Published in final edited form as:

Biometrics. 2015 June ; 71(2): 529–537. doi:10.1111/biom.12259.

A Penalized Likelihood Approach for Investigating Gene-Drug Interactions in Pharmacogenetic Studies

Megan L. Neely^{1,*}, Howard D. Bondell^{2,**}, and Jung-Ying Tzeng^{2,3,***}

¹Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, 27705, U.S.A

²Department of Statistics, North Carolina State University, Raleigh, North Carolina, 27695, U.S.A

³Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, 27695, U.S.A

Summary

Pharmacogenetics investigates the relationship between heritable genetic variation and the variation in how individuals respond to drug therapies. Often, gene-drug interactions play a primary role in this response, and identifying these effects can aid in the development of individualized treatment regimes. Haplotypes can hold key information in understanding the association between genetic variation and drug response. However, the standard approach for haplotype-based association analysis does not directly address the research questions dictated by individualized medicine. A complementary post-hoc analysis is required, and this post-hoc analysis is usually under powered after adjusting for multiple comparisons and may lead to seemingly contradictory conclusions. In this work, we propose a penalized likelihood approach that is able to overcome the drawbacks of the standard approach and yield the desired personalized output. We demonstrate the utility of our method by applying it to the Scottish Randomized Trial in Ovarian Cancer. We also conducted simulation studies and showed that the proposed penalized method has comparable or more power than the standard approach and maintains low Type I error rates for both binary and quantitative drug responses. The largest performance gains are seen when the haplotype frequency is low, the difference in effect sizes are small, or the true relationship among the drugs is more complex.

Keywords

Association analysis; Haplotype; Individualized medicine; Multiple comparisons; Penalized regression; Pharmacogenetics

*megan.neely@duke.edu

**hdbondel@ncsu.edu

***jytzeng@stat.ncsu.edu

6. Supplementary Materials

Web Appendix A, referenced in Sections 3 and 4, and an R package, Haplo. CasGLM, for the penalized method are available with this paper at the Biometrics website on Wiley Online Library.

1. Introduction

Inter-individual variation in the efficacy and toxicity of drug therapies is common among patients. Although the differences among patients could be the result of many factors, pharmacogenetics focuses on the relationship between drug response and heritable genetic variation. Two research goals are often paramount when conducting pharmacogenetic studies: (1) For an individual, determine which drugs will perform ‘best/worst’ for them based on their genetic make-up, and (2) for a particular drug, determine the class of individuals that it is ‘best/ill’ suited for based on their genetic make-up. Employing haplotype-based methods in pharmacogenetic studies is an attractive approach. Haplotype-based association analysis evaluates the joint effects of closely linked genetic markers on a trait of interest. When compared to its single-marker counterparts, this multi-marker approach can be more powerful to detect associations when the causal variants are not genotyped (de Bakker et al., 2005; Zaitlen et al., 2007), have low frequency (de Bakker et al., 2005; Schaid, 2004), or exhibit cis-acting effects (Clark, 2004; Schaid, 2004). Moreover, haplotypes are expected to better capture many pharmacogenetic variants (Goldstein, Tate, and Sisodiya, 2003).

The standard approach for performing haplotype-based analysis is to regress the drug response on the haplotypes, treatments, and their corresponding interactions and test the significance of the regression parameters (Balding, 2006). However, determining whether individual coefficient estimates are significantly different from zero does not directly address the research goals of individualized medicine. To establish which drugs have similar effects for a particular individual, or conversely, which individuals react similarly to a particular drug, a large number of pair-wise comparisons must be performed. Thus, the standard approach requires a complementary post-hoc analysis in order to produce the desired personalized output. However in practice, this post-hoc analysis is usually under-powered after adjusting for multiplicity. Furthermore, the pair-wise comparisons can yield contradictory conclusions about the effect structure. That is, when the pair-wise comparisons are translated into a grouping structure on the effects, the groups overlap rather than falling into distinct, interpretable subsets. This occurred when analyzing the Scottish Randomized Trial in Ovarian Cancer (SCOTROC1) data (Vasey et al., 2004). The trial investigated the effects of two treatment regimes involving taxane/platinum-based chemotherapy. In a previous analysis, McWhinney-Glass et al. (2013) found that the gene *BCL2* was significantly associated with an increased risk of experiencing severe neurotoxicity. Focusing on this gene, the goal of our analysis was to investigate potential gene-treatment interactions. Specifically, we wanted to determine if individuals responded differently to the assigned treatment based on their genetic make-up. Using the standard haplotype-based approach of regression and subsequent pair-wise testing, we were able to detect significant differences after adjusting for multiplicity using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). However, in each case, the groupings led to contradictory conclusions about which genetic variants had similar effects for a given treatment. Such results are difficult to interpret clinically and prevent the development of coherent personalized treatments. For individualized medicine to be clinically relevant, these issues need to be addressed and overcome.

