = 0.5$. Figure 1 shows that the power of Q-C-PAT($c$) increases when λ increases and the power under FS1 is larger than the power under FS3 for the PSM model. This is consistent with our observation from Web Figures 3 and 4 (FS1 vs. FS2). However, note that the number of persons genotyped for FS1 is larger than that for FS3. In general, the power of Q-C-PAT($c$) decreases when τ increases. However, from the figure, we notice that the power of Q-C-PAT($c$) is somewhat higher when $\tau = 100\%$ compared with when $\tau = 90\%$. This could be because the homogeneous information based on only incomplete families ($\tau = 1$) may result in a lower variability than the heterogeneous information based on the combination of both complete and incomplete families ($\tau = 0.9$).

**Table 2.** Empirical Type I Error Rates (%) of Q-C-PAT($c$) for Different Incomplete-Family Rates τ, With $\beta = 0.5$

| τ | Population Stratification Model | | Assortative Mating Model | |
|---|---|---|---|---|
| | FS1[a] | FS3[b] | FS1 | FS3 |
| 0.0 | 4.80 | 4.92 | 4.87 | 4.64 |
| 0.1 | 4.89 | 5.16 | 4.95 | 4.82 |
| 0.2 | 4.75 | 4.92 | 4.31 | 4.87 |
| 0.3 | 5.03 | 4.85 | 5.02 | 5.37 |
| 0.4 | 4.48 | 5.25 | 5.15 | 5.12 |
| 0.5 | 5.28 | 4.90 | 5.01 | 4.52 |
| 0.6 | 4.64 | 4.78 | 5.11 | 4.59 |
| 0.7 | 4.92 | 4.72 | 5.35 | 5.02 |
| 0.8 | 5.14 | 4.67 | 4.73 | 4.81 |
| 0.9 | 5.14 | 4.53 | 5.05 | 4.94 |
| 1.0 | 4.93 | 4.68 | 4.78 | 5.00 |

Abbreviations: FS, family sample; PAT, parental-asymmetry test.
[a] 300 families with 1 child each.
[b] 150 families with 1 child each and 75 families with 2 children each.

### Power comparison with LRT of Kistner et al.

A handful of other tests have been developed in the literature, and it would be of interest to compare their performances. Note that most of these methods either made the normality assumption regarding the distribution of quantitative traits or did not consider the situation where there are missing parental genotypes. As such, we only make the power comparison of our methods with the LRT method of Kistner et al. (17). Here we use the population stratification model for family samples FS1 and FS4, the same simulation setting as in the paper by Kistner et al. (17). For each family sample, we consider 2 types of data: complete data and combined data (both complete and incomplete families). For the combined data, we fix $\beta = 0.5$. The incomplete-family rate is taken as $\tau = 1/3$ and 0.6 for family samples FS1 and FS4, respectively. Because of the running speed of the

