Conducting Pharmacogenomics Research: A Primer

2:30 p.m.–4:30 p.m.
Convention Center: 213 D

Moderator: J. Herbert Patterson, Pharm.D., FCCP
Professor of Pharmacy and Research Professor of Medicine, UNC
Eshelman School of Pharmacy, University of North Carolina, Chapel
Hill, North Carolina

2:30 p.m. Using the PBRN as a Research Tool for Pharmacogenomics
Research
Grace M. Kuo, Pharm.D., MPH
Director of ACCP PBRN, American College of Clinical Pharmacy
Research Institute, Lenexa, Kansas; Associate Professor,
University of California, San Diego, California; Skaggs School of Pharmacy and Pharmaceutical
Sciences; Associate Adjunct Professor of Family and Preventive
Medicine, UCSD School of Medicine; Director of San Diego
Pharmacist Resource & Research Network, La Jolla, California

3:00 p.m. Sorting the Wheat from the Chaff: Statistical Issues with
Pharmacogenomics
Alison A. Motsinger-Reif, Ph.D.
Assistant Professor, North Carolina State University, Raleigh,
North Carolina

3:30 p.m. Pro-Con Debate on Prospective Study Designs: Convenience
Cohorts vs. Randomized Samples
Michael A. Pacanowski, Pharm.D.
Clinical Pharmacologist, U.S. Food and Drug Administration,
Silver Spring, Maryland

Craig R. Lee, Pharm.D., Ph.D.
Assistant Professor of Pharmacy, University of North Carolina at
Chapel Hill, Eshelman School of Pharmacy, Chapel Hill, North
Carolina

Faculty Conflict of Interest Disclosures

Grace M. Kuo: consultant/member of advisory board for ACCP’s Practice Based Research Network (PBRN); clinical investigator for PharmGenEd™ Education Program; received grant funding/research
support from Centers for Disease Control and ACCP Research Institute (RI); received assistance to attend this meeting from ACCP’s RI.

Craig R. Lee: no conflicts to disclose.
Alison A. Motsinger-Reif: no conflicts to disclose.
Michael A. Pacanowski: no conflicts to disclose.

**Learning Objectives**

1. Provide examples of pharmacogenomics research that can be conducted within a practice-based environment such as a PBRN.
2. Identify challenges to conducting pharmacogenomics research within a practice-based environment.
3. Discuss benefits to conducting pharmacogenomics research within a practice-based environment.
4. Provide examples of appropriate statistical techniques for use with pharmacogenomics studies.
5. Discuss statistical challenges commonly encountered when conducting pharmacogenomics research.
6. Critically appraise statistical techniques used in published pharmacogenomics papers.
7. Identify factors to consider in choice of study design for a pharmacogenomics study.
8. Articulate the positive reasons for using a convenience cohort when conducting a pharmacogenomics study.
9. Discuss the logic against using a randomized sample in conducting a pharmacogenomics study.

**Self-Assessment Questions**

Self-assessment questions are available online at www.accp.com/sf
Using the PBRN as a Research Tool for Pharmacogenomics Research

April 24, 2010
Presenter: Grace M. Kuo, PharmD, MPH (UCSD)

Objectives

- Provide examples of pharmacogenomics research that can be conducted within a practice-based environment such as a PBRN.
- Identify challenges to conducting pharmacogenomics research within a practice-based environment.
- Discuss benefits to conducting pharmacogenomics research within a practice-based environment.

ACCP PBRN

**Mission Statement**

- The mission of the ACCP PBRN is to facilitate collaborative research that promotes the safe, efficacious, and cost-effective use and delivery of medications and clinical pharmacy services.

ACCP PBRN Members

- Total n=695
  - 416 individual members +
  - 257 members (from existing PBRNs and integrated health systems +
  - 22 members opted out of this research arm.
- Currently involved in clinical research as an investigator, sub-investigator or study coordinator.
  - 248 (60%) are currently involved
- Average of 9 years since terminal degree/training.