In this work, we propose a penalized likelihood method that directly addresses the pharmacogenetic research goals stated above. The method builds off the regression framework of the standard haplotype-based approach, but is able to overcome the drawbacks of having to do a secondary post-hoc analysis and yields the desired personalized output directly. In the literature, using penalized likelihood methods to identify important haplotypic effects has been popular (Guo and Lin, 2009; Li, Sung, and Liu, 2007; Li et al., 2010; Tzeng et al., 2010). These methods introduce a penalty on the regression coefficients and shrink the coefficient estimates of non-important covariates toward zero. Recently, modifications of classic penalized methods have also been developed to perform haplotype-based analysis and attempt to address issues specific to this type of analysis. For example, Souverein, Tanck, and colleagues (Souverein, Zwinderman, and Tanck, 2006; Souverein et al., 2008) use a modified version of Ridge regression to stabilize inference for rare haplotypes. Chen, Chatterjee, and Carroll (2009) develop an adaptive penalized likelihood framework to address the precision-efficiency tradeoff encountered in haplotype-based retrospective methods. Tzeng and Bondell (2010) modify the adaptive LASSO (Tibshirani, 1996; Zou, 2006) to allow for effect comparisons between all pairs of distinct haplotypes, rather than with respect to an arbitrary baseline haplotype, during the estimation process.

In our penalized likelihood method, we place an L_1 -penalty on all relevant pair-wise comparisons, which are dictated by the research questions, so that the estimates of group means with the same magnitude are collapsed to be equal. For example, if two drugs perform similarly for an individual, we would like our method to collapse their means and produce estimates that are equal to reflect this inter-relationship. In this way, our method simultaneously achieves the effect estimation and pair-wise comparisons necessary to address the research questions stated above, which eliminates the need to perform any post-hoc analysis. In addition, our method is able to collapse the effects into an overall group structure without leading to contradictory conclusions. As a result, the proposed method directly produces the personalized output desired when developing individualized treatments. Simulation studies show that the proposed penalized method has comparable or more power than the standard approach and maintains low Type I error rates for both binary and quantitative drug responses. The largest gain is seen when the haplotype frequency is low, the difference in effect sizes are small, and more complex effect structures are present. The utility of the proposed method is further demonstrated when used to analyze the SCOTROC1 data. Our method was able to provide findings that are not only significant, but are also easily interpretable unlike the standard approach which resulted in contradictory results.

2. Methods

For individual i , $i = 1, \dots, n$, let Y_i be the phenotype, $\mathbf{G}_i$ be the unphased genotype of the m SNPs, $\mathbf{E}_i$ be the environmental covariates, and $\mathbf{H}_i$ be the corresponding haplotype counts based on the m SNPs. For each subject, we observe $(Y_i, \mathbf{G}_i, \mathbf{E}_i)$ but not $\mathbf{H}_i$.

The relationship between the response and the covariates can be characterized by the conditional density function $f(Y|\mathbf{H}, \mathbf{E})$. The standard approach for haplotype-based association analysis is regression, and generalized linear models (GLMs) provide a flexible

and robust framework for the analysis. GLMs assume that there exists a monotone function $g(\cdot)$ that links the expected value of Y , denoted by $\mu = E(Y|\mathbf{H}, \mathbf{E})$, to the linear predictor η . That is, $\eta = g(\mu) = \gamma_0 + \boldsymbol{\beta}^T \mathbf{Z}(\mathbf{H}) + \boldsymbol{\delta}^T \mathbf{E} + \boldsymbol{\gamma}^T \mathbf{Z}(\mathbf{H}) \otimes \mathbf{E}$, where γ_0 represents a global intercept, $\boldsymbol{\beta}$ is the vector of haplotype effects, $\boldsymbol{\delta}$ is the vector of drug effects, $\boldsymbol{\gamma}$ is the vector of haplotype-drug interaction effects, $\mathbf{Z}(\mathbf{H})$ is a vector-valued function of the vector of haplotype counts, and $\mathbf{Z}(\mathbf{H}) \otimes \mathbf{E}$ is the Kronecker product of the vectors $\mathbf{Z}(\mathbf{H})$ and $\mathbf{E}$, the vector of drug indicators. In this work, we assumed an additive model on the haplotype main and interaction effects. This implies that the function $\mathbf{Z}(\cdot)$ is the identity, and that the linear predictor can be written in the following ANOVA-like representation $\eta = g(\mu) = \gamma_0 + \beta_h + \beta_{h'} + \delta_j + \gamma_{hj} + \gamma_{h'j}$, where the sums $\sum_h \beta_h$, $\sum_j \delta_j$, $\sum_h \gamma_{hj}$, and $\sum_j \gamma_{h'j}$ must be set to zero to account for the over parameterization $h = 1, \dots, l$, the number of distinct haplotypes, and $j = 1, \dots, d$, the number of drugs under study. The linear predictor above represents an individual with the haplotypes h and h' , or the diplotype hh' , who received drug j in the study. For quantitative responses, the distribution of Y is typically assumed to be normal with the identity link to lead to linear regression. For binary responses, the distribution of Y is typically assumed to be Bernoulli with the logit link to lead to logistic regression.

Recall the two research goals stated in Section 1. To address the first goal, all pair-wise comparisons between drugs would need to be performed for each distinct “genetic makeup”. In the haplotype setting, an individual is uniquely identified by their diplotype. The appropriate hypothesis tests of drug differences for a diplotype hh' are thus given by $H_0 : \eta_{hh'j} - \eta_{hh'j'} = 0$ for $j < j'$ reduces to $H_0 : \delta_j + \gamma_{hj} + \gamma_{h'j} - \delta_{j'} - \gamma_{hj'} - \gamma_{h'j'} = 0$. Following the idea of the penalized approach to ANOVA (Bondell and Reich, 2009), the proposed penalized method estimates the haplotype, drug, and interaction effects as

