Heredity (2007) 98, 85–91

www.nature.com/hdy

ORIGINAL ARTICLE

A simple method for detection of imprinting effects based on case–parents trios

J-Y Zhou[1,2], Y-Q Hu[1,3] and WK Fung[1]
[1]*Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China;* [2]*Department of Mathematics, Huaihua University, Huaihua, Hunan, China and* [3]*Department of Mathematics, Southeast University, Nanjing, China*

Using data from families in which marker genotypes are known for the father, the mother and the affected offspring, a simple statistic for testing for imprinting effects is developed. The statistic considers whether the expected number of families in which the father carries more copies of a particular marker allele than the mother is equal to the expected number of families in which the mother carries more copies of the allele than the father. The proposed parent-of-origin effects test statistic (POET) is shown to be normally distributed and can be employed to test for imprinting in situations where the marker locus need not be a disease susceptibility locus and where the female and male recombination fractions are sex-specific. A simulation study is conducted to characterize the power of the POET and other properties, and its results show that it is appropriate to employ the POET.
Heredity (2007) **98,** 85–91. doi:10.1038/sj.hdy.6800906; published online 11 October 2006

Keywords: association; genomic imprinting; genotypic relative risk; linkage analysis; power

Introduction

Genomic imprinting, also known as 'parent-of-origin' effects, exists in many mammalian genes. Imprinting has been illustrated in several genetic disorders such as Beckwith–Wiedemann, Prader–Willi and Angleman syndromes (Falls *et al*., 1999). Morison *et al.* (2001) have constructed an imprinted-gene database, which contained more than 230 records at the time of submission (http://igc.otago.ac.nz). Reviews of the mechanisms and function of genomic imprinting have been conducted by Pfeifer (2000), Reik and Walter (2001) and Wilkins and Haig (2003).

As the usual statistical methods for linkage analysis may be ineffective or fail to detect linkage for an imprinted gene, several studies have been performed recently to test for imprinting for qualitative and quantitative traits or to incorporate imprinting effects into existing tests for linkage. Based on case–parents trios, Weinberg (1999) proposed some test statistics to test for imprinting and maternal effects when a marker locus is a disease susceptibility locus (DSL). A likelihood-based method, allowing for incorporation of ascertainment and differential male and female ascertainment probabilities, has been developed for testing for imprinting effects in complex diseases (Haghighi and Hodge, 2002). For the affected sib pairs analysis, Wu *et al.* (2005) have proposed a robust generalized minimax test for imprinting and linkage, based on alleles shared identity-by-descent for the incorporation of imprinting effects. For the quantitative trait locus, mixture models have been used to perform quantitative tests for maternal effects and imprinting effects with incomplete case–parents trios (Van den Oord, 2000). Hanson *et al.* (2001) and Shete and Amos (2002) have used the statistics constructed for linkage between the trait and the marker locus in the variance-components (Amos, 1994) and Haseman and Elston (1972) models, respectively, to test whether the effect of the maternally derived locus is equal to that of the paternally derived locus with sex-specific recombination fractions. Shete *et al.* (2003) have extended the variance-components model to test for imprinting by using extended pedigrees.

Human recombination fraction can differ between males and females. Feenstra *et al.* (2004) used the lod scores to detect sex differences in male–female recombination fractions. The recombination rate for human females is on average 60% higher than that for human males (Fann and Ott, 1995; Broman *et al*., 1998; Ott, 1999; Feenstra *et al*., 2004). In linkage analysis, sex-specific recombination rates will be a consequence of imprinting, and Smalley (1993) suggested that this information could be used to try to identify the traits undergoing imprinting. Hence, it is particularly important in testing for imprinting and conducting linkage analysis with imprinting effects to incorporate the sex-specific recombination fractions into the analysis.

The paper is organized as follows. Based on the case–parents trios, the same data used in the transmission disequilibrium test (TDT) for linkage (Spielman *et al*., 1993), we first propose a novel parent-of-origin effects test (POET) to test for imprinting in situations where the marker locus need not be a DSL *per se* and where the female and male recombination fractions are sex-specific. The POET tests the equality of the expected numbers of two groups of families that are in symmetry in terms of

Correspondence: Professor WK Fung, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China.
E-mail: wingfung@hku.hk
Received 18 March 2006; revised 20 July 2006; accepted 25 August 2006; published online 11 October 2006

the mating type. The test statistic is shown to be asymptotically normal distributed. We then compare the type I error rates of the test statistic simulated by 10 000 replicates with the nominal level of 0.05 for a variety of parameter values, and find that they are highly consistent. The analytical power is derived and is compared with the power obtained by simulation. They are found to be almost the same. The simulation results illustrate that it is appropriate to use the POET.

