가족기반연구에서 형매정보를 이용한 유전적 연관성 분석방법의 비교

정지원^a, 이석훈^a, 김순애^b, 박미라^{1,c}

^a충남대학교 통계학과, ^b을지대학교 약리학교실, ^c을지대학교 예방의학교실

요약

가족기반 유전연구에서 단일염기다형성(SNP) 및 일배체형(haplotype)을 이용한 질병-유전자간 연관성 분석은 주요 관심부분 중 하나이다. 이진형질인 경우 사용되는 고전적인 TDT(transmission/disequilibrium test)는 부/모/자의 유전정보를 모두 알 수 있을 때에 쓸 수 있다. 그러나 경우에 따라서는 부모의 유전자형에 대한 정보를 사용할 수 없을 수가 있는데, 치매등과 같이 늦게 발병하는 질환에서 이런 경우가 자주 발생한다. 부모의 유전자형을 모를 때 이를 대체할 수 있는 방법으로 형매(sib)의 유전정보를 이용하는 방법이 있다. 본고에서는 가족기반연구에서 형매의 유전정보를 이용하는 분석방법에 대해 고찰하고 모의실험을 통해 이들의 검정력을 비교하였다.

주요용어: 유전적 연관성, TDT, 단일염기다형성(SNP), 모의실험.

1. 서론

유전학 연구에서 단일염기다형성(single nucleotide polymorphism; SNP) 및 일배체형(haplotype)을 이용한 질병-유전자간 연관성분석은 주요 관심부분 중 하나이다. SNP는 염기가 개인마다 달라질 수 있는 DNA 상에 위치한 유전변이로서, 대부분 두 개의 대립형질(allele)을 가지며, 발병 원인이나 치료제에 대한 반응 등 개인적인 차이를 가져오는 원인이 될 수 있다. 유전적 연관성을 분석하기 위한 방법으로 연쇄불균형(linkage disequilibrium; LD)이나 사례-대조 분석, 가계(pedigree) 분석, 부모자(trio) 분석 등에서의 여러 가지 척도와 방법이 사용되고 있으며, 관련 연구들이 활발히 수행되고 있다 (Laird와 Lange, 2008; Saito 등, 2006; Zhao, 2000; 김진홍과 장양수, 2008; 이효정 등, 2010). 유전적 연관성 분석은 사례-대조연구와 같이 친척관계가 없는 사람들을 대상으로 하는 집단기반연구(population-based study)와 질병을 가진 사람과 그의 가족에 대한 유전정보를 이용하여 질병과 유전자간의 연관성을 분석하는 가족기반연구(family-based study)로 나눌 수 있다. 집단기반연구는 아집단(subpopulation)의 형성으로 인해 위험 유전자의 빈도가 질환 취약성 유전자(disease susceptibility gene)와 무관하게 대조군과 질환군간의 두 집단간에 다르게 분포되어 나타나는 오류가 발생할 수 있으며, 이를 피하기 위해서 개발된 것이 가족기반 연구방법이다. 연관성 분석에서 질병의 유무 등 이진 형질(binary trait)인 경우에 가장 많이 사용되는 것은 TDT(transmission/disequilibrium test)로서 고전적인 TDT는 부/모/자의 유전정보를 모두 알 수 있을 때에 쓸 수 있다. 그러나 경우에 따라서는 부모의 유전자형에 대한 정보를 사용할 수 없을 수가 있는데, 특히 치매 등과 같이 환자가 나이가 든 후에 발병하는 이른바 late-onset 질병의 경우에 환자의 부모에 대한 정보를 얻을 수 없는 경우가 흔하다. 부모의 유전자형을 모를 때 이를 대체할 수 있는 방법으로 형매(sib)의 유전정보를 이용하는 방법을 고려할 수 있다.

이 논문은 2007년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-531-C00018).

¹ 교신저자: (301-832) 대전 중구 용두동, 을지대학교 예방의학교실, 교수. E-mail: mira@eulji.ac.kr

본고에서는 가족기반연구에서 형매정보를 이용한 주요 연관성 분석을 데이터의 특성별로 i) 부모의 유전자형 정보만 있을 때, ii) 형매의 유전자형 정보만 있을 때, iii) 부모/형매의 유전자형 정보가 있을 때의 세 가지 경우로 분류하여 방법론과 특징을 소개하였다. 또한 모의실험을 통하여 주요 방법의 검정력을 비교하였다.