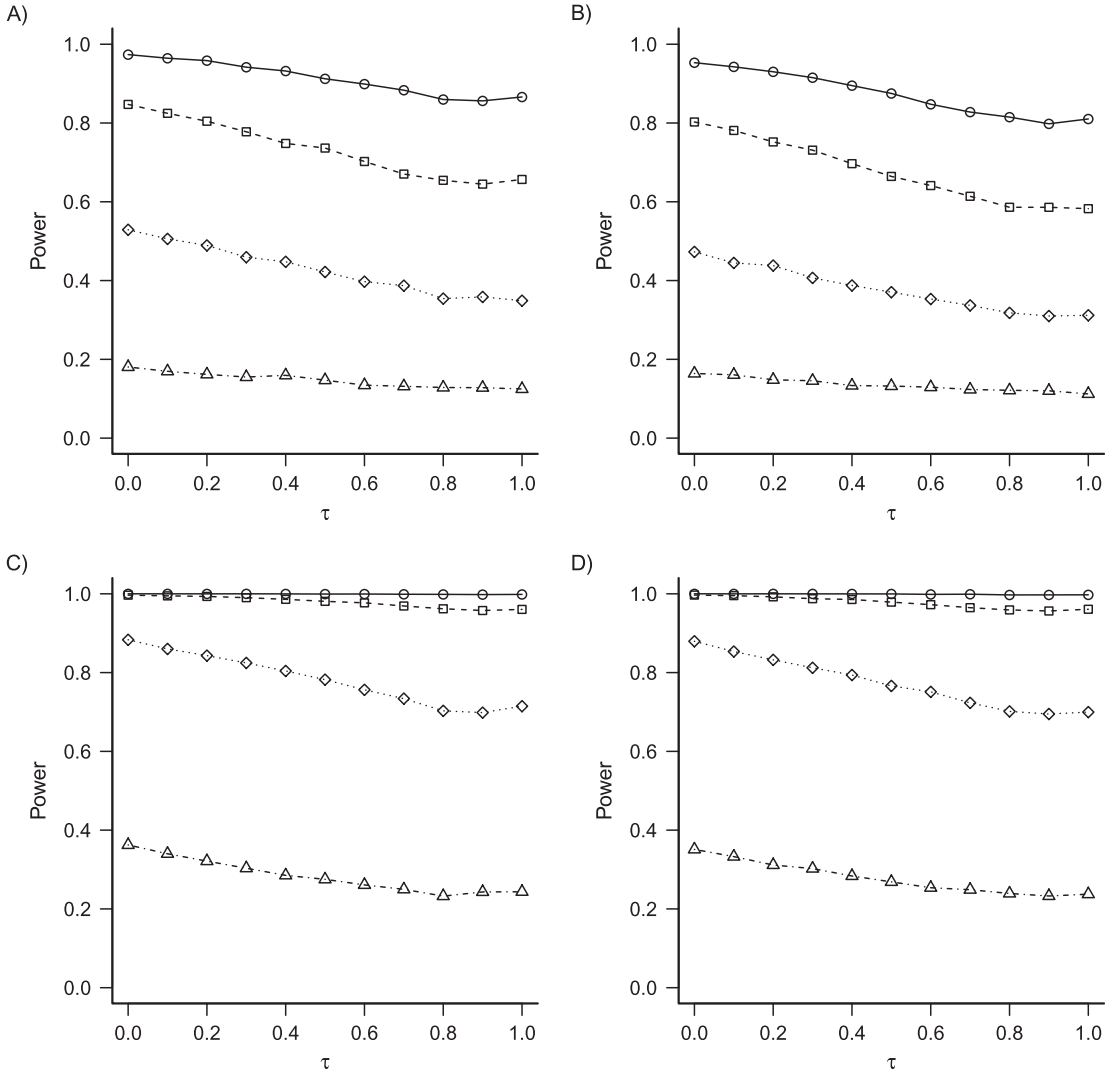


**Figure 1.** Power of the Q-C-PAT(*c*) parental-asymmetry test (PAT) for the incomplete-family rate $\tau$ under 2 types of family samples (FS), FS1 (300 families each with 1 child) and FS3 (150 families each with 1 child and 75 families each with 2 children), and 2 population models, with $\beta = 0.5$. The circles, squares, diamonds, and triangles represent the powers of Q-C-PAT(*c*) when $\lambda$ takes the values 1.2, 0.9, 0.6, and 0.3, respectively. A) FS1 and population stratification model (PSM); B) FS3 and PSM; C) FS1 and assortative mating model (AMM); D) FS3 and AMM.

LRT (program from the Web site of Kistner et al. (17) (http://www.niehs.nih.gov/research/atniehs/labs/bb/docs/maternal1.txt)), we carry out the simulations with 1,000 replicates only.

From the results shown in Table 3, we find that our methods work better in controlling the empirical type I error rates than the LRT. Power comparisons of Q-PAT($c$) and Q-C-PAT($c$) versus LRT are given in Figure 2, which shows that the proposed methods are more powerful than the LRT. Q-PAT($c$) and Q-C-PAT($c$) based on 300 families (FS1) have even better performance than the LRT based on 500 families (FS4). Generally speaking, Q-PAT($c$) and Q-C-PAT($c$) outperform the LRT. A similar finding can be seen in Figure 3 of the paper by Weinberg (21) for the power comparison

**Table 3.** Empirical Type I Error Rates (%) of Q-C-PAT(*c*) in Comparison With the Likelihood Ratio Test Under the Population Stratification Model, With $\beta = 0.5$

| Family Sample | Complete Data | | Combined Data | |
|---|---|---|---|---|
| | LRT | Q-PAT(*c*) | LRT | Q-C-PAT(*c*) |
| FS1[a] | 5.5 | 5.0 | 6.2[b] | 4.9[b] |
| FS4[c] | 5.4 | 5.3 | 6.3[d] | 5.3[d] |

Abbreviations: FS, family sample; LRT, likelihood ratio test; PAT, parental-asymmetry test.