ACCP PBRN Members *(n=416)*

*one dot denotes one zip code*

Clinical Research Experience Among PBRN Members *(n = 416)*

- No Exp: 5%
- PI: 51%
- CI: 67%
- Study Coord.: 19%
- Student: 73%
- Other: 5%
Types of Human Subjects
Research Conducted by PBRN Members (n = 416)

<table>
<thead>
<tr>
<th>PBR</th>
<th>Clinical</th>
<th>Industry</th>
<th>Phase I-IV</th>
<th>Survey</th>
<th>None</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>60%</td>
<td>60%</td>
<td>30%</td>
<td>38%</td>
<td>8%</td>
<td>4%</td>
<td></td>
</tr>
</tbody>
</table>

ACCP PBRN Registry: PRN Counts*

<table>
<thead>
<tr>
<th>CODE</th>
<th>PRN Name</th>
<th>COUNT</th>
<th>PRN Code</th>
<th>PRN Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>97</td>
<td>Adult Medicine</td>
<td>126</td>
<td>8402</td>
<td>Women's Health PRN</td>
</tr>
<tr>
<td>98</td>
<td>Ambulatory Care</td>
<td>183</td>
<td>8404</td>
<td>Health Outcomes</td>
</tr>
<tr>
<td>99</td>
<td>Cardiology</td>
<td>26</td>
<td>8408</td>
<td>Geriatrics PRN</td>
</tr>
<tr>
<td>100</td>
<td>Central Nervous System</td>
<td>32</td>
<td>8410</td>
<td>Pharmacokinetics/Pharmacodynamic sPKPD</td>
</tr>
<tr>
<td>101</td>
<td>Clinical Pharmacology</td>
<td>17</td>
<td>8412</td>
<td>Endocrine and Metabolism ENDO</td>
</tr>
<tr>
<td>102</td>
<td>Critical Care</td>
<td>26</td>
<td>8414</td>
<td>Pharmaceutical Industry INDU</td>
</tr>
<tr>
<td>103</td>
<td>Drug Information</td>
<td>22</td>
<td>8416</td>
<td>Emergency Medicine EMED</td>
</tr>
<tr>
<td>104</td>
<td>Education and Training</td>
<td>45</td>
<td>8418</td>
<td>Pediatrics</td>
</tr>
<tr>
<td>105</td>
<td>Emergency Medicine</td>
<td>26</td>
<td>8420</td>
<td>Pediatrics</td>
</tr>
<tr>
<td>106</td>
<td>Endocrine and Metabolism</td>
<td>10</td>
<td>8422</td>
<td>Pain and Palliative Care PAIN</td>
</tr>
<tr>
<td>107</td>
<td>Geriatrics</td>
<td>15</td>
<td>8424</td>
<td>Drug Information DINF</td>
</tr>
<tr>
<td>108</td>
<td>Nephrology</td>
<td>17</td>
<td>8426</td>
<td>Pediatrics</td>
</tr>
<tr>
<td>109</td>
<td>Critical Care</td>
<td>85</td>
<td>8428</td>
<td>Pediatrics</td>
</tr>
<tr>
<td>110</td>
<td>Pain and Palliative Care</td>
<td>17</td>
<td>8430</td>
<td>Pain and Palliative Care PAIN</td>
</tr>
<tr>
<td>111</td>
<td>Education and Training</td>
<td>36</td>
<td>8432</td>
<td>Pediatrics</td>
</tr>
</tbody>
</table>

*Note: An individual may belong to more than 1 PRN

ACCP PBRN Registry: Unique Clinical Sites (n=263)

- 45 States Represented
- 263 Sites Registered + 105 sites from existing PBRNs and integrated health systems
- 95% of sites in urban areas vs. 5% rural areas
- Ethnicity distribution of patients seen at sites
  - Hispanic or Latino: 20%
  - Not-Hispanic and Latino: 77%
  - Unknown: 3%
- Racial distribution of patients seen at sites
  - White, Caucasian: 58%
  - Black, African American: 28%
  - Asian: 8%
  - Native Hawaiian/Other Pacific Islander: 2%
  - American Indian/Alaska Native: 2%
  - Unknown: 2%

ACCP PBRN Registry: Unique Clinical Practice Sites (n=263)