2. 유전적 연관성 분석 방법

2.1. 부모의 유전자형 정보만 있을 때

이진 형질의 연관성 검정으로서 TDT는 환자와 환자의 부모에 대한 유전자형을 모두 알 수 있으며, 부모 중 한명은 이형접합(heterozygote)일 때 적용된다. 이 방법은 연관이 있을 때 연쇄(linkage)에 대한 검정방법으로 개발되었으며, 마커와 질병유전자간에 연쇄와 연관이 모두 없다면 마커의 형질은 부모에게서 자식으로 랜덤하게(즉, 50%의 확률로) 전이(transmit)된다는 전제에서 출발하였다 (Spielman 등, 1993). 이 방법은 환자에게로 전이되는 마커형질과 전이되지 않는 마커형질을 비교하여 검정하게 되며, M_1 , M_2 의 2개 대립형질이 있을 때 검정통계량은 다음과 같다.

$$TDT = \frac{(b - c)^2}{b + c},$$

여기서, b 는 유전자형 M_1/M_2 를 가진 부모에게서 환자에게로 형질 M_1 이 전이되는 수이고, c 는 형질 M_2 가 전이되는 수로서 이 통계량이 자유도 1인 카이제곱분포를 따른다는 것을 이용하여 검정한다. 이는 다중대립형질에 대해서도 확장된 바 있다 (Spielman과 Ewens, 1996). TDT는 집단기반의 사례-대조분석과는 달리 집단의 구조적인 문제를 일으키지 않으며, 유전자와 마커사이의 연쇄에 대해 검정력이 높은 방법으로 알려져 있다.

2.2. 형매의 유전자형 정보만 있을 때

부모의 유전자형은 알 수 없고 대신 또 다른 형매(sib)의 유전자형을 알 수 있을 때, 대부분의 경우 환자(affected sib)와 환자가 아닌 형매(unaffected sib)쌍, 즉 불일치 형매쌍(discordant sibpair)에 대한 유전정보를 필요로 한다. 주로 사용되는 방법은 다음과 같다.

2.2.1. S-TDT

Spielman과 Ewens (1998)는 S-TDT(sib transmission/disequilibrium test)를 제안하였다. 이 때 질환이 있는 형매와 질환이 없는 형매가 최소한 1명씩 있어야 하며, 각자의 유전자형이 동일해서는 안된다는 최소요건(minimal configuration)을 갖추어야 한다. 귀무가설은 “연쇄와 연관이 모두 없다”이다. 요건을 충족하는 n 개 가족이 있을 때 모든 환자에서 관찰된 M_1 형질의 총수를 Y 라고 하면, 이의 기댓값은 다음과 같이 구해진다.

$$A = \sum_{i=1}^n \frac{(2r_i + s_i)a_i}{t_i}$$

이다. 여기서 r_i 는 i 번째 가족에서 M_1/M_1 유전자형을 가진 형매의 수이고, s_i 는 M_1/M_2 유전자형을 가진 형매의 수이며, a_i 는 환자인 형매수, t_i 는 총 형매의 수이다. 또한 이의 분산은

$$V = \sum_{i=1}^n a_i u_i \frac{4r_i j_i + s_i(t_i - s_i)}{t_i^2(t_i - 1)}$$

이다. 여기서 j_i 는 i 번째 가족에서 M_2/M_2 유전자형을 가진 형매의 수이며, u_i 는 환자가 아닌 형매의 수이다. S-TDT의 검정 통계량은

$$Z = \frac{Y - A}{\sqrt{V}} \sim N(0, 1)$$

이 된다. Z 가 귀무가설하에서 근사적으로 표준정규분포를 따르는 것을 이용하여 가설검정한다. 각 가족에서 질환이 있는 형매와 없는 형매가 한 쌍만 있을 때에는 연쇄와 연관에 대해 모두 검정할 수 있지만, 여러 쌍이 있는 대가족에서는 연쇄에 대해서만 유효하다. Monks 등 (1998)은 S-TDT를 다중 대립형질로 일반화한 MS-TDT를 제안하였다.

2.2.2. SDT

Horvath와 Laird (1998)의 SDT(sibship disequilibrium test)는 S-TDT와 마찬가지로의 최소요건이 요구되며, 모든 형매가 동일한 유전자형을 갖지 않는다면 가족내 모든 형매의 정보를 이용할 수 있다. SDT는 형매 사이의 상관관계로 인해 생기는 문제를 피할 수 있으며, TDT와 유사한 형태의 검정통계량을 갖는다. r_i, s_i, a_i, u_i 를 S-TDT에서와 같이 정의하고, Y_i 를 i 번째 가족에서 환자에게서 관찰된 M_1 형질의 수라고 하자. i 번째 가족에서

$$\begin{aligned} m_a^1 &= \frac{Y_i}{a_i}, \\ m_u^1 &= \frac{(2r_i + s_i) - Y_i}{u_i}, \\ d^1 &= m_a^1 - m_u^1 \end{aligned}$$

로 정의하자. d^1 이 0보다 큰 형매의 수를 b_1 , d^1 이 0보다 작은 형매의 수를 c_1 이라 하고, d^1 이 0인 경우는 제외한다. SDT의 검정통계량은 다음과 같이 계산할 수 있다.

$$SDT = \frac{(b_1 - c_1)^2}{b_1 + c_1}.$$

연쇄가 없다는 귀무가설하에서 SDT가 자유도 1인 카이제곱 분포를 따른다는 것을 이용하여 검정한다. Horvath와 Laird (1998)는 이를 다중대립형질일때로 확장한 multiallelic SDT도 제안하였다.