[a] 300 families with 1 child each.

[b] $\tau = 1/3$.

[c] 500 families with 1 child each.

[d] $\tau = 0.6$.

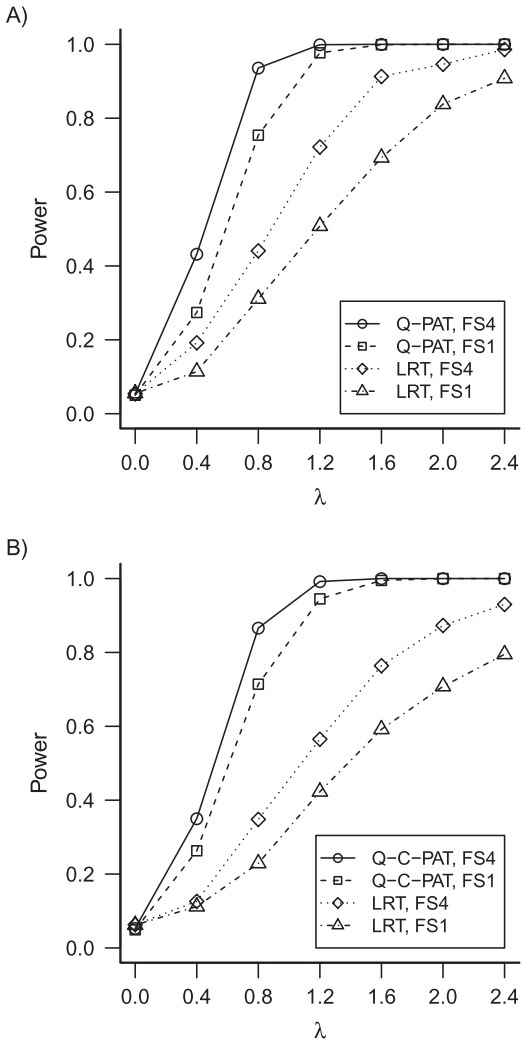


**Figure 2.** Power comparison of the Q-C-PAT(*c*) parental-asymmetry test (PAT) and the likelihood ratio test (LRT) for $\lambda$ based on 2 types of family samples (FS), FS1 (300 families each with 1 child) and FS4 (500 families each with 1 child), under the population stratification model, with $\beta = 0.5$. A) Complete data; B) combined data having $\tau = 1/3$ for FS1 and $\tau = 0.6$ for FS4.

of the transmission disequilibrium test and the LRT for association analysis.

## APPLICATION TO FRAMINGHAM HEART STUDY DATA

We applied Q-C-PAT(*c*) to data from the Framingham Heart Study (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000128.v3.p3), utilizing 2 blood pressure traits. In the Framingham Heart Study, the diastolic blood pressure (DBP) and systolic blood pressure (SBP) of the original cohort (the first generation) and the offspring cohort (the second generation) were measured at 4 different time points, although they were measured only once in the third generation. Therefore, in this application, we took the highest DBP (SBP) measurements among all available ones for each individual as the DBP (SBP) trait value. This choice of the phenotype was made to handle variable numbers of repeated measurements in different cohorts and as an attempt to minimize confounding with high blood pressure medication. Using these phenotypic and genotypic data, we were interested in identifying imprinted genetic variants that showed evidence of association with high blood pressures. After removing nuclear families in which the genotypes of both parents were missing (uninformative for imprinting), we randomly selected 1 nuclear family from each of the 3-generation pedigrees to be included in the analysis. This selection procedure led to approximately 300 nuclear families, with the number of children in each family ranging from 1 to 8. Based on the outcomes of an earlier analysis scanning chromosomes 1–6 for a dichotomized high blood pressure trait (utilizing both DBP and SBP measurements) (22), we decided to apply Q-C-PAT(*c*) to the 20 most significant single nucleotide polymorphisms (SNPs) identified there.

