

## Comparison of Haseman-Elston Linkage Tests with Age-of-Onset or Affection Trait

Kyoung-Hee Jung<sup>1)</sup> and Hae-Hiang Song<sup>2)</sup>

### Abstract

In this paper, we perform a simulation study of genetic model-free age-of-onset methods in linkage tests which has been proposed by Zhu et al. (1997). They performe. Haseman-Elston regression on a set of bipolar pedigree data using each of three dependent variables: a binary trait indicating disease concordance or discordance, a binary trait adjusted for age-of-onset, and the residuals from a survival analysis. We compare the powers of the proposed test statistics for various situations. Simulations that we have carried out show that the gains in power are observed when the residuals from a survival analysis are used in linkage tests.

*Keywords* : Haseman-Elston; Sib pairs; age-of-onset; binary trait; linkage.

### 1. 서론

유전학적 연관성이 있을 것이라고 짐작되는 인간의 질병에 대해 우리는 종종 그 유전 방식을 정확히 알 수 없기 때문에 질병에 영향을 미친다고 짐작되는 좌위(loci)를 확인하기 위해서 유전적 정보를 수집하여 유전연관성 검정(linkage test)을 하게 된다(Risch, 1990; Elston and Wilson, 1990; Suarez et al., 1990). 일반적으로 늦은 연령에서 발병하는 질병(late-onset diseases)에 대한 연구는 환자의 부모가 이미 사망한 경우가 대부분이므로 여러 세대에 걸친 가족구조(pedigree) 자료보다도 형제 쌍(sib pairs) 자료를 이용하게 된다.

이러한 질병 연구에서 단순한 형제 쌍으로부터 수집된 양적 형질(quantitative trait) 자료의 분석법으로는 이미 널리 사용되고 있는 Haseman-Elston (1972) 회귀 검정법과 이후 확장된 회귀 검정법이 있다(Forrest, 2001; Wang et al. 2004). 또한 형제의 연속형 형질 자료의 차이의 제곱을 종속변수로 두는 Haseman-Elston 회귀모형은 이 종속변수의 정규 분포 가정이 성립되지 않는 단점을 보완하고자 비모수 검정법도 활발히 개발되고 있다(Wan et al., 1997; Wang et al., 1998; Kim et al., 2006).

---

1) Graduate Student, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701.

2) Corresponding Author. Professor, Department of Biostatistics, The Catholic University of Korea, Seoul 137-701.

Correspondence : hhsong@catholic.ac.kr

연속형 형질에 못지 않게 의학 분야에서 이산형 형질(binary trait) 자료도 많다. 예를 들어서 어떤 질병의 발병 여부가 이에 해당한다. 이산형 형질 자료에 대해서 여러 분석법이 있어 질병에 걸린 형제 쌍(affected sib pair : ASP) 연구로 알려진 분석법들이 이에 포함된다. 예를 들어서 Weeks and Lange (1988)는 질병에 걸린 가족(affected pedigree members) 자료에 대한 분석법을 형제 쌍 자료 분석법으로 확장하였다. 한편 발병 여부에 덧붙여 발병 연령을 함께 감안한 이산형 형질 자료의 유전 연관성 연구가 오래전부터 관심을 끌여 왔으며 표준화된 LOD(likelihood log-odds) 점수에 근거한 유전 연관성 분석에서 발병 연령의 효과를 보정하는 것이 연관된 유전자를 발견하는데 효율적이라는 것을 밝혔다(Pericak-Vance et al., 1983; Hodge et al., 1980; Morton and Kidd, 1980; Haynes et al., 1986).

이러한 상황에서 Zhu et al. (1997)은 발병여부와 발병연령을 감안한 분석법으로서 Haseman-Elston (1972) 회귀 분석의 이용을 제안하였으며 각기 다른 종속변수를 가진 세 가지의 방법을 사용하여 양극성 장애(bipolar disorder)를 가진 가족 자료를 분석하였으며 이러한 연구의 목적은 발병에 영향을 미치는 주요 유전자를 발견하는 것이다. 또한 Dawson et al. (1990)은 이산형 형질 자료의 분석에서 Haseman and Elston (1972)의 형제 쌍 분석법에 초점을 맞춘 연구를 수행하였다. Zhu et al. (1997)은 첫째로 두 형제의 발병여부의 부합(concordance)을 나타내는 이항 형질에 대해 Haseman-Elston 회귀 분석을 수행하였고 나머지 두 방법은 발병연령의 효과를 보정한 후 두 형제의 발병여부의 부합 확률을 분석하거나 또는 생존분석의 잔차를 분석하였다. Zhu et al. (1997)은 일부 양극성 장애 자료에서 발병 연령 효과를 보정하지 않은 분석보다도 발병 연령 효과를 보정한 분석이 유전연관성에 대해 더욱 유의한 결과를 제시한다는 것을 보여주었다.

형제 자료에 대해 발병 연령의 효과를 보정한 유전연관성 분석이 단순한 이산형 형질 자료의 분석보다 바람직한 것으로 기대되지만, Zhu et al. (1997)은 모의실험의 검정력을 통해 여러 방법을 비교하지는 않았다. 이에 대해 본 논문은 Zhu et al. (1997)가 제안한 방법들에 대한 설명과 함께 모의실험을 통해 각 방법들의 검정력을 비교하고자 한다.

## 2. 방법

Haseman and Elston (1972)의 형제 쌍 분석법은 무작위 교배(random mating)와 형질과 표지유전자(marker) 좌위 사이에 상위성(epistasis)이 없다는 것을 가정한다. Haseman-Elston 방법의 기본 원리는 ‘만약 관심 있는 형질과 연관된 주요 유전자(major gene)가 표지유전자 좌위와 밀접하게 연관되어 있다면 두 형제의 표지 대립유전자(marker allele)의 공유 정도, 즉 공유 비율이 크면 클수록 형제 쌍의 형질 표현형(trait phenotype) 사이의 차이는 작을 것이다’라는 것이다. 유전학 분야에서는 부모로부터의 유전으로 두 형제가 대립유전자를 공유할 때 이러한 현상을 IBD(identical by descent)로 표현하며 장차 설명하게 되는 IBD 비율이 공유의 정도를 나타낸다. 만약 관심이 있는 형질이 이산형(발병 혹은 발병하지 않음)이면 한 개인이 병에 걸렸다면

1, 현시점까지 병에 걸리지 않았으면 0으로 표현할 때 만약 유전연관성 연구가 모두 병에 걸리지 않은 형제 자료를 포함하지 않게 되면 두 형제의 상태를 표현한 (0,0)의 자료가 없다. 또한 두 형제 사이에 구분이 없으므로 (0,1)과 (1,0)의 자료는 단순히 비부합 형제 쌍(discordant sib pair)이 되며 그리고 (1,1)로써 동일하게 두 형제 모두 질병 상태를 나타내는 부합 형제 쌍(concordant sib pair)은 비부합 형제 쌍보다 더욱 많은 대립유전자를 공유할 것이다. 반대로 주요 유전자가 존재하지 않는다면 형제 쌍이 가지고 있는 표지 대립유전자들의 비율과 형질의 유사성 사이에는 관련성이 없을 것이다.

