Published in final edited form as:

J Biopharm Stat. 2010 March ; 20(2): 301–314. doi:10.1080/10543400903572761.

Testing Gene-Treatment Interactions in Pharmacogenetic Studies

Min He¹ and Andrew Allen^{1,2}

¹Department of Biostatistics and Bioinformatics, Duke University, Durham NC

²Duke Clinical Research Institute, Duke University Durham NC

Summary

Drug related side effects are one of the leading causes of death and illness in the developed world. Finding genes that modify drug response has the potential to significantly improve drug delivery, by identifying both individuals that can benefit from therapy as well as those at increased risk of harm. We present a simple approach to testing gene-by-treatment interactions in case-control pharmacogenetic studies. The approach utilizes a retrospective model that seeks to increase power through a Hardy-Weinberg equilibrium assumption among the controls, but does not assume that the event of interest is rare in the target population. We conduct extensive simulations and find that the approach shows similar or improved power, compared to standard methods, in all cases considered. We present methods for both autosomal and X-linked markers and show how the methods can be easily implemented using standard statistical software.

Introduction

It is well known that the same medication can elicit radically different responses in different patients. Even those drugs that, on average, provide significant benefit can prove deadly some of the time. In fact, adverse reactions to drugs rank as one of the leading causes of death and illness in the developed world (Lazarou 1998). Differences in drug effect can be due to many factors including age, drug interactions, and the nature of the disease (Evans and McLeod 2003). However, even though these non-genetic factors are important in influencing medication response, there is growing evidence that genetic factors are among the most important (Evans and Relling 1999, Evans and Johnson 2001, McLeod and Evans 2001). In fact, it is estimated that genetics can account for 20 to 95 percent of variability in drug disposition and effects (Kalow 1998). Identifying variants associated with drug effects could lead to better guidelines for their use as well as valuable information about the pathways involved in their action. Thus there is great interest in identifying genetic variants associated with differential response to medication.

Much work has been done developing statistical methods for pharmacogenetics studies, especially methods that characterize genetic markers modifying pharmacokinetic and pharmacodynamics processes (see for example, Wu and Lin, 2009). Less attention, however, has focused on developing methods for genome-wide association mapping of genetic loci that influence medication response or the occurrence of drug induced side effects. Genome-wide association studies (GWAS) are the current “state-of-the-art” in genetic association studies and have been successful in identifying single nucleotide polymorphisms (SNPs) associated with a number of diseases (Wellcome Trust Case Control Consortium 2007). Though the application of GWAS in pharmacogenetic studies have, to date, been less

numerous, they have seen some successes in identifying genetic variants associated with differential response to drug therapy (Kindmark 2008). There are reasons for thinking that pharmacogenetic GWAS could be even more successful than their population based cousins. For many complex diseases, purifying selection makes it unlikely that genetic variants that strongly predispose an individual to disease will reach common frequency in a population. This explains why most GWAS of complex traits have only found common variants of modest effect (odds ratios in the 1.2–1.6 range; Hindorff et al. 2009), limiting their utility in disease prediction. However, pharmacogenetic GWAS represent a setting where purifying selection's role will be quite limited (or non-existent). In an evolutionary sense, humans will be naïve to most drug therapies, so that selection will not have had a chance to moderate the effects of even common variants on response or side effects. Therefore, large genetic effects on treatment response or side effects are quite plausible and individual SNPs may be good predictors of these outcomes.

Most primary analyses of GWAS conduct a simple test for disease/SNP association one SNP at a time. The complexity of these analyses comes in their totality: even simple tests can demand significant computational resources when scaled up to investigating hundreds-of-thousands or millions of SNPs. Further, the fact that a large number of tests are being conducted needs to be accounted for, and multiplicity adjustments often put additional computational demands on the analysis. Multiplicity adjustments can also lead to such stringent marginal significance levels that the power of each test becomes a key issue. Many of these concerns are amplified in whole genome pharmacogenetic studies, as the focus is not on the main effect of genotype, but on the *interaction* between genotype and treatment on response to medication.