Methods

Notations and data

Assume that at the marker locus there are two alleles M_1 and M_2 with population frequencies g and $g' = 1-g$, respectively. Note that M_1 and M_2 may represent two groups of alleles. It is convenient to denote the marker genotypes M_2M_2, M_1M_2, and M_1M_1 by the numbers 0, 1, and 2, respectively (i.e. the number of copies of the marker allele M_1 carried by an individual). At a DSL, there are two alleles D and d with population frequencies p and $q = 1-p$, respectively. In order to take imprinting effects into account, it is necessary, for a heterozygote Dd, to specify the parental origin of the disease allele D. We therefore let D/D, D/d, d/D and d/d denote the four ordered genotypes at the DSL, where the left allele of the slash / is paternal and the right allele is maternal. We assume that there is no maternal effect (e.g. intrauterine effect, mitochondrial effect), so the risks for an individual with two copies, one paternal copy, one maternal copy, and no copies of the disease allele D at the DSL are denoted by $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$ and $\phi_{d/d}$, respectively. The population disease prevalence is then $\phi = p^2\phi_{D/D} + pq\phi_{D/d} + pq\phi_{d/D} + q^2\phi_{d/d}$. Without loss of generality, we assume that the risk for an individual with two copies of the disease allele D is greater than for an individual with one copy, either paternal or maternal, and similarly that the risk for an individual with one copy is greater than for one with no copies. $\phi_{D/d} = \phi_{d/d}$ indicates that the paternally derived allele is completely silenced and $\phi_{d/D} = \phi_{d/d}$ indicates that the maternally derived allele is completely silenced. The condition $\phi_{D/d} = \phi_{d/D}$ holds if and only if there is no imprinting effect. $\phi_{D/d} > \phi_{d/D}$ indicates that a paternally derived allele is associated with a greater increase in risk than a maternally derived allele. Similarly, $\phi_{d/D} > \phi_{D/d}$ indicates that a maternally derived allele is associated with a greater increase in risk than a paternally derived allele. Imprinting effects can be measured by the degree of imprinting $I = (\phi_{D/d} - \phi_{d/D})/2$ (Strauch *et al*., 2000), the half of the difference between the expression of the paternally derived allele and that of the maternally derived allele, which is the negative value of the 'delta' measure (Haghighi and Hodge, 2002). It follows that $(\phi_{d/d} - \phi_{D/D})/2 \leqslant I \leqslant (\phi_{D/D} - \phi_{d/d})/2$. Note that here we allow genes to be partially imprinted, although genes are believed to be either completely imprinted or not at all. The case of $I = (\phi_{d/d} - \phi_{D/D})/2$ corresponds to complete paternal imprinting or complete maternal expression and the case of $I = (\phi_{D/D} - \phi_{d/d})/2$ corresponds to complete maternal imprinting or complete paternal expression. So the null hypothesis of no imprinting effects is equivalently expressed as $I = 0$. The coefficient of linkage disequilibrium (LD) between the marker locus and the DSL is taken as $\delta = P_{DM_1} - pg$, where P_{DM_1} is the frequency of haplotype DM_1. Suppose that the marker locus is in LD with the DSL (i.e. $\delta \neq 0$). Let θ_f and θ_m denote, respectively, the female and male recombination fractions between the marker locus and the DSL.

Hardy–Weinberg equilibrium in the parental generation and Mendelian transmission are assumed throughout this paper. There are n independent case–parents trios, each with known marker genotypes FMC, where F, M and C, respectively, denote the genotypes of the father, mother and affected child at the marker locus. Notice that in each case–parents trio, the child's disease phenotype is known (affected) and the marker genotypes of the trio are known. These families are classified jointly by the mating type FM, the paired parental genotypes and the genotype C of the affected child. The mating symmetry is assumed in the sense that $P(F = i, M = j) = P(F = j, M = i)$ for all $i, j = 0, 1, 2$. There are theoretically $3^3 = 27$ types of combinations, but only 15 of these types are genetically possible. Let N_{FMC} denote the number of families that fall into genotype category FMC. These 15 numbers $\{N_{FMC}\}$ follow a multinomial distribution. We are interested in the distribution of such a trio when the trio has been selected because the child is affected. Bayes' theorem is employed to calculate the conditional probability for a trio, given that the child is a case. Table 1 lists all 15 types of family and the corresponding conditional probabilities. Their derivations are outlined in Appendix A. An essential issue is to find out the informative families relevant to genomic imprinting from these 15 types of family.

Proposed statistic and power

In order to effectively exploit the information from trios relevant to imprinting effects, the trios are partitioned into three groups by the mating types FM. The first, second and third groups are characterized by $F > M$, $M > F$ and $F = M$ respectively (see the third column of Table 1). When the father and the mother carry the same number of copies of the marker allele M_1, they have an equal chance of transmitting M_1 to the affected offspring. Let $N_{F>M} = N_{100} + N_{101} + N_{201} + N_{211} + N_{212}$ denote the number of families in the first group and $N_{M>F} = N_{010} + N_{011} + N_{021} + N_{121} + N_{122}$ the number of families in the second group. If there is no imprinting effect, the difference between the numbers of families belonging to these two groups is expected to be zero by symmetry. Furthermore, $N_{F>M} > N_{M>F}$ would provide a clue of paternal expression of the disease allele D and the reverse case $N_{F>M} < N_{M>F}$ would provide a clue of maternal expression of the disease allele D. In fact, we obtain $E(N_{F>M} - N_{M>F}) = 2n\delta(1 - g + g^2)I/\phi$ from Table 1 after tedious algebra. So in the presence of LD, the expected numbers of families belonging to the first two groups are equal if and only if there is no imprinting effect (i.e. $I = 0$), based on the fact that $1 - g + g^2$ is greater than 3/4 whatever the g value. So the test for $I = 0$ essentially tests for the equality of $E(N_{F>M})$ and $E(N_{M>F})$. Hence, we propose the following 'parent-of-origin effects test' (POET) statistic

$$\mathrm{POET} = \frac{N_{F>M} - N_{M>F}}{\sqrt{N_{F>M} + N_{M>F}}} \qquad (1)$$

Table 1 Classification of all 15 genetically feasible family types for nuclear families each with a single affected child, together with the conditional probability of each trio, given that the child is a case