Haseman-Elston 회귀 검정은 다음의 모형을 기본으로 한다.

$$E(Y_j|\hat{\pi}_j) = \alpha + \beta\hat{\pi}_j, \quad j = 1, \dots, n. \quad (1)$$

여기서  $n$ 은 전체 형제 쌍의 수이고,  $Y_j$ 는 총  $n$ 개의 형제 쌍 중에서  $j$ 번째 형제 쌍의 형질 값 사이의 차이의 제곱이며,  $\hat{\pi}_j$ 는 각 쌍에 대한 표지 대립유전자를 공유한 IBD의 추정된 비율로 정의한다. IBD 비율로 표현하지 않고 IBD의 추정된 비율로 표현한 이유는 일부 부모의 자료가 없는 경우에 두 형제의 공유 비율을 계산하기 때문이다(Haseman and Elston, 1972).

#### 방법 1: 이산형 형질의 회귀 분석법

이산형 표현형의 경우에 만약 두 형제 모두 발병하지 않은 자료가 제외되었다면 이러한 형제 자료는 부합과 비부합으로 요약됨을 앞에서 설명하였다. 이와 같은 자료를 질병에 걸린 형제 쌍 자료, 즉 ASP 자료라고 부르겠다. 이러한 이산형 표현형 자료에 적용한 Haseman-Elston 회귀 검정은 어떤 표지 유전자에 대해서 평균 IBD 비율이 부합과 비부합 형제 쌍에 대해 동일한 지에 대한 두 표본  $t$ -검정과 동일하며 여기서 두 표본이란 부합과 비부합의 두 군이다(Elston et al., 1973). 다시 말하면 Haseman-Elston 회귀 검정은 원래 공유한 표지 유전자의 IBD 비율이 독립변수이고 비부합(비부합 형제 쌍: 1, 부합 형제 쌍: 0)이 종속변수인 회귀 검정이었으나 두 표본  $t$ -검정에서는 결과적으로 독립변수와 종속변수가 바뀐 상태가 되었다.

이산형 형질의 회귀 모형은 구체적으로 다음과 같다. 각 개체의 이산형 형질이 질병에 걸린 경우를 1로 두고 질병에 걸리지 않은 경우를 0으로 둘 때 ASP 자료에서 두 형제의 수치의 차이의 제곱은 다시 1과 0으로 표현된다. 즉,  $j$ 번째 형제 쌍에서 만약 두 형제 모두 질병에 걸렸다면 두 형제 수치의 차이의 제곱인  $Y_j$ 는 0이 되고, 두 형제 중 한 명만이 질병에 걸렸다면  $Y_j$ 는 1이 된다. 따라서  $Y_j$ 는  $j$ 번째 형제가 질병에 걸린 상태에 대해 불일치하는 것을 나타낸다고 해석할 수 있다. 그러므로 Haseman-Elston 회귀모형 (1)에  $Y_j$ 를 직접 적용할 수 있다. 여기서 기울기가 0이라는 귀무가설은 유전연관성이 없다는 것과 일치한다. 그리고 기울기가 음이라는 단측 대립 가설은 비부합 형제 쌍 보다 부합 형제 쌍에서 대립유전자를 공유한 IBD 비율  $\hat{\pi}_j$ 의 평균값이 크다는 가설을 검정하는 것과 동일하여, 이는 형질과 표지유전자 좌위 사이의 유전연관성을 주장하는 증거가 된다(Elston et al., 1973).

방법 2: 발병 연령을 감안한 이산형 형질의 분석법

Dawson et al. (1990)에 의하면 Haseman and Elston (1988)은 각 형제 쌍 내에서 형제들의 발병 가능성(susceptibility)이 독립임을 가정했을 때 발병 연령을 감안할 수 있는 Haseman-Elston의 분석법에 대한 확장을 제안하였다고 한다. 여기서 발병 가능성의 확률을 간략히 발병률로 표현하겠다. 이러한 연구의 목적도 위의 방법 1에서와 마찬가지로 발병에 영향을 미치는 주요 유전자가 존재한다면 이러한 유전자를 발견하는 것이다. 이 방법에서의 가정은 충분히 오래 산다면 질병에 발병될 가능성을 지닌 개체는 결국 발병할 것이고, 가능성을 지니지 않은 개체는 결코 발병하지 않을 것이라는 사실이며, 그러나 환경적인 요인에 의해 발병의 가능성은 가정하지 않는다.

Zhu et al. (1997)은 부모의 표현형 정보인 발병 여부에 따라  $K$ 개의 다른 군으로 분류하여 분석하였는데, 이는 일반적으로 부모의 유전자를 받은 자녀의 발병률 또는 발병 연령이 부모의 발병 여부에 따라 다르기 때문이다. 즉 두 부모 모두가 발병하든가, 한 부모만이 발병하든가 또는 두 부모 모두 발병하지 않음에 따라 자녀의 발병률이나 발병 연령이 모두 다를 것이다. 여러 군으로 분류한 모형에서  $k$ 번째 군의 발병 연령의 분포가 평균이  $\mu_k$ , 표준편차  $\sigma_k$ 인 정규분포를 따른다고 가정한다. 그러나 본 논문의 모의실험에서는 부모의 표현형 정보를 모른다는 가정 하에서 여러 군으로의 분류를 시도하지 않았다. 발병 연령에 대한 확률밀도함수는 다음과 같다.

$$\phi\left(\frac{a-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{a-\mu}{\sigma}\right)^2\right]. \quad (2)$$

여기서  $a$ 는 각 형제, 즉 각 개체의 발병 연령이다.

