예방의학회지 제40권 제2호(2007년 3월)
J Prev Med Public Health 2007;40(2):114-121

특 집

# 유전체 연관 연구에서의 검정력 및 연구대상수 계산 고찰

박애경, 김 호

서울대학교 보건대학원

## A Review of Power and Sample Size Estimation in Genomewide Association Studies

Ae Kyung Park, Ho Kim

Graduate School of Public Health, Seoul National University

Power and sample size estimation is one of the crucially important steps in planning a genetic association study to achieve the ultimate goal, identifying candidate genes for disease susceptibility, by designing the study in such a way as to maximize the success possibility and minimize the cost. Here we review the optimal two-stage genotyping designs for genomewide association studies recently investigated by Wang et al(2006). We review two mathematical frameworks most commonly used to compute power in genetic association studies prior to the main study: Monte-Carlo and non-central chi-square estimates. Statistical powers are computed by these two approaches for case-control genotypic tests under one-stage direct association study design. Then we discuss how the linkage-disequilibrium strength affects power and sample size, and how to use empirically-derived distributions of important parameters for power calculations. We provide useful information on publicly available softwares developed to compute power and sample size for various study designs.

*J Prev Med Public Health 2007;40(2):114-121*

**Key words :** Research design, Sample size, Genomics, Genetic screening

## 서 론

유전 연관 연구(genetic association study)에서 적절한 연구 설계와 대상자수 (표본수)를 결정하여 충분한 검정력(power)을 확보하는 것은 연구의 성과를 좌우하는 기본이 되며 궁극적으로 특정 질환과 관련되어 있는 유전자를 찾는다는 최종 목적을 달성하기 위하여 매우 중요하다. 유전 연관 연구는 접근하는 방법에 따라 "직접적"인 방법과 "간접적"인 방법으로 나눌 수 있다 [1,2]. 직접적인 방법은 특정 질환과 관련되어 있을 것이라고 추정되는 원인 변이(causal variant)에 대한 정보를 바탕으로 원인 변이와 특정 질환과의 직접적인 연관을 검정하는 것이고, 간접적인 방법은 원인 변이에 대한 사전 정보 없이 연관불평형(linkage disequilibrium, LD)에 기반을 두고, 여러 일반 변이(common variant)를 표지자(marker)로 이용하여 실제 특정질환과 관련된 원인 변이에 대한 정보를 얻고자 하는 것이다. 이러한 간접적 연관 연구의 궁극적인 형태는 whole-genome scan이라고 할 수 있는데 이것은 원인 변이를 찾기 위해 전체 유전체(genome)를 조사하는 것이다. 이 때 표지자들이 얼마나 조밀하게 분포되어 있는지가 문제 된다. 현재 개발되어 있는 가장 고밀도인 SNP (Single Nucleotide Polymorphism) 칩의 경우 약 500,000개의 SNP 표지자를 사용하는데 이것은 유전체 6kb당 1개 표지자의 밀도를 갖는다. 그러나 이 경우에도 대부분의 연관은 원인 변이와의 직접적인 연관이 아니라 LD에 의한 간접적인 연관에 의해 나타나게 된다 [1]. 이러한 간접적인 방법을 사용하는 경우, 직접적인 방법을 사용할 때와 동일한 검정력을 얻기 위해서는 더 많은 표본수가 필요하다. 예를 들어 8개의 tag SNP가 평균 연관정도 $r^2 = 0.8$을 나타낼 때 자유도 8인 분석을 사용하는 경우에 간접적인 방법은 직접적인 방법에 비해 약 2배 정도의 표본수를 필요로 하게 된다(1단계 연구 설계의 경우) [1]. SNP genotyping 기술 발전으로 1개 SNP당 genotyping 비용이 획기적으로 낮아졌으나 genomewide 연관 연구에서 비용절감의 문제는 여전히 중요하다. 따라서 genomewide association scan과 관련하여 비용을 최소화 하면서 충분한 검정력을 확보하기 위해 제안된 2단계 연구 설계(two-stage design) [2,3]에 대하여 고찰해 보고자 한다. 이에 앞서 유전 연관 연구에서 검정력(또는 표본수)을 계산하기 위해 사용되는 일반적인 방법인 몬테-칼로 모의실험(Monte-Carlo simulation)과 점근적 비중심 모수(asymptotic non-central parameter) 방법에 대하여 가장 간단

이 연구는 질병관리본부 학술용역사업(2006-347-2400-2440-215)으로 수행된 내용에 근거한 것임.
책임저자 : 김 호(서울시 종로구 연건동 28, 전화 : 02-740-8864, 팩스 : 02-745-9104, E-mail : hokim@snu.ac.kr)

한 형태의 연관 분석(1단계, 단일 유전좌위(single locus), 직접적 방법)을 대상으로 고찰해 보고 [4] 이 두 방법을 이용하여 검정력을 계산하여 제시하였다. 나아가 표지자를 사용한 간접 방법을 사용할 경우 LD 강도에 따라 표본수를 얼마나 증가시켜야 하는지 [5,6], 또한 검정력 추정에 필요한 모수(parameters)를 어떻게 실제적인 값에 가깝게 보정할 수 있는지에 대하여 고찰하였다 [8]. 마지막으로 유전 연관 연구 설계를 위하여 검정력 혹은 표본수를 쉽게 계산할 수 있도록 고안된 소프트웨어를 소개하였다.

## 유전 연관 연구에서 검정력 추정을 위해 일반적으로 사용되는 계산법

검정력이란 대립가설($H_1$)이 참일 때, 귀무가설($H_0$)을 기각시키는 확률로서 실제 특정표지자가 질환과 관련되어 있을 경우 연관이 있다고 결론지을 확률을 말한다. 보통 유전 연관 연구에서 80% 이상의 검정력을 확보하는 것을 목표로 하고 있다. 검정력(또는 표본수)을 계산하기 위해서는 대립가설 하의 검정통계량의 분포를 알아야 한다. 이러한 대립가설 하의 검정통계량 분포를 계산하는 방법으로 흔히 사용되는 것으로 몬테-칼로 모의실험 및 비중심 모수 분포를 이용하는 방법이 있다 [4].

### 1. 단일 유전좌위에 대하여 직접적 연관 분석을 하는 경우의 검정력 계산

환자-대조군 유전 연관 분석에서는 질환과 관련된 여러 유전 요인 (genetic determinants)들이 환자 군에서 더 높은 확률로 발견될 것이라는 가정에 기반을 두고 질환의 유무와 특정 유전좌위와의 연관정도를 측정하게 되는데 이러한 연관 정도는 특정 유전좌위(locus)의 유전자형 빈도(genotypic frequency) 혹은 대립형질 빈도(allelic frequency)를 환자군과 대조군에서 비교 분석함으로써 추정하게 된다. Table 1에서와 같이 유전자형 빈도를 이용한 검정통계량을 중심으로 각 유전좌위가 질환과 관련된 원인 대립형질(causal allele)이거나 혹은 원인 대립형질과 완전한 LD (perfect LD, $r^2$=1) 관계에 있다는 가정 하에 단일 유전좌위를 대상으로 하는 유전 연관 분석에서의 검정력 계산을 알아보기로 한다 [4].

#### 1) 유전 모델

검정통계량을 $S$, 유의수준 $\alpha = P_{H_0}(S \geq t_\alpha)$, 검정력을 $\pi(\alpha) = P_{H_1}(S \geq t_\alpha)$라고 할 때 Hardy Weinberg 평형(HWE) 가정 하에서 유전자형 빈도 및 유전형 투과도(penetrance)는 Table 2와 같다. 이 때 상대위험도 $RRi = f_i / f_0$ ($i$ = 1 또는 2)이며 유전 모델에 따라 다음과 같다.

Recessive model : $RR_1 = 1$

Multiplicative model : $RR_1 = \sqrt{RR_2}$

Additive model : $RR_1 = (RR_2+1)/2$

Dominant model : $RR_1 = RR_2$

유병률을 $K_p$라 하면

$$K_p = r_0 f_0 + r_1 f_1 + r_2 f_2$$

$$f_0 = K_p / (r_0 + RR_1 r_1 + RR_2 r_2)$$

$$f_i = RR_i f_0 (i = 1 \text{ 또는 } 2)$$

이 때 인구수를 무한하다고 가정하면 환자군에서 유전자형 분포(genotype distribution) ($D_0$, $D_1$, $D_2$)와 대조군에서의 유전자형 분포 ($C_0$, $C_1$, $C_2$)는 다음과 같이 다항 분포(multinomial)가 된다.

$$(D_0, D_1, D_2) \sim M\left(n_D; \frac{f_0 r_0}{K_p}, \frac{f_1 r_1}{K_p}, \frac{f_2 r_2}{K_p}\right)$$

$$(C_0, C_1, C_2) \sim M\left(n_C; \frac{(1-f_0) r_0}{(1-K_p)}, \frac{(1-f_1) r_1}{(1-K_p)}, \frac{(1-f_2) r_2}{(1-K_p)}\right)$$

여기에서 귀무가설은 $H_0$:{$RR_2 = 1$}, 대립가설은 $H_1$:{$RR_2 \neq 1$}이 된다. 일단 대립가설이 결정되면 몬테-칼로 모의실험 혹은 점근적 비중심 카이제곱 분포를 이용하여 대립가설 하에서의 검정통계량 $S$의 분포를 계산할 수 있다.

