# Sample Size Calculations - Cochran-Armitage Test for Trend

Copyrighted © 2019 Suzanne M. Leal

Webpage for the exercises: http://ihg.helmholtz-muenchen.de/cgi-bin/hw/power2.pl

**Question 1**

For a complex disease study, you plan to collect 35,000 cases and 70,000 controls and wish to know if this is a sufficient sample size to detect an associations with disease susceptibility loci. The disease has a population prevalence of 5%. You wish to estimate the power for a genotypic relative risk of 1.2 and a disease allele frequency of 0.02. What is the power for $\alpha=5x10^{-8}$ under a under a multiplicative model ($\gamma_2 = \gamma_1^2$) a.)_____and dominant model ($\gamma_2 = \gamma_1$) b.) _____?

**Question 2**

For your study you hypothesize that you will try to replicate associations for 100 variants that are in linkage equilibrium and you want to reject the null hypothesis using a p-value of 0.05. What is the Bonferroni correction you should use a.)_____. Determine what your power would be if you used a Bonferroni correction to control for the Family Wise Error Rate (FWER). Using the parameters provided in question 1 but for a sample size of 20,000 cases and 20,000 controls what is the power under the multiplicative model b.)_____ and under a dominant model c.)_____?

**Question 3**

You determine that you can ascertain 50,000 cases and 50,000 controls what is the power using the same parameters as described in question 1 for the multiplicative model _____ and dominant model_____?

**Question 4**

The power of the Cochran-Armitage test for trend is dependent on the underlying genetic model. Using the parameters from question 1 which of the following underlying genetic models: multiplicative ($\gamma_2 = \gamma_1^2$), additive($\gamma_2 = 2\gamma_1 - 1$), dominant ($\gamma_2 = \gamma_1$) or recessive ($\gamma_1 = 1$) would you predict to be the most powerful a.)_____ and least powerful b.)_____?

**Question 5**

Will the greatest difference in power between the dominant and recessive model be when the variant minor allele frequency is high or low a.) _____?

**Question 6**

You have selected tagSNPs using a cut-off of a minor allele frequency (MAF)=0.2. You estimated the allele frequency for your disease allele to be 0.02. If the disease and a SNP are in complete linkage disequilibrium (LD) what is the possible maximum $r^2$ between the two loci a.) _____. Using the parameters in question 1 to have the same power how many cases b.) _____and controls c.) _____would you have to study? Hint: adjust the sample size using the maximum estimate of $r^2$.

**Question 7**

For study design with equal numbers of cases and controls a genotype relative risk of 1.5 under a recessive model for a disease with a population prevalence of 0.05 and disease allele frequency of 0.1. How many cases a.)_____ and controls b.) _____should you ascertain for $\alpha=5.0 \times 10^{-8}$ and $1-\beta=0.80$?

**Question 8**

You are performing a rare variant association study and you assume that that cumulative frequency of the causal variants in your gene region is 0.01 with every variant having an effect size of 1.4. The disease you are

studying has a prevalence of 5%. For a study with 0.8 power and an $\alpha=2.5 \times 10^{-6}$ under a dominant model for equal numbers of cases and controls what is the total sample size a.) _____ do you need to ascertain. What is the total sample size b.)_____you need to ascertain if the cumulative frequency of causal variants is only 0.005?

**Question 9**

You are performing a study using the UK Biobank and for your phenotype of interest you have 50,000 cases and 100,000 controls. For a disease with 10% prevalence, disease allele frequency of 0.01, where each variant has an effect size of 1.2 under a dominant model what would be the power for an aggregate test where the cumulative allele frequency is 0.01 _____and a single variant test _____? Clue use the appropriate alpha for each test.

**Question 10**

Using have a replication sample of 25,000 cases and 25,000 controls and you plan to try to replicate 15 genes and 100 variants. Using the same parameters as in question 9 what would be your power to replicate a.)_____? Note for alpha use a Bonferroni correction.

ANWSERS

1. a.) 0.74 b.) 0.69
2. a.) $5.0x10^{-4}$ a.) 0.69 b.) 0.655
3. b.) 0.77 c.) 0.73
4. a.) multiplicative b.) recessive
5. a.) high
6. a.) 0.082 b.) 426,830 c.) 853,659
7. a.) 170,910 b.) 170,910
8. a.) 42,925 b.) 84,243
9. a.) 0.78 b.) 0.48 Hint: use $\alpha=5x10^{-8}$ for single variant test and $\alpha=2.5x10^{-6}$ for the aggregate test
10. a.) 0.75 (Hint: use $\alpha=8.7x10^{-3}$)

Calculation of $r^2$ for question 6A
Frequencies for the marker $A_1=0.2$ and $A_2=0.8$ and disease and wild type allele $B_1=0.02$ and $B_2=0.98$
Frequencies of Haplotypes under complete linkage disequilibrium
$A_1B_1=0.02$
$A_1B_2=0.18$
$A_2B_1=0$
$A_2B_2=0.8$
D=0.02*0.8-0.18*0=0.016; product of the allele frequencies 0.2*0.8*0.02*0.98=0.003136 and their square root 0.056
r=0.016/0.056=0.286 $r^2=0.0816$