

Case-Control Association Testing with Related Individuals: A More Powerful Quasi-Likelihood Score Test

Timothy Thornton and Mary Sara McPeck

We consider the problem of genomewide association testing of a binary trait when some sampled individuals are related, with known relationships. This commonly arises when families sampled for a linkage study are included in an association study. Furthermore, power to detect association with complex traits can be increased when affected individuals with affected relatives are sampled, because they are more likely to carry disease alleles than are randomly sampled affected individuals. With related individuals, correlations among relatives must be taken into account, to ensure validity of the test, and consideration of these correlations can also improve power. We provide new insight into the use of pedigree-based weights to improve power, and we propose a novel test, the M_{QLS} test, which, as we demonstrate, represents an overall, and in many cases, substantial, improvement in power over previous tests, while retaining a computational simplicity that makes it useful in genomewide association studies in arbitrary pedigrees. Other features of the M_{QLS} are as follows: (1) it is applicable to completely general combinations of family and case-control designs, (2) it can incorporate both unaffected controls and controls of unknown phenotype into the same analysis, and (3) it can incorporate phenotype data about relatives with missing genotype data. The methods are applied to data from the Genetic Analysis Workshop 14 Collaborative Study of the Genetics of Alcoholism, where the M_{QLS} detects genomewide significant association (after Bonferroni correction) with an alcoholism-related phenotype for four different single-nucleotide polymorphisms: *tsc1177811* ($P = 5.9 \times 10^{-7}$), *tsc1750530* ($P = 4.0 \times 10^{-7}$), *tsc0046696* ($P = 4.7 \times 10^{-7}$), and *tsc0057290* ($P = 5.2 \times 10^{-7}$) on chromosomes 1, 16, 18, and 18, respectively. Three of these four significant associations were not detected in previous studies analyzing these data.

We focus on the problem of testing for association between a binary trait and a genetic marker when cases and/or controls are related, with the pedigree(s) assumed known. An advantage in using multiplex families in association studies is that affected individuals who have affected relatives have a higher expected frequency of the alleles that increase susceptibility for a genetic trait than do affected individuals who do not have affected relatives. As a result, the power to detect association is expected to increase when affected individuals with affected relatives are included in the study.

Family-based association tests, such as the transmission/disequilibrium test (TDT),¹ have the advantage that they are robust to population heterogeneity. However, such tests typically require genotype data for family members of an affected individual. Case-control designs are less restrictive than family-based designs, because they can allow but do not require genotype data for relatives of affected individuals, and they are generally more powerful than family-based designs.²

When related individuals are used in case-control studies, one must account for the fact that subjects who are biologically related have correlated genotypes. One approach is to use the standard χ^2 statistic with a correction factor that takes into account the pedigree information³ ($W_{\chi^2_{\text{corr}}}$) or with a correction factor that takes into account

the conditional probability of identity-by-descent (IBD) sharing, given both the observed genotype data and the pedigree information (the “posterior kinship coefficient”)⁴, giving a statistic that we refer to as “ W_{SS} .” Such approaches correct the type I error but still use equal weighting of individuals, among cases and among controls, which is expected to be suboptimal, in terms of power, when individuals are related. As an alternative approach, a quasi-likelihood score (W_{QLS}) test has been proposed.³ Like the $W_{\chi^2_{\text{corr}}}$, the W_{QLS} accounts for the correlations among related individuals, to obtain the proper type I error rates. In addition, for a given alternative model, optimal weights, depending on the pedigree information, are used in the W_{QLS} , in an effort to improve power. For the situation when controls are not related to cases, a method for testing association of a binary trait to a haplotype has been proposed,⁵ where this method uses a similar weighting scheme to that of the W_{QLS} .

We analyze the strengths and weaknesses of the $W_{\chi^2_{\text{corr}}}$, W_{QLS} , and W_{SS} tests, and we use this improved understanding to propose a new and more powerful test, the M_{QLS} test. (The M in M_{QLS} stands for “more powerful” or “modified.”) The M_{QLS} test is more widely applicable than the previously proposed tests, in two ways. First, it distinguishes between unaffected controls and controls of unknown phenotype (i.e., individuals on whom no direct

From the Departments of Statistics (T.T.; M.S.M.) and Human Genetics (M.S.M.), University of Chicago, Chicago
Received March 26, 2007; accepted for publication May 7, 2007; electronically published July 10, 2007.

Address for correspondence and reprints: Dr. Mary Sara McPeck, Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637. E-mail: mcpeek@galton.uchicago.edu

Am. J. Hum. Genet. 2007;81:321–337. © 2007 by The American Society of Human Genetics. All rights reserved. 0002-9297/2007/8102-0012\$15.00
DOI: 10.1086/519497

phenotype information is measured) and can incorporate both into the analysis. Unaffected and unphenotyped general-population controls are the two standard types of controls in case-control studies of disease. Association tests based on combined samples may include both types. Sample sizes of cases and of controls strongly influence the power of the test, so including all available controls is desirable. However, under the alternative hypothesis, individuals who are known to be unaffected have a lower expected frequency of a predisposing allele than do individuals of unknown phenotype (when other factors such as relatives' phenotypes are held constant). This poses the problem of how best to combine the two types of controls in the analysis without compromising power. The M_{QLS} method provides a solution to this problem.

A second way in which the M_{QLS} test is more widely applicable than are the previously proposed tests is that it incorporates phenotype data about relatives who have missing genotype data at the marker being tested. This information is used to optimize the weights given to relatives with nonmissing genotype data at the marker being tested, following the principle that there is enrichment for predisposing variants in affected individuals with affected relatives. This enrichment principle implies, for example, that an affected individual with no phenotyped relatives should be weighted differently from an affected individual with an affected sibling and that this should still hold true when the affected sibling happens to have missing genotype data at the marker being tested. At the same time, the genotypes of the two siblings are dependent, so there should be downweighting of the siblings when they are both genotyped, which does not occur when only one is typed. The M_{QLS} takes into account both the enrichment principle and the effects of dependence in setting the weights.

In addition to the differences just mentioned, we show that, compared with the W_{corr} and W_{SS} , the M_{QLS} improves power by providing a more efficient estimator of allele frequency under the null hypothesis. In large pedigrees, a further power difference between the M_{QLS} and W_{SS} would theoretically arise from the fact that the W_{SS} essentially corrects for the presence of linkage in a family when testing for association,⁵ whereas the M_{QLS} allows both linkage and association to contribute to the test statistic. A more relevant difference between the W_{SS} and M_{QLS} is that the M_{QLS} is computationally feasible in large pedigrees, whereas the W_{SS} is not. Improvement of the M_{QLS} over the W_{QLS} is obtained primarily by capitalizing on the property that there is an enrichment for predisposing variants in affected individuals with affected relatives. The M_{QLS} is remarkable in its computational simplicity and can be used for any set of individuals, regardless of the complexity of the relationships of the individuals.

We give a statistical argument that the M_{QLS} should be a powerful test: the M_{QLS} maximizes the noncentrality parameter over a general class of linear statistics for all two-allele (single-locus) disease models in outbred samples, as

the effect size tends to 0, where we allow environmental effects that do not have familial correlation (see the "Development and Justification of the M_{QLS} Test" section). We simulate various multilocus disease models and directly compare the type I error and the power of the M_{QLS} , W_{QLS} , and W_{corr} tests in samples of related individuals. Since the current implementation of the W_{SS} does not allow it to be applied to these particular simulated data sets, we instead use the true IBD-sharing information from the simulations to investigate the use of posterior kinship coefficients to compute the variances of any of the case-control statistics. We find that, even if correction for linkage is desired, the extra computation required to use posterior kinship coefficients is generally not worthwhile in small pedigrees and is not computationally practical in larger pedigrees. We apply our methods to the Genetic Analysis Workshop (GAW) 14 Collaborative Study of the Genetics of Alcoholism (COGA) data,⁶ to identify regions of the genome that are associated with alcohol dependence (MIM 103780).