<table>
<thead>
<tr>
<th>Inpatient (n)</th>
<th>Outpatient (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Hospital or Health System</td>
<td>81% (63)</td>
</tr>
<tr>
<td>University Hospital or Academic Center</td>
<td>74% (89)</td>
</tr>
</tbody>
</table>

ACCP PBRN Registry: Unique Clinical Sites (n=263)

- 172 (65%) site have EMR
- Patient chart characteristics:
  - 7% use paper charts
  - 23% are totally paperless
  - 70% use a hybrid system
- 37 (31%) of sites have a central IRB
ACCP PBRN Registry: Clinical Practice
Mean (+/- SD)
- 344 practices registered
- 83% Provide clinical pharmacy services
  - Half-days/week: Mean 5.4(4)
  - Number of patients seen/week: Mean 42(42)
- Patient distribution:
  - Adults: 76% 
  - Pediatric: 9%
  - (unspecified:15%)
- 35% have collaborative practice agreements
- 33% have scope of practice agreements

Tests Pharmacists Perform/Order (n=344)

Conditions Routinely Managed By Pharmacists (n=344)

Summary of ACCP PBRN
- ACCP PBRN includes over 600 pharmacists representing almost every state in the US.
- Most members have research experience.
- Majority of clinical practice sites are located in the urban areas serving patients with multiethnic backgrounds.
- Clinical services provided by network pharmacists include pharmacotherapy and chronic disease management (anticoagulation, diabetes, hypertension, hyperlipidemia, heart failure, etc.).

Practice-Based Research—"Blue Highways" on the NIH Roadmap

Examples
- Clinical trials
  - Case-series
  - Observational, population studies
  - Case-control or Cohort studies
  - Randomized Controlled trials
- Clinical practice
  - Practice patterns
  - Whatahow pharmacogenomics is integrated into clinical practice
- Bridging the gap
  - Health-system interventions and implementation
  - Training programs
  - Assessment and evaluation of clinical practice based on guidelines and evidence-based recommendations
## Challenges

- Multiple IRB approval processes and requirements
- Lack of resources to support multi-site studies
- Consistency in multiple investigator training and subject recruitment
- Lack of integration of informatics tools from multiple sites
- Need for central coordinating center and statistical cores

## Benefits

- Large sample sizes and study power
- Wide geographical distribution of sites
- Collaborative efforts among investigators and participants
- Community engagement
- Enhanced academic and community partnerships
- Increased generalizability

## Questions or Comments?

Contact:
Email: gmkuo@ucsd.edu
Phone: (858) 822-7751
Sorting the Wheat from the Chaff: Statistical Issues with Pharmacogenomics

Alison Motsinger-Reif, Ph.D.
Bioinformatics Research Center
Department of Statistics
North Carolina State University
motsinger@stat.ncsu.edu

Learning Objectives

• Provide examples of appropriate statistical techniques for use with pharmacogenomics studies.
• Discuss statistical challenges commonly encountered when conducting pharmacogenomics research.
• Critically appraise statistical techniques used in published pharmacogenomics papers.

Steps in “Gene Mapping”

• Defining a phenotype for association mapping
• Determining the genetic component of a trait
• Study designs and analytical tools
• Genotyping strategies
• Replication, validation, and interpretation of results

Steps in “Gene Mapping”

• Defining a phenotype for association mapping
• Determining the genetic component of a trait
• Study designs and analytical tools
• Genotyping strategies
• Replication, validation, and interpretation of results
• Statistics plays a crucial role in EVERY stage of gene mapping!

Defining a Phenotype

• Defining a phenotype is both a biological and statistical choice
• Especially true in PGX
  – Outcomes are often generated through modeling
  – ADME modeling, PK/PD modeling, etc.

