**The Pharmacogenomics Journal (2001) 1**, 157–158
© 2001 Nature Publishing Group All rights reserved 1470-269X/01 $15.00
www.nature.com/tpj



The Pharmacogenomics Journal

EDITOR
Julio Licinio, University of California, Los Angeles, USA

EDITORIAL BOARD
J Azuma, Osaka University, Japan
E Boerwinkle, University of Texas, Houston, USA
U Brinkmann, Epidauros, Germany
RGH Cotton, University of Melbourne, Australia
D Flockhart, Indiana University School of Medicine, USA
P Froguel, Institut Pasteur de Lille, France
LM Furness, Incyte Genomics, UK
T Gingeras, Affymetrix, USA
M Hashida, Kyoto University, Japan
W Kalow, University of Toronto, Canada
T Klein, Stanford University, USA
I Kola, Pharmacia, USA
K Lindpaintner, Roche Genetics, Switzerland
R Long, National Institute of General Medical Sciences, USA
P Nadkarni, Yale University, USA
D Nickerson, University of Washington, USA
L Peltonen, University of California, Los Angeles, USA
M Polymeropoulos, Novartis, USA
G Poste, Health Technology Networks, USA
A Rane, Karolinska Institutet, Sweden
M Ratain, University of Chicago, USA
AD Roses, GlaxoSmithKline, USA
MA Rothstein, University of Louisville School of Medicine, USA
JC Venter, Celera Genomics, USA
R Weinshilboum, Mayo Clinic, USA
S Weiss, Harvard Medical School, USA
J Woodcock, Food and Drug Administration, USA
R Wolf, University of Dundee, UK
M-L Wong, University of California, Los Angeles, USA
A Wright, Western General Hospital, UK

EDITORIAL OFFICES

The Pharmacogenomics Journal
Nature Publishing Group
Specialist Journals Division,
Houndmills, Basingstoke,
Hampshire RG21 6XS, UK

and

The Pharmacogenomics Journal
Nature Publishing Group
Specialist Journals Division,
345 Park Avenue South
New York 10010-1707, USA

ThePharmacogenomicsJournal@nature.com
http://www.nature.com/tpj

## Case-Control Association Studies in Pharmacogenetics

One of the most common study designs used to assess pharmacogenetic effects is that of the case-control association study. In the context of pharmacogenetics, the usual approach is to examine the active treatment arm of a clinical trial and divide subjects in the treatment arm into those with a positive response to the drug and those with a negative or no response. These two groups then constitute cases and controls who are genotyped for a particular candidate gene thought to be related to the treatment phenotype. Risch and Merikangas have shown that association studies are more powerful than linkage studies for finding genetic determinants of a complex phenotype such as treatment response.[1]

The problem with such studies is that, though they are easy to perform, they are but fraught with a host of potential biases or difficulties in interpretation. Investigators in my laboratory have previously identifed some of the methodologic concerns.[2] Table 1 expands on the concerns identified by Silverman and Palmer, listing a total of ten important design issues that must be considered in any genetic case-control association study. While all ten of these issues may be of great importance in terms of study design, four merit extra attention.

In evaluating any case-control association study, careful attention to how the case subjects were defined and whether they met appropriate criteria for the affectation phenotype is critical. Just as important is that the control subjects clearly represent the opposite end of the phenotypic expression of disease and that they are free of potential confounding variables that could influence or create a spurious association. Subjects drawn from ongoing cohort studies or clinical trials have the advantage of being selected from a defined population base that is well characterized.

An additional common concern for pharmacogenetics studies is sample size. The fundamental question of interest is whether there is a gene-by-drug interaction. There may be an inadequate number of subjects in the case and control groups to define treatment response and lack of response in a single arm of a clinical trial. The usual clinical trial randomizes fewer than 200 subjects to treatment and placebo groups. Hence, there may be only 100 or fewer subjects in the treatment arm from which to define treatment responders and nonresponders. This is perhaps the biggest single failing of most pharmacogenetic association studies. At a minimum, case and control groups of 200 are necessary to detect odds ratios for a main effect of a gene of around 4. To determine odds ratios in the 2–4 range for the interaction of a drug and a gene, 400 cases and 800 controls are necessary. Given the need for sample sizes in this range, almost all of the existing case-control association studies examining pharmacogenetic effects are underpowered, an even more apparent problem if the polymorphisms of interest have allele frequencies below 0.2.