Methods

Association Testing with a Biallelic Marker

Suppose that we have phenotype information about a binary trait for $n + m$ sampled individuals, with each individual coded as "affected," "unaffected," or "unknown." Consider a single biallelic marker with allele labels "0" and "1" (the extension to multiallelic markers is given in appendix A), and suppose that the first n of the $n + m$ listed individuals have nonmissing genotype data at the given marker, whereas the last m individuals have missing genotype data at the marker. For the first n individuals, let $Y = (Y_1, \dots, Y_n)^T$, where $Y_i = \frac{1}{2} \times$ (the number of alleles of type 1 in individual i), so the value of Y_i is 0, $\frac{1}{2}$, or 1. The $n + m$ individuals are assumed to be ascertained with respect to phenotype, and they may be arbitrarily related (including inbreeding), with the pedigree(s) that specify the relationships assumed to be known. Let p be the frequency of allele 1 in the general population, where $0 < p < 1$. Under the null hypothesis of no association between the given marker and the trait (and making the obvious assumption that ascertainment is conditionally independent of marker genotype, given phenotype and pedigree information), Y has mean $p\mathbf{1}$, where $\mathbf{1}$ is a column vector of length n with every entry equal to 1. To obtain the null variance of Y , more assumptions are required. Namely, we assume that, under the null hypothesis, (1) there is neither linkage nor association between the given marker and the trait and (2) the pedigree founders are drawn from a population in Hardy-Weinberg equilibrium (HWE) for the given marker. (Note that we do not require HWE in the founders under the alternative model.) In that case, the null variance is given by $\text{Var}_0(Y) = \Sigma = \frac{1}{2}p(1-p)\Phi$ (see, e.g., the work of Bourgain et al.³), where Φ is the kinship matrix of the nonmissing individuals, given by

$$\Phi = \begin{pmatrix} 1 + h_1 & 2\phi_{12} & \dots & 2\phi_{1n} \\ 2\phi_{12} & 1 + h_2 & \dots & 2\phi_{2n} \\ \vdots & \dots & \dots & \vdots \\ 2\phi_{n1} & 2\phi_{n2} & \dots & 1 + h_n \end{pmatrix}, \quad (1)$$

and h_i is the inbreeding coefficient of individual i , and ϕ_{ij} is the

Table 1. w_{test} , w_{null} , and $\text{Var}_0(\hat{p}_{\text{test}} - \hat{p}_{\text{null}})$ for $W_{\chi^2_{\text{corr}}}$, W_{SS} , W_{QLS} , and M_{QLS}

Statistic	w_{test}	w_{null}	$\text{Var}_0(\hat{p}_{\text{test}} - \hat{p}_{\text{null}})$
$W_{\chi^2_{\text{corr}}}$	$\mathbf{1}_c$	$\mathbf{1}$	$\frac{1}{2}p(1-p) \left(\frac{1}{n_c^2} \mathbf{1}'_c \Phi \mathbf{1}_c - 2 \frac{1}{nn_c} \mathbf{1}'_c \Phi \mathbf{1}_c + \frac{1}{n^2} \mathbf{1}' \Phi \mathbf{1} \right)$
W_{SS}	$\mathbf{1}_c$	$\mathbf{1}$	(See text)
W_{QLS}	$\Phi^{-1} \mathbf{1}_c$	$\Phi^{-1} \mathbf{1}$	$\frac{1}{2}p(1-p) [(\mathbf{1}'_c \Phi^{-1} \mathbf{1}_c) (\mathbf{1}'_c \Phi^{-1} \mathbf{1})^{-2} - (\mathbf{1}' \Phi^{-1} \mathbf{1})^{-1}]$
M_{QLS}	$A_N + \Phi^{-1} \Phi_{N,M} A_M$	$\Phi^{-1} \mathbf{1}$	$\frac{1}{2}p(1-p) [(w_{\text{test}}^T \Phi w_{\text{test}}) (w_{\text{test}}^T \mathbf{1})^{-2} - (\mathbf{1}' \Phi^{-1} \mathbf{1})^{-1}]$

kinship coefficient between individuals i and j , where $1 \leq i \leq n$ and $1 \leq j \leq n$.

Overview of the W_{QLS} , $W_{\chi^2_{\text{corr}}}$, W_{SS} , and M_{QLS} Tests

To our knowledge, it has not been previously pointed out that the case-control association test statistics W_{QLS} , $W_{\chi^2_{\text{corr}}}$, W_{SS} , and M_{QLS} can all be thought of as having the common form

$$\frac{(\hat{p}_{\text{test}} - \hat{p}_{\text{null}})^2}{\text{Var}_0(\hat{p}_{\text{test}} - \hat{p}_{\text{null}})}$$

where

$$\hat{p}_{\text{null}} = \frac{w_{\text{null}}^T Y}{w_{\text{null}}^T \mathbf{1}}$$

is an estimator of allele frequency calculated under the assumption of no association,

$$\hat{p}_{\text{test}} = \frac{w_{\text{test}}^T Y}{w_{\text{test}}^T \mathbf{1}}$$

is a contrasting estimator of allele frequency that should have a different expectation from \hat{p}_{null} when there is association, and $\text{Var}_0(\cdot)$ denotes variance calculated under the assumption that the null hypothesis is true. For each of the four statistics, w_{test} and w_{null} are column vectors of length n , which can be written as functions of the phenotype and pedigree information.

The W_{QLS} , $W_{\chi^2_{\text{corr}}}$, and W_{SS} tests assume that every genotyped individual is coded as either “case” or “control,” with no particular distinction made between unaffected controls and controls of unknown phenotype (e.g., general-population controls). Furthermore, these tests ignore the phenotype information on the additional m relatives with missing genotypes. In contrast, the M_{QLS} treats unaffected controls and controls of unknown phenotype differently and can handle samples that contain both types. It also takes into account the phenotype data for the additional m relatives with missing genotypes. For the W_{QLS} , $W_{\chi^2_{\text{corr}}}$, and W_{SS} , phenotype data are accordingly assumed to be coded as $\mathbf{1}_c$, a column vector of length n with i th entry 1 if individual i is a case and 0 if i is a control. We define $n_c = \mathbf{1}'^T \mathbf{1}_c$ as the number of cases among the n individuals with nonmissing genotypes. For the M_{QLS} test, the phenotype data are coded as A , a column vector of length $n + m$ having i th entry 1 if individual i is affected,

$$\frac{-k}{1-k}$$

if i is unaffected, and 0 if i is of unknown phenotype. Here $0 <$

$k < 1$ is a constant that must be specified (see the “Development and Justification of the M_{QLS} Test” section for details). Write $A^T = (A_N^T, A_M^T)$, where A_N and A_M are column vectors containing the first n elements and last m elements of A , respectively, where N and M stand for the sets of individuals with nonmissing and missing genotypes, respectively, at the particular marker. For the M_{QLS} test to use the additional phenotype information from the m individuals with missing genotype data, additional relationship information is needed in the form of an $n \times m$ matrix $\Phi_{N,M}$, giving the kinship coefficients between the nonmissing and missing individuals—that is, $\Phi_{N,M}$ has (i,j) th element equal to $2\phi_{i,n+j}$, where $\phi_{i,n+j}$ is the kinship coefficient between the i th nonmissing individual and the j th missing individual. We write $\Phi_{N,N \cup M}$ for the $n \times (n + m)$ matrix with first n columns equal to the Φ of equation (1) and last m columns equal to $\Phi_{N,M}$.

Table 1 gives w_{test} and w_{null} for $W_{\chi^2_{\text{corr}}}$, W_{SS} , W_{QLS} , and M_{QLS} , and $\text{Var}_0(\hat{p}_{\text{test}} - \hat{p}_{\text{null}})$ for $W_{\chi^2_{\text{corr}}}$, W_{QLS} , and M_{QLS} . When the W_{QLS} , $W_{\chi^2_{\text{corr}}}$, and M_{QLS} statistics are calculated from data, the p in the variance expression is estimated by $\hat{p}_{\text{null}} = w_{\text{null}}^T Y$. From table 1, we see that the weights and variances for W_{QLS} and M_{QLS} involve Φ^{-1} . It can be shown that Φ is invertible provided that the sample of individuals with nonmissing genotypes does not include both members of any MZ twin pair (see appendix B). We have also developed an extension of the W_{QLS} and M_{QLS} to the case when the genotyped sample does include both members of one or more MZ twin pairs. In that case, we calculate the W_{QLS} and M_{QLS} by coding each MZ twin pair as a single individual, where, for the W_{QLS} , we assign the phenotype to be the average of the phenotype values for the pair (so the phenotype would take value 0, 1, or $\frac{1}{2}$ if the members of the pair were both controls, both cases, or one of each, respectively) and, for the M_{QLS} , we assign the phenotype value to be the sum of the phenotype values for the MZ twin pair (so the phenotype would take value 0,

$$\frac{-2k}{1-k},$$

or 2 if the members of the pair were both of unknown phenotype, both unaffected, or both affected, respectively, and value

$$1 - \frac{k}{1-k}$$

if one is affected and the other unaffected, etc.). It turns out that this procedure is justified by the fact that it is mathematically equivalent to substituting the Moore-Penrose generalized inverse (see, e.g., the work of Schott⁷) of Φ in place of Φ^{-1} in the formulas of table 1. Under regularity conditions on how $\Phi_{N,N \cup M}$ behaves as $n + m \rightarrow \infty$, which are sufficient for a central-limit theorem to hold, each of the four statistics follows a χ^2 distribution with 1 df as-

ymptotically under the null hypothesis. A previous simulation study documented the accuracy of the χ^2 approximation for type I error of $W_{\chi^2_{\text{corr}}}$ and W_{QLS} .³