Broad Classes of Phenotypes

• Qualitative traits
  – Presence or absence
    • Toxicities: Affected or unaffected
    • PGX: responders vs. nonresponder
  – “Unaffected” does not equal absence of trait
  – Threshold-based diagnosis
    • Toxicities, response, etc
    • May lose information
Broad Classes of Phenotypes

- **Quantitative traits**
  - Continuous measurements
    - Stable dose, blood pressure, etc
  - Advantages of quantitative traits
    - Many complex traits have quantitative characteristics that are directly related to trait risk
    - Can provide more effective descriptions of complex diseases
    - Analyses on quantitative traits can be powerful alternatives to analyses directly on disease status

Summary Points on Phenotyping

- Well defined phenotypes are crucial for association mapping!
  - Becomes increasingly important as association studies grow in scale
- Any type of phenotype can be evaluated in any study design.
  - Just need to match up
- When planning or reviewing a study, need to think about the consequences of the phenotype definition choices.

Determining the Genetic Component of a Trait

- First step in a gene mapping study is determining whether a trait has a genetic component.
  - Characterizing the sharing
- Characterizing the genetic basis of a trait is important before starting a mapping study.
  - This can be a particular challenge in PGX outcomes!!!

Methods for Assessing the Genetic Component of a Trait

- Familial aggregation
- Twin Studies
- Segregation analysis
- Increased risk to relatives
- Animal models

Heritability of a Quantitative Trait

- \( h^2 \): Proportion of observed variance in phenotype explained by genetic factors
  - \( h^2 > 0 \) indicates the presence of genetic contributions to the trait
  - Magnitude indicates “how genetic” the trait is
- Decomposition of total variance:
  - \( \sigma^2_I = \sigma^2_G + \sigma^2_e \) (total = genetic + environmental)
  - \( \sigma^2_G = \sigma^2_A + \sigma^2_D + \sigma^2_I \) (genetic = additive + dominance + interaction among genes)
  - \( \sigma^2_e = \sigma^2_I + \sigma^2_R \) (environmental = familial/household + random/individual)
- Broad sense heritability \( h^2_B = \sigma^2_G / \sigma^2_I \)
- Narrow sense heritability \( h^2_N = \sigma^2_A / \sigma^2_I \) (more commonly used)

Estimation of Heritability

- Twin studies: \( h^2 = 2(r_{MZ} - r_{DZ}) \)
- Pedigree data: Estimate \( \sigma^2_G \) (or \( \sigma^2_I \)) using information on relationship, ascertainment criteria, and covariates (age, gender, etc) using variance components methods
  - MERLIN
    - http://www.sph.umich.edu/csg/abecasis/merlin/
  - SOLAR
    - http://solar.sfbrgenetics.org/
Estimation of Heritability

- Caveats:
  - Estimates are dependent on model assumptions
  - Estimates may be difference across populations even if the genetic contribution is the same
  - Over-estimation may result from failure to adjust for important covariates, failure to include important variance components, failure to correct for ascertainment
  - Under-estimation may result from the inclusion of too many covariates in the model

Genetic Component of Qualitative Traits

\[ \lambda_x = \frac{K_x}{K} \]

- Encompasses all genetic and shared environmental effects, not just those due to a single locus
- K for the general population is often estimated from previous studies
- Note: the magnitude of the estimate is very dependent on the frequency in the population

Summary Points on Establishing Genetic Components

- When planning or reviewing a study, evaluate “how genetic” a trait is.
- How much of the variation is explained by known genetic components?
  - Gap between heritability and known effects motivates follow up studies

Designs for Association Studies

- Population-based and family-based designs
  - Can evaluate any type of phenotype within any study design
  - Practical and theoretical advantages and disadvantages of each design, particularly for PGX outcomes
- Wide range of association tools available
  - Specifics depend on study design and types of independent and dependent variables
- Main goal is the same: correlate phenotypic and genotypic variability!