#### 2) 몬테-칼로 모의실험에 의한 검정력 계산

대립가설이 참인 경우를 가정하고 환자-대조군 표본 ($D_0$, $D_1$, $D_2$)와($C_0$, $C_1$, $C_2$), 즉

$$X = \left\{ \begin{matrix} D_0, D_1, D_2 \\ C_0, C_1, C_2 \end{matrix} \right\} \quad (H_1 \text{가정 하})$$

를 생성해 낼 수 있으면 검정력을 구할 수 있다. $N$개의 표본 $x^{(i)}$ ($i$ 번째 표본)를 얻었다고 하면 $N$개의 검정통계량 $S^{(1)}, \cdots, S^{(N)}$이 얻어진다. 이 검정통계량으로부터 계산되는 검정력 $\pi(\alpha)$는 다음과 같다.

$$\hat{\pi}(\alpha) = \frac{\#\{s^{(i)} \geq t_\alpha\}}{N}$$

#### 3) 점근적 비중심 카이제곱 분포를 이용한 검정력 계산

$2 \times c$ 분할표에 사용되는 카이-제곱 빈도 분석의 점근적 분포(asymptotic distribution)는 대립가설 하에서 비중심 카이제곱 분포 $X'^2(k, \lambda)$를 따른다. 여기에서 $k$는 자유도이며 $\lambda$는 비중심 모수 (non-centrality parameter)로서 다음과 같이 계산된다.

$$\lambda = N_1 N_2 \times \sum_{j=1}^{c} \frac{(p_{1j} - p_{2j})^2}{N_1 p_{1j} + N_2 p_{2j}}$$

($p_{ij}$는 case ij 의 빈도, $N_1$, $N_2$는 분할표에서 1행과 2행의 총 합)

이 때 이 분석의 점근적 검정력은

$$\pi(\alpha) \xrightarrow[\infty]{} 1 - \chi'^2_{1-t_\alpha}(k, \lambda)$$

로 유도된다. 따라서 이러한 검정력 계산은 비중심 모수가 주어지면 표본수가 충분히 큰 경우 귀무가설 하에서 카이제곱 분포를 따르는 어떤 검정통계량에든지 적용할 수 있다. 예를 들어 Table 1에서와 같이 환자군

**Table 1.** The genotypic contingency table

| | *aa* | *aA* | *AA* | Total |
|---|---|---|---|---|
| Diseased | $D_0$ | $D_1$ | $D_2$ | $n_D$ |
| Control | $C_0$ | $C_1$ | $C_2$ | $n_C$ |
| Total | $n_0$ | $n_1$ | $n_2$ | n |

과 대조군에서 유전자형 빈도를 비교하는 분석에서 피어슨 카이제곱 통계량 $S_G$와 비중심 모수 $\lambda_G$는 다음과 같이 주어진다.

$$S_G = \sum_{i=0}^{2} \frac{\left(D_i - \frac{n_D \times n_i}{n}\right)^2}{\frac{n_D \times n_i}{n}} + \frac{\left(C_i - \frac{n_C \times n_i}{n}\right)^2}{\frac{n_C \times n_i}{n}} \quad H_0 \chi^2(2)$$

$$\lambda_G = n_D n_C \times \sum_{i=0}^{2} \frac{\left(\frac{f_i r_i}{K_p} - \frac{(1-f_i) r_i}{1-K_p}\right)^2}{n_D \frac{f_i r_i}{K_p} + n_C \frac{(1-f_i) r_i}{1-K_p}}$$

이 때 대립가설 하의 검정통계량의 분포는 다음과 같이 자유도 2, 비중심 모수 $\lambda_G$ 인 카이제곱 분포를 따르게 된다.

$$S_G \; H_1 \; \chi^2(2, \lambda_G)$$

## 2. Genomewide 연관 연구에서 Bonferroni 보정

위에서 살펴본 검정력 계산은 단일 유전좌위를 대상으로 한 것으로 이것은 SNP 수에 따라 제1종 오류율을 보정함으로써 genomewide 연관 연구로 확대될 수 있다. 전체 제1종 오류율을 0.05로 유지하기 위해서는 Table 3과 같이 SNP 수에 따라 Bonferroni 보정 $p$값을 사용 할 수 있다. 이 경우 각 유전좌위가 독립적이라는 가정이 전제되는데 실제로 각 유전좌위의 검정통계량은 서로 연관 되어 있는 경우가 많아 Bonferroni 보정은 너무 엄격(conservative)하다는 비판을 받기도 한다 [1]. Table 4 와 5는 Bonferroni 보정 $p$값을 사용하여 SNP 수, 원인 대립형질 빈도 및 유전 모델에 따

**Table 2.** Genetic model under Hardy-Weinberg equilibrium(HWE)

| Genotype | Frequency in population under HWE | Conditional probability(penetrance) | |
|---|---|---|---|
| | | Affected | Unaffected |
| *AA* | $r_2 = p^2$ | $f_2$ | $1-f_2$ |
| *Aa* | $r_1 = 2p(1-p)$ | $f_1$ | $1-f_1$ |
| *aa* | $r_0 = (1-p)^2$ | $f_0$ | $1-f_0$ |

**Table 3.** Type I error rates for genomewide association studies involving different numbers of SNP markers when a 3 billion-bp genome is assumed

| Interval(kb) | No. of SNP markers | p |
|---|---|---|
| 3 | 1,000,000 | $5\times10^{-8}$ |
| 6 | 500,000 | $1\times10^{-7}$ |
| 10 | 300,000 | $1.67\times10^{-7}$ |
| 20 | 150,000 | $3.33\times10^{-7}$ |
| 40 | 75,000 | $6.67\times10^{-7}$ |
| 80 | 37,500 | $1.33\times10^{-6}$ |
| 160 | 18,750 | $2.67\times10^{-6}$ |

**Table 4.** Power (1-$\beta$) estimation for case-control genotypic tests according to numbers of SNPs, allele frequencies ($p$), case sample sizes ($n_D$) and genetic models. Power is computed by Monte-Carlo simulation for prevalence $Kp$=0.05, $n_D$=$n_C$, odds ratio $RR_2$=1.5, and $\alpha$=0.05/No. of SNPs (Bonferroni-corrected type I error rate). Each Monte-Carlo estimate of power is carried out on the basis of N=1,000 simulations