Case-control studies are ubiquitous throughout medical research and, given the costs of genotyping, tend to be the dominant design for GWAS. Even pharmacogenetic GWAS, which are most often imbedded within a prospective clinical trial, will often employ a nested case-control study design in an attempt to maximize power while minimizing genotyping costs. Case-control studies are retrospective designs, as individuals are sampled based on their outcome and then their exposures are assessed retrospectively. However, the predominant analytic approach for case-control studies, logistic regression, is prospective as it conditions on exposure and prospectively models disease risk. This is due to the famous result due to Prentice and Pyke (1979) that showed that prospective logistic models yield valid inferences for odds ratio parameters under retrospective sampling. A less well known result from this paper is the fact that these inferences will tend to be less efficient than those based on a retrospective model whenever the retrospective model constrains the exposure distribution. This is usually not an issue as there is often very little information concerning the distribution of exposures and the fact that the prospective model ignores this information is seen as a pro and not a con. However, the exposures in a case-control genetic association study are genotypes and there are very good biologically motivated models for how genotypes are likely to be distributed. Recently, Chen and Chatterjee (2007) showed that power could be increased over the standard prospective analyses by incorporating a Hardy-Weinberg Equilibrium (HWE) assumption into a retrospective analysis framework. These results were limited to tests of the main effects of genotypes on disease risk. Kwee et al. (2007) presented a similar retrospective model for haplotype by environment interactions and showed that this model was valid even with finely stratified data. The Kwee model is made tractable by making two assumptions: (1) that the disease is rare and (2) that haplotypes and environmental covariates are independent in the target population. Since the events in a pharmacogenetic study (response, side effects etc.) are not necessarily rare, the Kwee model has limited utility in this case. Our focus here is on developing a retrospective model for testing the interaction between genotype and treatment in a case-control pharmacogenetic study that does not make a rare disease assumption. Further, as methods

for genome-wide data should be able to analyze the whole genome, we extend the retrospective model to handle X-linked markers. This is important as there are often large differences between males and females in drug metabolism and response, suggesting that variants on the X chromosome may play a role.

The structure of the remainder of the paper is as follows: in the next section (§2), we begin by setting notation and present a retrospective model that can be used for testing treatment-by-genotype interactions for autosomal and X-linked markers; in section 3 we investigate the performance of this approach and compare it to a standard prospective logistic regression model via simulation; in the final section we discuss possible extensions of our approach and make recommendations for dealing with large scale pharmacogenetic GWAS data sets.

Models for analyzing case-control pharmacogenetic studies

We begin by setting up notation. We assume individuals are sampled from a prospective clinical trial on the basis of a post-baseline event of interest (e.g., treatment response or side effect). Individuals with the event are indicated by $E=1$, while those without are indicated by $E=0$. For simplicity, we assume a two arm trial in which individuals are randomized at baseline to one of two treatment groups indicated by $T=0$ or $T=1$. At a given locus, denote genotype by G and assume that it is coded as the number of minor alleles present, so that $G=1$ indicates a heterozygote while $G=2$ indicates a homozygote for the minor allele. Lower case letters will denote realizations of random variables. Since the analysis of X-linked markers requires the sex of the individual, we indicated it by S , with $S=f$ for females and $S=m$ for males. We begin, however, by considering the autosomal case.

Autosomal markers

The natural likelihood for a case-control pharmacogenetic study is one that models the distribution of G and T conditional on E . This likelihood can be factored as

$$\prod_{i=1}^n \Pr(G=g_i|T=t_i, E=e_i) \Pr(T=t_i|E=e_i) \quad (1)$$

where i indexes individuals and n is the total sample size. We will discuss each factor of likelihood (1) in turn.