전체 형제 쌍의 수가  $n$ 이고 각 쌍에서 두 명의 형제가 존재하므로 총  $N(=2n)$ 명의 개체 중에서  $a_{i,j}$ 는  $j$ 번째 형제 쌍의  $i$ 번째 형제의 발병 연령을 나타낸다. 이제 각 형제의 발병 가능성, 즉 발병률을  $\gamma$ 로 표현할 때 한 개체가 연령  $a_{i,j}$ 에서 질병에 걸릴 확률을 표현한 우도함수 부분은 다음과 같다.

$$\gamma\phi\left(\frac{a_{i,j}-\mu}{\sigma}\right). \quad (3)$$

그리고 한 개체가 관측시점  $a_{i,j}$ 에서 발병되지 않을 확률을 표현한 우도함수 부분은 다음과 같다.

$$1 - \gamma\Phi\left(\frac{a_{i,j}-\mu}{\sigma}\right). \quad (4)$$

따라서 각 개체들이 독립이라고 가정하고,  $\delta_{i,j}$ 를 연령  $a_{i,j}$ 에서 질병에 걸린 경우 1이고 질병에 걸리지 않을 경우 0이라 할 때 전체 우도 함수는 다음과 같다.

$$L = \prod_{j=1}^n \prod_{i=1}^2 \left[ \gamma\phi\left(\frac{a_{i,j}-\mu}{\sigma}\right) \right]^{\delta_{i,j}} \left[ 1 - \gamma\Phi\left(\frac{a_{i,j}-\mu}{\sigma}\right) \right]^{1-\delta_{i,j}}. \quad (5)$$

따라서 전체 표본에 대한 로그 우도함수는 다음과 같다.

$$\ln L = \sum_{j=1}^n \sum_{i=1}^2 \delta_{ij} \ln[\gamma \phi(\frac{a_{i,j} - \mu}{\sigma})] + \sum_{j=1}^n \sum_{i=1}^2 (1 - \delta_{ij}) \ln[1 - \gamma \Phi(\frac{a_{i,j} - \mu}{\sigma})]. \quad (6)$$

Zhu et al. (1997)은 위의 우도함수를 이용하여 모수를 추정하기 위해 S.A.G.E.의 AGEON 프로그램을 사용하였고 Dawson et al. (1990)은 MAXLIK(Kaplan and Elston, 1978)을 사용하였다. 본 논문에서는 SAS IML 프로그램에서 'call nlpnea' 명령어를 사용하여 평균  $\mu$ , 표준편차  $\sigma$ , 발병률  $\gamma$ 의 최대우도 추정량(Maximum Likelihood Estimation: MLE)을 구하였다.

이제 위에서 구한 추정량을 이용하여 구한 각 형제의 발병 확률  $X$ 는 다음과 같으며 발병하였다면 확률은 1이고, 발병하지 않았을 때는 조건부 확률로 표현된다.

$$X = \begin{cases} 1, & \text{발병하였을 때,} \\ \frac{\hat{\gamma} - \hat{\gamma} \Phi(\frac{a - \hat{\mu}}{\hat{\sigma}})}{1 - \hat{\gamma} \Phi(\frac{a - \hat{\mu}}{\hat{\sigma}})}, & \text{발병하지 않았을 때.} \end{cases} \quad (7)$$

그러면 전체 형제 쌍의 수가  $n$ 이고  $j$ 번째 형제 쌍의 각 형제의 발병 확률을  $X_{1j}$ ,  $X_{2j}$ 라 할 때 이 형제 쌍의 비부합 확률(probability of discordance)  $Y_j$ 는 다음과 같이 정의된다.

$$Y_j = X_{1j}(1 - X_{2j}) + (1 - X_{1j})X_{2j}, \quad j = 1, \dots, n. \quad (8)$$

이와 같이 구한 발병 연령의 효과를 보정한 비부합일 확률  $Y_j$ 를 종속변수로 두고, 각 형제 쌍의 IBD 비율  $\hat{\pi}_j$ 를 독립변수로 두어 Haseman-Elston 회귀식 (1)을 적용하여 유전연관성 분석을 한다.

### 방법 3: 생존분석으로부터 구해진 잔차의 회귀 분석법

이 방법은 우선 모든 개체들이 동일한 발병률을 가진다고 가정하여 발병률을 고려치 않는다. 그리고 발병 시점  $a$ 에서 질병에 걸릴 확률이 누적 생존 함수  $S(a)$ 를 따른다고 가정한다. 발병 연령을 관측할 수 있었던 개체와 발병 연령이 절단된(censored) 경우의 개체를 구분하여 다음과 같이 정의된 잔차를  $R$ 이라 한다. 이 때 관측시기  $a'$ 은 질병에 걸리지 않은 개체의 절단시간(censoring time)이 된다.

$$R(a) = \begin{cases} S(a), & \text{발병하였을 때,} \\ S(a') - 1, & \text{발병하지 않았을 때.} \end{cases} \quad (9)$$

구해진 잔차를 양적 형질(quantitative trait)처럼 취급하여 유전연관성 분석을 한다. 그러므로 형제 쌍의 형질 값 차이의 제곱은  $Y_j = [R(a_{1j}) - R(a_{2j})]^2$ 로 정의하며,  $Y_j$ 는 구체적으로 다음과 같다.

$$Y_j = \begin{cases} [S(a_{1j}) - S(a_{2j})]^2, & \text{두 형제가 모두 발병하였을 때,} \\ [S(a'_{1j}) - S(a'_{2j})]^2, & \text{두 형제가 모두 발병하지 않았을 때,} \\ [1 + S(a_{1j}) - S(a'_{2j})]^2, & \text{한 형제만 발병하고 한 형제는 발병하지 않았을 때.} \end{cases} \quad (10)$$

생존함수는 모수적 혹은 비모수적 생존모형들을 사용하여 추정할 수 있는데 Zhu et

al. (1997)은 다음의 와이블 모형을 적합시켰다.

$$S(a) = \exp(-\exp(\frac{a-\mu}{\sigma})). \quad (11)$$

방법 2에서와 마찬가지로 발병 연령의 효과가 보정된 식 (10)의 잔차의 차이  $Y_j$ 를 종속변수로 두고, 각 형제 쌍의 IBD 비율  $\hat{\pi}_j$ 를 독립변수로 두어 Haseman-Elston 회귀식 (1)을 적용하여 유전연관성 분석을 한다. 스코아(score) 통계량을 이용한 위와 유사한 방법이 Commenges (1994)에 의해 제안되었다. 잔차는 SAS의 proc lifereg 프로그램을 사용하여 구하였다.