| $p$ | Number of SNPs | Additive model | | | | | Multiplicative model | | | | | Dominant model | | | | | Recessive model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | |
| | | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 |
| 0.1 | 100 | 0.07 | 0.30 | 0.57 | 0.78 | 0.90 | 0.05 | 0.21 | 0.46 | 0.67 | 0.84 | 0.60 | 0.96 | 1.00 | 1.00 | 1.00 | 0.00 | 0.02 | 0.03 | 0.06 | 0.06 |
| | 1,000 | 0.03 | 0.13 | 0.33 | 0.55 | 0.77 | 0.01 | 0.10 | 0.28 | 0.46 | 0.62 | 0.36 | 0.89 | 0.99 | 1.00 | 1.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 |
| | 10,000 | 0.01 | 0.05 | 0.21 | 0.39 | 0.58 | 0.00 | 0.03 | 0.12 | 0.27 | 0.45 | 0.18 | 0.78 | 0.98 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| | 100,000 | 0.00 | 0.03 | 0.09 | 0.23 | 0.42 | 0.00 | 0.01 | 0.06 | 0.13 | 0.28 | 0.10 | 0.63 | 0.95 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.2 | 100 | 0.22 | 0.65 | 0.91 | 0.99 | 1.00 | 0.16 | 0.57 | 0.85 | 0.96 | 0.99 | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 | 0.03 | 0.18 | 0.44 | 0.63 | 0.79 |
| | 1,000 | 0.09 | 0.44 | 0.79 | 0.92 | 0.99 | 0.04 | 0.37 | 0.68 | 0.89 | 0.98 | 0.66 | 0.99 | 1.00 | 1.00 | 1.00 | 0.02 | 0.06 | 0.24 | 0.37 | 0.59 |
| | 10,000 | 0.02 | 0.25 | 0.58 | 0.86 | 0.96 | 0.02 | 0.18 | 0.52 | 0.75 | 0.92 | 0.44 | 0.97 | 1.00 | 1.00 | 1.00 | 0.00 | 0.02 | 0.10 | 0.23 | 0.39 |
| | 100,000 | 0.01 | 0.14 | 0.44 | 0.74 | 0.90 | 0.01 | 0.08 | 0.32 | 0.59 | 0.82 | 0.27 | 0.91 | 1.00 | 1.00 | 1.00 | 0.00 | 0.01 | 0.04 | 0.13 | 0.25 |
| 0.3 | 100 | 0.30 | 0.80 | 0.97 | 1.00 | 1.00 | 0.28 | 0.76 | 0.97 | 1.00 | 1.00 | 0.84 | 1.00 | 1.00 | 1.00 | 1.00 | 0.22 | 0.66 | 0.92 | 0.98 | 1.00 |
| | 1,000 | 0.14 | 0.61 | 0.90 | 0.99 | 1.00 | 0.14 | 0.55 | 0.89 | 0.97 | 1.00 | 0.65 | 0.99 | 1.00 | 1.00 | 1.00 | 0.09 | 0.45 | 0.78 | 0.93 | 0.99 |
| | 10,000 | 0.06 | 0.41 | 0.77 | 0.97 | 0.99 | 0.04 | 0.38 | 0.72 | 0.92 | 0.99 | 0.45 | 0.96 | 1.00 | 1.00 | 1.00 | 0.04 | 0.28 | 0.60 | 0.86 | 0.96 |
| | 100,000 | 0.03 | 0.25 | 0.64 | 0.88 | 0.98 | 0.02 | 0.20 | 0.55 | 0.84 | 0.96 | 0.27 | 0.93 | 1.00 | 1.00 | 1.00 | 0.01 | 0.13 | 0.41 | 0.74 | 0.91 |
| 0.4 | 100 | 0.36 | 0.85 | 0.98 | 1.00 | 1.00 | 0.33 | 0.83 | 0.98 | 1.00 | 1.00 | 0.73 | 0.99 | 1.00 | 1.00 | 1.00 | 0.48 | 0.93 | 0.99 | 1.00 | 1.00 |
| | 1,000 | 0.18 | 0.67 | 0.92 | 0.99 | 1.00 | 0.15 | 0.65 | 0.92 | 0.99 | 1.00 | 0.53 | 0.97 | 1.00 | 1.00 | 1.00 | 0.27 | 0.82 | 0.98 | 1.00 | 1.00 |
| | 10,000 | 0.08 | 0.49 | 0.85 | 0.98 | 1.00 | 0.07 | 0.46 | 0.82 | 0.96 | 0.99 | 0.34 | 0.94 | 1.00 | 1.00 | 1.00 | 0.14 | 0.69 | 0.95 | 0.99 | 1.00 |
| | 100,000 | 0.04 | 0.33 | 0.72 | 0.93 | 0.98 | 0.04 | 0.30 | 0.68 | 0.92 | 0.99 | 0.17 | 0.83 | 0.99 | 1.00 | 1.00 | 0.07 | 0.48 | 0.89 | 0.99 | 1.00 |
| 0.5 | 100 | 0.34 | 0.85 | 0.98 | 1.00 | 1.00 | 0.34 | 0.83 | 0.98 | 1.00 | 1.00 | 0.56 | 0.97 | 1.00 | 1.00 | 1.00 | 0.71 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.15 | 0.65 | 0.92 | 0.99 | 1.00 | 0.16 | 0.68 | 0.94 | 0.99 | 1.00 | 0.31 | 0.87 | 0.99 | 1.00 | 1.00 | 0.48 | 0.96 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.07 | 0.46 | 0.82 | 0.96 | 1.00 | 0.07 | 0.45 | 0.82 | 0.97 | 1.00 | 0.16 | 0.74 | 0.97 | 1.00 | 1.00 | 0.31 | 0.89 | 0.99 | 1.00 | 1.00 |
| | 100,000 | 0.02 | 0.31 | 0.72 | 0.92 | 0.99 | 0.02 | 0.29 | 0.70 | 0.93 | 0.99 | 0.07 | 0.57 | 0.93 | 1.00 | 1.00 | 0.16 | 0.81 | 0.99 | 1.00 | 1.00 |
| 0.6 | 100 | 0.28 | 0.79 | 0.97 | 1.00 | 1.00 | 0.29 | 0.81 | 0.97 | 1.00 | 1.00 | 0.30 | 0.78 | 0.97 | 1.00 | 1.00 | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.12 | 0.58 | 0.87 | 0.98 | 1.00 | 0.15 | 0.57 | 0.91 | 0.98 | 1.00 | 0.14 | 0.60 | 0.90 | 0.98 | 1.00 | 0.63 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.06 | 0.34 | 0.74 | 0.94 | 0.99 | 0.07 | 0.39 | 0.77 | 0.95 | 1.00 | 0.05 | 0.41 | 0.79 | 0.95 | 0.99 | 0.42 | 0.96 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.01 | 0.21 | 0.57 | 0.87 | 0.97 | 0.02 | 0.24 | 0.62 | 0.90 | 0.98 | 0.01 | 0.21 | 0.62 | 0.88 | 0.97 | 0.28 | 0.90 | 1.00 | 1.00 | 1.00 |
| 0.7 | 100 | 0.18 | 0.65 | 0.90 | 0.98 | 1.00 | 0.25 | 0.68 | 0.93 | 0.99 | 1.00 | 0.10 | 0.40 | 0.72 | 0.87 | 0.97 | 0.83 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.08 | 0.40 | 0.74 | 0.92 | 0.99 | 0.08 | 0.46 | 0.81 | 0.96 | 0.99 | 0.02 | 0.21 | 0.49 | 0.75 | 0.89 | 0.66 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.01 | 0.21 | 0.55 | 0.83 | 0.95 | 0.05 | 0.31 | 0.66 | 0.88 | 0.97 | 0.01 | 0.10 | 0.31 | 0.56 | 0.78 | 0.44 | 0.97 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.01 | 0.11 | 0.37 | 0.70 | 0.89 | 0.02 | 0.16 | 0.46 | 0.79 | 0.93 | 0.00 | 0.04 | 0.15 | 0.38 | 0.60 | 0.31 | 0.90 | 1.00 | 1.00 | 1.00 |
| 0.8 | 100 | 0.11 | 0.40 | 0.67 | 0.85 | 0.96 | 0.12 | 0.49 | 0.75 | 0.92 | 0.98 | 0.01 | 0.08 | 0.19 | 0.36 | 0.52 | 0.74 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.04 | 0.21 | 0.47 | 0.70 | 0.88 | 0.04 | 0.26 | 0.59 | 0.81 | 0.92 | 0.01 | 0.03 | 0.09 | 0.14 | 0.30 | 0.53 | 0.97 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.01 | 0.09 | 0.30 | 0.49 | 0.75 | 0.01 | 0.10 | 0.36 | 0.63 | 0.81 | 0.00 | 0.01 | 0.03 | 0.07 | 0.14 | 0.35 | 0.92 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.01 | 0.03 | 0.13 | 0.35 | 0.57 | 0.00 | 0.06 | 0.22 | 0.46 | 0.67 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.19 | 0.83 | 0.99 | 1.00 | 1.00 |
| 0.9 | 100 | 0.02 | 0.10 | 0.27 | 0.45 | 0.59 | 0.03 | 0.17 | 0.31 | 0.55 | 0.71 | 0.00 | 0.00 | 0.01 | 0.03 | 0.03 | 0.40 | 0.88 | 0.99 | 1.00 | 1.00 |
| | 1,000 | 0.00 | 0.04 | 0.11 | 0.20 | 0.36 | 0.01 | 0.07 | 0.19 | 0.33 | 0.54 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.19 | 0.75 | 0.95 | 1.00 | 1.00 |
| | 10,000 | 0.00 | 0.02 | 0.05 | 0.11 | 0.19 | 0.00 | 0.02 | 0.06 | 0.17 | 0.32 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 | 0.53 | 0.89 | 0.98 | 1.00 |
| | 100,000 | 0.00 | 0.00 | 0.02 | 0.05 | 0.10 | 0.00 | 0.01 | 0.03 | 0.10 | 0.15 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.37 | 0.77 | 0.96 | 1.00 |

**Table 5.** Power (1-$\beta$) estimation for case-control genotypic tests according to numbers of SNPs, allele frequencies ($p$), case sample sizes ($n_D$) and genetic models. Power is computed by non-central chi-square approach for prevalence $Kp$=0.05, $n_D$=$n_C$, odds ratio $RR_2$=1.5, and $\alpha$=0.05/No. of SNPs (Bonferroni-corrected type I error rate)

