Power Analysis for Individual Variants and Rare Variant Aggregate Association Analysis

Suzanne M. Leal
Baylor College of Medicine
sleal@bcm.edu

Copyrighted © S.M. Leal 2017

Why Estimate Sample Sizes and/or Power?

1.) Not wasting your time and money
   Carrying out a study for which you will never find a true association due to inadequate sample sizes

2.) Almost always necessary for grant proposals
   Usually will be denied funding if cannot demonstrate planned study has adequate power

Power and Sample Size Estimation for Case-Control Data

- The correct $\alpha$ must be used for sample size estimation/power analysis
- Type I ($\alpha$) the probability of rejecting the null hypothesis of no association when it is true
- Due to multiple testing a more stringent value than $\alpha=0.05$ is used in order to control the Family Wise Error Rate

Power and Sample Size Estimation for Case-Control Data

- GWAS of common variants where each variant is tested separately
  - $\alpha=5 \times 10^{-8}$ (Bonferroni Correction for testing 1,000,000 variant sites)
  - Shown to be a good approximation for the effective number of tests
    - Valid even when more than 1,000,000 variant sites tested
    - Effective number of tests is dependent of the LD structure
- Analysis of individual variants for whole genome or exome sequence data
  - $\alpha$ yet to be determined

Selecting Alpha value for Power Calculations

- Various criterion have been suggested for controlling the family wise error rate (FWER)
- The Wellcome Trust study
  - First study tested approximately 500,000 markers
  - Used a genome wide significance level of $5.0 \times 10^{-7}$
  - Less conservative than a Bonferroni correction (i.e. $1 \times 10^{-7}$)
- Is this criterion stringent enough when 1 million SNPs are tested?

Selecting Alpha for Power Calculations

- New SNP chips with 2.5M and 5M SNP loci
- Due LD between markers after a certain threshold although N increases the effective number of tests $N'$ does not increase
### Determining Genome-wide Significance Levels

- Using genotypes from the Wellcome Trust Case-Control Consortium
- Dudbridge and Gusnato, Genet Epidemiol 2008
- Estimated a genome wide significance threshold for the UK Caucasian population

### Determining Genome-wide Significance Levels
- By sub-sampling the genotypes at increasing densities and using permutation to estimate the nominal p-value for 5% family-wise error
- Then extrapolating to infinite density
- The genome wide significance threshold was estimated to be ~7.2x10^{-8}
- Should a more or less stringent GWAS significance criterion be used for an African population?

### In the End
- GWAS studies generally use a criterion of \( p \leq 5 \times 10^{-8} \)
- Bonferroni correction for testing 1,000,000 loci

### Power and Sample Size for Aggregate Rare Variant Tests
- For gene based methods a Bonferroni correction for the number of genes/regions tested is used
  - e.g. 20,000 genes significance level \( \alpha = 2.5 \times 10^{-6} \)
  - Can usually use a less stringent criteria
    - Not all genes have two or more variants
      - Divide 0.05 by number of genes tested
- Little LD between variants in separate genes
  - Little to no correlation between tests

### Case-Control Data
- Can use a variety of statistical packages to calculate sample size.
  - Free Programs
    - Epiinfo
    - PS Power and Sample Size Calculator
- These programs were not developed for genetic data
  - Can be useful for
    - Quantitative traits
    - Mitochondrial inheritance

### Estimating Power/Sample Sizes For Individual Variants
- Can be obtained Analytically
- Information necessary
  - Prevalence
  - Risk allele frequency
  - Effective size (odds ratio-for case control data)
  - Genetic model for the susceptibility variant
    - Recessive \( (\gamma_1=1) \)
    - Dominant \( (\gamma_2=\gamma_1) \)
    - Additive \( (\gamma_2=2\gamma_1-1) \)
    - Multiplicative \( (\gamma_2=\gamma_1^2) \)
Estimating Power/Sample Sizes For Individual Variants

- Usually information on disease prevalence is known from epidemiological data
- A range of risk allele allele frequencies and effect sizes are used
- A variety of genetic models are also used
  - Dominant
  - Additive
  - Multiplicative