2.2.3. DAT

Boehnke와 Langefeld (1998)의 DAT(discordant alleles test) 역시 S-TDT와 마찬가지로의 최소요건이 요구된다. DAT는 질병과 마커가 연관이 없다는 귀무가설에 대하여 환자와 환자가 아닌 형매의 대립형질의 차이를 통해 검정을 한다. N 개의 환자와 환자 아닌 형매로 이루어진 가족이 있다고 하자. M_1, M_2 의 2개 대립형질이 있을 때, n_{1j} 를 N 명의 환자 중 대립형질 M_j 의 수라 하고, n_{2j} 를 N 명의 환자 아닌 형매 중 대립형질 M_j 의 수라 하자. DAT의 검정통계량은 다음과 같다.

$$DAT = \sum_{j=1}^2 \frac{(n_{1j} - n_{2j})^2}{n_{1j} + n_{2j}}.$$

귀무가설하에서 DAT가 자유도가 1인 카이제곱분포를 따른다는 것을 이용하여 검정한다. 이는 다중대립형질로 확장가능하다.

2.2.4. AS-TDT

위의 방법들은 현재 시점까지의 발병 여부로 분석을 하기 때문에 차후에 생길지도 모르는 상황이 무시된다. 이러한 점을 고려하여 Ghosh와 Reich (2004)는 발병 시점의 나이를 보정한 AS-TDT(age adjusted sib TDT)를 제시하였다. AS-TDT에서는 Cox의 비례 위험 모델을 사용하여 발병나이를 추정하고, 환자가 아닌 형매의 수인 u_i 에서 생존 함수의 추정치를 빼서 보정한다. 이 값을 u_i^* 라고 하고, 환자의 수는 각 가족의 형매의 총 수인 t_i 에서 u_i^* 를 뺀 값으로 a_i^* 로 정한다. S-TDT에서와 같이 전체 평균

$$A = \sum_{i=1}^n \frac{(2r_i + s_i)a_i^*}{t_i}$$

와 전체분산

$$V = \sum_{i=1}^n a_i^* u_i^* \frac{4r_i j_i + s_i(t_i - s_i)}{t_i^2(t_i - 1)}$$

를 구한다. 여기서 Y, r_i, s_i, a_i, u_i 는 S-TDT에서와 같다. AS-TDT는 다음과 같은 검정통계량

$$Z = \frac{Y - A}{\sqrt{V}}$$

이 연쇄가 없다는 귀무가설하에서 표준 정규 분포를 따르는 것을 이용하여 검정한다.

2.2.5. 기타

이 밖에 Curtis (1997)의 경우에는 각 가족에서 환자인 형매(case)를 랜덤하게 선택하고 이와 가장 유전자형이 차이가 나는 환자가 아닌 형매(control)를 한 쌍으로 선택한다. 이렇게 함으로써 여러 형매의 사용으로부터 오는 상관문제를 피한다. 그러나 가족내 여러 명의 환자가 있을 경우에 정보의 손실을 갖고 오게 되며, 가족내에 환자와 환자가 아닌 형매가 한 쌍만 있을 때에는 S-TDT와 동일하게 된다. Rieger 등 (2001)은 여러 쌍의 형매가 있을 때 이들 중에서 불일치쌍을 랜덤하게 재추출(resampling)하고 결과를 평균내는 방법인 WSPR(within sibship paired resampling)방법을 제안하였고, 최근 Xu와 George (2007)은 환자인 형매쌍의 공동전이상태(joint transmission status)를 이용하는 ASP-TDT를 제시한 바 있다.

2.3. 부모/형매의 유전자형 정보가 있을 때

2.3.1. 1-TDT

Sun 등 (1999)은 부모 중 한 명의 유전자정보만 있고, 환자가 아닌 형매는 없을 때 사용가능한 1-TDT를 제안하였다. 1-TDT를 하기 위해서는 i) 같은 유전자형을 갖고 있는 남녀는 동일한 교배 우선권(mating preference)을 가지며, ii) 둘 중 하나가 결측일 때 부와 모가 결측일 확률은 1/2로 동일하다는 2개의 가정을 전제로 한다. A_{ij} 를 i 번째 유전자형을 가진 환자의 부모가 j 번째 유전자형을 갖는 수라고 할 때 다음의 T_1 값을 계산한다.

$$T_1 = \frac{b - c}{\sqrt{b + c}},$$

여기서 $b = A_{01} + A_{12}$, $c = A_{10} + A_{21}$ 이며, T_1 은 연쇄와 연관이 없다는 귀무가설하에서 근사적으로 표준 정규 분포를 따르게 된다. 만약 위의 가정이 성립하지 않을 때는 T_1 대신 검정력이 약간 떨어지는 T_2 를

사용한다. T_2 의 귀무가설은 “연쇄가 없다”이다. P 와 M 을 각각 아버지, 어머니가 활용 가능한 가족의 수라고 하고, P_{ij} 와 M_{ij} 를 각각 환자가 i 번째 유전자형, 아버지나 어머니가 j 번째 유전자형을 갖는 환자 수라고 하자.

$$T_2 = \frac{M(pb_1 - pc_1) + P(mb_1 - mc_1)}{\sqrt{M^2 \times P(pb_1 + pc_1) + P^2 \times M(mb_1 + mc_1)}}$$

로 계산된다. 여기서 $pb_1 = P_{01} + P_{12}$, $pc_1 = P_{10} + P_{21}$, $mb_1 = M_{01} + M_{12}$, $mc_1 = M_{10} + M_{21}$ 이다. T_2 도 근사적으로 표준 정규 분포를 따르게 된다.