| $p$ | Number of SNPs | Additive model | | | | | Multiplicative model | | | | | Dominant model | | | | | Recessive model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample | | | | |
| | | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 |
| 0.1 | 100 | 0.07 | 0.30 | 0.57 | 0.78 | 0.91 | 0.05 | 0.22 | 0.46 | 0.67 | 0.82 | 0.58 | 0.97 | 1.00 | 1.00 | 1.00 | 0.00 | 0.01 | 0.03 | 0.05 | 0.08 |
| | 1,000 | 0.02 | 0.14 | 0.35 | 0.59 | 0.78 | 0.01 | 0.09 | 0.25 | 0.45 | 0.65 | 0.36 | 0.90 | 0.99 | 1.00 | 1.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 |
| | 10,000 | 0.01 | 0.06 | 0.19 | 0.39 | 0.60 | 0.00 | 0.03 | 0.12 | 0.27 | 0.45 | 0.20 | 0.78 | 0.98 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| | 100,000 | 0.00 | 0.02 | 0.09 | 0.24 | 0.43 | 0.00 | 0.01 | 0.05 | 0.14 | 0.28 | 0.10 | 0.63 | 0.94 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 0.2 | 100 | 0.21 | 0.65 | 0.90 | 0.98 | 1.00 | 0.17 | 0.56 | 0.85 | 0.96 | 0.99 | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 | 0.05 | 0.20 | 0.42 | 0.63 | 0.79 |
| | 1,000 | 0.09 | 0.43 | 0.77 | 0.94 | 0.99 | 0.06 | 0.35 | 0.68 | 0.89 | 0.97 | 0.63 | 0.99 | 1.00 | 1.00 | 1.00 | 0.01 | 0.08 | 0.22 | 0.41 | 0.60 |
| | 10,000 | 0.03 | 0.25 | 0.60 | 0.85 | 0.96 | 0.02 | 0.19 | 0.49 | 0.76 | 0.91 | 0.44 | 0.96 | 1.00 | 1.00 | 1.00 | 0.00 | 0.03 | 0.10 | 0.24 | 0.40 |
| | 100,000 | 0.01 | 0.13 | 0.42 | 0.72 | 0.90 | 0.01 | 0.09 | 0.32 | 0.60 | 0.82 | 0.27 | 0.91 | 1.00 | 1.00 | 1.00 | 0.00 | 0.01 | 0.04 | 0.12 | 0.24 |
| 0.3 | 100 | 0.31 | 0.80 | 0.97 | 1.00 | 1.00 | 0.27 | 0.75 | 0.95 | 0.99 | 1.00 | 0.83 | 1.00 | 1.00 | 1.00 | 1.00 | 0.22 | 0.66 | 0.91 | 0.98 | 1.00 |
| | 1,000 | 0.15 | 0.61 | 0.90 | 0.98 | 1.00 | 0.12 | 0.55 | 0.87 | 0.97 | 1.00 | 0.65 | 0.99 | 1.00 | 1.00 | 1.00 | 0.09 | 0.44 | 0.79 | 0.94 | 0.99 |
| | 10,000 | 0.06 | 0.41 | 0.79 | 0.95 | 0.99 | 0.05 | 0.35 | 0.73 | 0.93 | 0.99 | 0.45 | 0.97 | 1.00 | 1.00 | 1.00 | 0.03 | 0.26 | 0.62 | 0.86 | 0.96 |
| | 100,000 | 0.02 | 0.25 | 0.64 | 0.89 | 0.98 | 0.02 | 0.20 | 0.56 | 0.84 | 0.96 | 0.28 | 0.91 | 1.00 | 1.00 | 1.00 | 0.01 | 0.14 | 0.44 | 0.74 | 0.91 |
| 0.4 | 100 | 0.35 | 0.84 | 0.98 | 1.00 | 1.00 | 0.33 | 0.82 | 0.98 | 1.00 | 1.00 | 0.74 | 0.99 | 1.00 | 1.00 | 1.00 | 0.49 | 0.93 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.17 | 0.67 | 0.93 | 0.99 | 1.00 | 0.16 | 0.64 | 0.92 | 0.99 | 1.00 | 0.53 | 0.97 | 1.00 | 1.00 | 1.00 | 0.28 | 0.83 | 0.98 | 1.00 | 1.00 |
| | 10,000 | 0.08 | 0.48 | 0.84 | 0.97 | 1.00 | 0.07 | 0.45 | 0.82 | 0.96 | 1.00 | 0.34 | 0.92 | 1.00 | 1.00 | 1.00 | 0.14 | 0.67 | 0.95 | 1.00 | 1.00 |
| | 100,000 | 0.03 | 0.31 | 0.71 | 0.93 | 0.99 | 0.03 | 0.28 | 0.68 | 0.91 | 0.98 | 0.19 | 0.82 | 0.99 | 1.00 | 1.00 | 0.06 | 0.50 | 0.88 | 0.98 | 1.00 |
| 0.5 | 100 | 0.34 | 0.83 | 0.98 | 1.00 | 1.00 | 0.35 | 0.83 | 0.98 | 1.00 | 1.00 | 0.55 | 0.96 | 1.00 | 1.00 | 1.00 | 0.71 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.16 | 0.65 | 0.93 | 0.99 | 1.00 | 0.17 | 0.66 | 0.93 | 0.99 | 1.00 | 0.33 | 0.88 | 0.99 | 1.00 | 1.00 | 0.50 | 0.96 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.07 | 0.46 | 0.83 | 0.97 | 1.00 | 0.07 | 0.47 | 0.84 | 0.97 | 1.00 | 0.17 | 0.74 | 0.97 | 1.00 | 1.00 | 0.31 | 0.90 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.03 | 0.29 | 0.69 | 0.92 | 0.98 | 0.03 | 0.30 | 0.70 | 0.92 | 0.99 | 0.08 | 0.58 | 0.92 | 0.99 | 1.00 | 0.17 | 0.79 | 0.99 | 1.00 | 1.00 |
| 0.6 | 100 | 0.28 | 0.76 | 0.96 | 0.99 | 1.00 | 0.31 | 0.80 | 0.97 | 1.00 | 1.00 | 0.30 | 0.79 | 0.96 | 1.00 | 1.00 | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.13 | 0.56 | 0.88 | 0.98 | 1.00 | 0.14 | 0.61 | 0.90 | 0.98 | 1.00 | 0.14 | 0.59 | 0.89 | 0.98 | 1.00 | 0.63 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.05 | 0.37 | 0.75 | 0.93 | 0.99 | 0.06 | 0.41 | 0.79 | 0.95 | 0.99 | 0.06 | 0.40 | 0.78 | 0.95 | 0.99 | 0.44 | 0.96 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.02 | 0.21 | 0.58 | 0.85 | 0.96 | 0.02 | 0.25 | 0.64 | 0.89 | 0.98 | 0.02 | 0.24 | 0.62 | 0.88 | 0.97 | 0.27 | 0.91 | 1.00 | 1.00 | 1.00 |
| 0.7 | 100 | 0.19 | 0.62 | 0.89 | 0.98 | 1.00 | 0.23 | 0.69 | 0.93 | 0.99 | 1.00 | 0.11 | 0.40 | 0.70 | 0.88 | 0.96 | 0.83 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.08 | 0.40 | 0.74 | 0.92 | 0.98 | 0.10 | 0.47 | 0.81 | 0.95 | 0.99 | 0.04 | 0.21 | 0.49 | 0.74 | 0.89 | 0.65 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.03 | 0.23 | 0.56 | 0.82 | 0.95 | 0.04 | 0.29 | 0.65 | 0.88 | 0.97 | 0.01 | 0.10 | 0.30 | 0.56 | 0.76 | 0.46 | 0.97 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.01 | 0.12 | 0.39 | 0.68 | 0.88 | 0.01 | 0.16 | 0.47 | 0.77 | 0.93 | 0.00 | 0.04 | 0.17 | 0.38 | 0.61 | 0.29 | 0.92 | 1.00 | 1.00 | 1.00 |
| 0.8 | 100 | 0.10 | 0.38 | 0.68 | 0.87 | 0.95 | 0.13 | 0.47 | 0.77 | 0.92 | 0.98 | 0.02 | 0.09 | 0.20 | 0.35 | 0.50 | 0.74 | 0.99 | 1.00 | 1.00 | 1.00 |
| | 1,000 | 0.03 | 0.19 | 0.46 | 0.71 | 0.87 | 0.04 | 0.26 | 0.57 | 0.81 | 0.93 | 0.00 | 0.03 | 0.08 | 0.17 | 0.29 | 0.53 | 0.97 | 1.00 | 1.00 | 1.00 |
| | 10,000 | 0.01 | 0.09 | 0.28 | 0.52 | 0.73 | 0.01 | 0.13 | 0.38 | 0.65 | 0.84 | 0.00 | 0.01 | 0.03 | 0.07 | 0.15 | 0.34 | 0.92 | 1.00 | 1.00 | 1.00 |
| | 100,000 | 0.00 | 0.04 | 0.15 | 0.35 | 0.57 | 0.00 | 0.06 | 0.22 | 0.47 | 0.70 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.19 | 0.82 | 0.99 | 1.00 | 1.00 |
| 0.9 | 100 | 0.03 | 0.11 | 0.25 | 0.42 | 0.59 | 0.04 | 0.16 | 0.34 | 0.54 | 0.71 | 0.00 | 0.01 | 0.01 | 0.02 | 0.04 | 0.39 | 0.87 | 0.99 | 1.00 | 1.00 |
| | 1,000 | 0.01 | 0.04 | 0.11 | 0.22 | 0.37 | 0.01 | 0.06 | 0.17 | 0.32 | 0.50 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.20 | 0.72 | 0.95 | 1.00 | 1.00 |
| | 10,000 | 0.00 | 0.01 | 0.04 | 0.10 | 0.20 | 0.00 | 0.02 | 0.07 | 0.17 | 0.31 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.53 | 0.88 | 0.98 | 1.00 |
| | 100,000 | 0.00 | 0.00 | 0.01 | 0.04 | 0.10 | 0.00 | 0.01 | 0.03 | 0.08 | 0.17 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.36 | 0.77 | 0.95 | 0.99 |

라 주어진 표본수 에서의 검정력을 구한 것으로 *R* 프로그램 (*R* 2.4.1)을 이용하여 계산하였다. Table 4는 몬테-칼로 모의실험에 의하여 계산된 것이고 Table 5는 점근적 비중심 모수 분포를 이용하여 계산된 것으로 두 방법에 의하여 계산된 결과는 거의 동일함을 알 수 있다. Table 6과 7은 500,000개의 SNP을 사용할 때 OR (odds ratio), 원인 대립형질 빈도 및 유전 모델에 따라 계산된 검정력으로 *R* 프로그램을 이용하여 계산하였다.