For both the $W_{\chi^2_{\text{corr}}}$ and W_{SS} , \hat{p}_{test} is the sample mean based on cases and \hat{p}_{null} is the sample mean based on everyone. The distinction between the two statistics is in the variance calculation. Previous work⁴ gives the variance calculation for W_{SS} , for the case of outbred individuals, as a function of variances and covariances of genotype indicators for sampled individuals. The variances of the indicators are then calculated as a function of genotype probabilities and are valid under violations of HWE. However, the covariances are calculated with use of an ITO method that is valid only under HWE and with use of posterior IBD probabilities that are also calculated assuming HWE. Arguably the most substantial distinction between the variance calculations for $W_{\chi^2_{\text{corr}}}$ and W_{SS} is in the use of so-called prior versus posterior kinship coefficients or, equivalently, the probability of IBD sharing for a pair based on the pedigree alone versus conditional on genotype data as well as on the pedigree. When the data are in HWE and all genotyped individuals are outbred, the variance expression for W_{SS} is the same as that for $W_{\chi^2_{\text{corr}}}$ in table 1, except that Φ is replaced by $\Phi_{\text{posterior}}$ where $\Phi_{\text{posterior}}$ has (i, j) th element equal to

$$\xi_{ij}(2) + \frac{\xi_{ij}(1)}{2},$$

where $\xi_{ij}(k) = P(i \text{ and } j \text{ share } k \text{ alleles IBD at the given marker} | \text{all genotype and pedigree information})$. We call $\Phi_{\text{posterior}}$ the posterior kinship matrix, and, in the “Assessment of Use of Φ versus $\Phi_{\text{posterior}}$ in Variance Calculations” section, we consider the effect on power of the use of Φ versus $\Phi_{\text{posterior}}$ in the variance calculations for each of the statistics.

For W_{QLS} , \hat{p}_{null} is the best linear unbiased estimator (BLUE)⁸ based on everyone, and, in the special situation when the controls are not related to the cases, \hat{p}_{test} is the BLUE based on cases. The original presentation³ of the W_{QLS} was as a quasi-likelihood score statistic based on the mean model

$$E(Y | \mathbf{1}_c) = \mu = (\mu_1, \dots, \mu_i, \dots, \mu_N)^T,$$

with $\mu = p\mathbf{1} + r\mathbf{1}_c$ —that is,

$$\mu_i = \begin{cases} p + r & \text{if } i \text{ is a case} \\ p & \text{if } i \text{ is a control} \end{cases}, \quad (2)$$

where $0 < p + r < 1$. Under the null hypothesis of no association, $r = 0$. An equivalent expression³ for W_{QLS} is

$$W_{\text{QLS}} = (Y - \hat{\mu}_0)^T \hat{\Sigma}_0^{-1} \mathbf{1}_c \left[\mathbf{1}_c^T \hat{\Sigma}_0^{-1} \mathbf{1}_c \right. \\ \left. - \mathbf{1}_c^T \hat{\Sigma}_0^{-1} (\mathbf{1}^T \hat{\Sigma}_0^{-1} \mathbf{1})^{-1} \mathbf{1}^T \hat{\Sigma}_0^{-1} \mathbf{1}_c \right]^{-1} \mathbf{1}_c^T \hat{\Sigma}_0^{-1} (Y - \hat{\mu}_0), \quad (3)$$

where $\hat{p}_{\text{null}} = (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Phi^{-1} Y$ is the quasi-likelihood estimator of p when $r = 0$, $\hat{\Sigma}_0$ is the covariance matrix Σ evaluated at $r = 0$ and $p = \hat{p}_{\text{null}}$ —that is, $\hat{\Sigma}_0 = \frac{1}{2} \hat{p}_{\text{null}} (\mathbf{1} - \hat{p}_{\text{null}}) \Phi$ —and $\hat{\mu}_0$ is the mean vector μ evaluated at $r = 0$ and $p = \hat{p}_{\text{null}}$ —that is, $\hat{\mu}_0 = \hat{p}_{\text{null}} \mathbf{1}$. The W_{QLS} test has been shown³ to have certain optimality properties based on this model.

For the M_{QLS} , \hat{p}_{null} is the BLUE based on everyone. In the special case of no missing genotype data at the marker, \hat{p}_{test} for the M_{QLS} is calculated by taking a weighted average of the elements of Y

in which each affected individual’s Y value gets a weight of 1, each unaffected individual’s Y value gets a weight of

$$\frac{-k}{1 - k},$$

and Y values of individuals of unknown phenotype get weight of 0. When some individuals have missing genotype data, then $w_{\text{test}, i}$ the weight of the i th typed individual’s Y value in \hat{p}_{test} has an additional term added on, whose value is a function of the phenotypes of individuals with missing genotype data that occur in the same pedigree with individual i . In that case, $w_{\text{test}, i} = A_i + (\Phi_f^{-1} \Phi_{f, N, M} A_{f, M})_i$ where f is the pedigree containing individual i , and Φ_f^{-1} , $\Phi_{f, N, M}$ and $A_{f, M}$ are, respectively, Φ^{-1} , $\Phi_{N, M}$ and A_M restricted to members of f . In the “Development and Justification of the M_{QLS} Test” section, we justify the form of the M_{QLS} , given in table 1, on theoretical grounds, and, in the “Power Comparison of $W_{\chi^2_{\text{corr}}}$, W_{QLS} , and M_{QLS} ” and the “Assessment of Use of Φ versus $\Phi_{\text{posterior}}$ in Variance Calculations” sections, we demonstrate that it generally improves power over previously proposed tests.

Development and Justification of the M_{QLS} Test

The model on which W_{QLS} is based (eq. [2]) has the advantage of being simple and intuitive, and it works well in samples of unrelated individuals. However, when the sample consists of related individuals and the trait is complex, this model does not capture certain features of allele-frequency differences associated with a genetic trait. Situations in which the model in equation (2) would be expected to hold more or less exactly, even in a sample of related individuals, include (1) testing for an allele-frequency difference between two distinct populations (e.g., Swedes and Japanese) without admixture and (2) testing for association with a trait within a single population when the true genetic model is a rare, fully penetrant dominant allele. In such cases, the W_{QLS} enjoys certain optimality properties,³ which can be verified by simulation (see the “Results” section). One might hope that the simple model would be robust enough to maintain power with complex traits, but we find that power can be improved substantially by modifying the model while still retaining the computational simplicity of the original method.

The motivation for development of the M_{QLS} is to improve the power of the quasi-likelihood score test W_{QLS} by explicitly taking into account the fact that affected individuals who have affected relatives have a higher expected frequency of the alleles that increase susceptibility for a genetic trait than do individuals who do not have affected relatives. The simple model in equation (2) ignores this fact. Furthermore, when case individual i and control individual j are related, the model in equation (2) specifies that $E(Y_i - Y_j) = r$, regardless of how closely related i and j are, whereas, for example, if i and j were MZ twins, then, in reality, $Y_i - Y_j = 0$, and, more generally, the expected difference in Y values between two individuals depends on their relationship as well as on their phenotypes.