Armitage Trend Test

- Power and Sample size
  - Calculated under different models
    - Where $\gamma$ is the genotypic odds ratio
      - Multiplicative
        - $\gamma_2 = \gamma_1^2$
      - Additive
        - $\gamma_2 = 2\gamma_1 - 1$
      - Dominant
        - $\gamma_2 = \gamma_1$
      - Recessive
        - $\gamma_1 = 1$

Armitage Trend Test

Power Calculations

- Information need
  - Population prevalence
  - Genetic Model
  - Risk allele frequency
- Tool
  - http://ihg.gsf.de/cgi-bin/hw/power2.pl
  - Reference Slager and Schaid 2001

Genetic Power Calculator

- S Purcell & P Sham
  - http://statgen.iop.kcl.ac.uk/gpc/
- Uses the methods described in Sham PC et al. (2000) Am J Hum Genet 66:1616-1630

Genetic Power Calculator

- VC QTL linkage for sibships
- VC QTL association for sibships
- VC QTL linkage for sibships conditional on the trait
- TDT for discrete traits
- Case-Control for discrete traits
- TDT for quantitative traits
- Case-Control quantitative traits
### PAWE

- **Power Association With Errors**
  - Will give same results for case-control studies of discrete traits as Genetic Power Calculator when calculations are done without errors.
- Four different error models can be used
  - See online documentation for complete explanation

### PAWE

- Can either perform:
  - Power calculations for a fixed sample size
  - Sample size calculations for a fixed power
- The genotype frequencies can be generated either using a:
  - Genetic model free method or
  - Genetic model based method

### Quanto

Provides sample size and power calculations for testing
- Genetic and environmental main effects
- Interactions
  - Gene x gene
  - Gene x environment

### Quanto

- Sample & power calculations can be carried for:
  - Case-control
    - Unmatched
    - Matched
  - Case-sibling
  - Case-parent (trios)
    - Quantitative
    - Qualitative
  - Independent sample of individuals
    - Quantitative traits
      - Assumption sampled from a random population

### Linkage Disequilibrium (LD)

- Power will be reduced if causal variant is not in perfect LD ($r^2=1$) with causal variant.
- Can adjust sample size when $r^2$ is <1 to increase power to the same level as when $r^2=1$

### Power Analysis for Rare Variant Aggregate Association Tests

- More complex than power analysis for individual variants
  - Many unknown parameters must be modeled
  - Usually must be estimated empirically
- Allelic architecture within a genetic region, e.g. gene
  - Allelic architecture varied across genes and populations
- Effects of variants within a region
  - Fixed or varied effect sizes of causal variants
  - Bidirectional effect of variants
  - Proportion of non-causal variants
**Generating Variant Data**

- Using population demographic models
  - European
    - Gazave et al. 2013 *Proc Natl Acad Sci*
  - African
    - Excoffier et al. 2013 *PLoS Genetics*
- Using sequence data from a specific population
  - E.g. NHLBI-Exome sequencing project
    - European-Americans N=4,332
    - African-Americans N=*
  - Resampling with replacement
  - Allele frequencies

**Comparison of Simulating Sequence Data using Whole Exome Sequence Data and Population Demographic Models**

- Data generated on 16,568 genes
- Simulating variant data using reference sequence data a European population demographic model
  - Gazave et al. 2013
  - Haplotype pool generated for each gene
    - Each pool contains 1,308,000 haplotypes
- Simulations performed using variant data obtained from 4,332 European-Americans from NHLBI-ESP
  - Resampling with replacement
  - Allele frequencies

**Golden Standard**

- Sequence data obtained from 4,332 European-Americans from NHLBI-ESP
  - 16,568 genes
  - Sampled without replacement

**Does Generating Variant Data Using the European Population Demographic Model Perform Well?**

- **Simulated Data**
- **ESP Data**

**Simulating Data Using Sequence Data (ESP)*

*Only appropriate to generate small data sets, e.g. <1,000 samples

- Number of Variant Sites
- Proportion of Variant Sites that are Singletons, Doubletons and Tripletons

- Distribution of cumulative MAF per gene
- Proportions of Singletons, Doubletons and Tripletons
Simulating Data: Using Population Demographic Models (PDM)*

*Resample or using MAF to generate data from large haplotype pools

Number of Variant Sites  Proportion of Variant Sites that are Singletons, Doubletons and Tripletons

Simulation Studies for Evaluating Power

Variant data simulated for 16,568 genes
Five-thousand genes randomly sampled which have >1 variant site and are <8,000bp
Variants Increase Risk based using a variable effects model, e.g. odds ratio (OR)=1.2-3.0
A sample of 3,000 cases and 3,000 controls are generated for each gene
Only missense, nonsense and splice sites with a MAF of <1% are analyzed
Power Evaluated for $\alpha=2.5 \times 10^{-6}$

SEQPower

http://www.bioinformatics.org/spower/

What is the most Powerful Rare Variant Association Method?