**Table 4.** Q-C-PAT(*c*) Statistics and Corresponding *P* Values for Identified Single Nucleotide Polymorphisms Associated With Blood Pressure Traits in Framingham Heart Study Data[a]

| Variable and Location | Single Nucleotide Polymorphism | Q-C-PAT(*c*) | *P* Value |
|---|---|---|---|
| Diastolic blood pressure | | | |
| Chromosome 2p21 | rs10167883 | 1.955 | 0.050 |
| Chromosome 4q26 | rs3987 | 2.262 | 0.024 |
| Chromosome 4q26 | rs1459543 | 2.183 | 0.029 |
| Chromosome 6p22.3 | rs1951923 | −2.008 | 0.044 |
| Systolic blood pressure | | | |
| Chromosome 3p14.1 | rs1499499 | 2.095 | 0.036 |
| Chromosome 3p24.1 | rs7612518 | 2.095 | 0.036 |
| Chromosome 6p22.3 | rs1951923 | −2.847 | 0.004 |

Abbreviation: PAT, parental-asymmetry test.
[a] Unpublished data from the Framingham Heart Study (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000128.v3.p3).

For the DBP trait, at the 5% significance level, 3 SNPs (rs10167883, rs3987, and rs1459543) were identified as having a significant maternal imprinting effect and 1 SNP (rs1951923) as having a significant paternal imprinting effect. On the other hand, for the SBP trait, we found 2 SNPs (rs1499499 and rs7612518) with a significant maternal imprinting effect, which was different from that found for the DBP trait. The SNP rs1951923 was also found to be paternally imprinted for the SBP trait. The Q-C-PAT(*c*) statistics and corresponding *P* values for the SNPs showing statistical significance for the DBP and SBP traits are given in the upper and lower portions of Table 4, respectively. The findings on the 5 SNPs with maternal imprinting are consistent with those in the paper by Yang and Lin (22) based on the binary high blood pressure trait. However, it appears that the SNP with paternal imprinting is the consequence of maternal effects, as it is known that maternal effects can mimic the pattern of paternal imprinting (9), and in light of the

finding of Yang and Lin (22). This false-positive finding highlights the importance of checking the assumption of no maternal effects before using PAT-type tests. On the other hand, although all of the SNPs identified have been previously implicated as being associated with high blood pressure in humans (see the Genetic Association Studies of Complex Diseases and Disorders section of the Genome Browser (http://genome.ucsc.edu/)), the findings of the imprinting effects are novel in conjunction with those of Yang and Lin (22), confirming the effectiveness of our method.

## DISCUSSION

In this paper, we have successfully extended Weinberg's (10) PAT test for qualitative traits to Q-PAT($c$) to detect parent-of-origin effects for quantitative traits based on families with data on both parents. We further developed the Q-1-PAT($c$) statistics for situations where information is available on only 1 parent in each family. Finally, we proposed the Q-C-PAT($c$) statistics by combining families with data available on both parents and families with data on 1 parent. We assessed the validity and power of our methods under several family sample types, various incomplete-family rates and father-missing rates, and 2 population models. Simulation results showed that the proposed tests control the empirical type I error rates well under the null hypothesis of no parent-of-origin effects and have good statistical power under the alternative hypothesis. Power comparison also demonstrated that our methods are more powerful than the existing LRT. Further, we have successfully applied Q-C-PAT($c$) to the DBP and SBP traits using Framingham Heart Study data to demonstrate the utility of the proposed methods for detecting imprinted genetic variants. However, caution should be exercised in interpreting a paternal imprinting effect, since such findings may not be truly due to paternal imprinting but rather a consequence of maternal effects due to confounding. Our software, Q-C-PAT, implemented in R (R Foundation for Statistical Computing, Vienna, Austria), is freely available at http://www.echobelt.org/web/UploadFiles/qcpat.html.