$$\hat{\boldsymbol{\phi}}_{\lambda} = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \left\{ -l_n(\boldsymbol{\phi}, \boldsymbol{\xi}; \mathbf{Y}, \mathbf{G}, \mathbf{E}) + \lambda \sum_{hh'} \sum_{j < j'} w_{hh'jj'} |\delta_j + \gamma_{hj} + \gamma_{h'j} - \delta_{j'} - \gamma_{hj'} - \gamma_{h'j'}| \right\}$$

subject to $\sum_h \beta_h = \sum_j \delta_j = \sum_h \gamma_{hj} = \sum_j \gamma_{h'j} = 0$, where $\boldsymbol{\phi} = (\gamma_0, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}^T, l_n(\boldsymbol{\phi}, \boldsymbol{\xi}, \mathbf{Y}, \mathbf{G}, \mathbf{E}))$ is the log-likelihood, $\boldsymbol{\xi}$ is a (possible) set of nuisance parameters (e.g. haplotype frequencies), λ is the non-negative regularization parameter that controls the amount of shrinkage, and $w_{hh'jj'}$ are data-dependent weights. By placing an L_1 -norm penalty on each of the pair-wise comparisons, the penalized method can set these differences in linear predictors to be exactly zero if the value of λ is large enough. It is this feature that allows the procedure to perform parameter estimation while simultaneously considering the overall drug effect structure for each diplotype. The linear predictors of drugs that perform similarly for a particular diplotype will be shrunk towards each other. As a result, the proposed method will be able to collapse some linear predictors to exact equality and thus yield the overall drug effect structure for each diplotype directly, without leading to any contradictory groupings, which is the desired personalized output.

Zou (2006) suggests setting the weights as $w_{hh'jj'} = |\tilde{\delta}_j + \tilde{\gamma}_{hj} + \tilde{\gamma}_{h'j} - \tilde{\delta}_{j'} - \tilde{\gamma}_{hj'} - \tilde{\gamma}_{h'j'}|^{-\nu}$, where $\tilde{\boldsymbol{\phi}}$ is an initial root-n consistent estimator of $\boldsymbol{\phi}$ and $\nu > 0$ is an additional tuning parameter. In our analysis, we chose $\nu = 1$ and $\tilde{\boldsymbol{\phi}}$ be the maximum likelihood estimate (MLE) of $\boldsymbol{\phi}$

obtained from `haplo.glm` in R (Lake et al., 2003). It has been shown that these adaptive weights have desirable asymptotic properties in this setting (Bondell and Reich, 2009). The penalized solution ($\hat{\phi}_\lambda$) also depends on the value of λ . Because the goal of this analysis is more aligned with selecting the true model structure than minimizing prediction error, we use BIC for tuning which can achieve consistent model selection (Yang, 2005). BIC is defined as $BIC = -2l_n(\hat{\phi}_\lambda, \hat{\xi}; \hat{Y}, \mathbf{G}, \mathbf{E}) + df_\lambda \cdot \log(n)$, where $l_n(\hat{\phi}_\lambda, \hat{\xi}; \hat{Y}, \mathbf{G}, \mathbf{E})$ is the log-likelihood evaluated at the estimated regression coefficients and maximized over $\hat{\xi}$ for a given λ and df_λ is the degrees of freedom based on $\hat{\phi}_\lambda$. The quantity df_λ is given by $\sum_{hh} \text{unique}_j \{ \eta_{hhj} \}$, which equals the number of unique estimated linear predictors among all drugs within each homozygous diplotype. The λ that minimizes BIC is chosen as the regularization parameter, and its corresponding $\hat{\phi}_\lambda$ is the penalized estimate.

It should be noted that the data likelihood is a function of $\mathbf{G}$ and not $\mathbf{H}$. Haplotypes are not directly observable and must be inferred from the genotype data during estimation. This requires an iterative procedure, such as the EM Algorithm (e.g. Lake et al., 2003), to optimize the likelihood. However, this can be computationally intensive when fitting the penalized method as the optimization must be done at every grid point λ . For computational convenience, the objective function created via the least squares approximation (LSA) method was used to calculate the penalized solution. The LSA method replaces the objective function of the original penalized problem with a least squares objective function (Weng and Leng, 2007). The method is motivated by a standard Taylor series expansion of $-l_n(\phi, \xi)$ about $(\tilde{\phi}, \tilde{\xi})$, the function's unpenalized minimizer, and Weng and Leng show that the original penalized estimate has the exact same asymptotic distribution as $\hat{\phi}_\lambda = \text{argmin}_\phi \{ (\phi - \tilde{\phi})^T \tilde{\Sigma}^{-1} (\phi - \tilde{\phi}) + P(\lambda, \phi) \}$, where $\tilde{\Sigma}$ is the estimated covariance matrix of $\tilde{\phi}$ and $P(\lambda, \phi)$ is the penalty term. Using the LSA method eliminates the need for an iterative procedure to optimize the data likelihood; it only requires one unpenalized fit of the original objective function and then a grid search to determine λ .