For the M_{QLS} , we develop a new mean model that is a function of the relationships and the phenotypes of all $n + m$ individuals and is as follows: $E(Y | A) = \mu = (\mu_1, \dots, \mu_i, \dots, \mu_n)^T$, with $\mu = p\mathbf{1} + r\Phi_{N, N, M} A$ —that is,

$$\mu_i = p + r \sum_{j=1}^{n+m} 2\phi_{ij} A_j, \quad (4)$$

where we constrain $0 < p + r(\Phi_{N,N \cup M}A)_i < 1$ for $i = 1, \dots, n$. For a better understanding of this mean model, let us consider, for example, a case-control study in which all of the individuals in the study are outbred. If a sampled individual i has no relatives in the study or if the phenotypes of all of i 's relatives are unknown, then, under the model of equation (4), i 's expected allele frequency is $\mu_i = p$ if i is of unknown phenotype, $\mu_i = p + r$ if i is affected, and

$$\mu_i = p - \left(\frac{k}{1-k}\right)r$$

if i is unaffected. For each affected sibling i has, $\frac{1}{2}r$ is added to the baseline value of μ_i , whereas, for each unaffected sibling,

$$-\frac{1}{2}\left(\frac{k}{1-k}\right)r$$

is added. For example, if i is affected and has three affected siblings and no other relatives of known phenotype in the study, then $\mu_i = p + \frac{5}{2}r$. Note that, under model (4), MZ twins would always have the same expected allele frequency, even if one is a case and the other is a control.

The theoretical justification for this new mean model is as follows: assume, for the moment, that the trait is caused by the marker according to an arbitrary two-allele (single-locus) disease model, in which we allow environmental effects that do not have familial correlation. This model is specified by the population-allele frequency p and the penetrances of the three possible genotypes. Set the constant k (used in the calculation of A) to be K_p , the population prevalence of the trait. Consider an arbitrary set of possibly related outbred individuals, and calculate the true value, under the model, of the expected allele frequency in individual i conditional on all available phenotypes of i and i 's relatives, $\mu_i^* = E(Y_i|A)$, for any individual i in the sample. If we consider the ratio

$$\frac{\mu_i^* - p}{\mu_a^* - p},$$

where a is any affected individual with no phenotyped relatives, then, as the effect size (or differences among penetrance probabilities) tend to zero, this ratio tends to $\sum_{j=1}^{n+m} 2\phi_{ij}A_j$ (see appendix C for the proof). Thus, the mean model in equation (4) is asymptotically the correct one for any two-allele disease model as the effect size goes to zero. Moreover, if the individuals are inbred, then the model is asymptotically correct for an additive or multiplicative, two-allele disease model as the effect size goes to zero (but not for a general, two-allele disease model in the inbred case). Although the assumptions under which this mean model is derived are somewhat simplistic, the model captures the important feature that individuals with affected relatives are likely to be enriched for the predisposing allele relative to individuals without affected relatives.

The M_{QLS} statistic given in table 1 is derived as the quasi-likelihood score statistic based on the model in equation (4). This is obtained by substituting $\Phi_{N,N \cup M}A$ in place of $\mathbf{1}_c$ in equation (3). The resulting formula for M_{QLS} is

$$M_{QLS} = \hat{\sigma}_0^{-2}(Y - \hat{\mu}_0)^T \alpha \Gamma^{-1} \alpha^T (Y - \hat{\mu}_0), \quad (5)$$

where $\alpha = A_N + \Phi^{-1}\Phi_{N,M}A_M$,

$$\Gamma = \alpha^T(\Phi A_N + \Phi_{N,M}A_M) - (\mathbf{1}^T \alpha)^2 (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1}, \quad (6)$$

$\hat{\mu}_0 = \hat{p}_{\text{null}} \mathbf{1}$, $\hat{\sigma}_0^{-2} = [\frac{1}{2} \hat{p}_{\text{null}} (1 - \hat{p}_{\text{null}})]^{-1}$, and $\hat{p}_{\text{null}} = (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Phi^{-1} Y$. Following the same reasoning³ as for the W_{QLS} , the M_{QLS} has maximal noncentrality parameter, against the alternative specified in equation (4), among a general class of linear statistics of the form $W = S^T [\text{Var}_0(S)]^{-1} S$, where $S = V^T Y$ and $V \neq \mathbf{0}$ where $\mathbf{0}$ is a column vector of 0's of length n . Note that $W_{x_{\text{corr}}^2}$, W_{QLS} , and M_{QLS} are all of this form. As a result, under suitable regularity conditions, which are not discussed here, the M_{QLS} would be asymptotically locally most powerful against the alternative specified in equation (4). Simulation studies are undertaken in the "Results" section, to assess the usefulness of this test for complex traits, where two-allele models are not expected to hold.

To use the M_{QLS} test, a value for the constant k must be specified. We emphasize that the test will be valid for any value of k satisfying $0 < k < 1$. The value of k affects the power of the test, and, under a two-allele model, for outbred individuals, optimal power is attained when k is the population prevalence of the trait. For complex traits, we recommend setting k equal to an estimate of the population prevalence from previous studies or registry data from the population. In the "Results" section, we demonstrate, through simulation, that power is in fact quite robust to misspecification of k .

M_{QLS} Power-Improvement Diagnostic

We propose a method that uses only pedigree information and phenotype data (without genotype data) to determine whether the analyses using $W_{x_{\text{corr}}^2}$, W_{QLS} , and M_{QLS} would be expected to give similar or dissimilar results. (If they are expected to give dissimilar results, then, in most cases, the M_{QLS} is expected to have higher power.) One possible advantage of applying this diagnostic could be to avoid having to correct association P values for having performed multiple analyses unnecessarily, in the case when the results are predicted to be similar on the basis of the diagnostic. Because the diagnostic is based only on phenotype and pedigree information, and not on genotype data, such an approach does not create any bias in the results.

The idea is to calculate the weights assigned to the observations for each of the statistics and compare them. For the $W_{x_{\text{corr}}^2}$, the weight of an observation depends only on the individual's case-control status. In contrast, the weights can vary among cases and among controls in W_{QLS} and M_{QLS} , depending on the relationship configurations, as well as on the phenotypes of relatives. The total weights vary slightly from locus to locus, depending on missing data patterns. However, on the basis of the study design, weights for the genotyped individuals could be computed to get an idea of how different the analyses are likely to be under W_{QLS} , M_{QLS} , and $W_{x_{\text{corr}}^2}$. Details of the power improvement diagnostic are given in appendix D, where we find that, to compare the $W_{x_{\text{corr}}^2}$ weights with the M_{QLS} weights, we need only calculate the coefficients of variation of total M_{QLS} weights among cases and total M_{QLS} weights among controls. If these values are close to zero, then the M_{QLS} test results should be similar to the $W_{x_{\text{corr}}^2}$ test results. The same principle holds for comparison of $W_{x_{\text{corr}}^2}$ weights with W_{QLS} weights. Large differences in association-testing results between the $W_{x_{\text{corr}}^2}$ and either the M_{QLS} or the W_{QLS} tests would be expected to occur when one or both of the relevant coefficients of variation are far from zero. Our experience, in the context of our simulations, is that a coefficient of variation of ~ 1 or more in absolute

value for the cases and/or controls will result in an improvement in power for the M_{QLS} over the W_{Xcorr} .

Use of Prior versus Posterior Kinship Coefficients

Kinship coefficients (and, for an inbred population, inbreeding coefficients) are used in two different ways in the construction of the four statistics: W_{Xcorr} , W_{SS} , W_{QLS} , and M_{QLS} . First, kinship coefficients can be used in the calculation of the total weight vector, V , for each statistic (as they are in W_{QLS} and M_{QLS} but not in W_{Xcorr} and W_{SS}). Second, they are used in the calculation of the variance term $\text{Var}_0(V^T Y) = \text{Var}_0(\hat{p}_{\text{test}} - \hat{p}_{\text{null}})$ for each statistic. The question arises^{3,5} as to whether it is preferable to use the unconditional (i.e., prior) kinship coefficients,³ given by Φ , or the posterior kinship coefficients, $\Phi_{\text{posterior}}$, which are the conditional probabilities of IBD sharing given both the observed genotype data and the pedigree information,⁴ keeping in mind that Φ is vastly easier to compute than $\Phi_{\text{posterior}}$. In fact, we generally caution against the use of $\Phi_{\text{posterior}}$ in the calculation of V for W_{QLS} and M_{QLS} . To understand why, first consider the simpler problem of allele-frequency estimation from a sample of related individuals. The BLUE for that problem⁸ is the same as

$$\hat{p}_{\text{null}} = (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Phi^{-1} Y.$$

When $\Phi_{\text{posterior}}$ is substituted for Φ in the formula for \hat{p}_{null} , the resulting estimator is, in general, biased and even inconsistent as $n \rightarrow \infty$ when there is partial IBD information. This occurs because the random variables $\Phi_{\text{posterior}}$ and Y are, in general, dependent. (Note that, in principle, these difficulties could be avoided in certain situations when $\Phi_{\text{posterior}}$ is estimated in such a way that $E(Y|\Phi_{\text{posterior}}) = EY$; for example, if the true IBD sharing is known or if it is estimated from markers not in linkage disequilibrium [LD] with the marker of interest.) From these considerations, it seems clear that, if $\Phi_{\text{posterior}}$ is used in the calculation of V for the W_{QLS} and M_{QLS} , the resulting test statistic could be badly behaved. This might explain the difficulties encountered in previous work,⁵ which considered the use of $\Phi_{\text{posterior}}$ in the calculation of V for a haplotype-based method with a weighting scheme similar to that of the W_{QLS} .