For a qualitative trait of interest, if we assign a trait value 1 for the affected persons and 0 for the unaffected persons and the constant $c$ is taken as zero, then our proposed tests Q-PAT, Q-1-PAT, and Q-C-PAT reduce to PAT (10), 1-PAT (11), and C-PAT (11), respectively. Therefore, the approaches in this paper unify all of the PAT-type tests irrespective of quantitative or qualitative traits, which is a nice feature of our proposed tests. In addition, although all of the PAT-type tests were developed on the basis of PAT, note that the original PAT is only suitable for qualitative traits and child-parents trios with information available on both parents and a single child, while the proposed Q-C-PAT($c$) is a versatile tool that accommodates both quantitative and qualitative traits and families with both parents, families with only a single parent, and families with arbitrary numbers of children.

A handful of other methods for detecting parent-of-origin effects have been developed in the literature. Most of them make the assumption that the quantitative trait is normally distributed. Moreover, note that we do not make any assumption about the distribution of the trait values. In addition, Kistner et al. (17) proposed the LRT for parent-of-origin effects. However, the LRT is less powerful than the proposed methods when there are no maternal effects. On the other hand, the LRT is valid in the presence of maternal effects, but it can only accommodate families with 1 child. In contrast, our proposed Q-C-PAT($c$) is applicable to a combination of complete and incomplete families with an arbitrary number of children. However, it is not valid as a test for parent-of-origin effects when there are maternal effects, and as such, there may be spurious significance if the assumption of no maternal effects is violated. Finally, the LRT needs the expectation-maximization algorithm to incorporate incomplete trio data and, consequently, it is limited computationally to just a few hundred child-parents trios, whereas our methods are noniterative and thus are more computationally efficient with incomplete data.

The constant $c$ is taken as the mean quantitative trait value of all offspring in the sample, which is also used to contrast with the population mean of the quantitative trait. Other methods for choosing the $c$ value have also been discussed in the literature in the context of linkage analysis (23). Covariates, such as physiologic and environmental variables, may also influence the quantitative trait. As such, a function of such covariates rather than a constant $c$ may be more appropriate in removing the nongenetic effects. However, the underlying functional form is usually unknown, and therefore a reasonable approximation is a regression model on the covariates. To explore whether correcting for the covariate effects adequately may lead to increased power in detecting imprinting, we carried out a small simulation study wherein the underlying covariate model may or may not be correctly specified in our analysis. Our results indeed confirmed that power gains are possible without sacrificing the type I error rates. The results are given in Web Appendix 8.

For the population stratification model in our simulation study, we assume 2 distinct subpopulations with different allele frequencies at the locus in question and different population mean quantitative trait values. Our simulation results show that the size of our tests is correct when both parents from a given family come from the same ancestral subpopulation. Therefore, our method is robust to population stratification. However, when there is population admixture in the sense that the 2 parents are from 2 different subpopulations, the type I error rates may be inflated. To this end, it would be advisable to remove admixed families prior to analysis so that the proposed tests can be safely used.

## ACKNOWLEDGMENTS

Author affiliations: Department of Statistics and Actuarial Science, Faculty of Science, University of Hong Kong, Hong Kong, China (Feng He, Wing Kam Fung); Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, China

(Ji-Yuan Zhou); Institute of Biostatistics, School of Life Sciences, Fudan University, Shanghai, China (Yue-Qing Hu); Molecular and Computational Biology Program, Department of Biological Sciences, Dana and David Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, California (Fengzhu Sun); and Department of Statistics, College of Natural and Mathematical Sciences, Ohio State University, Columbus, Ohio (Jingyuan Yang, Shili Lin).

Feng He and Dr. Ji-Yuan Zhou contributed equally to this work.

The Framingham Heart Study research was supported by National Heart, Lung, and Blood Institute contract 2 N01-HC-25195-06 and by a National Heart, Lung, and Blood Institute contract with Affymetrix, Inc., for genotyping services (contract N02-HL-6-4278). The Genetic Analysis Workshop (GAW) was supported by National Institutes of Health grant R01 GM31575. The GAW16 Framingham data used for the analyses described in this paper were obtained through the Database of Genotypes and Phenotypes (dbGaP numbers 7325-3, 7640-2, 7664-1, 7724-2, and 2243-1). This work was supported by the Competitive Earmarked Research Grant of the Research Grants Council of Hong Kong (grant HKU 702207P); the Research Output Prize Funding of The University of Hong Kong; the National Natural Science Foundation of China (grant 81072386); and the Scientific Research Foundation of Fudan University.