### 3. 모의실험

#### 3.1 모의실험의 계획

본 연구는 앞에서 언급한대로 무작위 교배 및 여러 좌위 사이의 상위성이 없다는 가정하에서 출발하며, 또한 귀무가설 하에서 유전연관성이 균형(linkage equilibrium)임을 가정하며 단일 좌위(single locus)에 대한 모의실험 자료를 생성하였다. 여러 방법의 비교에서 여러 좌위들을 고려하는 것은 복잡성만을 초래하므로 단일 좌위를 선택하였다. 먼저 모의실험 자료의 생성과정을 설명하기 위해 우선  $x_{1j}$ 와  $x_{2j}$ 을  $j$ 번째 형제 쌍에서 첫 번째와 두 번째 형제의 발병 연령 또는 절단 시점이라 하자. 다음과 같은 일반적인 모형을 가정할 수 있다.

$$x_{1j} = \mu + g_{1j} + e_{1j}, \quad x_{2j} = \mu + g_{2j} + e_{2j}. \quad (12)$$

여기서  $\mu$ 는 전체 평균으로서 후에 본 연구의 모의실험에서 30으로 정하였으며, 그 이유를 설명하게 된다.  $g_{ij}$ 와  $e_{ij}$ 는 유전적 효과와 환경적 효과를 나타낸다. 또한 대립유전자 빈도(allele frequency)  $p$ 와  $q(=1-p)$ 를 가지는 두 대립유전자  $B$ 와  $b$ 를 가정한다. 유전자형에 따른 값(genotypic value)을 나타내는  $g_{ij}$ 는 유전자형이  $BB$ 일 경우  $a$ ,  $Bb$ 일 경우  $d$ ,  $bb$ 일 경우  $-a$ 로 정의할 수 있다( $a$ 와  $d$ 의 결정은 후에 설명하게 된다). 곧  $e_{ij}$ 의 생성을 설명할 것이며 이와 같이 생성한  $x_{1j}$ 와  $x_{2j}$ 로부터  $Y_j = (x_{1j} - x_{2j})^2$ 가 구해진다.

모의실험 자료의 첫 단계에서 두 형제의 형질 IBD가 0, 1, 2가 되는 비율을 0.25, 0.5, 0.25가 되도록 다항분포(multinomial distribution) 자료를 랜덤하게 생성한다. 두 형제의 유전자형의 조합으로서 9 가지가 가능하며 대립유전자 빈도와 IBD 비율의 값에 따라 이러한 9 가지의 가능성의 확률을 Haseman and Elston (1972)은 논문의 <표 1>에 제시하였으며, 설명에 도움이 되도록 <표 1>을 아래에 수록하였다. 따라서 두 번째 단계에서는 유전자형의 조합이 다른 형제 쌍 자료를 이 <표 1>의 확률을 가지는 다항분포로 랜덤하게 생성하였다. 셋째 단계에서는 9 가지의 유전자형의 조합에 따라 형제 쌍의 형질 즉, 두 형제의 발병 여부와 발병 연령을 각각 생성한다. 이 때 유전연관성을 감안하여 여러 대립가설에 대한 발병 여부와 발병 연령을 생성하는 것

이며 여러 대립가설에 대한 구체적인 차이는 <표 3>에 제시되었다. 우선 발병 연령에 대해 설명하면 고연령에 이르기까지 질병에 걸리지 않았다면 발병에 대한 위험이 낮을 것이므로 발병의 위험이 균일하고 사망이 드문 연령대인 12세부터 48세까지로 결정하여, 발병 연령의 분포는 평균 30세, 표준편차 4의 정규분포로 생성하였다. 따라서 생성된 분포의 (-3SD, 3SD) (여기서 SD는 standard deviation을 뜻한다)의 범위는 모의실험에서 우리가 원하는 대로 12세~48세에 이르게 된다. 더욱 자세히는, 두 형제의  $x_{1j}$ 와  $x_{2j}$ 에서  $e_{1j}$ 과  $e_{2j}$ 는 각각 평균이 0이고 표준편차 4의 정규분포가 되면서 동시에 서로간의 상관성은  $\rho$ 가 되도록 생성하였다.

<표 1>  $Y_j$ 의 조건부 분포

형제 쌍	$Y_j$	조건부 확률		
		$\pi_j = 0$	$\pi_j = \frac{1}{2}$	$\pi_j = 1$
$BB-BB$	$e_j^2$	$p^4$	$p^3$	$p^2$
$bb-bb$	$e_j^2$	$q^4$	$q^3$	$q^2$
$Bb-Bb$	$e_j^2$	$4p^2q^2$	$pq$	$2pq$
$BB-Bb$	$(a-d+e_j)^2$	$2p^3q$	$p^2q$	0
$Bb-BB$	$(-a+d+e_j)^2$	$2p^3q$	$p^2q$	0
$Bb-bb$	$(a+d+e_j)^2$	$2pq^3$	$pq^2$	0
$bb-Bb$	$(-a-d+e_j)^2$	$2pq^3$	$pq^2$	0
$BB-bb$	$(2a+e_j)^2$	$p^2q^2$	0	0
$bb-BB$	$(-2a+e_j)^2$	$p^2q^2$	0	0

두 형제의 상관성이 있는 이산형 변수, 즉 발병 여부의 자료는 Park et al. (1996)이 제시한 포아송 확률변수를 이용한 방법을 채택하였다. 각 형제의 발병률은 유전자형에 따라 각각  $p_{BB}$ ,  $p_{Bb}$ ,  $p_{bb}$ 의 순서로 낮아져 가는, 동일 간격의 확률로 지정하였다. 만약  $p_{BB} = 0.5$ ,  $p_{Bb} = 0.3$ ,  $p_{bb} = 0.1$ 을 가정하게 되면 위의 <표 1>에 근거하여 형제 쌍의 유전자형이  $BB-Bb$ 일 경우 형제의 발병비율이 각각 0.5, 0.3인 상관된 이산형 변수, 즉 발병 여부 자료를 생성하게 된다. 모의실험에서 발병률 간격은 0.05, 0.2, 0.3, 0.4를 선택하였다.

방법 1과 방법 2는 형제 쌍이 모두 발병되지 않은 경우 표본에서 제외하고 분석한다. 즉 ASP 자료에 근거한 방법이다. 그러나 방법 3은 모든 자료를 분석한다. 그러므로 방법 1과 방법 2의 표본크기가 방법 3의 표본크기와 다르기 때문에 검정력을 그대로 비교하기에는 무리가 있다. 따라서 이 논문에서는 형제 쌍 모두 발병되지 않은 자료를 표본에서 제외한, ASP 자료에만 근거한 방법 3을 추가로 시도하였고 이 방법을 방법 3.1로 부르겠으며 모든 표본 자료에 대해 방법 3을 적용한 것을 방법 3.2로 구분하여 검정력을 비교하였다. 그리고 방법 2에서 우도함수를 이용하여 모수를 추정하기 위해 SAS IML의 'call nlpnra'를 사용하였고 매번 500 형제 쌍 표본 자료로부터  $\mu$ ,

$\sigma$ ,  $\gamma$ 를 반복 추정하여 모의실험의 검정 과정에 사용하였다.