There remains the possibility of using $\Phi_{\text{posterior}}$ in the calculation of the variance term. Intuitively, this would be expected to correct the statistic for the presence of linkage.⁵ In a context in which one would be willing to detect either linkage or association or a combination of the two, then the use of $\Phi_{\text{posterior}}$ instead of Φ in the variance calculation would be expected to result in lower power, as well as more onerous computation. However, in a context in which linkage has already been established and one wishes to correct for it in testing for association, the use of $\Phi_{\text{posterior}}$ instead of Φ in the variance calculation would be expected to result in better control of type I error. In the "Assessment of Use of Φ versus $\Phi_{\text{posterior}}$ in Variance Calculations" section, we describe simulations to test the validity of this intuition.

To use $\Phi_{\text{posterior}}$ in the variance calculation for W_{Xcorr} , one need only replace Φ with $\Phi_{\text{posterior}}$ in the variance formula in table 1. We call the resulting statistic $W_{\text{Xcorr}}^{\text{COND}}$, where "COND" refers to

conditional kinship coefficients. To calculate $W_{\text{QLS}}^{\text{COND}}$ and $M_{\text{QLS}}^{\text{COND}}$, we replace the corresponding variance formulas in table 1 with

$$\begin{aligned} & \frac{1}{2} p(1-p) [(w_{\text{test}}^T \Phi_{\text{posterior}} w_{\text{test}}) \\ & \times (w_{\text{test}}^T \mathbf{1})^{-2} - 2(w_{\text{test}}^T \Phi_{\text{posterior}} \Phi^{-1} \mathbf{1}) \\ & \times (w_{\text{test}}^T \mathbf{1})^{-1} (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1} \\ & + (\mathbf{1}^T \Phi^{-1} \Phi_{\text{posterior}} \Phi^{-1} \mathbf{1}) (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-2}], \end{aligned}$$

where w_{test} is defined differently for W_{QLS} and M_{QLS} , with the definitions given in table 1. Software such as Merlin⁹ can be used to calculate $\Phi_{\text{posterior}}$ in small-to-moderate-sized pedigrees.

GAW 14 COGA Data

We analyze a COGA data set⁶ that was previously analyzed in the Genetics Analysis Workshop (GAW) 14. There are a total of 1,614 individuals from 143 pedigrees, with each pedigree containing at least three affected individuals. We include in our analysis only those individuals who are coded as "white, non-Hispanic." We designate as cases those individuals who are affected with ALDX1 or who have symptoms of ALDX1, where ALDX1 is defined to be DSM-III-R alcohol dependence with the Feighner Alc Definite phenotype. By these criteria, there are 830 cases with available SNP data. We designate as "unaffected controls" those individuals who are labeled as "pure unaffected," and we designate as "controls of unknown phenotype" those individuals who are labeled as "never drank alcohol." Among individuals with available SNP data, these criteria result in 187 unaffected controls and 13 unknown controls. Note that the M_{QLS} makes a distinction between the two control types, whereas the W_{Xcorr} and W_{QLS} do not. The data set includes 10,810 autosomal SNPs. We exclude 403 SNPs that are not polymorphic (minor-allele frequency <0.01). We analyze the remaining 10,407 SNPs using the W_{Xcorr} , W_{QLS} , and M_{QLS} tests. We could not use the W_{SS} software package to analyze these data because, at the time of our analysis, to the best of our knowledge, there was no implementation available that would handle the situation in which controls are related to cases (although, in principle, it could be extended to that situation).

Results

Simulation Studies

We perform simulation studies to (1) assess the type I error of the M_{QLS} ; (2) compare power of the W_{Xcorr} , W_{QLS} , and M_{QLS} ; (3) assess the practical impact of the use of Φ versus $\Phi_{\text{posterior}}$ in the variance calculations for the statistics; and (4) assess the robustness of power of M_{QLS} to the choice of the parameter k . For each simulation described below, 5,000 replicates were performed.

We consider three different study designs. In the first, affected and unaffected individuals from 60 outbred, three-generation pedigrees are sampled. Each pedigree has a total of 16 individuals, related as in figure 1, with the pattern of affected and unaffected individuals varying randomly according to one of the trait models described in the next paragraph. Pedigrees are sampled conditional on obtaining exactly 20 pedigrees with 4 affected individuals, 20 with 5, and 20 with 6. In each sampled pedigree, pheno-

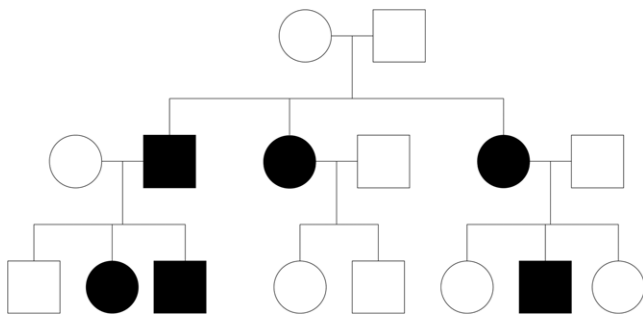


Figure 1. Example pedigree for study design 1 consisting of 60 outbred, three-generation pedigrees, where the overall structure of each pedigree is as depicted, but the pattern of affected and unaffected individuals in each pedigree varies randomly according to the specified model and ascertainment scheme.

types for all 16 individuals are observed. For each individual in a sampled pedigree, the individual's genotypes are observed if and only if at least 30% of the individual's siblings, parents, and offspring in the sampled pedigree are affected. The second study design is similar to the first, with two differences: (1) an additional 200 unrelated, unaffected controls are included in the study; and (2) for each individual in a sampled pedigree, the individual's genotypes are observed if and only if at least half of the individual's siblings, parents, and offspring in the sampled pedigree are affected. In the third study design, individuals from three extended pedigrees are sampled, with each pedigree consisting of 154 individuals over five generations. The pedigrees are sampled conditional on having at least 50 affected individuals. In each sampled pedigree, phenotypes for all 154 individuals are observed. For each individual in a sampled pedigree, the individual's genotypes are observed if and only if at least half of the individual's siblings, parents, and offspring in the sampled pedigree are affected.

We consider four different classes of multigene trait models.¹⁰ Model I has two unlinked causal SNPs, with epistasis between them and both of them acting dominantly. In model I, the frequencies of allele 1 at SNPs 1 and 2 are p_1 and p_2 , respectively. Individuals with at least one copy of allele 1 at SNP 1 and at least one copy of allele 1 at SNP 2 have a penetrance of f_1 . All other individuals have a penetrance of $f_2 < f_1$. We consider three different parameter settings for model I, which are listed as models I-a, I-b, and I-c in table 2. Model II also consists of two unlinked causal SNPs with epistasis between them, with SNP 1 acting recessively and SNP 2 following a general two-allele model. There are four penetrance parameters for this model, with $f_1 > f_2 > f_3 > f_4$. Individuals with two copies of allele 1 at SNP 1 and two copies of allele 1 at SNP 2 have a penetrance of f_1 . Individuals with two copies of allele 1 at SNP 1 and one copy of allele 1 at SNP 2 have a penetrance of f_2 . Individuals with two copies of allele 1 at SNP 1 and no copies of allele 1 at SNP 2 have a penetrance of

f_3 . All other individuals have a penetrance of f_4 . We consider one parameter setting for this class of model, which is listed as model II-a in table 2. Model III has three unlinked causal SNPs with epistasis between them and with each SNP acting dominantly. Individuals with at least one copy of allele 1 at SNP 1 and at least one copy of allele 1 at SNP 2 and/or SNP 3 have a penetrance of f_1 . All other individuals have a penetrance of $f_2 < f_1$. We consider two different parameter settings for this class of model, which are listed as models III-a and III-b in table 2. Model V is the same as model I except that, in model V, the two causal SNPs are tightly linked and in linkage equilibrium, whereas, in model I, the two causal SNPs are unlinked and in linkage equilibrium. We consider one parameter setting for model V, which is listed as model V-a in table 2. In addition to the multigene models, we also consider a single-gene dominant model, model IV, in which individuals with at least one copy of allele 1 at SNP 1 have a penetrance of f_1 and all other individuals have a penetrance of $f_2 < f_1$. We consider one parameter setting for model IV, which is listed as model IV-a in table 2 and which represents a rare dominant trait that is almost fully penetrant. In addition to the allele frequencies and penetrance parameters for each model, table 2 contains the resulting population prevalence K_p , prevalence K_s conditioned on having an affected sibling, and the sibling risk ratio

$$\lambda_s = \frac{K_s}{K_p};$$

these last 3 are calculated in outbreds. A broad range of models were chosen for our simulation studies: from highly penetrant disease models to disease models with low penetrance and models with high heritability to models with low heritability.