2.3.2. RC-TDT

Knapp (1999)은 부모 중 하나 혹은 두 명의 유전자형을 알 수 없을 때 자식들의 유전자형을 이용하여 부모의 유전자형을 재구성하여 연쇄가 없다는 귀무가설에 대해 검정하는 방법인 RC-TDT(reconstruction combined TDT)를 제안하였다. 그는 각 가족을 i) 부모 유전자형 정보가 모두 있고 최소 한 명의 부모가 이형접합인 상황, ii) 부모 중 한 명의 유전자형 정보가 있지만 결측된 부모의 유전자형 정보를 재구성할 수 있고 그 중 한 명의 부모가 이형접합인 상황, iii) 부모의 유전자형 정보가 모두 없지만 모두 재구성 가능하고, 그 중 한 명의 부모가 이형접합인 상황, iv) 최소 1명의 부모가 결측이고 재구성할 수 없지만 S-TDT의 조건을 만족하는 상황, v) 위 4가지 범주에 모두 속하지 않는 상황의 5가지 경우로 구분하여 v)의 상황은 분석에서 제외하고 나머지 상황에 대해서 환자인 자식에서 대립형질 M_1 의 수 T_i 의 기대값 e_i 와 분산 v_i 의 계산식을 표로 제시하였다. RC-TDT의 검정 통계량은 i)에서 iv)까지의 범주에 드는 모든 가족에 대해 합산한

$$Z = \frac{\sum(T_i - e_i)}{\sqrt{\sum v_i}}$$

로 주어진다. 귀무가설하에서 이 검정 통계량이 표준 정규 분포를 따르게 됨을 이용한다.

2.3.3. TDTDS

Deng 등 (2002)의 TDTDS(transmission disequilibrium test with discordant sib pairs)는 부모 자료와 형매 자료를 모두 이용하는 방법이다. 먼저 환자에게 M_1 형질이 전이된 수를 n_{11} 이라 하고, 환자가 아닌 형매에게 M_1 형질이 전이된 수를 n_{21} 이라 한다. 그리고 n_{11} 과 n_{21} 을 더한 값, 즉 전체 자료에서 M_1 형질의 전이 수를 n_{+1} 이라 한다. M_2 형질이 환자에게 전이된 수를 n_{12} , M_2 형질이 환자가 아닌 형매에게 전이된 수를 n_{22} 라 하고, n_{12} 와 n_{22} 를 합한 값, 즉 전체 자료에서 M_2 형질이 전이된 수를 n_{+2} 라고 한다. 그리고 환자의 수를 n_{1+} , 환자가 아닌 형매의 수를 n_{2+} 이라 하고 전체 형매의 수를 n_{++} 라고 한다. 이때 TDTDS의 검정통계량은 다음과 같다.

$$\text{TDTDS} = \frac{n_{++}(n_{11} \times n_{22} - n_{21} \times n_{12})^2}{n_{1+} \times n_{2+} \times n_{+1} \times n_{+2}}.$$

검정 통계량은 연쇄가 없다는 귀무가설하에서 자유도가 1인 카이 제곱 분포를 따르게 된다.

3. 모의 실험 및 결과

3.1. 모의 실험의 구조

위에서 소개한 방법 중 많이 사용되는 방법 위주로 8가지 방법을 선택하여 검정력을 비교하였다. 질병 모형으로 열성모형(recessive model)과 우성모형(dominant model)을 고려하였고, 각 모형마

다 완전투과(complete penetrance)와 감소투과(reduced penetrance)를 고려하였다. 여기서 열성완전투과모형은 유전자형이 M_1/M_1 인 경우에 무조건 발병하는 경우이고, 열성감소투과모형은 유전자형이 M_1/M_1 인 경우 50%의 확률로 발병하는 경우를 의미한다. 우성완전투과모형은 유전자형이 M_1/M_1 , M_1/M_2 즉, M_1 대립 형질을 가지면 무조건 발병하는 경우이고, 우성감소투과는 M_1/M_1 의 경우 80%의 확률로, M_1/M_2 인 경우에는 50%의 확률로 발병하는 경우로 하였다. 각 모형마다 대립형질빈도와 LD의 수준을 (0.5, 0.2), (0.5, 0.1), (0.3, 0.1), (0.2, 0.1)로 설정하여 각각의 유의수준과 검정력을 비교하였다. 모두 1000개의 데이터셋을 생성하였고, 각 데이터 셋마다 다음의 절차를 따라서 데이터가 생성되었다. 하나의 데이터셋마다 500개의 가족을 생성한 후, 환자 1명과 환자가 아닌 형제 1명이 있는 최소요건을 만족하는 가족만을 남겨 최종 데이터셋으로 결정하였다. 발병 나이의 생성은 Ghosh와 Reich (2004)의 방법과 같이 실시하였다.