Family type	*FMC*[a]	*Mating type*	*Family number*	*Conditional probability*[b]
1	100	1	N_{100}	$(\phi_{D/D}w_3w_6+\phi_{D/d}w_4w_6+\phi_{d/D}w_3w_8+\phi_{d/d}w_4w_8)/\phi$
2	101	1	N_{101}	$(\phi_{D/D}w_3w_5+\phi_{D/d}w_4w_5+\phi_{d/D}w_3w_7+\phi_{d/d}w_4w_7)/\phi$
3	201	1	N_{201}	$(\phi_{D/D}w_1w_3+\phi_{D/d}w_1w_4+\phi_{d/D}w_2w_3+\phi_{d/d}w_2w_4)/\phi$
4	211	1	N_{211}	$(\phi_{D/D}w_1w_{10}+\phi_{D/d}w_1w_{12}+\phi_{d/D}w_2w_{10}+\phi_{d/d}w_2w_{12})/\phi$
5	212	1	N_{212}	$(\phi_{D/D}w_1w_9+\phi_{D/d}w_1w_{11}+\phi_{d/D}w_2w_9+\phi_{d/d}w_2w_{11})/\phi$
6	010	2	N_{010}	$(\phi_{D/D}w_3w_{10}+\phi_{D/d}w_3w_{12}+\phi_{d/D}w_4w_{10}+\phi_{d/d}w_4w_{12})/\phi$
7	011	2	N_{011}	$(\phi_{D/D}w_3w_9+\phi_{D/d}w_3w_{11}+\phi_{d/D}w_4w_9+\phi_{d/d}w_4w_{11})/\phi$
8	021	2	N_{021}	$(\phi_{D/D}w_1w_3+\phi_{D/d}w_2w_3+\phi_{d/D}w_1w_4+\phi_{d/d}w_2w_4)/\phi$
9	121	2	N_{121}	$(\phi_{D/D}w_1w_6+\phi_{D/d}w_2w_6+\phi_{d/D}w_1w_8+\phi_{d/d}w_2w_8)/\phi$
10	122	2	N_{122}	$(\phi_{D/D}w_1w_5+\phi_{D/d}w_2w_5+\phi_{d/D}w_1w_7+\phi_{d/d}w_2w_7)/\phi$
11	000	3	N_{000}	$(\phi_{D/D}w_3^2+\phi_{D/d}w_3w_4+\phi_{d/D}w_3w_4+\phi_{d/d}w_4^2)/\phi$
12	110	3	N_{110}	$(\phi_{D/D}w_6w_{10}+\phi_{D/d}w_6w_{12}+\phi_{d/D}w_8w_{10}+\phi_{d/d}w_8w_{12})/\phi$
13	111	3	N_{111}	$(\phi_{D/D}w_5w_{10}+\phi_{D/d}w_5w_{12}+\phi_{d/D}w_7w_{10}+\phi_{d/d}w_7w_{12})/\phi+$ $(\phi_{D/D}w_6w_9+\phi_{D/d}w_6w_{11}+\phi_{d/D}w_8w_9+\phi_{d/d}w_8w_{11})/\phi$
14	112	3	N_{112}	$(\phi_{D/D}w_5w_9+\phi_{D/d}w_5w_{11}+\phi_{d/D}w_7w_9+\phi_{d/d}w_7w_{11})/\phi$
15	222	3	N_{222}	$(\phi_{D/D}w_1^2+\phi_{D/d}w_1w_2+\phi_{d/D}w_1w_2+\phi_{d/d}w_2^2)/\phi$

[a]*F*, *M* and *C* denote the genotypes of the father, mother and child, respectively.
[b]$w_1=(pg+\delta)g$, $w_2=(qg-\delta)g$, $w_3=(pg'-\delta)g'$, $w_4=(qg'+\delta)g'$, $w_5=pgg'+g'\delta-\delta\theta_m$, $w_6=pgg'-g\delta+\delta\theta_m$, $w_7=qgg'-g'\delta+\delta\theta_m$, $w_8=qgg'+g\delta-\delta\theta_m$, $w_9=pgg'+g'\delta-\delta\theta_f$, $w_{10}=pgg'-g\delta+\delta\theta_f$, $w_{11}=qgg'-g'\delta+\delta\theta_f$, $w_{12}=qgg'+g\delta-\delta\theta_f$, $\phi=p^2\phi_{D/D}+pq\phi_{D/d}+pq\phi_{d/D}+q^2\phi_{d/d}$.

where $N_{F>M}+N_{M>F}$ is the minimum variance unbiased estimator for the variance of $N_{F>M}-N_{M>F}$. Notice that the (square of the) POET is a McNemar's test and has a similar form as the TDT (Spielman *et al*., 1993).

Based on the theorem in Appendix B, we can derive the asymptotic distribution of the POET. This is $N(\mu,\sigma^2)$, where

$$\mu=\sqrt{n}\,\frac{e_1-e_2}{\sqrt{e_1+e_2}} \tag{2}$$

$$\sigma^2=1-\frac{(e_1-e_2)^2}{4(e_1+e_2)}-\frac{3(e_1-e_2)^2}{4(e_1+e_2)^2} \tag{3}$$

e_1 and e_2 are the sums of the conditional probabilities of mating types 1 and 2 (see the last column of Table 1), respectively, $e_1-e_2=2\delta(1-g+g^2)I/\phi$, $e_1+e_2=2g(1-g)(2-3g+3g^2)+2\delta\Delta(1-2g)(1-3g+3g^2)-2R\delta^2(1-3g+3g^2)/\phi$, $R=\phi_{D/D}-\phi_{D/d}-\phi_{d/D}+\phi_{d/d}$ is the difference between the sum of two homozygote risks and the sum of two heterozygote risks and $\Delta=(p(2\phi_{D/D}-\phi_{D/d}-\phi_{d/D})+q(\phi_{D/d}+\phi_{d/D}-2\phi_{d/d}))/(2\phi)$ is the difference between two ratios $P(D\,|\,\text{affected child})/P(D\,|\,\text{random man})$ and $P(d\,|\,\text{affected child})/P(d\,|\,\text{random man})$, where $P(D\,|\,\text{affected child})$ represents the probability that a chromosome of an affected child has an allele D at the DSL, and the other three probabilities $P(D\,|\,\text{random man})$, $P(d\,|\,\text{affected child})$ and $P(d\,|\,\text{random man})$ are similarly defined.