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-Sectional</td>
<td>Genotype and Phenotype collected across a random sample of the population; quantitative or qualitative traits</td>
</tr>
<tr>
<td>Cohort</td>
<td>Genotype subsection of population and follow disease incidence for a specific time period</td>
</tr>
<tr>
<td>Case-Control</td>
<td>Genotype collection of individuals with trait/phenotype, matched with samples without the trait</td>
</tr>
<tr>
<td>Extreme Values</td>
<td>Genotype collection of individuals at the upper and lower extremes of quantitative trait distribution</td>
</tr>
<tr>
<td>Trios, Sibling Pairs</td>
<td>Genotype affected individuals plus their parents or</td>
</tr>
<tr>
<td>Case-parent-grandparent sibsets</td>
<td>Genotype affected individuals plus their parents and grandparents</td>
</tr>
<tr>
<td>General Pedigrees</td>
<td>Genotype and phenotype random sample or trait selected sample of families from the general population</td>
</tr>
<tr>
<td>Case-only</td>
<td>Genotype only affected individuals</td>
</tr>
</tbody>
</table>
### Study Design

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross Sectional</td>
<td>Inexpensive; provides estimates of disease prevalence</td>
<td>Few affected individuals if the disease is rare</td>
</tr>
<tr>
<td>Cohort</td>
<td>Provides estimate of disease prevalence</td>
<td>Expensive to follow-up; drop-out can cause issues</td>
</tr>
<tr>
<td>Case-Control</td>
<td>No need to follow-up; provides estimates of exposure effects</td>
<td>Requires careful selection of controls; potential for confounding (population stratification, etc)</td>
</tr>
<tr>
<td>Extreme Values</td>
<td>Genotype only the most informative individuals so rare on genotyping costs</td>
<td>No estimate of true genetic effect sizes</td>
</tr>
<tr>
<td>Trios, Sibling Pairs</td>
<td>Robust to population stratification; can estimate maternal and imprinting effects</td>
<td>Less powerful than case-control design</td>
</tr>
<tr>
<td>Grandparent-grandparent septets</td>
<td>Robust to population stratification; can estimate maternality and imprinting effects</td>
<td>Grandparents rarely available</td>
</tr>
<tr>
<td>General Pedigrees</td>
<td>Higher power with large families; samples may already exist from linkage studies</td>
<td>Expanse to genotype; many missing individuals</td>
</tr>
<tr>
<td>Case-only</td>
<td>Very powerful design for detection of interactive effects</td>
<td>Very sensitive to population stratification</td>
</tr>
</tbody>
</table>

### Statistical Analysis Method

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Statistical Analysis Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-Sectional</td>
<td>Logistic regression, Chi-square tests of association; linear regression; nonparametric</td>
</tr>
<tr>
<td>Cohort</td>
<td>Survival analysis methods</td>
</tr>
<tr>
<td>Case-Control</td>
<td>Logistic regression; Chi-square tests of association; nonparametric</td>
</tr>
<tr>
<td>Extreme Values</td>
<td>Linear regression, non-parametric, or permutation approaches</td>
</tr>
<tr>
<td>Trios, Sibling Pairs</td>
<td>Transmission/disequilibrium test; conditional logistic regression, log-linear models</td>
</tr>
<tr>
<td>Case-parent-grandparent septets</td>
<td>Linear models</td>
</tr>
<tr>
<td>General Pedigrees</td>
<td>Pedigree Disequilibrium Test (PDT); Family-based Association tests (FBAT); Quantitative Transmission/disequilibrium test (Q2DT)</td>
</tr>
<tr>
<td>Case-only</td>
<td>Logistic regression; Chi-Square tests of association; nonparametric</td>
</tr>
</tbody>
</table>

### Specific Concerns In PGX

- Family based samples are rarely available
- Rare adverse events can limit sample size
- Nesting within clinical trials can limit study design and sample size
  - Consent
  - Treatment arms

### Study Design Conclusions

- Lots of options for study design
  - Sample collection and genotyping
- Choices depend on resources
  - Sample availability, budget
- Analytical methods depend on details of study design

### Statistical Methods for Data Analysis

- The study design, type of phenotype, and distributional assumptions make a decision tree for the choice of statistical test
  - Parametric vs. nonparametric tests
- Genotypes enter the statistical model as categorical variables
  - Encoding makes genetic assumptions
  - Dominance, additivity, etc.