(단계 1) 부모 각각의 마커 유전자형은 $M_1/M_1 = 0.25$, $M_1/M_2 = 0.5$, $M_2/M_2 = 0.25$ 의 확률을 갖도록 trinomial분포의 난수 생성을 통하여 500개씩 생성하였다. 각 가족에서 한 명씩의 형제에 대해 질병 인자(disease mutation)의 유무를 유병률(lifetime prevalence)의 확률을 갖는 베르누이 분포에 따라서 생성을 한다. 유병률은 질병 모형에 따라 변하게 되는데, 이는 다음과 같다.

$$\Delta = p^2 f_1 + 2pq f_2 + q^2 f_3,$$

여기서, Δ 는 유병률이고, p 와 q 는 각각 M_1 형질과 M_2 형질의 빈도이다. f_1 , f_2 , f_3 는 각각 M_1/M_1 , M_1/M_2 , M_2/M_2 유전자형의 발병률(lifetime penetrance)이다.

(단계 2) 한 명의 형제 자료의 질병 인자 유무 여부에 따라 각각 다른 확률로 나머지 한 명의 형제에 대한 질병 인자 유무 여부를 결정한다.

(단계 3) 각 형제의 유전자형을 결정한다. 부모의 유전자형이 동형 접합이면 같은 대립 형질이 전이될 것이다. 환자인 경우 부모의 유전자형이 이형 접합인 경우에 질병 인자 유무 여부에 따라 전이되는 대립 형질을 결정하는데, Ghosh와 Reich (2004)의 방법으로 결정한다.

(단계 4) 형제 자료에 대한 현재 나이는 각 가족의 형제 자료 중 한명의 나이를 균일 분포 [51, 70]에서 생성하고, 이를 R 이라고 한다. 나머지 한 명의 나이는 균일 분포 $[R - 10, R + 10]$ 에서 생성한다. 그리고 형제 자료에서 발병 나이를 생성하기 위해 Z_1 , Z_2 , Z_3 3개의 공변량을 생성한다. Z_1 은 Bernoulli($p = 0.5$)에서 생성한다. 만약 Z_1 이 0이면 Z_2 , Z_3 는 평균 0, 분산 1, 상관이 0.7인 이변량 정규분포를 따르는 난수를 생성하고, Z_1 이 1인 경우의 Z_2 , Z_3 는 평균 0, 분산 1, 상관이 0.5인 이변량 정규 분포에서 난수를 생성을 한다. 이 자료를 사용하여 형태모수 $p = 4$ 인 와이블분포를 가정하고, $[R - 10]$ 이하를 절단한 Cox의 비례 위험 모형을 사용하여 발병 나이를 생성한다. 이때 회귀 계수 beta는 (1, 0.5, 0.5)로 고정한다.

(단계 5) 각 형제 자료의 질병 여부는 질병 인자를 갖고 있는 형제 자료 중 발병 나이가 현재 나이보다 작은 경우에 질병을 갖는 것으로 하고, 나머지는 발병이 되지 않은 것으로 한다.

3.2. 모의 실험의 결과 및 해석

모의실험 결과는 표 1에서 표 2에 있는데, 유의수준 5%에서 양측검정을 1000회 시험하였을 때의 기각횟수이다. 표에서 Group I(TDT)은 부모자의 유전자형 정보가 있는 경우, Group II(S-TDT, SDT, DAT)는 형제의 유전자형 정보만 있는 경우, Group III(1-TDT, TDTDS, RC-TDT1, RC-TDT2)는 부

표 1: 모 형과 방법에 따른 유의수준($\theta = 0.5$)

Group			I		II			III		
Model	P	LD	TDT	S-TDT	SDT	DAT	1-TDT	TDTDS	RC-TDT1	RC-TDT2
REC CP	0.5	0.2	0.045	0.032	0.044	0.060	0.072	0.060	0.036	0.078
	0.5	0.1	0.057	0.039	0.057	0.077	0.057	0.077	0.051	0.076
	0.3	0.1	0.059	0.045	0.063	0.087	0.061	0.087	0.066	0.232
	0.2	0.1	0.047	0.025	0.048	0.069	0.050	0.069	0.047	0.044
REC RP	0.5	0.2	0.047	0.034	0.047	0.067	0.053	0.067	0.049	0.108
	0.5	0.1	0.049	0.037	0.048	0.067	0.063	0.067	0.046	0.102
	0.3	0.1	0.059	0.032	0.056	0.070	0.059	0.070	0.049	0.362
	0.2	0.1	0.041	0.016	0.044	0.064	0.046	0.064	0.039	0.564
DOM CP	0.5	0.2	0.062	0.045	0.052	0.079	0.059	0.079	0.052	0.049
	0.5	0.1	0.052	0.049	0.058	0.073	0.046	0.073	0.056	0.040
	0.3	0.1	0.047	0.042	0.047	0.079	0.053	0.079	0.052	0.051
	0.2	0.1	0.058	0.037	0.047	0.070	0.063	0.070	0.050	0.059
DOM RP	0.5	0.2	0.046	0.041	0.043	0.063	0.042	0.063	0.043	0.047
	0.5	0.1	0.054	0.037	0.037	0.064	0.052	0.064	0.044	0.046
	0.3	0.1	0.051	0.037	0.045	0.069	0.053	0.069	0.045	0.051
	0.2	0.1	0.049	0.027	0.041	0.057	0.056	0.057	0.038	0.062