We begin by specifying a model for $\Pr(G|T, E)$. Define π_t the frequency of the minor allele among individuals in treatment group t , without the event, by $\pi_t = \exp(\gamma_t) / \{1 + \exp(\gamma_t)\}$. Assume that the odds of having an event are given by

$$\frac{\Pr(E=1|G=j, T=t)}{\Pr(E=0|G=j, T=t)} = \exp(\alpha + X_t^T \xi + X_j^T \beta + X_{tj}^T \eta), \quad (2)$$

where X_j , X_t , and X_{tj} are design vectors that code the genetic, treatment, and gene-by-treatment interaction models, respectively. Then, assuming HWE among those without the event, the distribution of genotype at an autosomal locus, conditional on event occurrence, and treatment group, can be written as

$$\Pr(G=j|E=e, T=t) = \left(\frac{\exp\{j\gamma_t + I(j=1)\ln(2)\}}{\sum_{j^*=0}^2 \exp\{j^*\gamma_t + I(j^*=1)\ln(2)\}} \right)^{I(e=0)} \times \left(\frac{\exp\{X_j^T\beta + X_{jt}^T\eta + j\gamma_t + I(j=1)\ln(2)\}}{\sum_{j^*=0}^2 \exp\{X_{j^*}^T\beta + X_{j^*t}^T\eta + j^*\gamma_t + I(j^*=1)\ln(2)\}} \right)^{I(e=1)} \quad (3)$$

where I is an indicator function. Note that both the intercept parameter α and the treatment main effects $X_t^T\xi$ cancel, implying that $\Pr(G|T, E)$ contains no information on the intercept α or treatment main effects ξ .

We now consider the second factor within (1), $\Pr(T|E)$, and show that when X_t is saturated (which is always the case when there are only two treatment arms; L arms would require $L-1$ free parameters), then $\Pr(T|E)$ contains no information concerning the genetic main effects β and gene-by-treatment interaction effects η . To see this, begin by noting that (2) implies

$$\theta_t \equiv \frac{\Pr(E=1|T=t)}{\Pr(E=0|T=t)} = \exp(\alpha + \xi_t + \varphi_t)$$

where

$$\varphi_t = \ln \left[\sum_j \exp(X_j^T\beta + X_{jt}^T\eta) \Pr(G=j|E=0, T=t) \right],$$

and ξ_t is the parameter corresponding to treatment t with the convention that $\xi_0=0$ (otherwise the model would be over parameterized). The distribution of $\Pr(T|E)$ can be written as a function of θ_t i.e.,

$$\Pr(T=t|E=e) = \frac{[\theta_t/(1+\theta_t)]^e p_t}{\sum_t [\theta_t/(1+\theta_t)] p_t}$$

where p_t are known, fixed, treatment randomization probabilities. Thus $\Pr(T|E)$ depends only on $\tilde{\xi}_t \equiv \xi_t + \varphi_t - \varphi_0$. Since the model for ξ is already saturated, we conclude that $\Pr(T|E)$ contains no information concerning the genetic main effects β and gene-by-treatment interaction effects η . Thus, when testing for gene-by-treatment interaction effects, we can ignore the second factor in (1) and base inference solely on (3). This model can be fit using software that fits a polytomous logistic regression and allows an offset. We give example code using SAS proc CATMOD in the appendix.

X-linked markers

When developing a whole genome testing procedure, one needs to consider the entire genome, including the X chromosome. Since males only have one copy of the X chromosome, while females have two, analyses of X-linked markers need to further consider the sex of the individual. Here we extend the model presented for autosomal loci by conditioning on the sex of the individual and modifying our notation accordingly. Assume that the odds of having the event are given by

$$\frac{\Pr(E=1|G=j, T=t, S=s)}{\Pr(E=0|G=j, T=t, S=s)} = \exp(\alpha_s + X_{ts}^T \xi + X_{js}^T \beta + X_{tjs}^T \eta), \quad (4)$$

where X_{js} , X_{ts} , and X_{tjs} are design vectors that code the genetic, treatment (allowing for both treatment main effects and interactions between treatment and sex), and gene-by-treatment interaction models, respectively. Note that we are not assuming that X_{tjs} codes a three way interaction involving treatment, genotype, and sex, only that sex may change the coding of the treatment by genotype interaction. Assuming this model and HWE (HWE at an X-linked marker assumes HWE among the females and similar allele frequencies between males and females) among those individuals that remain event free, the distribution of genotype at an X-linked locus, conditional on event occurrence, treatment group, and sex can be written as

$$\Pr(G=j|E=e, T=t, S=s) = \left(\frac{\exp\{j\gamma_t + I(j=1, s=f)\ln(2)\}}{\sum_{j^*=0}^2 \exp\{j^*\gamma_t + I(j^*=1, s=f)\ln(2)\}} \right)^{I(e=0)} \\ \times \left(\frac{\exp\{X_{js}^T \beta + X_{tjs}^T \eta + j\gamma_t + I(j=1, s=f)\ln(2)\}}{\sum_{j^*=0}^2 \exp\{X_{j^*s}^T \beta + X_{tj^*s}^T \eta + j^*\gamma_t + I(j^*=1, s=f)\ln(2)\}} \right)^{I(e=1)}$$