It is obvious that under the null hypothesis of no imprinting, we have $\mu=0$ and $\sigma^2=1$. So the corresponding rejection region about no genomic imprinting is $|\text{POET}|>z_{\alpha/2}$, where α is the significance level and $z_{\alpha/2}$ is the upper $\alpha/2$ point of a standard normal distribution. In this paper, the significance level is taken as 0.05 and $z_{0.025}=1.96$. It follows immediately that the asymptotic power of the POET is

$$\text{Power}=\Phi\left(\frac{-z_{\alpha/2}-\mu}{\sigma}\right)+1-\Phi\left(\frac{z_{\alpha/2}-\mu}{\sigma}\right) \tag{4}$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable. The asymptotic normality of the POET and the accuracy of the power Equation (4) are assessed via simulations.

Remarks

It is obvious from the power expression (4) that there are numerous parameters that affect the power of the POET. Let $\gamma_2=\phi_{D/D}/\phi_{d/d}$, $\gamma_{1p}=\phi_{D/d}/\phi_{d/d}$, $\gamma_{1m}=\phi_{d/D}/\phi_{d/d}$ denote the genotypic relative risks (Risch and Merikangas, 1996), and $\gamma_1=(\gamma_{1p}+\gamma_{1m})/2$ denote the average of the two heterozygous genotypic relative risks, then $\Delta=(p\gamma_2+(q-p)\gamma_1-q)/(p^2\gamma_2+2pq\gamma_1+q^2)$, $I/\phi=(\gamma_{1p}-\gamma_{1m})/(2(p^2\gamma_2+2pq\gamma_1+q^2))$, $R/\phi=(\gamma_2-2\gamma_1+1)/(p^2\gamma_2+2pq\gamma_1+q^2)$. It is concluded from Equation (4) that the power depends on the four risks $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$, $\phi_{d/d}$ only through γ_2, γ_{1p} and γ_{1m}. So we fix the background risk $\phi_{d/d}$ at 0.05 in our simulation study. Notice that $\gamma_{1p}=1$ means complete silencing of the paternally derived allele and $\gamma_{1m}=1$ means complete silencing of the maternally derived allele.

It is noted that the power of the POET is independent of the designation of the two marker alleles M_1 and M_2. In fact, if we interchange M_1 and M_2, the coefficient of LD changes from δ to $-\delta$. So, from the expressions of μ and σ^2 in Equations (2) and (3), we find that the mean of the POET changes from μ to $-\mu$ while σ^2 remains unchanged. Therefore, the power of the POET is invariant when the two marker alleles M_1 and M_2 are interchanged. As a result, we need only to consider the case where the marker allele frequency g is less than 0.5 in the simulation study.

Notice from Equations (2) and (3) that the analytical power of the POET is symmetric about the degree of imprinting I. Interpreted precisely, for the parameter vector $(\phi_{D/d},\ \phi_{d/D})$, the power evaluated at $\phi_{D/d}=a$, $\phi_{d/D}=b$ is just the same as that evaluated at $\phi_{D/d}=b$, $\phi_{d/D}=a$ while all other parameters remain unchanged. In particular, we conclude that the silencing of the paternally derived allele and the complementary silencing of the maternally derived allele are equally easy or equally difficult to detect. We therefore need only to assess the performance of the POET in the case of $I>0$ in our simulation study.

Simulation methods

In order to assess the performance of the POET in finite samples, a number of parameter values are chosen for the simulation study, based on the following principles. The parameters γ_2 and p are chosen as follows: $\gamma_2 = 2, 4, 8, 16$ and $p = 0.01, 0.1, 0.5, 0.8$ (Risch and Merikangas, 1996; Knapp, 1999; Deng and Chen, 2001). When γ_2 is given, we take $\gamma_1 = (3+\gamma_2)/4$, $(1+\gamma_2)/2$, $(1+3\gamma_2)/4$, equally spaced in the range of 1 and γ_2. For a given γ_1, the range of the degree of imprinting I according to the diamond of inheritance (Strauch *et al*., 2000) can be derived as follows: when $\gamma_1 = (3+\gamma_2)/4$ or $(1+3\gamma_2)/4$, $\phi_{d/d}(1-\gamma_2)/4 \leqslant I \leqslant \phi_{d/d}(\gamma_2-1)/4$; when $\gamma_1 = (1+\gamma_2)/2$, $\phi_{d/d}(1-\gamma_2)/2 \leqslant I \leqslant \phi_{d/d}(\gamma_2-1)/2$. The coefficient of LD δ is taken as $0.9\delta_{\max}$ (Deng and Chen, 2001), where $\delta_{\max} = \min(p(1-g), g(1-p))$. Remember that the analytical power of the POET is independent of the designation of the two marker alleles. So we choose $g = 0.2$ or 0.4. For illustrative purposes, we adopt $\theta_f = 0.146$ and $\theta_m = 0.084$, which are the female and male recombination fractions between ABO and the locus for the nail–patella syndrome (NPS1) respectively (Ott, 1999). Once the parameter values and the sample size n are chosen, we randomly generate n trios according to the probability distribution listed in the last column of Table 1. For a given number of replicates (e.g. 10 000 replicates), the n trios are generated with 10 000 replicates, and so it follows 10 000 values of the POET. For the given significance level α, the actual power/type I error rate is then estimated as the proportion of rejecting the null hypothesis (i.e. $|\mathrm{POET}| > z_{\alpha/2}$), in 10 000 replicates performed when the alternative/null hypothesis holds.