P : allele frequency

LD : linkage disequilibrium

REC CP : recessive complete penetrance

REC RP : recessive reduced penetrance

DOM CP : dominant complete penetrance

DOM RP : dominant complete penetrance

 θ : recombination fraction

모/형매의 유전자형 정보가 있는 경우이다. Group II의 경우에는 분석에 사용되는 데이터가 완전히 동일하여 직접적인 비교도 가능하다. Group III의 경우에는 방법에 따라 데이터가 갖는 정보에 차이가 있다. Knapp이 제안한 RC-TDT에 대해서는 다음 두 가지 경우로 나눠서 비교하였다. 부모 중 1명을 모르는 경우 재건한 방법을 RC-TDT1이라 하였고, 부모 2명 다 모르는 경우 재건한 방법을 RC-TDT2라고 하였다. 1-TDT의 경우는 Bernoulli ($p = 0.5$)의 분포를 사용하여 부와 모를 선택하여 선택된 부나 모의 자료와 환자의 자료를 사용하였다. TDTDS의 경우에는 부모와 형매의 자료를 모두 사용하였다.

표 1은 각 모형의 유의수준을 보여준다. 이는 재조합율이 0.5로 질병유전자와 마커가 독립인 경우로서 연관 또는 연쇄가 없다는 귀무가설을 만족하는 경우의 기각확률을 의미한다. 먼저 열성완전투과(REC CP)의 상황에서는 TDT와 SDT, RC-TDT1 방법이 모든 수준에서 어느 정도 유의 수준을 맞추고, S-TDT는 유의수준보다 약간 작게 나타났다. DAT와 1-TDT, TDTDS와 RC-TDT2의 경우에는 유의 수준보다 약간 높게 나타났다. 부모 두 명을 모두 추정하는 RC-TDT2에서 ($P = 0.3, LD = 0.1$)수준인 경우는 유의수준이 큰 값으로 나타났다. 우성완전투과(DOM CP)에서도 DAT와 TDTDS는 약간 높은 유의수준을 보임을 확인할 수 있다.

표 2에서는 모든 모형에서의 각 수준별 검정력을 보여준다. 여기서는 대립가설을 재조합율이 0.01로서 거의 재조합이 발생되지 않는 비율로 정하였다. 우성모형과 열성모형을 비교해보면, 대립형질의 빈도가 0.3과 0.2인 경우에는 완전투과와 감소투과 각각에서 모두 우성모형의 검정력이 모든 방법에서 높지만, 대립형질의 빈도가 0.5로 높은 경우에는 열성모형의 검정력이 더 높은 경향이 있는 것으로 나타났다. 이는 대립형질의 빈도에 따라 모형별 검정력이 달라질 수 있음을 의미한다고 볼 수 있을 것이다. 한편 열성모형의 경우에는 완전투과 모형이 감소투과모형보다 항상 높은 검정력을 갖는 것을 확인할 수 있다. 그러나 우성모형의 경우에는 대립형질의 빈도가 0.2, 0.3인 경우에는 열성 모형과 같

표 2: 모형과 방법에 따른 검정력($\theta = 0.01$)

Group			I		II			III		
Model	P	LD	TDT	S-TDT	SDT	DAT	1-TDT	TDTDS	RC-TDT1	RC-TDT2
REC CP	0.5	0.2	0.997	0.995	0.998	0.999	0.961	0.999	0.999	0.525
	0.5	0.1	0.742	0.721	0.753	0.798	0.395	0.798	0.765	0.162
	0.3	0.1	0.786	0.686	0.761	0.798	0.482	0.798	0.743	0.182
	0.2	0.1	0.792	0.628	0.729	0.791	0.522	0.791	0.727	0.320
REC RP	0.5	0.2	0.960	0.953	0.968	0.972	0.839	0.973	0.966	0.303
	0.5	0.1	0.506	0.437	0.483	0.544	0.249	0.544	0.483	0.132
	0.3	0.1	0.528	0.374	0.474	0.541	0.300	0.541	0.476	0.266
	0.2	0.1	0.535	0.327	0.456	0.580	0.310	0.580	0.442	0.450
DOM CP	0.5	0.2	0.947	0.803	0.817	0.871	0.855	0.871	0.828	0.241
	0.5	0.1	0.455	0.280	0.294	0.370	0.345	0.370	0.300	0.088
	0.3	0.1	0.848	0.720	0.736	0.808	0.721	0.808	0.747	0.219
	0.2	0.1	0.953	0.921	0.926	0.946	0.795	0.946	0.934	0.303
DOM RP	0.5	0.2	0.976	0.961	0.963	0.978	0.900	0.978	0.969	0.489
	0.5	0.1	0.544	0.466	0.483	0.552	0.367	0.552	0.492	0.144
	0.3	0.1	0.736	0.697	0.715	0.787	0.457	0.787	0.727	0.193
	0.2	0.1	0.853	0.826	0.836	0.895	0.583	0.895	0.861	0.231