## 3. 연관불평형(LD)과 검정력과의 관계

지금까지는 연관 분석의 대상이 되는 유전좌위가 질환과 직접 관계된 원인 변이라는 가정 하에 검정력을 계산하였으나 실제 연구에서는 대부분 표지자를 이용하게 된다. 즉 분석의 대상이 되는 유전좌위는 질환과 직접 관련된 원인 변이가 아니라 원인 변이와 LD 관계에 있게 된다. 이때 LD 강도(LD strength)에 따라 검정력이 변하는 데 동일한 표본수일 경우 LD 강도가 커질수록 검정력이 커지게 된다. 표지자를 사용하는 경우 원인 변이를 직접 분석하는 경우와 동일한 검정력을 확보하기 위해서는 $1/r^2$ 만큼 표본수를 증가시켜야 한다고 알려져 있다 [5,6]. 예를 들어 $r^2$=0.8인 표지자의 경우 그 유전좌위가 원인 변이인 경우와 동일한 검정력을 얻으려면 약 1.25배의 표본수를 사용하여야 한다는 것이다. $r^2$=0.1인 표지자의 경우에는 약 10배의 표본수를 사용하여야 한다는 것인데 이것은 현실적이지 못하다. 따라서 $r^2$이 작은 경우 이 표지자 유전좌위에서 검정력은 매우 낮아지므로 연관을 알아낼 확률은 매우 적어진다.

## 4. 검정력 계산에서 모수의 확률분포 이용

유전 연관 연구에서 검정력 혹은 표본수를 계산할 때 많은 모수를 미리 알고 있어야 한다. 예를 들면 원인 대립형질 빈도, 표지자들의 유전좌위와 원인 유전좌위와의 LD, 원인 대립형질의 투과도, 표지자 유전좌위의 유전형 또는 대립형질 빈도 등을 알고 있어야 한다. 이미 잘 알려진 특정 유전체 부분을 대상으로 하는 경우 이러한 모수를 이미 알고 있는 경우도 있으나 대부분의 연관 연구나 특히 genomewide scan에 있어서 이러한 모수에 대하여 미리 알고 있는 경우는 거의 없다. 따라서 이러한 모수들을 특정 값이라고 가정하고 검정력 혹은 표본수를 계산하게 되는 데 이러한 경우 얻어진 결과가 실제값과 다를 것이라고 추측할 수 있다. 따라서 이미 알려져

**Table 6.** Power (1-$\beta$) estimation for case-control genotypic tests for 500,000 SNPs according to odds ratios ($RR_2$), allele frequencies ($p$), case sample sizes ($n_D$) and genetic models. Power is computed by Monte-Carlo simulation for prevalence $K_P$=0.05, $n_D$=$n_C$ and $\alpha$=$1 \times 10^{-7}$ (Bonferroni-corrected type I error rate). Each Monte-Carlo estimate of power is carried out on the basis of N=1,000 simulations

| $p$ | $RR_2$ | Additive | | | | | Multiplicative | | | | | Dominant | | | | | Recessive | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | |
| | | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 |
| 0.1 | 1.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.10 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1.50 | 0.00 | 0.01 | 0.06 | 0.15 | 0.31 | 0.00 | 0.01 | 0.03 | 0.09 | 0.16 | 0.05 | 0.51 | 0.89 | 0.99 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1.75 | 0.02 | 0.18 | 0.54 | 0.83 | 0.96 | 0.01 | 0.08 | 0.31 | 0.63 | 0.83 | 0.49 | 0.99 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 |
| | 2.00 | 0.10 | 0.64 | 0.95 | 0.99 | 1.00 | 0.04 | 0.36 | 0.80 | 0.97 | 1.00 | 0.94 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.09 |
| 0.2 | 1.25 | 0.00 | 0.00 | 0.01 | 0.01 | 0.01 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.00 | 0.04 | 0.12 | 0.31 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1.50 | 0.00 | 0.08 | 0.30 | 0.61 | 0.84 | 0.00 | 0.05 | 0.22 | 0.50 | 0.74 | 0.17 | 0.86 | 0.99 | 1.00 | 1.00 | 0.00 | 0.01 | 0.03 | 0.06 | 0.16 |
| | 1.75 | 0.10 | 0.60 | 0.96 | 1.00 | 1.00 | 0.05 | 0.46 | 0.86 | 0.98 | 1.00 | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.08 | 0.29 | 0.61 | 0.84 |
| | 2.00 | 0.40 | 0.97 | 1.00 | 1.00 | 1.00 | 0.23 | 0.90 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.04 | 0.38 | 0.81 | 0.97 | 1.00 |
| 0.3 | 1.25 | 0.00 | 0.00 | 0.00 | 0.02 | 0.05 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.00 | 0.03 | 0.16 | 0.37 | 0.60 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 |
| | 1.50 | 0.01 | 0.19 | 0.55 | 0.82 | 0.95 | 0.01 | 0.14 | 0.45 | 0.73 | 0.93 | 0.19 | 0.87 | 0.99 | 1.00 | 1.00 | 0.00 | 0.09 | 0.32 | 0.63 | 0.86 |
| | 1.75 | 0.15 | 0.80 | 0.99 | 1.00 | 1.00 | 0.11 | 0.72 | 0.97 | 1.00 | 1.00 | 0.83 | 1.00 | 1.00 | 1.00 | 1.00 | 0.08 | 0.65 | 0.96 | 1.00 | 1.00 |
| | 2.00 | 0.57 | 0.99 | 1.00 | 1.00 | 1.00 | 0.44 | 0.98 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.41 | 0.98 | 1.00 | 1.00 | 1.00 |
| 0.4 | 1.25 | 0.00 | 0.00 | 0.00 | 0.03 | 0.10 | 0.00 | 0.00 | 0.01 | 0.03 | 0.08 | 0.00 | 0.02 | 0.10 | 0.23 | 0.48 | 0.00 | 0.00 | 0.02 | 0.05 | 0.13 |
| | 1.50 | 0.01 | 0.20 | 0.62 | 0.90 | 0.98 | 0.02 | 0.18 | 0.59 | 0.86 | 0.97 | 0.11 | 0.75 | 0.98 | 1.00 | 1.00 | 0.03 | 0.40 | 0.82 | 0.96 | 1.00 |
| | 1.75 | 0.19 | 0.83 | 0.99 | 1.00 | 1.00 | 0.16 | 0.82 | 1.00 | 1.00 | 1.00 | 0.65 | 1.00 | 1.00 | 1.00 | 1.00 | 0.36 | 0.96 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.57 | 1.00 | 1.00 | 1.00 | 1.00 | 0.54 | 0.99 | 1.00 | 1.00 | 1.00 | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 0.84 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.5 | 1.25 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.00 | 0.00 | 0.01 | 0.02 | 0.08 | 0.00 | 0.01 | 0.04 | 0.11 | 0.26 | 0.00 | 0.01 | 0.06 | 0.18 | 0.35 |
| | 1.50 | 0.02 | 0.20 | 0.57 | 0.86 | 0.97 | 0.01 | 0.22 | 0.59 | 0.88 | 0.98 | 0.05 | 0.48 | 0.87 | 0.99 | 1.00 | 0.11 | 0.73 | 0.97 | 1.00 | 1.00 |
| | 1.75 | 0.16 | 0.82 | 0.99 | 1.00 | 1.00 | 0.18 | 0.84 | 0.99 | 1.00 | 1.00 | 0.34 | 0.96 | 1.00 | 1.00 | 1.00 | 0.67 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.52 | 1.00 | 1.00 | 1.00 | 1.00 | 0.55 | 0.99 | 1.00 | 1.00 | 1.00 | 0.77 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.6 | 1.25 | 0.00 | 0.01 | 0.01 | 0.02 | 0.06 | 0.00 | 0.00 | 0.02 | 0.03 | 0.07 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.00 | 0.03 | 0.13 | 0.32 | 0.55 |
| | 1.50 | 0.02 | 0.14 | 0.48 | 0.77 | 0.93 | 0.01 | 0.19 | 0.52 | 0.81 | 0.95 | 0.01 | 0.14 | 0.46 | 0.81 | 0.94 | 0.18 | 0.86 | 1.00 | 1.00 | 1.00 |
| | 1.75 | 0.12 | 0.69 | 0.97 | 1.00 | 1.00 | 0.14 | 0.78 | 0.98 | 1.00 | 1.00 | 0.12 | 0.70 | 0.97 | 1.00 | 1.00 | 0.81 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.36 | 0.96 | 1.00 | 1.00 | 1.00 | 0.46 | 0.98 | 1.00 | 1.00 | 1.00 | 0.34 | 0.98 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.7 | 1.25 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.03 | 0.14 | 0.36 | 0.57 |
| | 1.50 | 0.00 | 0.06 | 0.30 | 0.61 | 0.81 | 0.01 | 0.10 | 0.36 | 0.66 | 0.88 | 0.00 | 0.02 | 0.11 | 0.27 | 0.48 | 0.19 | 0.87 | 0.99 | 1.00 | 1.00 |
| | 1.75 | 0.04 | 0.45 | 0.86 | 0.99 | 1.00 | 0.08 | 0.60 | 0.94 | 0.99 | 1.00 | 0.01 | 0.18 | 0.56 | 0.85 | 0.97 | 0.84 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.18 | 0.82 | 0.99 | 1.00 | 1.00 | 0.29 | 0.94 | 1.00 | 1.00 | 1.00 | 0.04 | 0.51 | 0.91 | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.8 | 1.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.09 | 0.28 | 0.48 |
| | 1.50 | 0.00 | 0.02 | 0.09 | 0.26 | 0.43 | 0.00 | 0.03 | 0.16 | 0.36 | 0.58 | 0.00 | 0.00 | 0.01 | 0.01 | 0.03 | 0.14 | 0.73 | 0.98 | 1.00 | 1.00 |
| | 1.75 | 0.01 | 0.18 | 0.51 | 0.82 | 0.96 | 0.03 | 0.30 | 0.71 | 0.93 | 0.99 | 0.00 | 0.01 | 0.03 | 0.14 | 0.28 | 0.65 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.04 | 0.48 | 0.89 | 0.99 | 1.00 | 0.11 | 0.71 | 0.98 | 1.00 | 1.00 | 0.00 | 0.02 | 0.15 | 0.39 | 0.63 | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.9 | 1.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.05 | 0.13 |
| | 1.50 | 0.00 | 0.00 | 0.01 | 0.02 | 0.06 | 0.00 | 0.00 | 0.01 | 0.04 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.25 | 0.67 | 0.93 | 0.99 |
| | 1.75 | 0.00 | 0.01 | 0.05 | 0.18 | 0.37 | 0.00 | 0.02 | 0.14 | 0.39 | 0.62 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.20 | 0.85 | 0.99 | 1.00 | 1.00 |
| | 2.00 | 0.00 | 0.05 | 0.19 | 0.49 | 0.73 | 0.01 | 0.14 | 0.49 | 0.80 | 0.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.50 | 0.99 | 1.00 | 1.00 | 1.00 |