Table 2. Allele Frequencies and Penetrance Parameters for Simulation Models

Model	Allele Frequencies			Penetrance Parameters						
	p_1	p_2	p_3	f_1	f_2	f_3	f_4	K_p	K_s	λ_s
I-a	.1	.52	.05071	.087	1.224
I-b	.5	.418	.07122	.132	1.077
I-c	.4	.254	.2256	.268	1.047
II-a	.3	.25	.45	.32042	.135	3.198
III-a	.1	.2	.3	.4	.03078	.156	2.001
III-b	.05	.4	.2	.5	.01047	.196	4.209
IV-a	.05	1	.05143	.417	2.926
V-a	.1	.36	.15194	.232	1.202

NOTE.—I, II, III, IV, or V indicates the model type, and the letter (a, b, or c) that follows the model type is used to distinguish among the different settings of allele frequencies and penetrance parameters used for that model type. p_i is the frequency of allele 1 at SNP i , K_p is the population prevalence of the disease, K_s is the probability that a non-inbred individual is affected, given that the individual has a sibling who is affected, and $\lambda_s = \frac{K_s}{K_p}$ is the sibling risk ratio.

Table 3. Empirical Type I Error of the M_{QLS} Test, Based on 5,000 Simulated Replicates

p	N_{cases}	$N_{controls}$	Empirical Type I Error (SE) with Nominal Type I Error of	
			.05	.01
.4	77.2	405.5	.050 (.003)	.011 (.0015)
.2	77.0	405.9	.052 (.003)	.010 (.0014)
.05	76.9	405.4	.053 (.003)	.010 (.0014)

NOTE.—Association is tested with a biallelic marker having minor-allele frequency p . N_{cases} is the average number of cases sampled. $N_{controls}$ is the average number of controls sampled.

Assessment of Type I Error of M_{QLS}

Previous simulation studies³ verified that the use of the χ^2 approximations to the null distributions of $W_{\chi^2_{corr}}$ and W_{QLS} give the appropriate type I error for the tests. We perform a similar verification for the M_{QLS} by simulating under the null hypothesis of no association and no linkage. We compare the proportion of simulations in which the statistic exceeds the $(1 - \alpha)$ th quantile of the χ^2 distribution to the nominal type I error level α , for $\alpha = .01$ and $.05$. Simulations are performed on the basis of the second study design, which consists of individuals from 60 moderate-sized pedigrees plus 200 unrelated unaffected individuals. The phenotype is simulated from model I-a (table 2). We test at an unlinked, unassociated SNP with three different allele-frequency settings, which are given in table 3.

Table 3 gives the empirical type I error of the M_{QLS} , estimated from 5,000 simulations, for nominal levels $.05$ and $.01$. For each simulation scenario, the empirical type I error is not significantly different from the nominal. These results verify that the use of the χ^2 approximation results in an accurate assessment of significance for the M_{QLS} .

Power Comparison of $W_{\chi^2_{corr}}$, W_{QLS} , and M_{QLS}

To compare the power of $W_{\chi^2_{corr}}$, W_{QLS} , and M_{QLS} , we perform simulations on the basis of the second study design, which consists of individuals from 60 moderate-sized pedigrees plus 200 unrelated unaffected individuals. Five thousand replicates from each of models I-a, I-b, I-c, II-a, III-a, and III-b are simulated. The test is performed at SNP 2 for each model, with the significance threshold set to $.05$.

Estimated power with the SE for the $W_{\chi^2_{corr}}$, W_{QLS} , and M_{QLS} tests is given in figure 2. Recall that the numbers of cases and controls in each replicate are randomly determined. The average number of cases for a given simulation setting has a range of 70.2–121.6, and the average number of controls has a range of 390.2–417.4. The average coefficient of variation for the total weights of cases in the M_{QLS} has a range of 0.3–0.4, and, for controls, it has a range from -2.6 to -1.6 . Because the average coefficient of variation of total weights of controls is >1 in absolute value for the M_{QLS} for every model considered, we expect an improvement in power of the M_{QLS} over the $W_{\chi^2_{corr}}$. As shown

in figure 2, the M_{QLS} is more powerful than both the $W_{\chi^2_{corr}}$ and W_{QLS} in our simulation studies. The increase in power for the M_{QLS} is substantial (a difference in power of at least 0.20) for models I-a, I-b, and III-a. Our theoretical results indicate that the M_{QLS} should be powerful for two-allele disease models as effect size tends to zero. Our simulation studies indicate that, in fact, the M_{QLS} performs well for a range of more-complex disease models.

In the “Development and Justification of the M_{QLS} Test” section, we note that an example in which the W_{QLS} has certain theoretical optimality properties is the case of a rare, fully penetrant dominant trait. To demonstrate by simulation that there can be cases in which the W_{QLS} has higher power than the other statistics, we perform simulations that are based on the first study design (60 moderate-sized pedigrees) with the phenotype simulated from model IV-a, which approximates a rare, fully penetrant dominant. We perform the tests at the causal SNP and also at a tightly linked SNP that has allele frequency 0.5 and is associated with the causal SNP, with $D' = .2$. When the causal SNP was tested, all three tests had power close to 1 (results not shown). Table 4 compares the power of the three tests when they are performed at the tightly linked SNP that has allele frequency 0.5 and is associated with the causal SNP with $D' = .2$. As expected, W_{QLS} has higher power than the other statistics in the case of a rare, fully penetrant dominant. This is because the conditional expected frequency of the allele in an individual given the phenotype information on everyone depends on only whether the individual is affected or unaffected, so the model on which the W_{QLS} is based holds.

Assessment of Use of Φ versus $\Phi_{posterior}$ in Variance Calculations

We assess by simulation two predicted consequences of the use of $\Phi_{posterior}$ in place of Φ in the variance calculations of the statistics. To avoid the extra burden of calculating $\Phi_{posterior}$, we instead use the true IBD-sharing information in place of $\Phi_{posterior}$. This corresponds to the best-case scenario for the use of $\Phi_{posterior}$, in which the markers provide complete IBD information.

The first predicted consequence is that use of $\Phi_{posterior}$ in

Table 4. Example in Which W_{QLS} Is More Powerful than $W_{\chi^2_{corr}}$ and M_{QLS}

Statistic	Estimated Power (SE)
W_{QLS}	.47 (.007)
$W_{\chi^2_{corr}}$.37 (.007)
M_{QLS}	.36 (.007)

NOTE.—The tested SNP is associated ($D' = .2$) with the causal SNP for a rare, fully penetrant dominant trait (model IV-a). Power is based on 5,000 simulated replicates.