형질의 표현형의 어느 정도가 다음 세대에 유전되는지를 표현하는 유전율 (heritability)  $h^2$ 는 본 논문에서 0.2로 가정하였다. 유전율은 다음과 같이 정의된다.

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$$

여기서 분산  $\sigma_e^2$ 는 16으로 가정하였고,  $\sigma_g^2 = \sigma_a^2 + \sigma_d^2$ 이 된다. 우리는 가법(additive) 모형을 가정하며  $\sigma_d^2 = 4p^2q^2d^2 = 0$ 이 된다. 즉  $\sigma_g^2 = \sigma_a^2$ 이 된다. 여기서  $\sigma_a^2 = 2pq \times [a - d(p - q)]^2$ 이 되는데, 이때  $p$ 는 대립유전자 빈도이며  $q = 1 - p$ 이고, 결과적으로 가법 모형에서  $d$ 는 0이 된다(Haseman-Elston, 1972). 다시 말하면  $d = 0$ 인 모형을 가법 모형이라 하며 유전학의 수많은 연구의 모의실험에서 채택되는 모형이다. 반면에 우성 모형의 경우는  $a = d$ , 열성 모형의 경우는  $d = -a$ 의 형태가 된다. 따라서 가정된 유전율  $h^2 = 0.2$ 에 근거하여 가법 모형하에서 유전적 효과  $a$ 의 값을 구체적으로 계산할 수 있으며 더불어 식 (12)에서  $g_{1j}$ 와  $g_{2j}$ 가 결정된다.

모의실험의 마지막 단계로서 대립유전자 빈도  $p$ 는 0.1(rare allele), 0.5로 정하였다. 표본의 크기는 여러 문헌에 근거하여 500개의 형제 쌍 자료를 생성하였고 모의실험의 반복회수는 1,000번으로 정하였다. 위에서 언급하였던 각 사항에 대해 가법 모형 하에서 대립유전자 빈도, 유전자형에 따른 발병률, 유전율에 따라 세 가지 다른 분석법의 검정력을 비교하였다.

### 3.2 모의실험의 결과

<표 2>는 유의수준 5%에서 1,000번의 모의실험으로부터 유전연관성이 없다는 귀무가설 하에서의 검정력, 즉 추정된 유의수준을 나타낸 것이다. 모든 상황에서 5%에 비슷한 값들을 나타내었다.

<표 2> 모의실험으로부터 추정된 유의수준

대립유전자 빈도	상관관계	방법1	방법2	방법3.1	방법3.2
0.1	0.2	0.059	0.059	0.052	0.046
	0.5	0.047	0.041	0.049	0.052
	0.8	0.052	-	0.047	0.042
0.5	0.2	0.049	0.050	0.048	0.046
	0.5	0.045	-	0.046	0.045
	0.8	0.053	-	0.055	0.052

- : 최적화 횟수 < 10

<표 3>과 <표 4>는 유전율이  $h^2 = 0.2$ 이고 대립유전자 빈도가 0.1, 0.5로 다를 때 세 방법을 사용하여 대립가설 하에서 형제 쌍의 발병 여부 또는 발병 연령 자료를 분



석한 결과를 나타낸 것이다. 가법 모형에서 대립유전자 빈도에 따른 각 방법의 검정력은 비슷한 경향을 나타내었는데 이는 대립유전자 빈도가 0.1과 0.5로 다를 때 생성된 자료와 모수의 추정값들이 크게 다르지 않았기 때문이기도 하다. 모수의 추정값에 대해 곧 살펴보게 된다.

<표 3>에 제시된 방법 중 동일 ASP 표본수 자료에 근거한 방법 1과 방법 3.1의 검정력을 살펴보면 방법 1은 발병률 간격이 0.3이상으로 유전연관성이 뚜렷할 때 비교적 높은 검정력을 나타내며 나머지 대부분의 경우, 특히 형제 쌍의 상관관계가 높을 때 방법 3.1의 검정력이 높게 나타난다. 또한 모든 방법이 형제 쌍의 상관관계가 클수록 검정력이 높다. 모든 표본 자료를 사용할 수 있는 방법 3.2는 대립유전자 빈도가 0.5일 때 몇몇 경우를 제외한 대부분의 상황에서 방법 1과 방법 3.1에 비해 높은 검정력을 나타내었다. 그러나 대립유전자 빈도가 0.1이면서 두 형제 쌍의 발병 연령의 상관계수가 0.5 또는 0.8인 경우에 발병률이 낮은 경우에서 방법 3.2는 방법 3.1보다 검정력이 낮았다. 두 방법의 결과 차이에 대한 이유로는 모든 표본 자료를 사용하는 방법 3.2는 특히 두 형제 모두 발병하지 않은 자료를 포함하여 검정하며, 이와 같은 자료의 포함이 방법 3.2의 검정력이 낮아지는 결과를 초래하였다. 방법 3.2의 표본수가 방법 3.1의 표본수보다 크다는 점이 특징인데, 이러한 사실에도 불구하고 방법 3.2가 방법 3.1보다 검정력이 낮았던 경우의 모든 표본 자료 중 두 형제 모두 발병하지 않은 자료를 제외한 나머지 형제의 표본수( $n_{ASP}$ )를 살펴보면 250쌍 이하이었다. 즉 절반 이상의 자료가 두 형제 모두 발병하지 않은 자료이며 방법 3.2에서 이러한 자료는 절단 자료로 취급되며 이와 같은 절단 자료의 정도가 심할 때 검정력이 결코 높지 않은 것이다.

방법 2의 모의실험 결과에서는 각 표본 자료의 경우 모수의 최적화를 얻어내기 어려웠으며 이는 SAS의 비선형 최적화(nonlinear optimization)의 여러 서브루틴(routine) 중의 선택을 달리 하여도 변함이 없었다. 따라서 방법 2를 방법 1, 방법 3의 모의실험 결과와 직접 비교하는 것이 무리이므로 아래의 <표 4>에 따로 제시하게 되었다. 그러나 검정력 계산에서는 1,000번의 반복횟수 중 최적화가 이루어진 경우의 수 중에서 기각횟수의 비율로 계산하였다. <표 4>에 모수의 최적화가 이루어진 횟수를 추가로 제시하면서 비교적 최적화된 횟수가 높았던 경우만의 검정력 결과를 제시하였다. 결과를 살펴보면 상관관계가 낮은 0.2에서 대립유전자 빈도가 0.5일 때가 대립유전자 빈도가 0.1일 때에 비해 높은 검정력을 나타내었다. 또한 유전자형에 따른 발병률이 높고 발병률 간격이 클 때 약간 검정력이 높은 것으로 나타났다. 한편 <표 3>에 제시된 방법 1과 방법 3.1과 3.2의 결과와 비교해 볼 때 방법 2의 검정력이 이들 방법보다 검정력이 높지 않았다.