This model can also be fit using SAS proc CATMOD and we give example code implementing an X-linked marker analysis in the appendix.

Design matrices for autosomal and X-linked models

In order to fully parameterize the above models we need to specify models for the genetic main effects and gene-by-treatment interaction effects. We begin by considering an autosomal locus and partitioning the genetic effect into dominant and recessive components. Define the recessive effect by $X_r(G)=I(G=2)$ and the dominant effect by $X_d(G)=I(G>0)$. For both the genetic main effect X_j and the gene-by-treatment interaction effects X_{jt} we define 4 basic models: recessive, dominant, additive, and general. Design elements for these models can be found in table 1. When testing for gene-treatment interaction we recommend saturating the model for the genetic main effects (i.e., use the general model) to avoid having misspecification of this model confound the interaction test.

For an X-linked locus the situation is more complex. In females, the above models translate without modification, as females continue to have two alleles. The difficulty is with males, as they only have one copy of the X chromosome. One option is to allow for separate effects in males and females. Note that among males, G is either 0 or 1 and X becomes an indicator

of whether the minor allele is present. Thus X_d completely captures the genetic effect in males and can be used to specify a separate male effect. To condense notation, let $\delta_f = I(S = f)$ and $\delta_m = I(S = m)$. Table 2 presents design elements for models that allow for a separate effect among males, but specify a particular genetic model among females.

Having a separate parameter for males may be reasonable, but it uses a degree of freedom and there may be more restrictive models that, in some situations, better reflect the expected biology. For example, if a mutation leads to a loss of gene function (which subsequently leads to an event) then the risk of an event among males with one copy of the mutation should be similar to the risk of an event among females with two copies of the mutation (all other things being equal). This scenario suggests that the genetic effect among males share a parameter with the recessive genetic effect among females. For example, a two degree of freedom genetic model that allows for both dominant and recessive effects but which maps the male's effect onto the female's recessive effect parameter is given by $X_{js} = (\delta_f X_r(j) + \delta_m X_d(j), \delta_f X_d(j))^T$. A similar model for the gene-by-treatment interaction parameter is given by $X_{jst} = tX_{js}$. (We illustrate how this model can be fit using PROC CATMOD in the appendix). Many other parameterizations are possible (a comprehensive discussion is beyond the scope of this work) and the methods presented here are able to capture this richness.

Simulation Experiment

Simulation of autosomal locus

Here we study the performance of the proposed tests via simulation. We simulate the data prospectively, i.e., first genotype and treatment, then event status (given genotype and treatment). We focus on two autosomal scenarios: one that assumes HWE in the population and the other that does not. Both of these scenarios are encompassed within the following model, which is used to simulate genotypes,

$$\Pr(G=g) = \begin{cases} (1-\pi)F_{st} + (1-\pi)^2(1-F_{st}), & \text{if } g=0 \\ \pi(1-\pi)(1-F_{st}), & \text{if } g=1, \\ \pi F_{st} + \pi^2(1-F_{st}), & \text{if } g=2 \end{cases}$$

where π is the frequency of the 1 (risk) allele and F_{st} is the inbreeding coefficient (for a complete discussion of this model see Cavalli-Sforza and Bodmer, 1971). When $F_{st}=0$, this model corresponds to assuming HWE. Once an individual's genotype is generated, they are randomly assigned one of two treatment arms ($t=0,1$). Given their genotype and treatment assignment, events ($e=0,1$) are simulated according to equation (2) where $\alpha = -3$ (corresponds to ~5% event rate among individuals with $t=0$), $\xi = \ln(2)$, and $\beta = 0$ (we also investigated scenarios with nonzero genetic main effects but found little effect on interaction tests; we report only the no genetic main effect case here). This process is repeated until the dataset has 500 individuals with and 500 individuals without an event. We simulated a great number of scenarios, varying the interaction effect size, risk allele frequency, and F_{st} value.