Results

For a given significance level $\alpha = 0.05$, Table 2 exhibits the actual type I error rates of the POET with 10 000 replicates for a variety of parameter values, where the sample size $n = 100$, $g = 0.2$, $\theta_f = 0.146$ and $\theta_m = 0.084$. The entries in Table 2 show that the actual type I error rates are highly consistent with the nominal value of 0.05, which demonstrates the high accuracy of the asymptotic normality result of the POET under the null hypothesis of no imprinting. We also investigate the performance of the POET when the female and male recombination fractions take other values, when $g = 0.4$, and when the sample size $n = 200$ or 400. It is found that the corresponding type I error rates of the POET are again very close to the nominal value of 0.05 (results not shown here for brevity). All these results indicate that the POET controls the size very well, even though the sample size is not large.

Based on the power calculation formula (4) of the POET, we first obtain the sample size needed for the test to gain 80% analytical power, when there is imprinting. For this particular sample size, we then evaluate the actual power by simulation with 10 000 replicates. Table 3 lists the required sample sizes and the actual powers associated with them. It is observed that the sample sizes needed for the test to gain 80% power are very large in some scenarios but quite modest in others. However, for common diseases with large γ_2, the sample sizes needed to gain 80% power are small to moderate and so it is practical to use this approach. We also conducted the simulation with marker allele frequency $g = 0.4$ and with the other parameter values the same as those given in Table 3 (results omitted for brevity). All the simulation results show a very strong agreement between the analytical powers and the actual powers. So the power calculation Equation (4) appears very accurate.

The sample sizes are chosen respectively as 100, 200 and 400 to exhibit the effect of the degree of imprinting

Table 2 Type I error rates of the POET at a significance level of 0.05 for simulation with no imprinting, having the marker allele frequency 0.2 and the sample size 100

		P			
γ_2	γ_1	*0.01*	*0.1*	*0.5*	*0.8*
2	1.25	0.050	0.045	0.051	0.046
	1.50	0.051	0.052	0.050	0.051
	1.75	0.048	0.051	0.050	0.045
4	1.75	0.050	0.048	0.052	0.049
	2.50	0.051	0.048	0.051	0.051
	3.25	0.048	0.050	0.051	0.045
8	2.75	0.051	0.047	0.051	0.050
	4.50	0.047	0.046	0.047	0.051
	6.25	0.052	0.049	0.050	0.051
16	4.75	0.055	0.048	0.048	0.049
	8.50	0.051	0.051	0.048	0.052
	12.25	0.052	0.051	0.048	0.050

Table 3 Sample sizes required to gain 80% power of the POET according to the asymptotical power calculation formula and the actual powers associated with these sample sizes, having the marker allele frequency 0.2

γ_2	I	γ_1	P			
			0.01	*0.1*	*0.5*	*0.8*
2	0.0125	1.25	422680 (0.804)	4741 (0.804)	5228 (0.802)	50223 (0.796)
	0.0250	1.50	106954 (0.800)	1312 (0.795)	1557 (0.801)	13697 (0.797)
	0.0125	1.75	433004 (0.802)	5805 (0.799)	7340 (0.799)	59571 (0.798)
4	0.0375	1.75	48125 (0.805)	662 (0.803)	1430 (0.802)	19029 (0.801)
	0.0750	2.50	12465 (0.800)	213 (0.800)	493 (0.801)	5468 (0.808)
	0.0375	3.25	51647 (0.804)	1092 (0.799)	2620 (0.804)	24937 (0.795)
8	0.0875	2.75	9272 (0.800)	178 (0.797)	781 (0.799)	12842 (0.801)
	0.1750	4.50	2509 (0.799)	70 (0.803)	297 (0.804)	3799 (0.799)
	0.0875	6.25	10850 (0.797)	427 (0.797)	1708 (0.803)	17768 (0.797)
16	0.1875	4.75	2213 (0.797)	70 (0.793)	576 (0.797)	10707 (0.797)
	0.3750	8.50	647 (0.804)	34 (0.816)	233 (0.792)	3216 (0.803)
	0.1875	12.25	3014 (0.803)	246 (0.798)	1398 (0.801)	15240 (0.800)