<표 3> 방법 1, 방법 3의 모의실험 결과

$\rho$	발병률 간격	$p_{BB}$	$p_{Bb}$	$p_{bb}$	대립유전자빈도 0.1			대립유전자빈도 0.5				
					$n_{ASP}$	방법 1	방법3.1	방법3.2	$n_{ASP}$	방법 1	방법3.1	방법3.2
0.2	0.05	0.95	0.90	0.85	478	0.061	0.110	0.107	486	0.060	0.148	0.148
		0.65	0.60	0.55	379	0.051	0.099	0.088	396	0.066	0.099	0.096
		0.55	0.50	0.45	329	0.057	0.074	0.084	350	0.060	0.084	0.085
		0.45	0.40	0.35	273	0.053	0.087	0.079	296	0.058	0.078	0.083
		0.15	0.10	0.05	52	0.068	0.068	0.077	86	0.058	0.067	0.058
	0.20	0.90	0.70	0.50	368	0.077	0.109	0.111	431	0.152	0.207	0.230
		0.80	0.60	0.40	317	0.086	0.089	0.099	393	0.122	0.158	0.180
		0.70	0.50	0.30	258	0.083	0.094	0.100	347	0.121	0.149	0.175
		0.60	0.40	0.20	191	0.096	0.101	0.079	293	0.158	0.166	0.203
		0.50	0.30	0.10	117	0.109	0.106	0.087	231	0.163	0.155	0.185
	0.30	0.90	0.60	0.30	270	0.127	0.121	0.135	390	0.366	0.348	0.431
		0.80	0.50	0.20	205	0.138	0.154	0.167	344	0.374	0.329	0.439
0.70		0.40	0.10	132	0.189	0.161	0.160	290	0.391	0.316	0.451	
0.40	0.90	0.50	0.10	145	0.365	0.263	0.318	339	0.746	0.614	0.769	
0.5	0.05	0.95	0.90	0.85	460	0.053	0.139	0.145	473	0.076	0.207	0.209
		0.65	0.60	0.55	342	0.043	0.072	0.069	360	0.058	0.109	0.118
		0.55	0.50	0.45	292	0.048	0.086	0.088	312	0.058	0.101	0.104
		0.45	0.40	0.35	238	0.037	0.074	0.080	260	0.047	0.084	0.094
		0.15	0.10	0.05	44	0.063	0.062	0.064	73	0.065	0.073	0.078
	0.20	0.90	0.70	0.50	332	0.089	0.130	0.141	404	0.243	0.291	0.326
		0.80	0.60	0.40	282	0.076	0.112	0.116	360	0.203	0.255	0.299
		0.70	0.50	0.30	226	0.092	0.118	0.121	312	0.189	0.204	0.255
		0.60	0.40	0.20	166	0.097	0.120	0.108	260	0.209	0.261	0.287
		0.50	0.30	0.10	100	0.121	0.133	0.119	202	0.222	0.237	0.266
	0.30	0.90	0.60	0.30	238	0.147	0.183	0.177	362	0.594	0.586	0.678
		0.80	0.50	0.20	178	0.165	0.177	0.190	312	0.487	0.461	0.554
0.70		0.40	0.10	112	0.271	0.266	0.225	257	0.528	0.455	0.545	
0.40	0.90	0.50	0.10	123	0.527	0.431	0.367	307	0.919	0.820	0.886	
0.8	0.05	0.95	0.90	0.85	442	0.053	0.213	0.229	461	0.181	0.474	0.495
		0.65	0.60	0.55	304	0.054	0.118	0.122	323	0.054	0.149	0.164
		0.55	0.50	0.45	255	0.052	0.105	0.117	275	0.058	0.115	0.143
		0.45	0.40	0.35	203	0.059	0.103	0.104	224	0.060	0.121	0.138
		0.15	0.10	0.05	36	0.106	0.113	0.099	60	0.152	0.169	0.138
	0.20	0.90	0.70	0.50	297	0.260	0.347	0.368	380	0.804	0.856	0.890
		0.80	0.60	0.40	247	0.257	0.363	0.360	332	0.756	0.794	0.847
		0.70	0.50	0.30	195	0.245	0.313	0.333	281	0.710	0.728	0.782
		0.60	0.40	0.20	140	0.319	0.374	0.332	229	0.708	0.714	0.742
		0.50	0.30	0.10	83	0.396	0.384	0.332	173	0.725	0.688	0.707
	0.30	0.90	0.60	0.30	206	0.533	0.578	0.576	335	0.973	0.976	0.990
		0.80	0.50	0.20	152	0.569	0.559	0.540	284	0.968	0.952	0.970
0.70		0.40	0.10	92	0.619	0.565	0.499	226	0.940	0.922	0.940	
0.40	0.90	0.50	0.10	103	0.803	0.747	0.650	280	0.996	0.998	0.995	

$n_{ASP}$  : ASP 연구인 방법 1과 방법 3.1에서 형제 쌍의 수

<표 4> 방법 2의 모의실험 결과

$\rho$	$p_{BB}$	$p_{Bb}$	$p_{bb}$	대립유전자빈도 0.1		대립유전자빈도 0.5	
				최적화 횟수	검정력	최적화 횟수	검정력
0.2	0.15	0.10	0.05	237	0.072	362	0.111
	0.50	0.30	0.10	192	0.099	362	0.127
	0.90	0.50	0.10	259	0.143	50	0.320
0.5	0.15	0.10	0.05	101	0.079	88	0.068
	0.50	0.30	0.10	88	0.102	147	0.122
	0.90	0.50	0.10	100	0.180	-	-