Employing similar logic, a penalty term can be developed to address the second research goal by considering all pair-wise comparisons among diplotypes for each drug. Thus, the penalty terms for the first and second research goals are given by

$$P_1(\lambda, \phi) = \lambda \sum_{hh'} \sum_{jj'} w_{hh'jj'} |\delta_j + \gamma_{hj} + \gamma_{h'j} - \delta_{j'} - \gamma_{hj'} - \gamma_{h'j}|$$

and

$$P_2(\lambda, \phi) = \lambda \sum_j \sum_{hh' < kk'} w_{hh'jj'} |\beta_h + \beta_{h'} + \gamma_{hj} + \gamma_{h'j} - \beta_k - \beta_{k'} - \gamma_{kj} - \gamma_{k'j}|,$$

respectively. In this way, the penalized method can be written succinctly as $\hat{\phi}_\lambda = \text{argmin}_\phi \{ -l_n(\phi, \hat{\xi}; \hat{Y}, \mathbf{G}, \mathbf{E}) + P(\lambda, \phi) \}$, where $P(\lambda, \phi)$ can be either $P_1(\lambda, \phi)$ or $P_2(\lambda, \phi)$ depending on the research goals.

3. Simulation Studies

We performed simulation studies to compare the performance of the proposed penalized likelihood method against the standard approach of regression followed by a post-hoc analysis to address the research goals described in Section 1. Two post-hoc analysis procedures were considered – the ‘unadjusted’ and the ‘FDR-adjusted’ methods. The ‘unadjusted’ method performs the pair-wise testing without adjusting for multiple comparisons. The ‘FDR-adjusted’ method uses the Benjamini-Hochberg (Benjamini and Hochberg, 1995) adjustment procedure to control the false discovery rate (FDR) among the multiple comparisons. In our simulations, we controlled the FDR at 5%.

3.1 Simulation Designs

We considered two simulation designs. The first simulation design (Simulation I) aimed to evaluate the performance under the first research goal. Specifically, we examined each method’s ability to identify the correct drug structure for each diplotype (i.e., individual). The simulation study was based on a haplotype distribution (given in Table 1) studied by Lin and Huang (2008). The distribution is based on the common haplotypes formed by five SNPs found on chromosome 18 in the CEU sample of the HapMap data. Using this haplotype distribution, we considered two simulation studies – a quantitative drug response and a binary drug response. We considered 5 drugs. To investigate the performance of the three methods, we varied the drug effect structure among the diplotypes (given in Web Table 1). Specifically, we considered three increasingly complex drug effect structures – Null, One Best Drug (1BD) and Two Best Drugs (2BD). Under the Null structure, all drugs perform the same with respect to each other. Under the 1BD structure, there is one drug that outperforms the other four drugs, meaning it has a higher or lower linear predictor than the others. Under the 2BD structure, there are two drugs that outperform the other three drugs, but are equivalent to each other. We set the values of n and θ (the effect size difference of the ‘best’ drug) so that the power of identifying the effect structures fell within a reasonable range. For the quantitative response simulation, we set $n = 500$ and $\theta = \{1.5, 2.0\}$. For the binary response simulation, we set $n = 750$ and $\theta = \{2.0, 2.5\}$. We assumed a balanced design among the different drugs.

We generated the data prospectively by first randomly selecting n diplotypes which were then used to compute the linear predictor $\phi^T X$, where $X = (H, E, H \otimes E)^T$ and $\phi = (\gamma_0, \beta, \delta, \gamma)^T$ are the regression coefficients (given in Web Table 2) used to induce the drug effect structure given in Web Table 1. A quantitative drug response was generated under a normal distribution with a mean of $\phi^T X$ and a standard deviation of $\sigma_\varepsilon = 1$. A binary drug response was generated under a Bernoulli distribution with $p = \text{logit}^{-1}(\phi^T X)$.

For each simulation setting, 200 replicate data sets were generated. For each data set, analysis began by calculating the unpenalized MLEs of the regression coefficients ($\hat{\phi}$) and their estimated covariance matrix ($\hat{\Sigma}$) using `haplo.glm` in R (Lake et al.; 2003). To determine which drugs perform similarly (i.e., which pairs of drugs should be grouped together or not) for each diplotype, the tests $H_0: \eta_{hh'j} - \eta_{hh'j'} = 0$ need to be performed for $j < j'$ and for all diplotypes hh' . For the unadjusted and the FDR-adjusted methods, the unpenalized MLEs and their covariance matrix were used directly to perform the tests using the asymptotic

normal distribution of $\eta_{hh'j} - \eta_{hh'j'}$ described by Lake et al (2003). For the penalized method, the final penalized estimates of the regression coefficients ($\hat{\phi}$) were computed using the unpenalized MLEs and their covariance matrix in the penalized likelihood with $P_1(\lambda, \phi)$ and BIC tuning. If the estimated linear predictors from the penalized method are not identical we would conclude that those drugs performed differently for that particular diplotype. Based on the testing results, true positive rates were calculated over the 200 simulated runs to compare the performance of the three methods.

The second simulation design (Simulation II) aimed to examine the performance under the second research goal. Specifically, we examined each method's ability to identify the correct diplotype structure under each drug for a quantitative trait. We used the haplotype distribution and effect structure observed in the real data example (Section 4; haplotype distribution given in Web Table 7 and diplotype effect structure given in Web Table 8) and examined the three methods' performance under 3 drug scenarios – one with no differences across diplotypes (like the effect structure observed under Treatment B), one with one “best” haplotype (like the effect structure observed under Treatment A) where the haplotype had high frequency, and one with one “best” haplotype where the haplotype had lower frequency. This design allowed us to determine whether the pattern of results fully studied in Simulation I held under the framework of the real data analysis. We set the values of $n = 100$ and $\theta = 0.5$ to be consistent with the values observed in the real data example. We assumed a balanced design among the different drug.