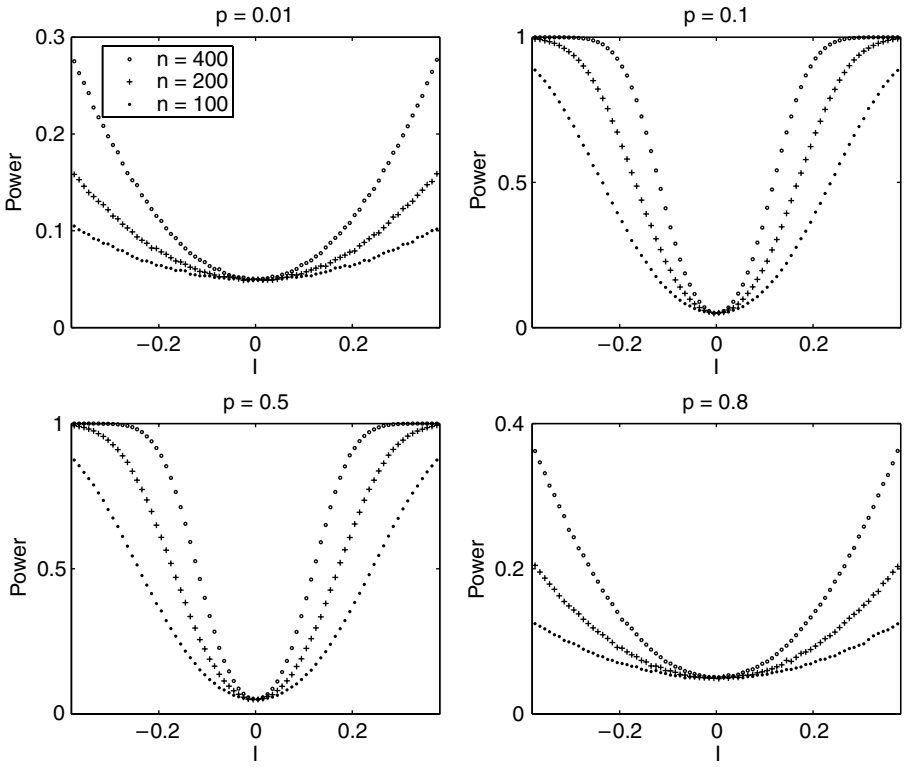

Figure 1 The power function of the POET obtained by simulation with 100 000 replicates, having $\gamma_2 = 16$, $\gamma_1 = (1+\gamma_2)/2$, $\theta_f = 0.146$, $\theta_m = 0.084$, $\delta = 0.9\delta_{\max}$, $g = 0.4$ and $\phi_{d/d} = 0.05$ with different choices of $p = 0.01$, 0.1, 0.05, 0.8 and the sample size $n = 100$, 200, 400.

I on the actual power of the POET via simulation with 100 000 replicates. Figure 1 depicts the actual power function of the POET when the degree of imprinting changes from the leftmost point $\phi_{d/d}(1-\gamma_2)/2$ to the rightmost point $\phi_{d/d}(\gamma_2-1)/2$. The power at $I=0$ (i.e. under the null hypothesis) is very close to 0.05, as expected. The symmetry of the power function about $I=0$ (as mentioned earlier) is confirmed in Figure 1.

As expected, we observe that the power is a monotonically increasing function of the sample size n or the absolute value of the degree of imprinting $|I|$. The rightmost point of I, $\phi_{d/d}(\gamma_2-1)/2$, corresponding to $\gamma_{1p}=\gamma_2$ and $\gamma_{1m}=1$, or equivalently $\phi_{D/d}=\phi_{D/D}$ and $\phi_{d/D}=\phi_{d/d}$, means that the disease allele D is complete maternal imprinting or complete paternal expression. In this situation (and also at the leftmost point of I), the difference between the expression of the paternally derived allele and that of the maternally derived allele reaches its greatest extent. This kind of differential expression should be the easiest one to detect. The upper right part of Figure 1 shows that in our case, for moderate disease allele frequency $p=0.1$, the modest imprinting effects are easy to detect.

In addition to the simulation studies given above, we would also like to investigate the power of the POET for a small to medium sample size of 100 or 200, with parameters taking the following values: the disease allele frequency $p\in[0.05, 0.35]$, the marker allele frequency $g\in[0.1, 0.5]$, $\phi_{d/d}\in[0, 0.1]$, $\phi_{D/D}\in[0.8, 1]$, and the degree of imprinting $I=(\phi_{D/D}-\phi_{d/d})/2$. More precisely, we calculate the power of the POET for p with increments of 0.01, g with 0.1, $\phi_{d/d}$ with 0.02, and $\phi_{D/D}$ with 0.04. Thus, we have $31\times 5=155$ powers of the POET for given $\phi_{d/d}$ and $\phi_{D/D}$. It is observed from the results that the bigger the difference between $\phi_{D/D}$ and $\phi_{d/d}$, the more powerful the POET, assuming all other parameter values are fixed. The mean and standard deviation of those 155 powers with given $\phi_{d/d}$ and $\phi_{D/D}$ can be obtained accordingly. For a sample size $n=100$, when $\phi_{d/d}=0.1$ and $\phi_{D/D}=0.8$, the corresponding mean and standard deviation are 80.13 and 20.08% respectively. When $\phi_{d/d}=0.0$ and $\phi_{D/D}=0.8$, the corresponding values are 98.70 and 4.90%. For a bigger sample size $n=200$, the power increases considerably. For example, when the sample size $n=200$, $\phi_{d/d}=0.1$ and $\phi_{D/D}=0.8$, the mean and standard deviation of the corresponding powers of the POET are 93.70 and 12.08%, respectively. In summary, for the parameter values listed in the beginning of this paragraph, the power of the POET is relatively high. Notice that based on our Equation (4), it is possible to know the power of the POET for a given sample size and set of parameter values.

Discussion

The POET was developed to test for imprinting when association between the marker locus and a DSL is present. Furthermore, from the expression of $E(N_{F>M}-N_{M>F})$, we know that it is also appropriate to use the POET to test for association if there is imprinting. The POET can therefore be thought of as a test for imprinting in the presence of known association, or as a test for both imprinting and association.

The primary benefit of the POET is that it is appropriate in situations where the marker locus need not be the DSL *per se* and where the female and male

recombination fractions are sex-specific. One attractive feature of the POET is that it uses the same marker genotype data (from affected children and their parents) collected for testing for linkage by the TDT, and so no extra data are needed. In fact, the TDT uses the families in which there is at least one heterozygous parent at the marker locus and tests for the equality of the expected number of times of the marker allele M_1 transmitted and the expected number of times not transmitted from the heterozygous parents to the affected children. The POET uses the families in which the two parents' marker genotypes are different and tests for the equality of the expected number of families in which the father carries more copies of the marker allele M_1 than the mother and the expected number of families in which the mother carries more copies of the marker allele M_1 than the father. Our POET uses data of 10 family types of Table 1, namely types 1–10, while the TDT uses 11 types (1, 2, 4, 5, 6, 7, 9, 10, 12, 13, and 14). The TDT has been commonly used and is a powerful tool for testing for linkage when the case–parents trios are available. We believe that the POET would be an equally effective way of testing for imprinting. These two tests are employed for different purposes, but both are simple and easy to use. Moreover, the POET, like the TDT, is also applicable to the population not in Hardy–Weinberg equilibrium (simulation results are omitted for brevity), for example, the population stratification demographic model.