- : 최적화 횟수 < 10

이제 각 방법에서의 검정 결과에 영향을 미치는 모수의 추정값들에 대한 결과를 살펴 보겠다. 아래의 <표 5-1>는 형제 쌍의 상관관계가 0.2이고 유전자형에 따른 형제의 발병률이  $p_{BB} = 0.5$ ,  $p_{Bb} = 0.3$ ,  $p_{bb} = 0.1$ 일 때 대립유전자 빈도에 따라 추정값들을 나타낸 것이다. 500개의 형제 쌍 자료를 생성하여 1,000번 반복하였으나 모수의 최적화에 어려움이 있어 발병률이 1 이상이고 표준편차가 극도로 작게 추정되었을 때 수렴이 이루어지지 않은 것을 프로그램의 세부과정으로 확인하여 최적화가 이루어지지 않은 것으로 결론짓고 이러한 결과는 모의실험에서 제외하였다. 추정된 발병률, 발병 연령의 평균, 표준편차의 기술 통계량을 보면 우선 발병 연령의 평균은 참값이 30인데 추정된 발병 연령은 대립유전자 빈도가 0.5일 때 참값에 가깝고 더욱 신뢰성이 있으며, 표준편차도 참값인 4에 더욱 가깝다. 즉 대립유전자 빈도가 0.5일 때보다도 대립유전자 빈도가 0.1일 때 발병률이 낮게 추정된 반면 발병 연령의 평균과 표준편차의 편의(bias)가 크게 추정됨(overestimated)을 알 수 있다.

방법 3의 검정력 또한 방법 2와 마찬가지로 추정값에 큰 영향을 받을 수 있는데, 아래 <표 5-2>은 형제 쌍의 상관관계가 0.2이고 유전자형에 따른 형제의 발병률이  $p_{BB} = 0.5$ ,  $p_{Bb} = 0.3$ ,  $p_{bb} = 0.1$ 일 때 대립유전자 빈도에 따라 방법 3에서의  $\mu$ 와  $\sigma$ 의 추정값을 나타낸 것이다. 방법 3에서는 와이블 분포 모형에 적합시켜  $\mu$ 와  $\sigma$ 를 추정하게 된다.

<표 5-1>의 결과를 보면 모든 표본 자료를 사용하여 분석하는 방법 3.2가 ASP 표본수 자료에 근거한 방법 3.1보다 모수 추정값이 모두 높게 나타났다. 또한 대립유전자 빈도가 작은 0.1일 때 대립유전자 빈도가 0.5일 때 보다  $\mu$ 가 높은 값으로 추정되었으며  $\sigma$ 는 낮은 값으로 추정되었다. 이는 대부분의 상황에서 대립유전자 빈도가 0.5일 때의 검정력이 대립유전자 빈도가 0.1일 때의 검정력보다 높은 값을 나타낸 결과와 관련이 있다고 할 수 있다.

&lt;표 5-1&gt; 방법 2의 모수 추정값 - 발병률과 발병 연령의 평균 및 표준편차

대립유전자 빈도	모수	평균(표준편차)	최소값	최대값
0.1	$\gamma$	0.970 (0.020)	0.915	0.999
	$\mu$	35.000 (0.465)	33.775	36.163
	$\sigma$	5.331 (0.361)	4.458	6.295
0.5	$\gamma$	0.982 (0.012)	0.937	0.999
	$\mu$	31.012 (0.306)	29.809	31.735
	$\sigma$	4.943 (0.203)	4.400	5.697

&lt;표 5-2&gt; 방법 3의 모수 추정값

대립유전자 빈도	방법	모수	평균(표준편차)	최소값	최대값
0.1	방법 3.1	$\mu$	3.612 (0.011)	3.579	3.659
		$\sigma$	0.130 (0.008)	0.102	0.158
	방법 3.2	$\mu$	3.838 (0.018)	3.785	3.909
		$\sigma$	0.137 (0.008)	0.110	0.164
0.5	방법 3.1	$\mu$	3.503 (0.009)	3.473	3.535
		$\sigma$	0.141 (0.006)	0.124	0.162
	방법 3.2	$\mu$	3.640 (0.012)	3.605	3.693
		$\sigma$	0.152 (0.006)	0.132	0.171

방법 3.1: ASP 자료의 분석, 방법 3.2: 전체 자료의 분석

여러 방법 간의 결과를 요약하면 대부분의 경우에서 모든 표본 자료를 사용한 방법 3.2가 절단 자료가 많지 않은 경우에 높은 검정력을 나타내었다. 동일한 표본수에 근거한 방법 3.1과 방법 1, 방법 2와의 비교에서도 방법 3.1이 높은 검정력을 나타내었다. 이는 대부분의 상황에서 구체적인 발병 연령에 근거한 방법 3을 사용하는 것이 유전연관성 검정에 효율적이라는 것을 보여주는 증거라 할 수 있다.

#### 4. 토의 및 결론

인간의 질병에 대한 역학적 연구에서 발병 연령이 꾸준한 관심의 대상이었으며 따라서 유전학적 연관성 연구에서도 발병 여부 또는 발병 연령과 관련된 유전자를 찾는 것은 당연한 일이다. 인간의 장수와 관련하여 치매의 발병 연령과 관련된 유전자를 찾는 연구, 또는 아직 확실한 치료가 알려지지 않은, 예를 들어서 다발성 경화증(multiple sclerosis)의 연구 등에서 발병 연령과 관련된 유전자를 찾는 일이 중요한 일로 인식되고 있다. 그 이유는 이 같은 연구로서 다발성 경화증의 발병을 10년 늦게 발병하도록 지연할 수만 있다면 대사회적으로 의료비의 경감이 클 것임을 짐작할 수 있기 때문이다.

발병의 위험과 발병 연령의 관계는 실제로 매우 복잡한 것으로 짐작된다. 심장 또는 관상 동맥병의 경우 발병의 위험이 크다고 해서 반드시 발병 연령이 낮지 않아 발병의 위험과 발병 연령이 무관할 수 있음을 말하고 있다(Fraser et al., 1995). 최근에는 DeStefano et al. (2002)은 파킨슨병 환자의 게놈 스캔(genome scan) 자료를 분석하여 발병 여부에 관련된 유전자와 발병 연령과 관련된 유전자는 다를 수 있음을 설명하였다. 이러한 사실들로부터 발병 여부와 발병 연령에 대한 각각의 연구가 모두 중요함을 알 수 있으며 따라서 두 가지 다른 유전 표현형 자료의 분석법에 대해 함께 살펴보아야 한다.