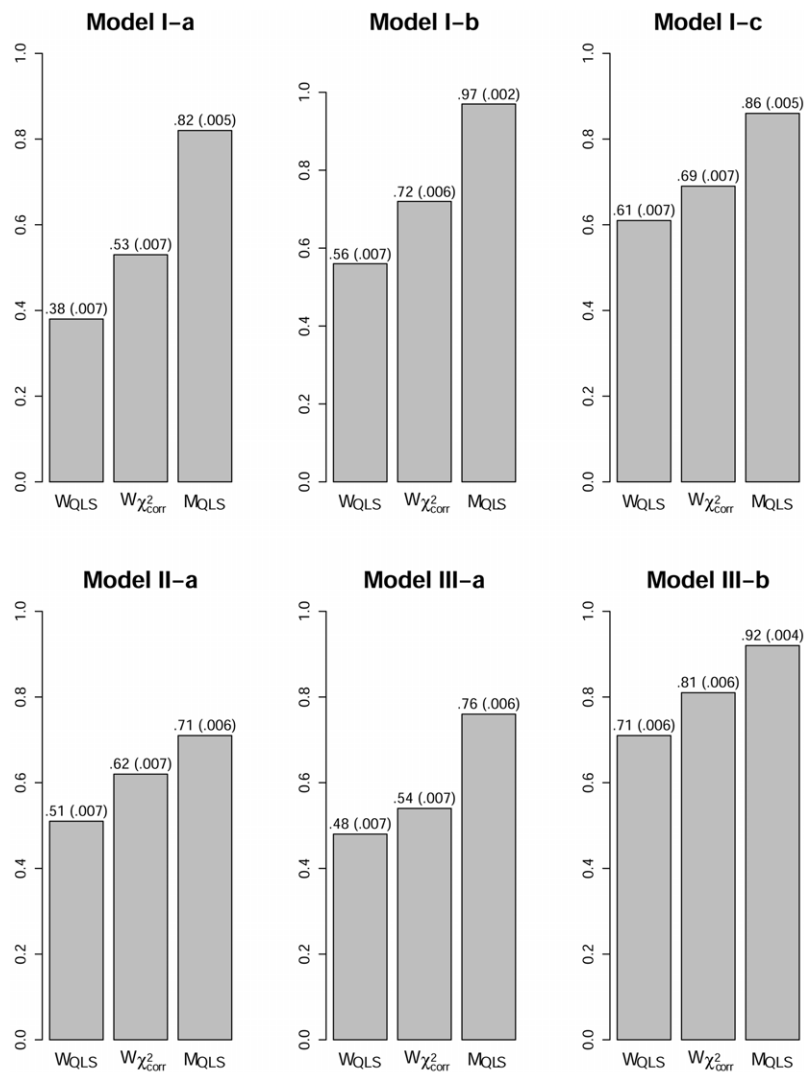


Figure 2. Estimated power with SE for the W_{QLS} , $W_{\chi^2_{corr}}$ and M_{QLS} tests, on the basis of 5,000 simulated realizations for each of six different models.

place of Φ in the variance calculations would result in lower power in a context in which one would be willing to detect either linkage or association or a combination of the two; that is, when the null hypothesis is no association and no linkage. To test this, we simulate on the basis of the first (60 moderate-sized pedigrees) and third (three extended pedigrees) study designs, with the phenotype simulated from model V-a. We perform the tests at SNP 2, which is both linked and associated with the phenotype.

The second predicted consequence is that use of $\Phi_{posterior}$ would give better control of type I error in the context in which linkage has already been established and one wishes to correct for it in testing for association; that is, when the null hypothesis is no association. To test this, we use the same simulation scenario as for the first predicted consequence except that, instead of testing at SNP

2, which is both linked and associated with the phenotype, we tested at SNP 3, with allele frequency 0.5, which is tightly linked to both SNPs 1 and 2 but is not associated with either of them.

Table 5 demonstrates that, for moderate-size pedigrees, there is almost no difference between the use of Φ and $\Phi_{posterior}$ in the variance calculations for the statistics. This lack of difference between the two approaches holds for a SNP that is linked but not associated, as well as for a SNP that is both linked and associated. This is good news, because the calculation of Φ is vastly simpler than that of $\Phi_{posterior}$.

For extended pedigrees, on the other hand, table 5 demonstrates that our predictions were correct—namely, that use of $\Phi_{posterior}$ in place of Φ in the variance calculations results in (1) lower power to test the joint null hypothesis of no association and no linkage and (2) better control of

Table 5. Power Comparison for the Use of Φ versus $\Phi_{\text{posterior}}$ in Variance Calculations, for Simulations under Model V-a, When Testing at a Tightly Linked Marker, Based on 5,000 Simulated Replicates

Setting	Associated	Pedigree Size	Estimated Power (SE)					
			W_{QLS}	$W_{\text{QLS}}^{\text{COND}}$	W_{Xcorr}^2	$W_{\text{Xcorr}}^{\text{COND}}$	M_{QLS}	$M_{\text{QLS}}^{\text{COND}}$
1	Yes	Large	.78 (.006)	.64 (.007)	.84 (.005)	.73 (.006)	.96 (.003)	.88 (.005)
2	Yes	Moderate	.86 (.005)	.84 (.005)	.94 (.003)	.93 (.004)	.96 (.003)	.96 (.003)
3	No	Large	.13 (.005)	.05 (.003)	.11 (.004)	.05 (.003)	.14 (.005)	.05 (.003)
4	No	Moderate	.06 (.003)	.05 (.003)	.05 (.003)	.05 (.003)	.05 (.003)	.05 (.003)

NOTE.— $W_{\text{QLS}}^{\text{COND}}$, $W_{\text{Xcorr}}^{\text{COND}}$, and $M_{\text{QLS}}^{\text{COND}}$ are the statistics that use $\Phi_{\text{posterior}}$ for the variance calculation. W_{QLS} , W_{Xcorr}^2 , and M_{QLS} are the statistics that use Φ for the variance calculation. “Large” refers to the study design consisting of three extended pedigrees, each containing 154 individuals. “Moderate” refers to the study design consisting of 60 pedigrees, each containing 16 individuals. The average numbers of cases and controls for the three large extended pedigrees study are 70.4 and 106.3, respectively. The average numbers of cases and controls for the 60 moderate-sized pedigrees study are 164.9 and 360.5, respectively.

type I error to test the null hypothesis of no association in the presence of linkage. This is because, when Φ is used in the variance calculation, linkage is allowed to contribute to the signal, whereas, when $\Phi_{\text{posterior}}$ is used in the variance calculation, linkage is not allowed to contribute to the signal. In extended pedigrees, the calculation of $\Phi_{\text{posterior}}$ can present substantial difficulties. Particularly in the context of whole-genome association, use of $\Phi_{\text{posterior}}$ might be unfeasible in extended pedigrees, and, as we have seen, it makes almost no difference in moderate-size pedigrees. Therefore, as a practical matter, it seems to make sense to use Φ , with the understanding that, in extended pedigrees, this provides a test of the joint null hypothesis of no association and no linkage.

Robustness of Power of M_{QLS} to Choice of k

The M_{QLS} test is valid for any choice of the parameter k . In outbreds, when k is equal to the population prevalence of the disease, we have argued that the M_{QLS} is asymptotically locally most powerful for all two-allele disease models as the effect size tends to zero. In reality, the trait will usually be complex, and the prevalence will be estimated. To see how the power of the M_{QLS} test is affected by different choices of k , we perform a simulation study on the basis of the second study design, with the phenotype simulated from model III-a. For model III-a, the true population prevalence is $K_p = .078$. We perform the M_{QLS} test with different settings of the parameter k , which are given in table 6.

Table 6 gives power results for the M_{QLS} test for values for k ranging from one-quarter of the true value to 6 times the true value. In this case, choosing k to be within a factor of 3 or 4 of the population prevalence appears to give high power, suggesting that the procedure is quite robust to choice of k .

GAW 14 COGA Data

The National Institute on Alcohol Abuse and Alcoholism has estimated¹¹ that the prevalence of alcohol dependence in the United States is ~5%. For the M_{QLS} , we accordingly

set $k = 0.05$ in the analysis. The average coefficient of variation of the weights given by the W_{QLS} for the cases is 1.568 and for the controls is -0.655 . For the M_{QLS} , they are 2.510 for the cases and -0.668 for the controls. The fact that the coefficients of variation for the cases are >1 for both the M_{QLS} and W_{QLS} suggests that these tests have the potential to give different results from those given by the W_{Xcorr}^2 test, and, in particular, we might expect that there is some advantage to be gained by applying the M_{QLS} . Table 7 gives the results of the analyses for those SNPs for which at least one of the tests has a nominal $P < 4.0 \times 10^{-5}$. For 13 of these 15 SNPs, the M_{QLS} test has the smallest P value among the three tests used. After Bonferroni correction to adjust for three different tests of association at each of 10,407 SNPs, the M_{QLS} test is significant at the 5% level for four SNPs: *tsc1177811* on chromosome 1 ($P = 5.9 \times 10^{-7}$ uncorrected; .018 corrected), *tsc1750530* on chromosome 16 ($P = 4.0 \times 10^{-7}$ uncorrected; .012 corrected), *tsc0046696* on chromosome 18 ($P = 4.7 \times 10^{-7}$ uncorrected; .015 corrected), and *tsc0057290* on chromosome 18 ($P = 5.2 \times 10^{-7}$ uncorrected; .016 corrected). Note that the two significant SNPs on chromosome 18 are 71 cM apart. The W_{Xcorr}^2 test is significant at the 5% level for an additional SNP, *tsc0571038* on chromosome 11 ($P = 6.2 \times 10^{-7}$ uncorrected; .019 corrected).