있는 정보로부터 모수의 분포를 추정하고 이를 이용하여 검정력을 계산하는 방법이 고안되었다 [7,8]. 즉 모수의 분포를 이용하여 가장 높은 확률을 가지는 값을 모수로 선택하거나 혹은 더 나아가 이 모수가 특정값을 가질 확률에 따라 가중치를 부여하여 검정력을 계산하는 방법도 사용될 수 있다 [7]. 이러한 방법은 Bayesian 방법과 매우 유사하다. 즉 정확한 모수의 값을 모를 때 모수의 분포를 이용하는 것이다. 그렇다면 모수의 분포는 어디에서 얻을 수 있는가 하는 문제가 대두된다. 이것은 기존 자료로부터 얻게 된다. 예를 들면 기존 SNP 자료로부터 대립형질 빈도 분포, SNP들 간의 LD 같은 모수값의 확률분포를 계산하여 이를 이용하는 것이다. 즉 기존 자료로부터 모수의 분포가 얻어지면 검정력은 모수의 확률분포를 이용하여 수학적으로 이 모수에 대하여 적분하여 계산할 수 있다. 실제 계산에 있어서는 모수 구간을 불연속 구간으로 나누어 적분식(integrals)을 합(summation)으로 바꾸어 계산하게 된다.

## Genomewide 연관 연구를 위한 최적 2단계 연구 설계

지금까지 살펴본 검정력 계산은 1단계 연구 설계(one-stage study design)에 적용되는 것으로서 2단계(two-stage) 혹은 다단계(multi-stage) 연구 설계에 비하여, 동일한 검정력을 얻기 위해서는 훨씬 더 많은 표본수를 필요로 한다 [1]. 따라서 genomewide 연관 연구를 위해서는 충분한 검정력을 확보하면서 표본수와 소요비용을 최소화하기 위하여 여러 단계에 걸쳐 genotyping 하는 방법이 더 적합하다 [1-3]. 2단계 혹은 다단계 설계는 다음과 같은 절차로 이루어진다. 먼저 첫 번째 단계 에서는 표본의 일부만을 대상으로 모든 SNP 표지자들을 genotyping하여 연관 분석을 하고 이 중에서 적절한 유의수준 이상의 결과를 나타낸 SNP들을 선택한다. 이 후 단계에서는 나머지 표본을 이용하여 선택된 SNP만을 genotyping하여 연관 분석을 다시 하게 된다. 이러한 2단계 혹은 다단계로 genotyping 하는 연구 설계가 갖는 장점은 첫째 검정력의 손실 없이 비용 절감을 할 수 있다는 점이다. 그 이유는 두 번째 단계에서는 첫 번째 단계에서 어느 정도 이상의 유의수준을 나타낸 SNP만을 대상으로 genotyping을 하게 되므로 원인 표지자가 아닌 다른 표지자들로 인한 비용손실을 없애므로 제한된 비용으로 검정력을 높이는 효과를 가져 올 수 있기 때문이다. 이 연

**Table 7.** Power (1-$\beta$) estimation for case-control genotypic tests for 500,000 SNPs according to odds ratios ($RR_2$), allele frequencies ($p$), case sample sizes ($n_D$) and genetic models. Power is computed by non-central chi-square approach for prevalence $Kp$=0.05, $n_D$=$n_C$ and $\alpha=1\times10^{-7}$ (Bonferroni-corrected type I error rate)

| $p$ | $RR_2$ | Additive | | | | | Multiplicative | | | | | Dominant | | | | | Recessive | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | | Sample size($n_D$) | | | | |
| | | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 | 1,000 | 2,000 | 3,000 | 4,000 | 5,000 |
| 0.1 | 1.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.03 | 0.09 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1.50 | 0.00 | 0.01 | 0.05 | 0.15 | 0.31 | 0.00 | 0.00 | 0.03 | 0.09 | 0.19 | 0.05 | 0.51 | 0.90 | 0.99 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1.75 | 0.01 | 0.18 | 0.54 | 0.84 | 0.96 | 0.00 | 0.02 | 0.32 | 0.62 | 0.84 | 0.50 | 0.99 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 |
| | 2.00 | 0.02 | 0.65 | 0.96 | 1.00 | 1.00 | 0.03 | 0.04 | 0.79 | 0.96 | 1.00 | 0.92 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.10 |
| 0.2 | 1.25 | 0.05 | 0.00 | 0.00 | 0.01 | 0.02 | 0.00 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | 0.03 | 0.13 | 0.32 | 0.55 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| | 1.50 | 0.00 | 0.08 | 0.31 | 0.62 | 0.84 | 0.00 | 0.03 | 0.22 | 0.49 | 0.73 | 0.18 | 0.85 | 0.99 | 1.00 | 1.00 | 0.00 | 0.00 | 0.02 | 0.07 | 0.16 |
| | 1.75 | 0.00 | 0.63 | 0.95 | 1.00 | 1.00 | 0.04 | 0.15 | 0.87 | 0.98 | 1.00 | 0.81 | 1.00 | 1.00 | 1.00 | 1.00 | 0.00 | 0.08 | 0.31 | 0.61 | 0.83 |
| | 2.00 | 0.00 | 0.97 | 1.00 | 1.00 | 1.00 | 0.23 | 0.36 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.03 | 0.38 | 0.81 | 0.97 | 1.00 |
| 0.3 | 1.25 | 0.01 | 0.09 | 0.00 | 0.00 | 0.08 | 0.39 | 0.00 | 0.01 | 0.08 | 0.36 | 0.00 | 0.05 | 0.46 | 0.89 | 0.60 | 0.00 | 0.00 | 0.00 | 0.01 | 0.02 |
| | 1.50 | 0.01 | 0.17 | 0.53 | 0.82 | 0.96 | 0.01 | 0.13 | 0.45 | 0.76 | 0.93 | 0.19 | 0.86 | 0.99 | 1.00 | 1.00 | 0.00 | 0.08 | 0.33 | 0.63 | 0.85 |
| | 1.75 | 0.16 | 0.81 | 0.99 | 1.00 | 1.00 | 0.11 | 0.72 | 0.98 | 1.00 | 1.00 | 0.80 | 1.00 | 1.00 | 1.00 | 1.00 | 0.09 | 0.64 | 0.95 | 1.00 | 1.00 |
| | 2.00 | 0.56 | 0.99 | 1.00 | 1.00 | 1.00 | 0.44 | 0.98 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.40 | 0.97 | 1.00 | 1.00 | 1.00 |
| 0.4 | 1.25 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.00 | 0.02 | 0.10 | 0.26 | 0.48 | 0.00 | 0.00 | 0.02 | 0.06 | 0.13 |
| | 1.50 | 0.01 | 0.21 | 0.60 | 0.88 | 0.97 | 0.01 | 0.19 | 0.57 | 0.86 | 0.97 | 0.12 | 0.74 | 0.98 | 1.00 | 1.00 | 0.03 | 0.38 | 0.81 | 0.97 | 1.00 |
| | 1.75 | 0.18 | 0.85 | 0.99 | 1.00 | 1.00 | 0.16 | 0.82 | 0.99 | 1.00 | 1.00 | 0.65 | 1.00 | 1.00 | 1.00 | 1.00 | 0.38 | 0.97 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.58 | 1.00 | 1.00 | 1.00 | 1.00 | 0.54 | 0.99 | 1.00 | 1.00 | 1.00 | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 | 0.85 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.5 | 1.25 | 0.00 | 0.00 | 0.01 | 0.03 | 0.08 | 0.00 | 0.00 | 0.01 | 0.03 | 0.08 | 0.00 | 0.01 | 0.04 | 0.12 | 0.26 | 0.00 | 0.01 | 0.06 | 0.18 | 0.36 |
| | 1.50 | 0.01 | 0.20 | 0.58 | 0.86 | 0.97 | 0.01 | 0.20 | 0.59 | 0.87 | 0.97 | 0.05 | 0.46 | 0.87 | 0.99 | 1.00 | 0.11 | 0.70 | 0.97 | 1.00 | 1.00 |
| | 1.75 | 0.16 | 0.81 | 0.99 | 1.00 | 1.00 | 0.17 | 0.83 | 0.99 | 1.00 | 1.00 | 0.36 | 0.96 | 1.00 | 1.00 | 1.00 | 0.68 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.51 | 0.99 | 1.00 | 1.00 | 1.00 | 0.55 | 0.99 | 1.00 | 1.00 | 1.00 | 0.76 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.6 | 1.25 | 0.00 | 0.00 | 0.01 | 0.02 | 0.06 | 0.00 | 0.00 | 0.01 | 0.03 | 0.06 | 0.00 | 0.00 | 0.01 | 0.03 | 0.07 | 0.00 | 0.03 | 0.13 | 0.32 | 0.55 |
| | 1.50 | 0.01 | 0.14 | 0.47 | 0.78 | 0.93 | 0.01 | 0.17 | 0.52 | 0.82 | 0.95 | 0.01 | 0.16 | 0.50 | 0.81 | 0.95 | 0.18 | 0.85 | 0.99 | 1.00 | 1.00 |
| | 1.75 | 0.10 | 0.69 | 0.97 | 1.00 | 1.00 | 0.13 | 0.77 | 0.98 | 1.00 | 1.00 | 0.10 | 0.70 | 0.97 | 1.00 | 1.00 | 0.81 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.36 | 0.96 | 1.00 | 1.00 | 1.00 | 0.46 | 0.98 | 1.00 | 1.00 | 1.00 | 0.35 | 0.96 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.7 | 1.25 | 0.00 | 0.00 | 0.00 | 0.01 | 0.03 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.03 | 0.15 | 0.37 | 0.61 |
| | 1.50 | 0.00 | 0.07 | 0.28 | 0.57 | 0.80 | 0.01 | 0.10 | 0.36 | 0.67 | 0.88 | 0.00 | 0.02 | 0.10 | 0.27 | 0.49 | 0.20 | 0.87 | 1.00 | 1.00 | 1.00 |
| | 1.75 | 0.04 | 0.45 | 0.86 | 0.98 | 1.00 | 0.07 | 0.59 | 0.94 | 1.00 | 1.00 | 0.01 | 0.18 | 0.56 | 0.85 | 0.96 | 0.81 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.18 | 0.84 | 0.99 | 1.00 | 1.00 | 0.30 | 0.94 | 1.00 | 1.00 | 1.00 | 0.05 | 0.51 | 0.90 | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.8 | 1.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.10 | 0.26 | 0.48 |
| | 1.50 | 0.00 | 0.02 | 0.09 | 0.25 | 0.45 | 0.00 | 0.03 | 0.14 | 0.36 | 0.60 | 0.00 | 0.00 | 0.00 | 0.01 | 0.04 | 0.12 | 0.74 | 0.98 | 1.00 | 1.00 |
| | 1.75 | 0.01 | 0.16 | 0.51 | 0.81 | 0.95 | 0.02 | 0.29 | 0.71 | 0.93 | 0.99 | 0.00 | 0.01 | 0.04 | 0.13 | 0.28 | 0.65 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 2.00 | 0.05 | 0.46 | 0.87 | 0.98 | 1.00 | 0.11 | 0.71 | 0.97 | 1.00 | 1.00 | 0.00 | 0.04 | 0.17 | 0.40 | 0.65 | 0.95 | 1.00 | 1.00 | 1.00 | 1.00 |
| 0.9 | 1.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.05 | 0.13 |
| | 1.50 | 0.00 | 0.00 | 0.01 | 0.02 | 0.06 | 0.00 | 0.00 | 0.01 | 0.04 | 0.11 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.25 | 0.67 | 0.91 | 0.99 |
| | 1.75 | 0.00 | 0.01 | 0.06 | 0.19 | 0.36 | 0.00 | 0.03 | 0.15 | 0.37 | 0.61 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.18 | 0.84 | 0.99 | 1.00 | 1.00 |
| | 2.00 | 0.00 | 0.05 | 0.22 | 0.49 | 0.73 | 0.01 | 0.15 | 0.48 | 0.79 | 0.94 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.51 | 0.99 | 1.00 | 1.00 | 1.00 |