The statistical methods for testing for linkage incorporating imprinting effects generally have greater power than the usual statistical methods ignoring imprinting effects, if the gene is actually imprinted (Wu *et al*., 2005). So it is important to test if the trait of interest is imprinted. Here we propose a simple test, using the case–parents trios, to detect imprinting effects. Many of the known imprinted genes influence fetal development, so incorporating imprinting effects into linkage analysis may be particularly useful for study of development-related traits such as birth weight (Bartolomei and Tilghman, 1997; Hanson *et al*., 2001). Notice that the POET is not valid when there are maternal effects (e.g. effects mediated through the intrauterine environment by the mother). In a future study, we will focus on the improvement of the TDT when there is a prior belief of imprinting for the trait of interest, and the development of test statistics for testing for imprinting effects in the presence of maternal effects.

Acknowledgements

We are grateful for the valuable comments made by the editor and two reviewers on an early draft of this paper. We thank David Wilmshurst for editorial assistance. This work was partially supported by a CRCG Seed Grant of The University of Hong Kong and funding from the National Natural Science Foundation of China (10329102, 10561008) and the Scientific Research Fund of Huaihua University.

References

Amos CI (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet* **54**: 535–543.

Bartolomei MS, Tilghman SM (1997). Genomic imprinting in mammals. *Annu Rev Genet* **31**: 493–525.

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* **63**: 861–869.

Deng HW, Chen WM (2001). The power of the transmission disequilibrium test (TDT) with both case–parent and control–parent trios. *Genet Res* **78**: 289–302.

Falls JG, Pulford DJ, Wylie AA, Jirtle RL (1999). Genomic imprinting: implications for human disease. *Am J Pathol* **154**: 635–647.

Fann CSJ, Ott J (1995). Parsimonious estimation of sex-specific map distances by stepwise maximum likelihood regression. *Genomics* **29**: 571–575.

Feenstra B, Greenberg DA, Hodge SE (2004). Using lod scores to detect sex differences in male–female recombination fractions. *Hum Hered* **57**: 100–108.

Haghighi F, Hodge SE (2002). Likelihood formulation of parent-of-origin effects on segregation analysis, including ascertainment. *Am J Hum Genet* **70**: 142–156.

Hanson RL, Kobes S, Lindsay RS, Knowler WC (2001). Assessment of parent-of-origin effects in linkage analysis of quantitative traits. *Am J Hum Genet* **68**: 951–962.

Haseman JK, Elston RC (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* **2**: 3–19.

Knapp M (1999). A note on power approximations for the transmission/disequilibrium test. *Am J Hum Genet* **64**: 1177–1185.

Morison IM, Paton CJ, Cleverley SD (2001). The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* **29**: 275–276.

Ott J (1999). *Analysis of Human Genetic Linkage*, 3rd edn. The Johns Hopkins University Press: Baltimore and London.

Pfeifer K (2000). Mechanisms of genomic imprinting. *Am J Hum Genet* **67**: 777–787.

Rao CR (1973). *Linear Statistical Inference and Its Applications*. John Wiley: New York.

Reik W, Walter J (2001). Genomic imprinting: parental influence on the genome. *Nat Rev Genet* **2**: 21–32.

Risch N, Merikangas K (1996). The future of genetic studies of complex human diseases. *Science* **273**: 1516–1517.

Shete S, Amos CI (2002). Testing for genetic linkage in families by a variance-components approach in the presence of genomic imprinting. *Am J Hum Genet* **70**: 751–757.

Shete S, Zhou X, Amos CI (2003). Genomic imprinting and linkage test for quantitative-trait loci in extended pedigrees. *Am J Hum Genet* **73**: 933–938.

Smalley SL (1993). Sex-specific recombination frequencies: a consequence of imprinting? *Am J Hum Genet* **52**: 210–212.

Spielman RS, McGinnis RE, Ewens WJ (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* **52**: 506–516.

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. *Am J Hum Genet* **66**: 1945–1957.

Van den Oord EJCG (2000). 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* **30**: 335–343.

Weinberg CR (1999). Methods for detection of parent-of-origin effects in genetic studies of case–parents triads. *Am J Hum Genet* **65**: 229–235.

Wilkins JF, Haig D (2003). What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* **4**: 359–368.

Wu CC, Shete S, Amos CI (2005). Linkage analysis of affected sib pairs allowing for parent-of-origin effects. *Ann Hum Genet* **69**: 113–126.

Appendix A

Derivation of the conditional probabilities in Table 1

The first family type 100 in Table 1 is taken as an example to illustrate how the conditional probabilities listed in the last column of Table 1 are derived. In order to find the conditional probability $P(FMC=100\,|\,\text{child is affected})$, it is necessary to calculate $P(F=M_1M_2,\ M=M_2M_2,\ C=M_2M_2,\ \text{child is affected})$, which is the sum of the following four terms by the law of total probability: $P(F=M_1M_2,\ M=M_2M_2,\ C=DM_2/DM_2)\phi_{D/D}$, $P(F=M_1M_2,\ M=M_2M_2,\ C=DM_2/dM_2)\phi_{D/d}$, $P(F=M_1M_2,\ M=M_2M_2,\ C=dM_2/DM_2)\phi_{d/D}$, $P(F=M_1M_2,\ M=M_2M_2,\ C=dM_2/dM_2)\phi_{d/d}$ where the left haplotype of the slash / is paternal and the right one is maternal. The first term (ignoring factor $\phi_{D/D}$) can be further written as a product of two probabilities, $P(F=M_1M_2,\ C_P=DM_2)P(M=M_2M_2,\ C_M=DM_2)$, where C_P and C_M denote respectively the paternal and maternal haplotypes at the marker locus and the DSL. For the first probability, considering all possible ordered genotypes at the marker locus and the DSL and seeing whether there is recombination between these two loci, we have