We generated the data using three different models for gene-by-treatment interaction corresponding to the recessive, dominant, and additive models presented in table 1. These models are specified by a scalar parameter which we varied to represent a spectrum of effects. For each simulated dataset we applied 2 tests of gene-by-treatment interaction corresponding to the additive and general models detailed in table 1. For each of these tests we saturated the model for the genetic main effects. We chose to test for interactions under the additive and general models as they are, by far, the most commonly tested in practice

and have been shown to have good power across a range of alternatives. We compared these tests with 2 similar tests (additive and general models) using a prospective logistic regression model corresponding to equation 2. Note that this approach requires the estimation of an intercept and treatment main effect parameters. Again, the genetic main effects were fit using a general (saturated) model. Type I error rates were estimated using 10,000 simulated datasets for each scenario investigated. Alternative simulations used 2,000 simulated datasets for each scenario to estimate the power of the various approaches.

Simulation of X-linked locus

As mentioned above, the possibility of separate effects (both main effects and interaction effects) leads to a large number of potential models that can be considered. A comprehensive examination of this model space is beyond the scope of this manuscript. Here we illustrate the proposed methodology via a limited simulation study focusing on a biologically plausible recessive model in which the effect of a male's having one copy of the risk allele is the same as recessive effect in a female. As above, we generate the data prospectively, i.e., first genotype, sex and treatment, then event status (given genotype, sex and treatment). The genotypes of females were generated under HWE while that of males was obtained as a Bernoulli random variable with the same allele frequency used in females. Once an individual's genotype and sex is generated, they are randomly assigned one of two treatment arms ($t=0,1$). Given their genotype and treatment assignment, events ($e=0,1$) are simulated according to equation (4) where $\alpha = -3$, $\xi = \ln(2)$, and $\beta = 0$. This process is repeated until the dataset has 500 individuals with and 500 individuals without an event. To each of dataset we applied the 2 degree-of-freedom test (2 separate genotype effects in females, male effect assumed the same as female recessive effect) given by the retrospective model detailed in the appendix and an analogous 2 degree-of-freedom prospective test given by logistic regression. Type I error rates were estimated using 10,000 simulated datasets for each scenario investigated. Alternative simulations used 2,000 simulated datasets for each scenario to estimate the power of the various approaches. Tables 3 and 4 presents the results of autosomal null simulations ($\alpha = 0.05$) conducted for a number of risk allele frequencies.

Table 3 summarizes results for simulations where the data generating model assumed HWE. As can be seen by this table all methods maintain the nominal level. Table 4 presents similar results for the situation when the true model exhibits a substantial departure for HWE. Nevertheless, all methods maintain the correct type I error rate. We note that this is a rather large inbreeding coefficient ($F_{st}=0.15$) which is, for example, consistent with the mixture of two distinct subpopulations whose minor allele frequencies differ by an order of magnitude. The fact that the retrospective model, which assumes HWE, maintains the correct size under this extreme of a departure from HWE is significant, and suggests that tests of gene-by-treatment interactions exhibit some robustness to misspecification of the genotype model.

Tables 5 and 6 present power results across a variety of true genetic models and allele frequencies. For the additive and dominant alternatives, the power of the retrospective and prospective approaches quite similar. However, under recessive alternatives the retrospective approach shows a clear advantage, both for when the simulated data are in HWE (table 5), and when the simulated data exhibit a significant departure from HWE (table 6). Figures 1, 2 and 3 (for the X-linked case) illustrate this further by looking at power across a spectrum of true effect sizes for three risk allele frequencies. The power of the retrospective approach is clearly higher over most of the range of effect sizes. We note that improving power under retrospective alternatives is quite desirable, as this type of effect is both biologically plausible and notoriously difficult to detect.