3.2 Results of Simulation I

We present true negative rates (for diplotype H5H5) and true positive rates (for all other diplotypes), in Figures 1 – 3. Each figure contains the results of the quantitative and binary simulations (across columns) at both levels of the effect difference θ (down rows). Because the performance of each method improved as diplotype frequency increased, for brevity, we only present the results for 7 diplotypes, which represent the other diplotypes with similar frequencies. The selected diplotypes are listed on the x-axis of each figure. For each drug effect structure, we selected a representative ‘low’ (unshaded) and ‘high’ (shaded) frequency diplotype. The performance of the three procedures was compared by examining their ability to correctly identify pre-specified drug effect structures for a given diplotype.

3.2.1 Correct Structure Results—Figure 1 presents the correct structure results (CSR), which is the proportion of runs a method was able to correctly identify the inter-relationship among all drugs for a given diplotype. For all three methods, the CSR is lower for the binary drug response than the quantitative drug response (except in the Null drug structure), but the relative patterns among the different methods are similar across the two simulation studies. Under the Null drug structure (diplotype H5H5), the unadjusted method yields a low true negative rate, while the penalized and FDR-adjusted methods yield similar true negative rates. A low true negative rate translates into a high overall Type I error rate, where the Type I error rate for the unadjusted method ranges from 0.27 to 0.30 versus 0.02 to 0.08 and 0.02 and 0.14 for the penalized and FDR-adjusted methods, respectively. For the remaining diplotypes in Figure 1, the unadjusted method performs comparably or better than the penalized method (i.e., dotted line is about the same level or higher than the solid line).

However, this gain in power comes at the expense of having a much higher Type I error rate than the penalized method and, hence, is not really comparable. On the other hand, the FDR-adjusted method maintains a similar Type I error rate as the penalized method, but its performance suffers as a result (i.e., the dashed line is about the same level or lower than the solid line). To examine whether the penalized method's advantage in identifying true negatives and true positives mainly arises from the dynamic selection of tuning parameters or from a different ordering of signal identification, we studied the relationship between the solution path (in terms of increasing λ) of the penalized approach and the order of significances of pair-wise tests. Figure 4 presents the pointwise-averaged ROC curve for the simulation using the quantitative trait with 'best' drug effect difference set to 1.5. The penalized path looks significantly better, particularly in the region to the right of 1-specificity = 0.05, which would be the region of interest. A vertical line is plotted at this point as reference as it is essentially equivalent to controlling the family-wise error rate at 5%, which would be extremely conservative given that there are 150 pairwise differences that are being performed in this multiple testing problem. These results suggest that the penalized method's advantage in true negatives and true positives arise from not only the tuning procedure, but also from a more useful ordering of the differences as well. From these findings, it's apparent that the standard approaches must compromise between identifying true negatives and true positives, and the penalized method finds a better balance between identifying these true signals. As a result, the penalized method can uncover the underlying drug effect structure more accurately for a given diplotype.

As expected, as the difference in effect size increases (i.e., top row of Figure 1 <bottom row of Figure 1) or as the diplotype frequency increases (i.e., unshaded results <shaded results) or as the effect magnitude increases (i.e., 1BD Half results <1BD Full results), the CSR increases. Thus, the power of a method to identify the correct drug structure for a particular diplotype depends on the underlying diplotype frequency, the effect size, and the complexity of the drug structure. The largest gain in performance of the penalized method occurs when the diplotype frequency is low, the effect size small, or the effect structure is more complex.

3.2.2 Best Drug Results: One Best Drug Structure—Figure 2 presents the best drug results (BDR) for the 1BD structure. Under this structure, we defined BDR as the proportion of runs a method was able to identify the single best drug as significantly better than all other drugs for a given diplotype. This performance definition is less strict than the definition of CSR. As a result, the BDR is higher than the CSR for a given diplotype with the 1BD structure across all simulation scenarios (i.e., curve in Figure 1 <curve in Figure 2). When comparing the BDR between the different methods, the results in Figure 2 mirror those in Figure 1. Again, as the diplotype frequency increases (i.e., unshaded results <shaded results) or as the effect magnitude increases (i.e., 1BD Half results <1BD Full results), the BDR increases. The BDR of the penalized method is comparable to or better than the BDR of the unadjusted and FDR-adjusted methods (i.e., solid line is about the same level or higher than the dotted or dashed lines). The penalized method's performance gain over the two standard methods is more pronounced for BDR than for CSR (i.e., solid line – dotted (dashed) line in Figure 1 <solid line – dotted (dashed) line in Figure 2). This gain diminishes as diplotype frequency increases. These results again suggest that the penalized

method outperforms the standard approaches and can more accurately determine which drug is the best for a given diplotype among a group of treatments even if the complete drug structure was not identified.

3.2.3 Best Drug Results: Two Best Drugs Structure—Figure 3 presents the best drug results (BDRs) for the 2BD structure. Under the 2BD structure, we defined two separate BDRs. The first is the proportion of runs a method was able to find the two best drugs as significantly better than all other drugs; this is referred to as *Both Found* BDR. Because this performance definition ignores the relationship between the two best drugs, we also defined a second BDR for this drug structure, which is the proportion of runs a method was able to find the two best drugs as significantly better than all other drugs and equivalent to each other; this is referred to as *Both & Exact* BDR. These performance definitions are less strict than the definition of CSR, but not as liberal as the BDR definition for 1BD drug structure. As a result, the BDRs are higher than the CSR for a given diplotype with the 2BD structure across all simulation scenarios (i.e., curve in Figure 1 < curve in Figure 3). As expected, the Both BDR is less than the Both & Exact BDR for a given diplotype (i.e., unshaded (shaded) Both results < unshaded (shaded) Both & Exact results). For both BDRs, as the diplotype frequency increases, the power increases (i.e., unshaded results < shaded results for Both Found (Both & Exact)). When comparing the BDRs between the different methods, the penalized method has uniformly higher BDRs (i.e., solid line above the dotted and dashed lines). These results again suggest that the penalized method outperforms the standard approaches and can more accurately determine which drugs are the best for a given diplotype among a group of treatments even if the complete drug structure was not identified.