### Analysis Method

<table>
<thead>
<tr>
<th>Analysis Method</th>
<th>Description</th>
<th>Software</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>Model log of odds of disease as a linear function of genotype</td>
<td>Standard statistical packages</td>
<td><a href="http://www.cancer.gov">www.cancer.gov</a> <a href="http://www.sas.com">www.sas.com</a> <a href="http://www.r-project.org">www.r-project.org</a></td>
</tr>
<tr>
<td>Chi-Square Test of Association</td>
<td>Test for independence of disease status and genotype status</td>
<td>Standard statistical packages</td>
<td><a href="http://www.linkage.rockefeller.edu/soft/gh/">www.linkage.rockefeller.edu/soft/gh/</a></td>
</tr>
<tr>
<td>Linear Regression</td>
<td>Model quantitative trait as a linear function of genotype</td>
<td>Standard statistical packages</td>
<td><a href="http://www.pgenet.org">www.pgenet.org</a></td>
</tr>
<tr>
<td>Survival Analysis</td>
<td>Model survivor function or hazard as a function of genotype</td>
<td>Standard statistical packages</td>
<td><a href="http://www.cimr.cam.ac.uk/clayton/software">www.cimr.cam.ac.uk/clayton/software</a></td>
</tr>
<tr>
<td>Transmission/Disequilibrium Test</td>
<td>Test for departure of allele transmission from heterogenous parents to affected offspring from null hypothesis of 1/2</td>
<td>Various (e.g. GenetHunter, GeneAssoc, Unphased)</td>
<td><a href="http://dbiologia.unifr.it/alg/">http://dbiologia.unifr.it/alg/</a> <a href="http://www.mrc-bsu.cam.ac.uk/personal/frank/softwar">http://www.mrc-bsu.cam.ac.uk/personal/frank/softwar</a> <a href="http://www.sas.com">www.sas.com</a> <a href="http://www.r-project.org">www.r-project.org</a> <a href="http://www.lisa.cam.ac.uk/personal/leem/software/">www.lisa.cam.ac.uk/personal/leem/software/</a></td>
</tr>
</tbody>
</table>
### Analysis Method | Description | Software | Links
---|---|---|---
Conditional Logistic Regression | Calculate conditional probability of affected offspring genotypes given parental genotypes | GenAssoc Unphased | http://www.gene.cimr.cam.ac.uk/clayton/software/stata/genassoc/ www.mrc-bsu.cam.ac.uk/personal/frank/software/unphased/
Log Linear Models | Model counts of genotype combinations for mother, father, and offspring | Standard statistical packages | See previous slide
Pedigree Disequilibrium Test | Test for departure of allele transmission to affected pedigree members from null | PDT | http://www.chg.duke.edu/software/pdt.html
Family-Based Association Tests | Test for linkage or association between traits and haplotypes using family based controls | FBAT | http://biosun1.harvard.edu/~fbat/fbat.htm
Quantitative TDT | LD analysis based on variance components | QTDT | http://www.sph.umich.edu/csg/abecasis/QTDT/

### Considerations in Data Analysis
- Are the methods applied appropriate?
  - Match the study design, etc.
  - Distributional assumptions were checked
- What were the genetic assumptions in the model and were they appropriate?
- Was the study well powered? How does this influence conclusions?

### Genotyping Strategies
- Genotyping technology is rapidly changing genotyping strategies in association studies
  - Association analyses are the same within each strategy
  - Scale is different
- Candidate Gene
- Genome-Wide Association Studies
- Next-generation Sequencing
  - Open methods questions.....

### Statistical Issues in Genotyping
- Just as statistics plays an important role in phenotyping, it plays a crucial role in genotyping as well.
- Important steps in quality control (QC) for genotype data:
  - Tests for Hardy-Weinberg Equilibrium
    - Can detect genotyping error
    - Chi-square, trend, or exact tests used to test observation versus expectation of genotype frequencies
  - Genotyping efficiency
    - Statistical algorithms for genotype calling
  - Population Stratification
    - Principle Component Analysis (PCA) can be used to detect association

### Candidate Gene Studies
- How are candidate genes chosen?
  - Biological/network knowledge, drug mechanism
  - Clinical knowledge
  - Previous studies
- How do you pick variants within genes?
  - Potential functional significance
  - Population frequencies
  - Linkage disequilibrium in the gene
- Higher power than GWAS when the candidates are correct
  - Fewer tests
- Limited in potential to identify “novel biology”