$$\begin{aligned}
&P(F=M_1M_2, C_P=DM_2)\\
&=P(F=DM_1/DM_2, C_P=DM_2)\\
&\quad+P(F=DM_2/DM_1, C_P=DM_2)\\
&\quad+P(F=DM_1/dM_2, C_P=DM_2)\\
&\quad+P(F=dM_1/DM_2, C_P=DM_2)\\
&\quad+P(F=DM_2/dM_1, C_P=DM_2)\\
&\quad+P(F=dM_2/DM_1, C_P=DM_2)\\
&=P_{DM_1}P_{DM_2}\left(\frac{\theta_m}{2}+\frac{1-\theta_m}{2}\right)+P_{DM_2}P_{DM_1}\left(\frac{\theta_m}{2}+\frac{1-\theta_m}{2}\right)\\
&\quad+P_{DM_1}P_{dM_2}\frac{\theta_m}{2}\\
&\quad+P_{dM_1}P_{DM_2}\frac{1-\theta_m}{2}+P_{DM_2}P_{dM_1}\frac{1-\theta_m}{2}\\
&\quad+P_{dM_2}P_{DM_1}\frac{\theta_m}{2}\\
&=gP_{DM_2}+\theta_m(P_{DM_1}P_{dM_2}-P_{dM_1}P_{DM_2})
\end{aligned}$$

which is just w_6 listed in the footnote of Table 1 after some transformations. The second probability is just w_3 by the same principle. So $P(F=M_1M_2,\ M=M_2M_2,\ C=DM_2/DM_2)=w_3w_6$. Similarly, we have $P(F=M_1M_2,\ M=M_2M_2,\ C=DM_2/dM_2)=w_4w_6$, $P(F=M_1M_2,\ M=M_2M_2,\ C=dM_2/DM_2)=w_3w_8$ and $P(F=M_1M_2,\ M=M_2M_2,\ C=dM_2/dM_2)=w_4w_8$. Finally, we have the expression of the conditional probability $P(FMC=100\,|\,\text{child is affected})$.

Appendix B

Proof for asymptotic normality of the POET

Theorem *Let* $u=(u_1,\ \ldots,\ u_m)^{\mathrm{T}}$, $v=(v_1,\ \ldots,\ v_m)^{\mathrm{T}}$, $u_iv_i=0$, $u_i^2=u_i$, $v_i^2=v_i$ (i.e. $u_i=0$ *or* 1 *and* $v_i=0$ *or* 1), *for any* i, $(N_1,\ \cdots,\ N_m,\ N_{m+1})^{\mathrm{T}}$ *be a multinomially distributed random variable of size* $n=\sum_{i=1}^{m+1}N_i$ *and with parameters* $(s_1,\ldots,\ s_m,\ s_{m+1})$, *where* $s_{m+1}=1-\sum_{i=1}^{m}s_i$. *Denote* $N=(N_1,\ \ldots,\ N_m)^{\mathrm{T}}$, *then the distribution of* $(u^{\mathrm{T}}N-v^{\mathrm{T}}N)/\sqrt{u^{\mathrm{T}}N+v^{\mathrm{T}}N}$ *tends to the normal distribution with mean* $\sqrt{n}(a_1-a_2)/\sqrt{a_1+a_2}$ *and variance* $1-(a_1-a_2)^2/(4(a_1+a_2))-3(a_1-a_2)^2/(4(a_1+a_2)^2)$, *where* $a_1=u^{\mathrm{T}}s$, $a_2=v^{\mathrm{T}}s$, *and* $s=(s_1,\ \ldots,\ s_m)^{\mathrm{T}}$.

Proof Note that N/n is the maximum likelihood estimator of the parameter vector s and from the asymptotic normality of the maximum likelihood estimator, we have in law

$$\sqrt{n}\left(\frac{N}{n}-s\right)\to N\left(0,\sum\right)$$

where $\sum=(\sigma_{ij})_{m\times m}$, $\sigma_{ii}=s_i(1-s_i)$ and $\sigma_{ij}=-s_is_j$ for $i\neq j$ (Rao, 1973). It follows immediately that

$$\sqrt{n}\left(\begin{pmatrix}\frac{u^{\mathrm{T}}N}{n}\\ \frac{v^{\mathrm{T}}N}{n}\end{pmatrix}-\begin{pmatrix}u^{\mathrm{T}}s\\ v^{\mathrm{T}}s\end{pmatrix}\right)\to N(0,(u,v)^{\mathrm{T}}\sum(u,v))$$

Let

$$g\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\frac{x_1-x_2}{\sqrt{x_1+x_2}}$$

then we have

$$\begin{aligned}
&\sqrt{n}\left(g\begin{pmatrix}\frac{u^{\mathrm{T}}N}{n}\\ \frac{v^{\mathrm{T}}N}{n}\end{pmatrix}-g\begin{pmatrix}u^{\mathrm{T}}s\\ v^{\mathrm{T}}s\end{pmatrix}\right)\\
&\quad\to N\left(0,\frac{\partial g}{\partial x^{\mathrm{T}}}(u,v)^{\mathrm{T}}\sum(u,v)\frac{\partial g}{\partial x}\right)
\end{aligned}$$

where the derivative $\partial g/\partial x$ is evaluated at $x_1=u^{\mathrm{T}}s$ and $x_2=v^{\mathrm{T}}s$. After some matrix multiplication, we complete the proof.