3.3 Results of Simulation II

We assessed the ability of all three methods to identify the correct diplotype effect structure for each drug as well as their ability to identify the diplotype that is the “best” match for each drug (results given in Web Table 9). The penalized method identifies the correct diplotype effect structure and “best” diplotype across more of the simulated data sets than either of the unpenalized methods. The Type I error rate (i.e. performance rates under Drug 3) is well-maintained for the penalized method. The frequency of the underlying haplotypes also influences the performance of all three methods under research goal 2 as it did under research goal 1. As the frequency of the causal haplotype decrease, so does each method’s ability to correctly identify the correct effect structure or the “best” diplotype (i.e., rate under Drug 1 is lower than the rate under Drug 2). This suggests that the penalized method might be missing true signals in the real data example if the causal haplotype has low frequency (and/or small effect size). However, the penalized method uniformly outperforms the other methods suggesting it will likely have the greatest ability to detect signals.

4. Real Data Example: Application to SCOTROC1 Data

The penalized method proposed in this work was applied to the pharmacogenetic data from the Scottish Randomized Trial in Ovarian Cancer (SCOTROC1) (McWhinney-Glass et al., 2013). The trial investigated the effects of two treatment regimes involving taxane/platinum-

based chemotherapy: Treatment A refers to docetaxel-carboplatin and Treatment B refers to paclitaxel-carboplatin. In a preliminary analysis, McWhinney-Glass et al. (2013) found that the gene *BCL2* was significantly associated with an increased risk of experiencing severe neurotoxicity. Focusing on this gene, the goal of our analysis was to determine if individuals responded differently to the treatment they received based on their genetic make-up.

We focused our analysis on the 105 SNPs genotyped within the *BCL2* gene for 808 patients. A sliding window analysis was used to investigate the existence of diplotype-treatment interactions. For each window, the resulting data was analyzed using the three procedures discussed in the Simulation Study section. All methods were performed assuming an additive haplotype-interaction model. For each method, all pair-wise comparisons between diplotypes were performed within each treatment arm. The sliding window analysis was performed using a window of size 6 SNPs. Note that although the FDR-adjusted method adjusts for multiple testing within a particular window, it does not adjust for the multiplicity that results from the sliding window analysis. Hence, the error rate is not controlled overall.

In the sliding window analysis of the gene *BCL2* for all three methods, the standard methods are able to detect significant differences in effect sizes between diplotypes within a treatment arm (Web Figure 1). However, in each case, the groupings led to contradictory conclusions about which diplotypes had similar effects for a given treatment. Such results are difficult to interpret clinically. For example, the pairwise testing results of both standard methods for Window 73 in the *BCL2* gene under Treatment A are presented in Web Table 6. For ease of discussion, consider grouping only the homozygous diplotypes under Treatment A. Based on the output of the standard approaches, two group structures would emerge: Group A = {H1H1, H2H2, H3H3} and Group B = {H1H1, H3H3, H4H4}. From the significant test of H2H2 vs. H4H4, they can infer that patients with the diplotype H2H2 will incur a higher risk of severe neurotoxicity than patients with the diplotype H4H4 because its estimated odds are larger (1.155 vs. 0.314, respectively). But what about patients with diplotypes H1H1 and H3H3 – should clinicians consider these patients to have elevated risk because they are in Group A or not because they are in Group B? In this context, overlapping group structures prevent the development of clear personalized treatments.

The penalized method is able to overcome these hurdles by combining the estimation and testing into a single procedure (i.e., the presence of plus marks at the “SDs & No CGs” position in Web Figure 1). Although the penalized method found significant differences in five windows (8, 31, 73, 74, and 100), specific results are only presented for Window 73 because the method found significant differences for two of adjacent windows. Web Table 3 lists the SNPs corresponding to this window and the resulting haplotypes; Table 2 and Web Table 6 present the specific results of the penalized, unadjusted, and FDR-adjusted analyses, respectively, for this window. The penalized method found three distinct groupings of diplotypes under Treatment A, while no significant differences were found among the diplotypes under Treatment B (i.e., three unique values for the estimated odds under Treatment A and only one under Treatment B in Table 2). It appears that H2 may be the causal haplotype, or at least in linkage disequilibrium (LD) with the causal variant. With each additional copy of H2, the risk of a patient undergoing Treatment A experiencing severe neurotoxicity increases. Similar results were observed for Window 74 (given in Web

Table 4 and 5). The standard approaches detected very few of the same significant differences found by the penalized method (see Web Table 4). Neither approach was able to separate out all diplotypes involving H2, or even separate out the homozygous diplotypes H2H2 from all other diplotypes not involving H2. This leads to contradictory groupings of the diplotypes, like the one described above, and making the output of the standard approaches hard to interpret clinically. This will never occur when using the penalized method. Because it automatically yields non-overlapping group structures, output from the penalized method (Table 2) is easily interpretable and can be used directly in the development of individualized medicine.