P : allele frequency

LD : linkage disequilibrium

REC CP : recessive complete penetrance

REC RP : recessive reduced penetrance

DOM CP : dominant complete penetrance

DOM RP : dominant complete penetrance

 θ : recombination fraction

이 완전투과상황이 감소투과상황보다 높은 검정력을 갖는데 반하여, 대립형질의 빈도가 0.5인 경우에는 감소투과상황이 완전투과상황보다 높은 검정력을 갖는 것을 알 수 있다. 이러한 결과는 감소투과율과 대립형질의 빈도의 조합에 따라 검정력이 달라진 것으로 추측된다.

방법별로 비교하자면 RC-TDT2의 경우를 제외하고, ($P = 0.5$, $LD = 0.2$) 수준에서의 모든 방법에서 검정력이 0.9이상으로 매우 좋은 것을 알 수 있다. 모든 수준에서 TDT와 S-TDT, SDT의 관계를 비교하자면 항상 TDT가 가장 좋고, SDT와 S-TDT는 비슷하지만 SDT가 약간 더 좋은 것을 확인할 수 있다. 또한 1-TDT는 S-TDT나 SDT의 검정력보다 떨어짐을 알 수 있다. 이로부터 부모자료 대신 형매자료를 사용하는 것도 어느 정도 경쟁력이 있다고 생각된다. 부모자료와 형매자료 모두를 사용한 TDTDS도 검정력은 높았지만, 유의수준을 잘 맞추지 못한다는 단점이 있다. RC-TDT의 경우 부모를 모두 추정(RC-TDT2)하게 되면 부모 중 한명의 정보를 가지고 하나를 추정하는 것(RC-TDT1)보다 현저히 검정력이 떨어짐을 확인할 수 있다. RC-TDT1의 경우 SDT와 비슷한 검정력을 갖는 것을 알 수 있었고, S-TDT보다 검정력이 높았다. 추가로 대립형질 빈도와 연쇄 불균형의 수준이 ($P = 0.5$, $LD = 0.2$)인 경우에서의 AS-TDT의 유의수준과 검정력을 구하였다. 유의수준은 5% 이내로 만족되었으며, 검정력은 열성완전투과, 열성감소투과, 우성완전투과, 우성감소투과에서 각각 0.995, 0.953, 0.809, 0.959로 S-TDT와 거의 비슷하게 나타났다.

4. 토의

가족기반 유전연구에서 단일염기다형성을 이용한 질병-유전자간 연관성 분석은 고전적인 TDT방법을 많이 사용하고 있다. 그러나 질병의 발생 나이가 늦어서 부모의 유전자형에 대한 정보를 구하기

그림 1: 가족 기반 연구에서 자료 형태에 따른 분석 방법

어려운 경우에는 형매의 유전정보를 이용한 분석방법을 쓸 수 있다. 또한 확보가능한 일부 부모의 정보를 이용하는 방법도 개발되었다. 그림 1은 각 상황에 맞춰 쓸 수 있는 분석방법을 정리한 것이다.

본 연구에서는 여러 방법들을 소개하고 이들에 대한 모의실험을 실시하여 검정력을 비교해보았다. 각 방법론에서 필요한 데이터의 형태가 서로 다르므로, 완전히 동일한 조건에서의 비교는 일부(Group II)만이 가능하지만, 가족의 데이터를 모의생성하고 경우에 따라 일부 데이터만 사용할 수 있는 실제의 경우에 각 방법론의 검정력을 비교하는 것도 의미가 있을 것으로 생각된다. 모의실험결과 TDT가 대부분의 경우에 좋은 검정력을 보였지만, 부모자의 데이터를 확보하기 어려울 때 사용할 수 있는 다른 대체 방법들도 상당한 검정력을 갖는 것을 확인하였다. SDT와 S-TDT의 경우 검정력이 비슷하며 상당히 높은 수준의 검정력을 보였다. DAT와 TDTDS의 경우에 검정력은 이들과 비슷하지만 유의수준을 맞추지 못하였으며, 1-TDT의 검정력은 이들과 현저히 낮게 나타났다. RC-TDT의 경우 한 부모의 정보가 있고 재건 가능할 때에는 SDT와 비슷한 수준의 검정력을 가지며 S-TDT보다 약간 높은 것으로 나타났지만, 부모의 자료를 모두 복원해야 하는 경우에는 검정력이 현저히 떨어짐을 보였다. 생존함수를 고려한 AS-TDT의 경우에는 비록 일부 결과만을 구하였으나, S-TDT와 유사한 높은 검정력을 보였다. 그러나 보정값을 구하기 위해서는 상당한 수준의 재표본이 추가로 수행되어야 하므로, 다른 방법론에 비해 계산에 많은 시간이 소요되었다. 그러나 그에 비해 보정되는 값의 크기가 미미하고, 검정력의 차이도 거의 없어서 효율성이 낮은 것으로 생각된다.