구 설계가 갖는 두 번째 장점은 두 번째 단계에서는 첫 번째 단계에서 선정된 영역 내의 모든 tag SNP를 genotyping 할 수 있다는 것이다. 즉 첫 번째 단계의 genotyping에 사용되는 fixed array는 모든 SNP를 커버하지 못하는 데 실제로는 tag SNP 조차 모두 포함되기 어렵다. 그러나 두 번째 단계에서는 첫 번째 단계에서 선정된 유전체 영역 내의 모든 tag SNP를 genotyping 할 수 있다. 세 번째 장점은 각 단계에서 서로 다른 genotyping 방법을 사용하게 되므로 첫 번째 단계에서 위양성(false-positive) 연관을 초래할 수 있었던 원인을 그 다음 단계에서는 제거할 수 있다는 것이다. 이와 같은 이유로 인하여 genomewide 연관 연구에서 소요비용을 최소화 하면서 검정력을 높일 수 있는 최적 2단계 연구 설계 방법(optimal two-stage design)이 제시되었다 [2]. 이 연구 설계의 특징은 먼저 첫 번째 단계에서 사용되는 SNP 칩과 두 번째 단계에서 사용되는 high-throughput genotyping에 소요되는 비용 비율(cost ratio)을 고려하여 최적화 하였다는 것이며 둘째 1종 오류율을 $1\times10^{-7}$(500,000개 SNP를 다중 분석(multiple testing) 했을 때 위양성율(false-positive rate)이 5%, 즉 Bonferroni 보정 후 1종 오류율이 0.05), 검정력을 90%로 유지하면서 비용을 최소화 할 수 있도록 하였다는 것이다.

## 1. 2단계 연구 설계의 최적화

인구기반(population-based) 환자-대조군 연구를 대상으로 한 2단계 연구 설계 최적화 과정은 다음과 같다. 먼저 첫 번째 단계에서 $n_1$명의 표본에 대하여 $m$개의 표지자를 genotyping 한다. 연관분석 결과 이 중에서 유의수준 $\alpha_1$이상을 보이는 표지자들을 선택하고 두 번째 단계에서는 첫 번째 단계에서 선택된 표지자들에 대해서만 나머지 $n_2$명의 표본에 대하여 genotyping을 하게 된다. 두 번째 단계에서의 검정통계량 계산에서는 ($n_1+n_2$) 표본수를 사용하며 유의수준 $\alpha_2$에 대하여 하게 된다.

Hardy-Weinberg 평형하에서 $p$를 질병위험을 증가시키는 위험 대립형질(risk allele)의 빈도라 하고 $\delta$를 한 사람이 가지는 위험 대립형질 수(allele dosage)라고 하면 $\delta$는 0,1,2 중 한 값을 가지게 된다. multiplicative 유전 모델을 가정하고 $\psi$를 위험 대립형질 수가 증가함에 따른 상대위험도(OR)라고 하자. 첫 번째 단계에서는 환자군과 대조군에서 위험 대립형질 수(risk allele dosage)의 평균의 차가 검정통계량 $S_1$으로 사용되며 이 값이 근사적으로 정규분포 하는 것을 이용 하여 연관을 분석하게 된다. 두 번째 단계에서는 두 단계에서 사용된 표본 전체로부터 얻어진 검정통계량 $S$를 사용

**Table 8.** Minimum expected cost for a one-to-one unmatched study and various allele frequencies, $p$, and odds ratios, $\psi$, and the set of parameters when the minimum is reached, for $m$ = 500,000, $T$ = 1, $t_1$ = \$0.002, $t_2$ = \$0.035 and $\alpha = 1\times10^{-7}$ (one-sided), 1-$\beta$ = 0.90 (From Wang et al, 2006)

| $p$ | $\psi$ | $\alpha_1$ | 1-$\beta_1$ | $\alpha_2$($\times10^{-7}$) | $n_1$ | $n_2$ | Total cost ($thousands) | Platforms |
|---|---|---|---|---|---|---|---|---|
| 0.1 | 1.35 | 0.00370 | 0.907 | 1.6 | 3,238 | 7,490 | 3,724 | 0.302 |
| | 1.5 | 0.00366 | 0.907 | 1.6 | 1,662 | 3,824 | 1,906 | 0.303 |
| | 1.75 | 0.00362 | 0.907 | 1.6 | 792 | 1,818 | 907 | 0.303 |
| | 2 | 0.00365 | 0.907 | 1.6 | 476 | 1,090 | 545 | 0.304 |
| 0.2 | 1.35 | 0.00373 | 0.907 | 1.7 | 1,920 | 4,458 | 2,211 | 0.301 |
| | 1.5 | 0.00374 | 0.907 | 1.6 | 1,004 | 2,314 | 1,156 | 0.303 |
| | 1.75 | 0.00372 | 0.907 | 1.7 | 494 | 1,146 | 568 | 0.301 |
| | 2 | 0.00372 | 0.907 | 1.6 | 306 | 704 | 351 | 0.303 |
| 0.4 | 1.35 | 0.00376 | 0.907 | 1.6 | 1,420 | 3,298 | 1,637 | 0.301 |
| | 1.5 | 0.00376 | 0.907 | 1.7 | 772 | 1,794 | 889 | 0.301 |
| | 1.75 | 0.00379 | 0.907 | 1.7 | 402 | 936 | 463 | 0.300 |
| | 2 | 0.00376 | 0.907 | 1.7 | 262 | 610 | 302 | 0.300 |
| 0.6 | 1.35 | 0.00378 | 0.907 | 1.6 | 1,568 | 3,648 | 1,809 | 0.301 |
| | 1.5 | 0.00382 | 0.907 | 1.6 | 880 | 2,054 | 1,017 | 0.300 |
| | 1.75 | 0.00382 | 0.907 | 1.7 | 482 | 1,126 | 557 | 0.300 |
| | 2 | 0.00377 | 0.907 | 1.7 | 328 | 766 | 378 | 0.300 |
| 0.8 | 1.35 | 0.00383 | 0.907 | 1.6 | 2,582 | 6,030 | 2,987 | 0.300 |
| | 1.5 | 0.00382 | 0.907 | 1.7 | 1,498 | 3,516 | 1,733 | 0.299 |
| | 1.75 | 0.00386 | 0.907 | 1.7 | 854 | 2,020 | 991 | 0.297 |
| | 2 | 0.00386 | 0.907 | 1.7 | 600 | 1,418 | 697 | 0.298 |
| 0.9 | 1.35 | 0.00379 | 0.907 | 1.7 | 4,814 | 11,272 | 5,562 | 0.299 |
| | 1.5 | 0.00383 | 0.907 | 1.7 | 2,828 | 6,652 | 3,274 | 0.298 |
| | 1.75 | 0.00386 | 0.907 | 1.7 | 1,648 | 3,884 | 1,910 | 0.298 |
| | 2 | 0.00388 | 0.907 | 1.7 | 1,176 | 2,770 | 1,365 | 0.298 |