5. Discussion

In this work, we introduce a penalized likelihood approach that is powerful and that can effectively address research goals pertinent in pharmacogenetics. Our method places an L_1 -penalty on all pair-wise effect comparisons dictated by the research goals. By combining effect estimation and testing into a single procedure, the penalized method can bypass the issues associated with multiple testing, namely low power and contradictory conclusions about the effect structure. As a result, our penalized method directly produces personalized output that is easily interpretable and clinically relevant. When compared with standard haplotype-based approaches, the penalized method finds a better balance between identifying true negatives and true positives. Thus, the penalized method has comparable or more power, while maintaining reasonable Type I error rates, to identify the underlying effect structure for a particular drug or genetic variant. The largest gain in performance for the penalized method occurs when the frequency of the genetic variant is low, the effect size small, or the effect structure is more complex.

In addition, the penalized method can be used to decipher the global signals from gene-based association studies of genetic variants that may influence drug response. Because gene-based methods consider all variants within a gene jointly, they are considered to be a promising tool for pharmacogenetic research. Drug response is believed to be influenced by both regulatory and structural polymorphisms within the gene and by a wide range of gene products (Goldstein et al., 2003). Marker-based methods might miss these signals; however, because gene-based methods aggregate information across all variants within a gene, they can better capture these effects. The compromise is that gene-based methods assign a single p-value to all variants within a gene. Our penalized method can be used to characterize which variants may be responsible for the signal.

We investigated the performance of the penalized method under a prospective design, but the framework can easily be extended to consider a retrospective design. Using a prospective likelihood when a retrospective likelihood is dictated by the sampling scheme can be detrimental when performing haplotype-based association analyses (Koehler, Bondell, and Tzeng, 2010). The penalized retrospective likelihood can be optimized directly through the use of an iterative procedure; however, we suggest taking advantage of the least square approximation to alleviate the computational burden. This approach requires an initial optimization of the unpenalized method to obtain effect estimates and their estimated covariance matrix; one example of freely available software for obtaining retrospective

estimates is HAPSTAT (Lin, Huang, and Milikan, 2005). Although this work focused on investigating gene-drug interactions, the penalized approach can also be used to investigate gene-gene interactions by considering the diplotypes of another gene in place of the drug therapies. The penalized method can also be extended to simultaneously collapse effects across both factors being studied (i.e., taking $P(\lambda, \phi) = P_1(\lambda, \phi) + P_2(\lambda\phi)$) so that the personalized output can be compared across rows and down columns, rather than in one direction only.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Howard McLeod and the Scottish Gynecological Clinical Trials Group (SGCTG) for the use of the SCOTROC1 data. MLN was supported by NIH T32GM081057. HDB and JYZ were supported by NIH R01 MH084022, and NIH P01-CA142538. HDB was also supported by NSF DMS-1308400.

References

- Balding DJ. A tutorial on statistical methods for population association studies. *Nature Reviews Genetics*. 2006; 7:781–791.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*. 1995; 57:289–300.
- Bondell HD, Reich BJ. Simultaneous factor selection and collapsing levels in ANOVA. *Biometrics*. 2009; 65:169–177. [PubMed: 18510652]
- Clark AG. The role of haplotypes in candidate-gene studies. *Genetic Epidemiology*. 2004; 27:321–333. [PubMed: 15368617]
- Chen YH, Chatterjee N, Carroll RJ. Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *Journal of the American Statistical Association*. 2009; 104:220–233. [PubMed: 19430598]
- de Bakker PW, Yelensky R, Pe'er I, Gabriel S, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nature Genetics*. 2005; 37:1217–1223. [PubMed: 16244653]
- Goldstein DB, Tate SK, Sisodyia SM. Pharmacogenetics goes genomic. *Nature Reviews Genetics*. 2003; 4:937–947.
- Guo W, Lin S. Generalized linear modeling with regularization for detecting common disease rare haplotype association. *Genetic Epidemiology*. 2009; 33:308–316. [PubMed: 19025789]
- Koehler ML, Bondell HD, Tzeng JY. Evaluating haplotype effects in case-control studies via penalized likelihood approaches: prospective or retrospective analysis? *Genetic Epidemiology*. 2010; 34:892–911. [PubMed: 21104891]
- Lake SL, Lyon H, Tantisira K, Silverman EK, Weiss ST, Laird NM, Schaid DJ. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Human Heredity*. 2003; 55:56–65. [PubMed: 12890927]
- Li M, Romero R, Wenjiang JF, Cui Y. Mapping haplotype-haplotype interactions with adaptive LASSO. *BMC Genetics*. 2010; 11:79. [PubMed: 20799953]
- Li Y, Sung WK, Liu JJ. Association mapping via regularized regression analysis of single nucleotide polymorphism haplotypes in variable-sized sliding windows. *American Journal of Human Genetics*. 2007; 80:705–715. [PubMed: 17357076]
- Lin DY, Huang BE. The use of inferred haplotypes in downstream analysis. *American Journal of Human Genetics*. 2008; 80:577–579. [PubMed: 17380613]
- Lin DY, Zeng D, Milikan R. Maximum likelihood estimation of haplotype effects in genetic association studies. *Journal of the American Statistical Association*. 2005; 101:89–104.