### Genome Wide Association Studies
- Genotype 100K to 3M SNPs per individual
- Two Major Platforms:
  - Affymetrix
    - http://www.affymetrix.com/
  - Illumina
    - http://www.illumina.com/
- Differences in design and coverage
Advantages of GWAS

- Compared to candidate gene studies
  - unbiased scan of the genome
  - potential to identify totally novel susceptibility factors
- Compared to linkage-based approaches
  - capitalize on all meiotic recombination events in a population
  - Localize small regions of the chromosome
  - enables rapid detection causal gene
  - Identifies genes with smaller relative risks

Concerns with GWAS

- Expense
- Power dependent on:
  - Allele frequency
  - Relative risk
  - Sample size
  - LD between genotyped marker and the risk allele
  - disease prevalence
  - .......
- Study Design
  - Replication
  - Choice of SNPs
- Analysis methods
  - IT support, data management
  - Variable selection
  - Multiple testing

Major Assumptions....

- Common-Disease Common Variant (CDCV) Hypothesis
  - predicts that common disease-causing alleles, or variants, will be found in all human populations which manifest a given disease
  - each variant at each gene influencing a complex disease will have a small additive or multiplicative effect on the disease phenotype
  - Assumes traits are evolutionary neutral in part because so many genes influence the traits
  - Has held true for many diseases
    - APOE ε4 and Alzheimer’s disease
  - Likely not true for many diseases
    - Schizophrenia

Major Assumptions....

- Alternative...
- Common-Disease Rare Variant (CDRV) Hypothesis
  - proposes that a significant proportion of the inherited susceptibility to relatively common human chronic diseases may be due to the summation of the effects of a series of low frequency dominantly and independently acting variants of a variety of different genes
  - each conferring a moderate but readily detectable increase in relative risk
  - will mostly be population specific because of founder effects resulting from genetic drift
- GWAS chips will not detect rare variants

Sequencing

- New technologies provide the ability to process millions of sequence reads in parallel
- Major platforms:
  - Roche(454) www.454.com
  - Illumina(Solexa) www.illumina.com
  - SOLID www.solid.appliedbiosystems.com
- Will be able to detect ALL variation in the genome
- Analysis to perform mutation detection, then association....

Sequencing

- Annotation of sequencing data is an important bioinformatics challenge
  - Statistics need to address the error in the data
- Rare variants present an important statistical challenge
  - How do you do association?
  - First approaches use collapsing approaches
  - Collapse by gene, by function, etc.
Replication and Validation

• After an association is identified.... Now what? How do you follow-up?

• Replication is gold standard
  — Detection of the same association in an independent sample

• Challenges
  — There are many negative replication studies even in the most replicated genetic associations
  — Additionally, when associations are replicated, it is often in different phenotypes, with different polymorphisms, or with different alleles

Initial Study Results

Replication Strategy

Replication Study Outcomes

<table>
<thead>
<tr>
<th>Same Trait</th>
<th>Same Gene</th>
<th>Same Variant</th>
<th>Same Risk Model</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Exact Replication</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Genetic Heterogeneity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Allelic heterogeneity</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Population Differences</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Phenotypic heterogeneity</td>
</tr>
</tbody>
</table>

Additional Challenges

• How do you collect well-powered independent samples?
  — Expanding phenotype introduces heterogeneity

• Potential resources
  — Other data types?
  — Similar phenotypes?
  — Meta-analysis?
  — Functional studies?