- It is difficult to compare power of rare variant association tests by reading the literature
  - Different methods are used to generate data
  - Variant and/or disease model is often generated in a manner to display a particular rare variant association test to appear to be superior to other methods

Almost Every Test will give the Most Significant results at Least Some of the Time

- Ideally a test should be powerful for a variety of genes and underlying genetic etiologies
- Ranking of a test varies due to
  - Number of variant sites
  - Distribution of MAF
  - Distribution of functional variants

Data generated using the variable effect model: ORs 1.2-3.0

Power Comparison for Rare Variant Association Methods
**Power Comparisons for Rare Variant Association Methods**

Proportion of Strongly Protective Variants (with Remain Proportion Increasing Disease Risk)

Variable effect size: Protective variants OR 0.1-0.3 & susceptibility variants ORs 1.2-3.0

**What is the Most Powerful Rare Variant Association Test?**

Proportion of Mildly Protective Variants (with Remain Proportion Increasing Disease Risk)

Variable effect size: Protective variants OR 0.33-0.83 & susceptibility variants ORs 1.2-3.0

---

**Example 2 Simulation Studies to Evaluate Power**

Using Bokyo et al. 2008 European demographic model variants simulated for a gene which is 1,500bp

A sample of 1,000 cases and 1,000 controls are generated for each gene

“Causal” variants are generated using variable and fixed effect models

Only non synonymous variants (based upon simulated selection coefficients) MAF of <1% are analyzed

P-values are obtained empirically using 3,000 permutations

Power Evaluated for α=0.05 using 2,000 replicates

---

**Power Analysis Impact of Non-Causal Variants**

- Fixed effect Model
  - OR=2.0
- Disease prevalence 1%
- Sample Sizes
  - 1000 cases & 1000 controls
- Empirical p-values
  - One-sided test
  - α=0.05

---

**Power Analysis Detrimental and Protective Variants**

- Variable effect Model
  - Detrimental OR=1.5_{min}, 5_{max}
  - Protective OR=0.6_{min}, 0.2_{max}
- Disease prevalence 1%
- Sample Sizes
  - 1000 cases & 1000 controls
- Empirical p-values
  - α=0.05
### Power Analysis

**Detrimental and Protective Variants**

- **Fixed effect Model**
  - Detrimental OR=2.0
  - Protective OR=0.5
- **Disease prevalence**
- **Sample Sizes**
  - 1000 cases & 1000 controls
- **Empirical p-values**
  - $\alpha=0.05$

### Reasons for Low Power for Burden Tests

- Reasons for low power for burden tests
  - Very few causal variants within gene region
  - Causal variants have very low frequencies
    - Singletons
  - High rates of contamination within region of non casual variants
- Small effect sizes of causal variants
- Causal variants have effect sizes in different directions

### Power Analysis Rare Variants (Burden Tests)

- Power will not only vary between traits greatly
- The power to detect an association will also vary drastically between genes
- For some genes even with hundreds of thousands of samples power will still be low, while for others a few thousand samples may be sufficient

### No Most Powerful Rare Variant Association Method

- Most powerful method dependent on the underlying genetic model, which is usually unknown
  - No clear winner
  - Differences in power usually modest
- Should use a method which is powerful for a variety of different genetic models
- Can chose to perform different tests which are powerful in different circumstances
  - However must correct for multiple testing
    - Therefore may lead to a loss of power instead of a gain in power

---

**Comparison of Complex Trait Rare Variant Association Methods**

- Various tests have superior power in different circumstances
  - However for many tests there is only small differences in their power
    - e.g. WSS, VT, KBAC, CMC & GRANVIL
    - If testing for an association with detrimental variants using a one-sided test can increase power
  - Most tests which were developed to detect associations with variants within a region which are bidirectional
    - Not very powerful when the variant effects are unidirectional