Discussion

In this work we presented methods for testing gene-treatment interactions based on a retrospective likelihood. We have shown that, for the autosomal case, all the information concerning the interaction parameters are contained in $\Pr(G|T, E)$ and that $\Pr(T|E)$ can be ignored without a loss of efficiency.

For X-linked markers we extended the autosomal approach by further conditioning on sex, i.e., we base inference on $\Pr(G|T, E, S)$. While this approach leads to valid tests of interaction parameters, it is not clear that it is fully efficient. There may, in fact, be information concerning interaction parameters contained in $\Pr(S|T, E)$. We plan to more fully explore the X-linked case in future work.

Our approach assumes that cases (those with an event) and controls (those without the event) are sampled from a prospective clinical trial but that cases are unmatched to controls. However, some studies may wish to match cases and controls by sampling a control for each case from the case's risk set (individuals at risk of having an event at the time the case did). This design requires methods that can handle finely stratified data. We note that if one assumes common genetic and interaction effects in each stratum, then the methods presented here can be extended to handle finely stratified data, as stratum specific intercept and treatment effect parameters will be naturally conditioned out of the retrospective model.

We have shown how standard statistical packages can be used to implement the proposed methodology. These packages, however, proceed by iteratively estimating model parameters. Though quite fast in the examples and simulations we considered, when applied to GWAS data, iterative estimation may become prohibitive, require substantial computing resources especially for newer genotyping platforms (where the number of marker can exceed a million). As an alternative, we note that score statistics will not require the estimation of interaction parameters and that there is reason to believe that the genetic main effect and genotype parameters may be available in closed form (Chen and Chatterjee, 2007). This suggests that closed-form score statistics for gene-by-treatment interaction may exist. A closed-form score statistic approach would be extremely fast, not only making for quick computations across the genome, but allowing the application of permutation-based multiple testing procedures. We plan to investigate this further in a future manuscript.

Acknowledgments

ASA acknowledges funding support through CTSA grant UL1-RR-024128

References

- Chen J, Chatterjee N. Exploiting Hardy-Weinberg equilibrium for efficient screening of single SNP associations from case-control studies. *Human Heredity*. 2007; 63:196–204. [PubMed: 17317968]
- Cavalli-Sforza, LL.; Bodmer, WF. *The Genetics of Human Populations*. W. H. Freeman; San Francisco: 1971.
- Evans WE, McLeod HL. Pharmacogenomics -- drug disposition, drug targets, and adverse effects. *N Engl J Med*. 2003; 348:538–549. [PubMed: 12571262]
- Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. *Science*. 1999; 286:487–491. [PubMed: 10521338]
- Evans WE, Johnson JA. Pharmacogenomics: the inherited basis for interindividual differences in drug response. *Annu Rev Genomics Hum Genet*. 2001; 2:9–39. [PubMed: 11701642]
- Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *PNAS*. 2009; 106(23):9362–9367. [PubMed: 19474294]

- Kalow W, Tang BK, Endrenyi I. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics*. 1998; 8:283–289. [PubMed: 9731714]
- Kindmark A, Jawaid A, Harbron CG, et al. Genome-wide pharmacogenetic investigation of a hepatic adverse event without clinical signs of immunopathology suggests an underlying immune pathogenesis. *The Pharmacogenomics Journal*. 2008; 8:186–195. [PubMed: 17505501]
- Kwee LC, Epstein MP, Manatunga AK, Duncan R, Allen AS, Satten GA. Simple methods for assessing haplotype-environment interactions in case-only and case-control studies. *Genetic Epidemiology*. 2007; 31:75–90. [PubMed: 17123302]
- Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. *JAMA*. 1998; 279:1200–1205. [PubMed: 9555760]
- McLeod HL, Evans WE. Pharmacogenomics: unlocking the human genome for better drug therapy. *Annu Rev Pharmacol Toxicol*. 2001; 41:101–121. [PubMed: 11264452]
- Prentice RL, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika*. 1979; 66:403–412.
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
- Wu, RL.; Lin, M. Statistical and computational pharmacogenomics. Chapman & Hall/CRC; London: 2008.