**Table 9.** Publicly available tools for calculations of power and sample size for genetic association study

| *Programs* |
|---|
| Genetic Power Calculator |
| Calculates power for QTL(quantitative-trait loci) mapping study under variance-components(VC) model |
| Provides power calculators for |
| QTL linkage for sibships under VC |
| QTL association for sibships under VC |
| TDT(transmission disequilibrium test) for discrete traits |
| Case-control for discrete traits |
| TDT for threshold-selected quantitative traits |
| Case-control for threshold selected quantitative trait |
| PAWE(Power for Association with Error) |
| Calculates power and sample sizes for genetic-control studies in the presence errors(genotyping and phenotyping error) |
| Provides 3D picture according to the range of parameters specified |
| Provides power calculators for |
| Genotypic test or linear trend test |
| Discrete trait or QTL |
| With genotyping error or phenotyping error |
| CaTS Power Calculator |
| Specially designed for the two-stage genomewide association studies |
| Only for qualitative(discrete) trait |
| User-friendly tool |
| Provides an optimized study design to reduce the cost |

하며, 최종적으로 벡터($S_1$, $S$)가 이변량 정규분포를 따르는 것을 이용하여 표본수($n_1$, $n_2$), 모델 모수($p$, $\psi$), 유의수준($\alpha_1$, $\alpha_2$)이 주어졌을 때 검정력 1-$\beta$와 1종 오류 $\alpha$를 구할 수 있다. 비용함수는 첫 번째 단계와 두 번째 단계에서 1개 genotyping 당 소요되는 비용비율($t_1/t_2$), 표본수, 두 번째 단계에서 사용될 표지자수의 기대치에 의해 결정된다. 즉 첫 번째 단계에서 사용되는 표지자의 수를 $m$이라고 할 때 이 중 $T$개 SNP만이 원인 유전자(혹은 원인 유전자와 완전한 LD 관계)라고 하고 $t_1$, $t_2$가 각각 첫 번째 단계 및 두 번째 단계에서 1개 SNP를 genotyping하는데 소요되는 비용이라고 하면 전체 필요한 비용의 기대값은 $t_1n_1m+t_2n_2m_2$이다. 이 때 $m_2$는 두 번째 단계에서 사용되는 표지자 수이며 $m_2$의 기대값은 $[(m-T)\alpha_1+T(1-\beta_1)]$가 된다. 최종적으로 1종 오류율과 검정력이 원하는 정도가 되도록 하면서 비용을 최소화 하는 표본수 및 유의수준을 격자탐색법(grid search)에 의하여 찾게 된다.

## 2. 최적 연구 설계의 예

Table 8은 500K SNP fixed array를 사용하였을 때, 대립형질 빈도와 OR가 주어지고 검정력은 90%, Bonferroni 보정 제1종 오류율을 $1\times10^{-7}$로 하였을 때 소요비용을 최소화 하는 최적 연구 설계의 예이다. 이 때 첫 번째 단계와 두 번째 단계에서의 1개 genotyping 당 소요비용은 각각 0.2, 3.5 센트로 계산하였고 단측검정을 가정한 것이다. Table 8은 최적 비용과 표본수가 변이 대립형질 빈도 $p$와 OR에 의해 크게 좌우되는 것을 보여준다. 대립형질 빈도와 OR이 각각 0.05-0.95, 1.35-2.0 범위에 있는 경우 두 번째 단계에서 필요한 표본수는 첫 번째 단계 표본수의 약 2.2배 정도인 것으로 나타나고 있다.

## 유전 연관 연구의 검정력 및 표본수 계산에 이용할 수 있는 소프트웨어

유전 연관 연구에서 검정력 혹은 표본수 계산에 사용할 수 있는 여러 소프트웨어가 개발되어 있다. Genetic Power Calculator [9,10]와 PAWE(Power for Association with Error) [11-13]는 1단계 연구 설계에서의 검정력과 표본수를 계산할 수 있는 대표적인 소프트웨어이고 최근 개발된 CaTS Power Calculator [14]는 2단계 연구 설계를 위한 검정력과 표본수를 계산할 수 있는 소프트웨어이다. Table 9에 각 소프트웨어의 특징을 요약하였다.

## 결 론

급속한 genotyping 기술의 개발과 고밀도 SNP 칩의 상용화로 인하여 genomewide 연관 연구는 다양한 질환과 관련하여 더 확대될 전망이다. 외국에서는 이미 많은 genomewide 연관 연구가 시행되고 있고, 새로이 시작되고 있으며 일부에서는 이미

그 결과가 나오고 있다 [1]. 우리나라에서도 수 년 전부터 대규모 유전체 코호트 구축의 필요성이 대두되어 왔고 [15], 이제 그 시작 단계에 와 있다고 할 수 있다. 이러한 대규모 유전체 연구에서 통계적으로 정밀하게 계획된 연구 설계는 연구의 성공을 위하여 매우 중요하며 연구의 타당성과 신뢰성 뿐 아니라 효율성을 위하여 필요불가결한 요건이라고 할 수 있다 [16]. 본 연구는 이러한 연구 설계의 가장 기본적인 단계로 소요 비용을 고려하면서 충분한 검정력을 확보하고자 할 때 실제적인 도움을 줄 수 있는 데이터를 제시하는 것을 목적으로 하였다. 그러나 genomewide 연관 연구를 위한 적절한 연구 설계 방법에 대해서는 계속 논란 중에 있으며 더 다양한 방법론이 제시될 수 있을 것이다. 나아가 유전자간의 상호작용이나 유전자와 환경과의 상호작용, 인종적인 유전 차이 등을 밝히기 위해서는 보다 많은 연구 대상자수가 필요하며 따라서 더 정밀한 연구 설계가 고려되어야 할 것이다.

## 참고문헌


1. Thomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. *Am J Hum Genet* 2005; 77(3): 337-345
2. Wang H, Thomas DC, Pe'er I, Stram DO. Optimal two-stage genotyping designs for genome-wide association scans. *Genet Epidemiol* 2006; 30(4): 356-368
3. Satagopan JM, Elston RC. Optimal two-stage genotyping in population-based association studies. *Genet Epidemiol* 2003; 25(2): 149-157
4. Guedj M, Della-Chiesa E, Picard F, Nuel G. Computing power in case-control association studies through the use of quadratic approximations: Application to meta-statistics. *Ann Hum Genet* 2007; 71(Pt 2): 262-270
5. Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 2001; 69(1): 1-14
6. Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 1999; 22(2): 139-144
7. Schork NJ. Power calculations for genetic association studies using estimated probability distributions. *Am J Hum Genet* 2002; 70(6): 1480-1489
8. Ambrosius WT, Lange EM, Langefeld CD. Power for genetic association studies with random allele frequencies and genotype distributions. *Am J Hum Genet* 2004; 74(4): 683-693
9. Purcell S, Cherny SS, Sham PC. Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; 19(1): 149-150
10. Sham PC, Cherny SS, Purcell S, Hewitt JK. Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 2000; 66(5): 1616-1630
11. Gordon D, Finch SJ, Nothnagel M, Ott J. Power and sample size calculations for case-control genetic association tests when errors are present: Application to single nucleotide polymorphisms. *Hum Hered* 2002; 54(1): 22-33
12. Edwards BJ, Haynes C, Levenstien MA, Finch SJ, Gordon D. Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet* 2005; 6(1): 18
13. Gordon D, Haynes C, Blumenfeld J, Finch SJ. PAWE-3D: Visualizing power for association with error in case-control genetic studies of complex traits. *Bioinformatics* 2005; 21(20): 3935-3937
14. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; 38(2): 209-213
15. Kang D, Lee KM. Current status of genomic epidemiology research. *Korean J Prev Med* 2003; 36(3): 213-222
16. Park S. Statistical issues in genomic cohort studies. *J Prev Med Public Health* (Korean)(in press)