열성모형의 경우에는 완전투과모형이 감소투과모형보다 항상 높은 검정력을 갖는 반면, 우성모형의 경우에는 대립형질의 빈도가 0.5인 경우에는 감소투과모형이 완전투과모형보다 검정력이 높아지는 특징도 발견하였다. 이러한 점을 고려하여, 각자의 상황에 맞는 검정법을 선택할 수 있을 것이다.

참고 문헌

- 김진흠, 장양수 (2008). 형제 및 자매의 유전자형 자료에 기초한 전달불균형 검정법에 관한 연구, <응용통계연구>, **21**, 81-94.
- 이효정, 김민지, 박미라 (2010). 집단 및 가족기반연구에서의 유전적 연관성 분석 고찰: 방법론과 소프트웨어, <응용통계연구>, **23**, 95-111.
- Boehnke, M. and Langefeld, C. D. (1998). Genetic association mapping based on discordant sib pairs: The discordant - Alleles test, *The American Society of Human Genetics*, **62**, 950-961.
- Curtis, D. (1997). Use of siblings as controls in case-control association studies, *Annals of Human Genetics*, **61**, 319-333.

- Deng, H. W., Chen, W. M. and Recker, R. R. (2002). Transmission disequilibrium test with discordant sib pairs when parents are available, *Human Genetics*, **110**, 451–461.
- Ghosh, S. and Reich, T. (2004). The Sib TDT adjusted for age of disease onset, *Annals of Human Genetics*, **68**, 249–256.
- Horvath, S. and Laird, N. M. (1998). A discordant-sibship test for disequilibrium and linkage: No need for parental data, *The American Society of Human Genetics*, **63**, 1886–1897.
- Knapp, M. (1999). The transmission/disequilibrium test and parental-genotype reconstruction: The reconstruction-combined transmission/disequilibrium test, *The American Society of Human Genetics*, **64**, 861–870.
- Laird, N. M. and Lange, C. (2008). Family-based methods for linkage and association analysis, *Advances in genetics*, **60**, 219–252.
- Monks, S. A., Kaplan, N. L. and Weir, B. S. (1998). A comparative study of sibship tests of linkage and/or association, *American Journal of Human Genetics*, **63**, 1507–1516.
- Rieger, R. H., Kaplan, N. L. and Weinberg, C. R. (2001). Efficient use of siblings in testing for linkage and association, *Genetic Epidemiology*, **20**, 175–191.
- Saito, Y. A., Talley, N., Andrade, M. and Petewrsen, G. (2006). Case-control genetic association studies in gastrointestinal disease: Review and recommendations, *American Journal of Gastrointorology*, **101**, 1379–1389.
- Spielman, R. S. and Ewens, W. J. (1996). The TDT and other family-based tests for linkage disequilibrium and association, *The American Society of Human Genetics*, **59**, 983–989.
- Spielman, R. S. and Ewens, W. J. (1998). A Sibship Test for linkage in the presence of association: The sib transmission/disequilibrium test, *The American Society of Human Genetics*, **62**, 450–458.
- Spielman, R. S., McGinnis, R. E. and Ewens, W. J. (1993). Transmission test for linkage disequilibrium: The Insulin gene region and insulin-dependent diabetes mellitus(DDM), *The American Society of Human Genetics*, **52**, 506–516.
- Sun, F., Flanders, W. D., Yang, Q. and Khoury, M. J. (1999). Transmission disequilibrium test(TDT) when only one parent is available: The 1-TDT, *American Journal of Epidemiology*, **150**, 97–104.
- Xu, H. and George, V. (2007). A new transmission test for affected sib-pair families, *BMC Proceedings*, **1(Suppl 1)**, S32.
- Zhao, H. (2000). Family-based association studies, *Statistical Methods in Medical Research*, **9**, 563–587.

Comparison of Genetic Association Analyses Using Sib Information in Family-Based Study

Ji Won Jung^a, Suk Hoon Lee^a, Soon Ae Kim^b, Mira Park^{1,c}

^aDepartment of Statistics, Chungnam National University

^bDepartment of Pharmacology Medicine, Eulji University

^cDepartment of Preventive Medicine, Eulji University

Abstract

Recently, disease-genetic association analyses using single nucleotide polymorphisms(SNPs) and haplotypes in family-based genetic study have come into the spotlight. In binary trait, the classic transmission disequilibrium test(TDT) can only be applied if genetic information of parents and their offspring is available. However, in case of diseases having a late age of onset such as dementia, the TDT cannot be applied due to the fact that parental genotype data are unavailable. For this reason, alternate methods using genetic sib information instead of parental genotype data are proposed. In this study, methods using genetic sib information are reviewed and power of analysis tests is also compared throughout simulation experiment.

Keywords: Genetic association, TDT, SNP, simulation.

This work was supported by the Korea Research Foundation Grant funded by the Korean Government(MOEHRD, Basic Research Promotion Fund)(KRF-2007-531-C00018).

¹ Corresponding author: Professor, Department of Preventive Medicine, Eulji University, Yongdu-Dong, Jung-Gu, Daejeon 301-832, Korea. E-mail: mira@eulji.ac.kr