Appendix

SAS code for autosomal marker

We give two examples of analyses involving autosomal markers. This example assumes the data is in a SAS dataset called “pharma” which contains three variables: g, e, and t. The variable g contains the marker’s genotype and is assumed to be coded as 0 (homozygous for major allele), 1 (heterozygous), or 2 (homozygous for the minor allele). The variable t contains treatment assignment and is assumed to be either 0 or 1. Similarly, the variable e indicates whether an individual has an event or not and is assumed to be either 0 (no event) or 1 (event). The first example implements a general 2 degree of freedom interaction test (in the contrast statement below).

```
proc catmod data=pharma;
population e t;
model g = ( 0 0 0 0 -2 0 0,
0 0 0 0 -1 0 0.6931472,
0 0 0 0 0 -2 0,
0 0 0 0 0 -1 0.6931472,
-1 -1 0 0 -2 0 0,
0 -1 0 0 -1 0 0.6931472,
-1 -1 -1 -1 0 -2 0,
0 -1 0 -1 0 -1 0.6931472)
(1='beta_d' , 2='beta_r' ,3='eta_d' ,4='eta_r' ,5='gamma_0' ,
6='gamma_1',7='offset') / design zero=samp;
restrict b7=1;
contrast "eta=0 general 2df test" all_parms 0 0 1 0 0 0 0,
all_parms 0 0 0 1 0 0 0;
run;
```

The second example tests for a 1 degree of freedom additive interaction.

```

proc catmod data=casecontrol;
population e t;
model g = ( 0 0 0 -2 0 0,
0 0 0 -1 0 0.6931472,
0 0 0 0 -2 0,
0 0 0 0 -1 0.6931472,
-1 -1 0 -2 0 0,
0 -1 0 -1 0 0.6931472,
-1 -1 -2 0 -2 0,
0 -1 -1 0 -1 0.6931472)
(1='beta_d',2='beta_r',3='eta',4='gamma_0',
5='gamma_1',6='offset') / design zero=samp;
restrict b6=1;
contrast "eta=0 additive test" all_parms 0 0 1 0 0 0;
run;

```

SAS code for X-linked marker

In the following example we compute a general 2 degree of freedom test for an X-linked marker. The model is parameterized so that males with the risk allele ($g=1$) contribute information to the recessive effect. This parameterization reflects the plausible biologic model that it may take two copies of a mutant allele to disrupt gene function in females but only one mutant copy is required to disrupt function in males. This example assumes the data is in a SAS dataset called “pharmaX” which, like the autosomal examples above, contains the three variables, g , e , and t , along with a new variable f which indicates whether the individual is female or not ($f=1$ if individual is female; $f=0$ otherwise).

```

proc catmod data=pharmaX;
population f e t;
model g = ( 0 0 0 0 -1 0 0,
0 0 0 0 0 0 0,
0 0 0 0 0 -1 0,
0 0 0 0 0 0 0,
0 -1 0 0 -1 0 0,
0 0 0 0 0 0 0,
0 -1 0 -1 0 -1 0,
0 0 0 0 0 0 0,
0 0 0 0 -2 0 0,
0 0 0 0 -1 0 0.6931472,
0 0 0 0 0 -2 0,
0 0 0 0 0 -1 0.6931472,
-1 -1 0 0 -2 0 0,
0 -1 0 0 -1 0 0.6931472,
-1 -1 -1 -1 0 -2 0,
0 -1 0 -1 0 -1 0.6931472)
(1='beta_d',2='beta_r',3='eta_d',4='eta_r',
5='gamma_0',6='gamma_1',7='offset') / design
zero=structural ml=nr;

```

```
restrict b7=1;  
contrast "eta=0 2df test" all_parms 0 0 1 0 0 0 0,  
all_parms 0 0 0 1 0 0 0;  
run;
```

Figure 1. Plots of power by interaction effect size for autosomal locus and three different risk alleles frequencies assuming HWE ($\alpha = 0.05$, $F_{st} = 0$)

Figure 2.
Plots of power by interaction effect size for autosomal locus and three different risk alleles frequencies assuming substantial departure from HWE ($\alpha = 0.05$, $F_{st} = 0.15$)

Figure 3. Plots of power by interaction effect size for X-linked locus and three different risk alleles frequencies ($\alpha = 0.05$)

Table 1

Design elements for autosomal locus with $G = j$ and $T = t$. Note that $X_r(j) = I(j = 2)$ and $X_d(j) = I(j > 0)$.