New Approaches for Analysis

• Traditional statistical approaches typically cannot model the complexity of PGX traits
  — Heterogeneity
  — Interactions
  — Pathways/networks
  — New technologies create an immense variable selection problem

• Many new data-mining approaches are now being developed and applied

Conclusions

• Statistics plays a crucial role in EVERY stage of gene mapping

• Careful consideration should be given to the appropriateness of the statistics used
Questions?

motsinger@stat.ncsu.edu
Pro-Con Debate on Prospective Study Designs: *Convenience Cohorts Versus Randomized Samples*

Craig R. Lee, Pharm.D., Ph.D.
Division of Pharmacotherapy & Experimental Therapeutics
UNC Eshelman School of Pharmacy
and
Michael Pacanowski, Pharm.D., M.P.H.
Office of Clinical Pharmacology
FDA
ACCP Spring Practice and Research Forum
Charlotte, NC
April 24, 2010

Limited Incorporation of PGx into Clinical Guidelines

- Pre-requisite for pushing PGx into practice
- The epitome of evidence-based medicine
- In cardiology:
  - 11% level A (multiple RCTs or meta-analyses)
  - 41% level B (a single RCT or nonrandomized studies)
  - 48% level C (expert opinion, case studies, or standards of care)

19% of class I recommendations, i.e., procedure/treatment is useful or effective, are based on level A evidence

Califf, et al. 2009 [PMID 19244190]

Personalized Medicine
*Is this really a new concept?*

**Definition**
- Integrating evidence generated at the population level (e.g., registries, RCT’s) into clinical decisions for individual patients.
- Diagnostics, pharmacotherapy

**The “Omics” Definition**
- Integrating “omics” technology into clinical decisions
  - Genomics, biomarkers (e.g., transcriptomics, proteomics, metabolomics)

Questions to Consider

- What evidence do you need to make effective clinical decisions?
- When is it necessary to prospectively evaluate pharmacogenetic markers?

Primary Objective

In order to validate the utility of pharmacogenetic markers in a manner which will facilitate implementation into clinical practice, we will debate the pros/cons of using:

- “convenience sample” studies versus prospective, randomized, controlled clinical trials

“Convenience Sample” Studies Definition

Obtaining DNA (or other biological specimens) from subsets of participants enrolled in:
- an observational/registry study
  OR
- a prospective, controlled trial
  without regard to time of enrollment and without any specific hypotheses
Key Study Design Issues to Discuss

- Scientific Rigor
- Practicality
- Potential for translation

Scientific Rigor

“Convenience Sample” Studies
Pros:
- Large numbers of samples / events can be obtained
  - Conducive to evaluation of rare events
  - Ability to assess clinical outcomes
- Opportunities for unbiased identification of the “best” pharmacogenetic marker (e.g., GWAS)

Prospective RCT’s
Cons:
- Must identify the “best” pharmacogenetic marker (and therapeutic strategy) \textit{a priori}
- Difficult to utilize clinical outcomes as endpoint
  - Typically surrogate measures

Practicality

“Convenience Sample” Studies
Pros:
- Opportunity for rapid validation of pharmacogenetic associations across multiple, independent studies
- Opportunity to evaluate the impact of new markers as they are discovered

Prospective RCT’s
Cons:
- Substantial cost and length
  - Fewer opportunities for replication

Case Study

KRAS mutations and efficacy of anti-EGFR therapy

Practicality

Prospective RCT’s
Pros:
- An adequate, well-controlled trial provides eliminates need for numerous, less-than-adequate studies
- Yields actionable information

“Convenience Sample” Studies
Cons:
- Limited to available data
- “Fishing expeditions”
Case Study

CYP2C19 genotype and clopidogrel hyporesponsiveness

Potential for Translation

“Convenience Sample” Studies

Pros:
- Generalizable findings
  - Broad sampling of patient populations
  - Data collection in a real-world clinical environment (registries)

Cons:
- Comparative effectiveness studies difficult

Prospective RCT’s

Pros:
- Opportunity to build in assessment of comparative treatment effects and cost-effectiveness
- Opportunity to evaluate a clinical strategy
  - Value to payers and clinicians

Cons:
- Less generalizable findings
  - Narrowly defined populations
  - “Control” conditions may not be representative of standard of care in the real-world clinical environment
  - Trial procedures (e.g., genetic testing, monitoring) may not be feasible in clinical environment

Questions Revisited

- What evidence do you need to make effective clinical decisions?
- When is it necessary to prospectively evaluate pharmacogenetic markers?

Case Study

VKORC1/CYP2C9 genotyping to guide warfarin dose selection