- McWhinney-Glass S, Winham SJ, Revolla JY, Paul J, Brown R, Motsinger-Reif A, McLeod HL. Cumulative genetic risk predicts platinum/taxane-induced neurotoxicity. *Clinical Cancer Research*. 2013; 19:57–69.
- Schaid DJ. Evaluating associations of haplotypes with traits. *Genetic Epidemiology*. 2004; 27:348–364. [PubMed: 15543638]
- Souverein OW, Zwinderman AH, Tanck MW. Estimating haplotype effects on dichotomous outcome for unphased genotype data using a weighted penalize log-likelihood approach. *Human Heredity*. 2006; 61:104–110. [PubMed: 16717475]
- Souverein OW, Zwinderman AH, Jukema JW, Tanck MW. Estimating effects of rare haplotypes on failure times using a penalized Cox proportional hazards regression model. *BMC Genetics*. 2008; 9:9. [PubMed: 18221501]
- Tibshirani R. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society B*. 1996; 58:267–288.
- Tzeng JY, Bondell HD. A comprehensive approach to haplotype-specific analysis by penalized likelihood. *European Journal of Human Genetics*. 2010; 18:95–103. [PubMed: 19584902]
- Tzeng JY, Lu W, Farmen MW, Liu Y, Sullivan PF. Haplotype-based pharmacogenetic analysis for longitudinal quantitative traits in the presence of dropout. *Journal of Biopharmaceutical Statistics*. 2010; 20:334–350. [PubMed: 20309762]
- Vasey PA, Jayson GC, Gordon A, Gabra H, Coleman R, Atkinson R, Parkin D, Paul J, Hay A, Kaye SB. Phase III randomized trial of docetaxel-carboplatin versus paclitaxel-carboplatin as first-line chemotherapy for ovarian carcinoma. *Journal of the National Cancer Institute*. 2004; 96:682–691.
- Weng H, Leng C. Unified LASSO estimation by least squares approximation. *Journal of the American Statistical Association*. 2007; 102:1039–1048.
- Yang Y. Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. *Biometrika*. 2005; 92:937–950.
- Zaitlen M, Kang H, Eskin E, Halperin E. Leveraging the HapMap correlation structure in association studies. *American Journal of Human Genetics*. 2007; 80:683–691. [PubMed: 17357074]
- Zou H. The adaptive LASSO and its oracle properties. *Journal of the American Statistical Association*. 2006; 101:1418–1429.

Figure 1. Correct Structure Results for Simulation I

the proportion of runs the complete drug effect structure was correctly identified for the given diplotype. Lines indicate the different estimation methods where *Penalized* denotes the proposed penalized method, *FDR-adjusted* denotes Benjamini-Hochberg adjusted tests with FDR controlled at 0.05, and *Unadjusted* denotes unadjusted tests with a set at 0.05. Shading of the points indicates low (unshaded) vs. high (shaded) diplotype frequency among the drug structure groups (Null, 1BD half, 1BD full, and 2BD; for definitions, see Simulation Studies section).

Figure 2. Best Drug Results for ‘One Best’ Drug Effect Structure for Simulation I
 the proportion of runs the ‘best’ drug found significantly better than all other drugs for the given diplotype. Lines indicate the different estimation methods where *Penalized* denotes the proposed penalized method, *FDR-adjusted* denotes Benjamini-Hochberg adjusted tests with FDR controlled at 0.05, and *Unadjusted* denotes unadjusted tests with a set at 0.05. Shading of the points indicates low (unshaded) vs. high (shaded) diplotype frequency among the drug structure groups (1BD half and 1BD full; for definitions, see Simulation Studies section).

Figure 3. Best Drug Results for ‘Two Best’ Drugs Effect Structure for Simulation I
 the proportion of runs (1) the two ‘best’ drugs found significantly better than all other drugs (*Both Found*) and (2) the two ‘best’ drugs found significantly better than all other drugs and equivalent to each other (*Both & Exact*) for the given diplotype. Lines indicate the different estimation methods where *Penalized* denotes the proposed penalized method, *FDR-adjusted* denotes Benjamini-Hochberg adjusted tests with FDR controlled at 0.05, and *Unadjusted* denotes unadjusted tests with a set at 0.05. Shading of the points indicates low (unshaded) vs. high (shaded) diplotype frequency among the 2BD drug structure group (for definition, see Simulation Studies section).

Figure 4. ROC Curves for Correct Structure Results for Quantitative Trait Simulation with Effect Size of 1.5 for Simulation I

By ordering the pairwise differences created by the penalized likelihood and by the ordering created by the multiple significance t-tests from the unpenalized methods, the ROC curve for each method was constructed by plotting the pointwise average curve over the 200 simulated data sets. A vertical line is plotted at $1 - \text{specificity} = 0.05$ as reference, which is essentially equivalent to controlling the family-wise error rate at 5%, which would be extremely conservative given that there are 150 pairwise differences that are being performed in this multiple testing problem.

Table 1

Haplotype frequency distribution used in Simulation I

Hap ID	Haplotype	Frequency
H1	00001	0.225
H2	01111	0.149
H3	10000	0.139
H4	10001	0.058
H5	00000	0.429

Table 2Results of penalized analysis for Window 73 in *BCL2* gene^{*}

	Treatment A	Treatment B
H2H2	0.970	0.139
H1H2	0.589	0.139
H2H3	0.589	0.139
H2H4	0.589	0.139
H1H1	0.357	0.139
H1H3	0.357	0.139
H1H4	0.357	0.139
H3H3	0.357	0.139
H3H4	0.357	0.139
H4H4	0.357	0.139
<i>df</i>	3	1

^{*} Values given are the estimated odds of each cell