이 논문에서는 이러한 발병 여부 또는 발병 연령의 분석법에 대한 비교를 모의실험으로 알아보았다. 검정력이 서로 다르다는 사실은 어떤 표현형의 자료를 어떤 분석방법으로 알아볼 때 더욱 신뢰성 있는 결론을 유도하기 쉬운지를 알려준다. 방법 1은 유전자형에 따른 발병률 간격이 0.3이상으로 유전연관성이 뚜렷할 때 비교적 높은 검정력을 나타내는 것을 제외하고는 발병 연령을 감안한 분석법, 특히 방법 3이 발병 여부에 대한 연구보다도 신뢰성 있는 결론을 유도하기가 쉽다 하겠다. 그러나 발병 여부만을 고려하는 검정법인 방법 1은 발병률 간격이 크지 않아서 검정력이 낮음에도 불구하고 그와 같은 표현형 자료에 근거하여 유전연관성 연구를 진행하고자 한다면 대표본 형제 자료를 수집해야 함을 시사한다. 방법론의 적용면에서 방법 2는 발병 연령을 보정한 후의 발병 여부 자료에 근거하여 유전연관성 검정을 한다는 점이 큰 장점으로 짐작되었으나, 최적화의 해결 특히 대립 유전자 빈도가 0.1로 낮을 때에 발생하는 과추정(overestimation) 및 검정력이 높지 않은 이유를 밝혀야 하는 문제가 남아 있다. 일반적으로 유전연관성이 뚜렷할 것인지 미약할 것인지 짐작할 수 없을 경우에는 발병 연령을 감안한 방법 3이 바람직하며 이 경우 두 형제 모두 발병되지 않은 자료, 즉 절단 자료가 절반 이상으로 많지만 않다면 방법 3.1과 3.2에 큰 차이가 없다 하겠다. 그러나 발병 연령 자료의 올바른 수집에 대한 제반 문제는 간과할 수 없이 중요한 문제이며, 이에 대해서는 의학, 역학 분야의 다른 논문들을 참고하기 바란다.

## 참고문헌

- [1] Commenges, D. (1994). Robust genetic linkage analysis based on a score test of homogeneity: the weighted pairwise correlation statistic. *Genetic Epidemiology*, Vol. 11, 189-200.
- [2] Dawson, D.V., Kaplan, E.B. and Elston, R.C. (1990). Extensions to sib-pair linkage tests applicable to disorders characterized by delayed onset. *Genetic Epidemiology*, Vol. 7, 453-466.
- [3] DeStefano, A.L., Lew, M.F., Golbe, L.I., Mark M.H., Lazzarini, A.M., Guttman, M., Montgomery, E., Waters, C.H., Singer, C., Watts, R.L., Currie, L.J., Wooten, G.F., Maher, N.E., Wilk, J.B., Sullivan, K.M., Slater, K.M., Saint-Hilaire, M.H., Feldman R.G., Suchowersky, O., Lafontaine, A.L., Labelle, N., Growdon, J.H., Vieregge, P., Pramstaller, P.P., Klein, C,

- Hubble, J.P., Reider, C.R., Stacy, M., MacDonald, M.E., Gusella, J.F. and Myers, R.H. (2002). PARK3 Influences age at onset in Parkinson disease: A genome scan in the genePD study. *American Journal of Human Genetics*, Vol. 70, 1089-1095.
- [4] Elston, R.C., Kringlen, E. and Namboodiri, K.K. (1973). Possible linkage relationships between certain blood groups and schizophrenia or other psychoses. *Behavior Genetics*, Vol. 3, 101-106.
- [5] Elston, R.C., Wilson, A.F. (1990). Genetic linkage and complex diseases: A comment. *Genetic Epidemiology*, Vol. 7, 17-19.
- [6] Forrest, W.F. (2001). Weighting improves the 'new Haseman-Elston' method. *Human Heredity*, Vol. 53, 59-67.
- [7] Fraser, G.E., Lindsted, K.D. and Beeson, W.L. (1995). Effect of risk factor values on lifetime risk and age at first coronary event. *American Journal of Epidemiology*, Vol. 142, 746-758.
- [8] Haseman, J.K. and Elston R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, Vol. 2, 3-19.
- [9] Haynes, C., Pericak-Vance, M.A. and Dawson, D.V. (1986). Genetic analysis workshop IV: Analysis of Huntington's disease linkage and age of onset curve. *Genetic Epidemiology*, Supplement 1, 235-239.
- [10] Hodge, S.E., Spence, M.A., Crandall, B.F., Sparkes, R.S., Sparkes, M.C., Crist, M. and Tideman, S. (1980). Huntington disease: Linkage analysis with age-of-onset corrections. *American Journal of Medical Genetics*, Vol. 5, 247-254.
- [11] Kim, M.K., Hong, Y.J., Song, H.H. (2006). Improved nonparametric trend statistic incorporating dispersion differences in sib pair linkage for quantitative traits. *Human Heredity*, To be published.
- [12] Morton, L.A., Kidd, K.K. (1980). The effects of variable age-of-onset and diagnostic criteria on the estimates of linkage. *Social biology*, Vol. 27, 1-10.
- [13] Park C.G, Park T.S and Shin D.W (1996). A simple method for generation correlated binary variates. *American Statistician*, Vol. 50, 306-310.
- [14] Penrose, L.S. (1938). Genetic linkage in graded human characters. *Annals of eugenics*, Vol 6, 133-138.
- [15] Pericak-Vance, M.A., Elston, R.C., Conneally, P.M. and Dawson, D.V. (1983). Age of onset heterogeneity in huntington disease families. *American Journal of Medical Genetics*, Vol. 14, 49-60.
- [16] Risch, N. (1990). Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genetic Epidemiology*, Vol. 7, 3-16.
- [17] Suarez, B.K., Reich, T., Rice, P.J. and Cloninger, C.R. (1990). Genetic linkage

- and complex diseases: A comment. *Genetic Epidemiology*, Vol. 7, 37-40.
- [18] Wan, Y., Cohen, J. and Guerra R. (1997). A permutation test for the robust sib-pair linkage method. *Annals of Human Genetics*, Vol. 61, 79-87.
- [19] Wang, J., Guerra, R. and Cohen, J. (1998). Statistically robust approaches for sib-pair linkage analysis. *Annals of Human Genetics*, Vol. 62, 349-359.
- [20] Wang, T. and Elston, R.C. (2004). A modified revisited Haseman-Elston method to further improve power. *Human Heredity*, Vol. 57, 109-116.
- [21] Weeks, D.E. and Lange, K. (1988). The affected-pedigree-member method of linkage analysis. *American Journal of Human Genetics*, Vol. 33, 315-329.
- [22] Zhu, X., Olson, J.M., Schnell, A.H. and Elston, R.C. (1997), Model-free age-of-onset methods applied to the linkage of bipolar disorder. *Genetic Epidemiology*, Vol. 14, 711-716.
- [23] SAS Release 8.1 (1998). *Cary SAS Institute*.

[Received August 2006, Accepted October 2006]