Positive results in a case-control association study may be due to a direct effect of the SNP in question, linkage disequilibrium, or population stratification, a spurious association due to differences in allele frequency between poorly matched cases and controls resulting from differences in ethnic origins. Matching of cases and controls for ethnicity or use of multiple unlinked markers to assess for the presence of this confounding variable are useful techniques to detect or eliminate the potential bias of population stratification.[3] An additional criterion for evaluation of the quality of the case-control study is assessment of Hardy–Weinberg equilibrium in the markers studied within the control group. Hardy–Weinberg equilibrium implies that the genotype frequencies can be determined directly from the allele frequencies and provides a check to ensure that genotyping errors, mutation, or population stratification do not explain observed results.

**Table 1 Evaluation of candidate gene case control association studies**

| *Issue* | *Key questions* | *Possible solutions* |
|---|---|---|
| (1) Selection of cases and controls | (a) How were cases and controls selected?<br>(b) Do cases meet appropriate criteria for disease status?<br>(c) Are controls free of disease and potential confounders?<br>(d) Do cases and controls have similar demographic and environmental factors? | (a) Population-based selection<br>(b) Family-based association design |
| (2) Sample size | (a) Is there an adequate number of cases and controls? | (a) Minimum number of cases and controls included |
| (3) Population stratification | (a) Are cases and controls matched? | (a) Matching on ethnicity<br>(b) Family-based association designs<br>(c) Negative results with multiple unlinked markers |
| (4) Selection of candidate gene polymorphism | (a) Biologically reasonable?<br>(b) Positional candidate?<br>(c) All variants identified? | (a) Demonstration of biologically functional effect<br>(b) Within linked region in human or syntenic from animal model<br>(c) Complete sequencing of the gene |
| (5) Observation bias | (a) How was the phenotyping and genotyping done? | (a) Blind assessment of genotype and phenotype |
| (6) Linkage disequilibrium | (a) Are there other genes?<br>(b) Are there other polymorphisms in this gene? | (a) Haplotypes<br>(b) Family-based association |
| (7) Allele or genotype-based analysis | (a) How are the heterozygotes treated in analysis? | (a) Use appropriate genetic model |
| (8) Mutlivariate analysis | (a) Are relevant covariates identified? | (a) Use appropriate genetic model |
| (9) Gene by environment interaction | (a) Is sample large enough to detect a gene environment? | (a) Stratification by environmental exposure<br>(b) Multivariate analysis interaction |
| (10) Multiple comparisons | (a) How many alleles were tested?<br>(b) How many phenotypes were tested?<br>(c) How many genetic loci were tested? | (a) Bonferroni correction<br>(b) Estimation of empirical *P* values |

The final key to these studies is replication, which can be performed with a second case-control association study or a family-based study. Replication performed with family-based designs can be used in conjunction with a case-control association study. Replication can be performed at the same time that other methodologic concerns are being addressed. In fact, replication without addressing the methodologic concerns in Table 1 will not be useful. While extremely difficult in the context of pharmacogenetics, replication is clearly the mode of choice for traditional gene-finding studies.

While all ten of the issues presented in Table 1 should be addressed in any case-control association study, careful attention to the four that we have highlighted would go a long way to improving the quality of existing studies and preventing false positive associations that will need to be refuted in future work.

## DUALITY OF INTEREST

None declared.

## REFERENCES

1 Risch N, Merikangas K. *Science* 1996; **273**: 1516–1517.
2 Silverman EK, Palmer LJ. *Am J Respir Cell Mol Biol* 2000; **22**: 645–648.
3 Pritchard JK, Rosenberg NA. *Am J Hum Genet* 1999; **65**: 220–228.

ST Weiss, EK Silverman, LJ Palmer
Harvard Medical School, Boston, MA, USA