Model	Genetic Main Effect X_j	Gene-by-Treatment Interaction Effect X_{jt}
Recessive	$X_r(j)$	$tX_r(j)$
Dominant	$X_d(j)$	$tX_d(j)$
Additive	$X_r(j) + X_d(j)$	$tX_r(j) + tX_d(j)$
General	$(X_r(j), X_d(j))^T$	$(tX_r(j), tX_d(j))^T$

Table 2

Design elements for X-linked locus with $G = j$, $T = t$, $S = s$. Note that $X_r(j) = I(j = 2)$, $X_d(j) = I(j > 0)$, $\delta_f = I(s = f)$, and $\delta_m = I(s = m)$.

Model (Females)	Genetic Main Effect X_{js}	Gene-by-Treatment Interaction Effect X_{jts}
Recessive	$(\delta_f X_r(j), \delta_m X_d(j))^T$	$(t\delta_f X_r(j), t\delta_m X_d(j))^T$
Dominant	$(\delta_f X_d(j), \delta_m X_d(j))^T$	$(t\delta_f X_d(j), t\delta_m X_d(j))^T$
Additive	$(\delta_f [X_r(j) + X_d(j)], \delta_m X_d(j))^T$	$(t\delta_f [X_r(j) + X_d(j)], t\delta_m X_d(j))^T$
General	$(\delta_f X_r(j), \delta_f X_d(j), \delta_m X_d(j))^T$	$(t\delta_f X_r(j), t\delta_f X_d(j), t\delta_m X_d(j))^T$

Table 3Type I error rates for autosomal locus assuming HWE ($\alpha = 0.05$, $F_{st} = 0$)

π	Retrospective general	Retrospective additive	Prospective general	Prospective additive
0.1	0.035	0.046	0.019	0.046
0.2	0.044	0.053	0.046	0.053
0.5	0.053	0.053	0.049	0.054

Table 4

Type I error rates for autosomal locus assuming substantial departure from HWE ($\alpha=0.05$, $F=0.15$)

π	Retrospective general	Retrospective additive	Prospective general	Prospective additive
0.1	0.039	0.053	0.029	0.054
0.2	0.053	0.057	0.048	0.049
0.5	0.056	0.055	0.052	0.048

Table 5

Power for autosomal locus assuming HWE ($\alpha = 0.05$, $F_{st} = 0$)

True Model	η	π	Retrospective general	Retrospective additive	Prospective general	Prospective additive
Additive	ln (2.2)	0.1	0.60	0.74	0.51	0.75
		0.2	0.85	0.91	0.86	0.92
		0.5	0.94	0.97	0.94	0.97
Dominant	ln (2.2)	0.1	0.59	0.67	0.51	0.67
		0.2	0.75	0.79	0.75	0.79
		0.5	0.59	0.47	0.57	0.48
Recessive	ln (3.4)	0.1	0.11	0.09	0.03	0.08
		0.2	0.51	0.30	0.39	0.24
		0.5	0.98	0.95	0.96	0.93

Table 6
Power for autosomal locus assuming substantial departure from HWE ($\alpha = 0.05$, $F_{st} = 0.15$)

True Model	η	π	Retrospective general	Retrospective additive	Prospective general	Prospective additive
Additive	$\ln(2.2)$	0.1	0.67	0.78	0.68	0.81
		0.2	0.89	0.94	0.91	0.95
		0.5	0.95	0.98	0.96	0.98
Dominant	$\ln(2.2)$	0.1	0.54	0.60	0.53	0.62
		0.2	0.71	0.73	0.71	0.75
		0.5	0.61	0.49	0.58	0.52
Recessive	$\ln(3.4)$	0.1	0.26	0.20	0.19	0.17
		0.2	0.67	0.46	0.55	0.40
		0.5	0.98	0.95	0.97	0.95