The authors are grateful to the many investigators within the Framingham Heart Study who have collected and managed the data.

Parts of this work were orally presented at the Joint Statistical Meetings in Vancouver, British Columbia, Canada, July 31–August 5, 2010, and at the Eighth ICSA International Conference, Guangzhou, China, December 19–22, 2010.

Conflict of interest: none declared.

## REFERENCES

1. Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res.* 2001; 29(1):275–276.
2. Falls JG, Pulford DJ, Wylie AA, et al. Genomic imprinting: implications for human disease. *Am J Pathol.* 1999;154(3): 635–647.
3. Ziegler A, König IR. *A Statistical Approach to Genetic Epidemiology: Concepts and Applications*. 1st ed. Weinheim, Germany: Wiley-VCH; 2006.
4. Chatkupt S, Lucek PR, Koenigsberger MR, et al. Parental sex effect in spina bifida: a role for genomic imprinting? *Am J Med Genet.* 1992;44(4):508–512.
5. Temple IK, James RS, Crolla JA, et al. An imprinted gene(s) for diabetes? *Nat Genet.* 1995;9(2):110–112.
6. Abel KM. Foetal origins of schizophrenia: testable hypotheses of genetic and environmental influences. *Br J Psychiatry.* 2004;184(5):383–385.
7. Dong C, Li WD, Geller F, et al. Possible genomic imprinting of three human obesity-related genetic loci. *Am J Hum Genet.* 2005;76(3):427–437.
8. Samaco RC, Hogart A, LaSalle JM. Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum Mol Genet.* 2005;14(4):483–492.
9. Hager R, Cheverud JM, Wolf JB. Maternal effects as the cause of parent-of-origin effects that mimic genomic imprinting. *Genetics.* 2008;178(3):1755–1762.
10. Weinberg CR. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am J Hum Genet.* 1999;65(1):229–235.
11. Zhou JY, Hu YQ, Lin S, et al. Detection of parent-of-origin effects based on complete and incomplete nuclear families with multiple affected children. *Hum Hered.* 2009; 67(1):1–12.
12. van den Oord EJ. The use of mixture models to perform quantitative tests for linkage disequilibrium, maternal effects, and parent-of-origin effects with incomplete subject-parent triads. *Behav Genet.* 2000;30(4):335–343.
13. Hanson RL, Kobes S, Lindsay RS, et al. Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *Am J Hum Genet.* 2001;68(4):951–962.
14. Shete S, Amos CI. Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. *Am J Hum Genet.* 2002;70(3):751–757.
15. Shete S, Zhou X, Amos CI. Genomic imprinting and linkage test for quantitative-trait loci in extended pedigrees. *Am J Hum Genet.* 2003;73(4):933–938.
16. Whittaker JC, Gharani N, Hindmarsh P, et al. Estimation and testing of parent-of-origin effects for quantitative traits. *Am J Hum Genet.* 2003;72(4):1035–1039.
17. Kistner EO, Infante-Rivard C, Weinberg CR. A method for using incomplete triads to test maternally mediated genetic effects and parent-of-origin effects in relation to a quantitative trait. *Am J Epidemiol.* 2006;163(3):255–261.
18. Kistner EO, Weinberg CR. Method for using complete and incomplete trios to identify genes related to a quantitative trait. *Genet Epidemiol.* 2004;27(1):33–42.
19. Weinberg CR, Wilcox AJ, Lie RT. 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. *Am J Hum Genet.* 1998;62(4):969–978.
20. Sun FZ, Flanders WD, Yang QH, et al. Transmission/disequilibrium tests for quantitative traits. *Ann Hum Genet.* 2000;64(6):555–565.
21. Weinberg CR. Allowing for missing parents in genetic studies of case-parent triads. *Am J Hum Genet.* 1999;64(4):1186–1193.
22. Yang J, Lin S. Detection of imprinting and heterogeneous maternal effects on high blood pressure using Framingham Heart Study data. *BMC Proc.* 2009;3(suppl 7):S125. doi: 10.1186/1753-6561-3-S7-S125.
23. Elston RC, Buxbaum S, Jacobs KB, et al. Haseman and Elston revisited. *Genet Epidemiol.* 2000;19(1):1–17.