Three of the five significant SNPs in table 7 are near genes: *tsc1750530* is 3 kb from the gene encoding HEAT repeat containing 3 (*HEATR3* [Affymetrix; EntrezGene]), located at 16q12.1. The same SNP is also 25 kb from the gene encoding transmembrane protein 188 (*TMEM188* [Affymetrix; EntrezGene]), located at 16q12.1. SNP *tsc0046696* is 238 kb from the gene encoding F-box protein 15 (*FBXO15* [MIM 609093; Affymetrix; EntrezGene]), located at 18q22.3. The other significant SNP on chromosome 18, *tsc0057290*, is 411 kb from the gene encoding VAMP (vesicle-associated membrane protein)-associated protein A (*VAPA* [MIM 605703; Affymetrix; EntrezGene]), 33 kDa, located at 18p11.22. To our knowledge, none of these genes are obvious candidates. The two significant SNPs that are not in close proximity to any known genes, *tsc1177811* and *tsc0571038*, are located at 1p31.1 and

Table 6. Robustness of M_{QLS} to Misspecification of K_p

Assumed K_p	Multiple of True K_p	Estimated Power (SE)
.019	$\frac{1}{4}$.74 (.006)
.026	$\frac{1}{3}$.75 (.006)
.039	$\frac{1}{2}$.74 (.006)
.078	1	.76 (.006)
.156	2	.76 (.006)
.234	3	.75 (.006)
.312	4	.72 (.006)
.390	5	.63 (.007)
.468	6	.43 (.007)

NOTE.—Simulation model is III-a, given in table 2.

11q21, respectively (Affymetrix). Among the other SNPs in the table, two are located in or near genes of potential interest: (1) *tsc1687605* is in the 3' UTR of the gene that encodes cytochrome P450, family 2, subfamily C, polypeptide 18 (*CYP2C18* [MIM 601131; Affymetrix; EntrezGene]) located at 10q24—*CYP2C18* is a member of the CYP2C subfamily of P450 enzymes that is involved with drug metabolism (EntrezGene; OMIM); and (2) SNP *tsc0768481* is in the same cytogenetic region (13q14-q21) as the gene that encodes 5-hydroxytryptamine receptor 2A (*HTR2A* [MIM 182135; Affymetrix; EntrezGene]). There is evidence that *HTR2A* is associated with alcohol dependence.¹²

Three previous analyses^{13–15} of these data used the ALDX1 phenotype and performed family-based association tests using FBAT.¹⁶ When only white individuals were analyzed, no SNPs were significant at the 5% level with use of FBAT with Bonferroni correction.¹³ When all individuals were analyzed, three SNPs were significant after Bonferroni correction,¹⁴ only one of which (*tsc1750530* on chromosome 16) is in the set of five SNPs we detect. These SNPs and their *P* values were reported¹⁴ as *tsc0515272* on chromosome 3 ($P = 3.8 \times 10^{-7}$ uncorrected), *tsc0029429* on chromosome 9 ($P = 2.0 \times 10^{-8}$ uncorrected), and *tsc1750530* on chromosome 16 ($P = 4.5 \times 10^{-7}$ uncorrected). False-discovery rates for these SNPs are reported¹⁵ as .0270, .0019, and .0094, respectively. The corresponding uncorrected *P* values for these SNPs by the M_{QLS} are .058, .061, and 4.0×10^{-7} , respectively. Note that the M_{QLS} detected four significant SNPs (and the $W_{\chi^2_{corr}}$ detected a fifth significant SNP) with a smaller sample than that used by the FBAT to detect three significant SNPs. This indicates that the M_{QLS} is a powerful test that provides additional results complementary to those provided by FBAT.

Analysis of 10,407 SNPs with three tests (M_{QLS} , W_{QLS} , and $W_{\chi^2_{corr}}$) took ~35 minutes on a Pentium 4 3-GHz machine with 1 GB RAM. The calculations should scale linearly with the number of SNPs. We have not made serious attempts to optimize the code, so this time could presumably be improved. The slow step is the Cholesky decomposition⁷ of Φ , which would need to be performed only once if, for

every SNP, the same individuals had missing genotype data. However, this is generally not the case, so, in our implementation, the Cholesky decomposition is recomputed for every SNP.

Discussion

Despite major advances in high-density genome scans, disappointing results in the mapping of many common diseases illustrate the need for more-powerful methods for detection of susceptibility loci. We specifically address the problem of genomewide association analysis of binary traits when some individuals in the sample are related with known kinship. This arises naturally, for instance, when families sampled for a linkage study are included in an association study. This can be desirable, because it is expected that affected individuals from multiplex families would have a higher expected frequency of the alleles that increase susceptibility for a genetic trait than would affected individuals who do not have affected relatives. As a result, the power to detect association is expected to increase when affected individuals from multiplex families are included in a study. However, analysis of such data presents statistical and computational challenges.

We have developed a new test, the M_{QLS} , which is applicable to association studies with completely general combinations of family and case-control designs. For instance, the M_{QLS} allows cases to be related to controls, and it is equally applicable to complex inbred pedigrees and to simpler study designs consisting of unrelated individuals and small outbred families. The M_{QLS} distinguishes between unaffected controls and controls of unknown phenotype and can incorporate both into the same analysis. Furthermore, it makes use of phenotype data about relatives who have missing genotype data at a given SNP, where this information is used to optimize the weights given to relatives with nonmissing data at the SNP. We also extend the test to multiallelic markers. Our method is computationally feasible to use for genomewide association studies with hundreds of thousands or millions of SNPs. Our simulations indicate that the M_{QLS} represents an overall—and, in many cases, substantial—improvement in power over competing methods for a broad range of multigene trait models, while controlling type I error. In a reanalysis of the GAW 14 COGA data, the M_{QLS} detected four SNPs with genomewide-significant association to alcoholism, three of which had not been identified as significant in previous analyses.

We suggest a simple diagnostic, based on only phenotype information, that determines whether the M_{QLS} , W_{QLS} , and $W_{\chi^2_{corr}}$ would be expected to give different results. This can allow one to avoid correcting for use of three different tests in situations in which they are expected to give similar results. In our simulation studies, when the diagnostic indicated that the tests would give different results, the M_{QLS} was generally the most powerful. In the GAW 14 COGA data, the diagnostic indicated that the tests would

Table 7. COGA Data Results

Marker	Chromosome	Position (cM)	N_{ca}	N_{co}	p	P		
						$W_{\chi^2_{corr}}$	W_{QLS}	M_{QLS}
<i>tsc1750530</i>	16	59.8297	644	145	.85	3.4e-4	1.6e-4	4.0e-7*
<i>tsc0046696</i>	18	104.665	459	118	.60	4.0e-1	7.6e-1	4.7e-7*
<i>tsc0057290</i>	18	33.9594	497	126	.71	5.8e-2	4.9e-2	5.2e-7*
<i>tsc1177811</i>	1	105.535	587	149	.68	2.7e-2	5.0e-2	5.9e-7*
<i>tsc0571038</i>	11	95.3968	581	122	.56	6.2e-7*	6.0e-5	4.5e-3
<i>tsc1637642</i>	5	95.4901	419	159	.84	3.5e-1	3.0e-1	3.2e-6
<i>tsc1189131</i>	19	68.94	478	121	.55	7.3e-1	1.7e-1	5.2e-6
<i>tsc1519933</i>	3	167.431	515	134	.64	9.5e-2	5.7e-2	8.0e-6
<i>tsc0808295</i>	6	47.1522	681	162	.76	9.4e-1	2.9e-1	8.2e-6
<i>tsc1034745</i>	14	41.7798	806	193	.74	8.8e-3	1.3e-5	5.3e-2
<i>tsc0569292</i>	11	6.78451	455	127	.74	5.9e-1	4.3e-1	1.3e-5
<i>tsc0175005</i>	3	158.199	594	152	.82	1.5e-2	2.9e-2	1.7e-5
<i>tsc0784937</i>	14	45.6372	561	127	.60	6.1e-1	6.8e-1	2.2e-5
<i>tsc0768481</i>	13	56.4194	568	138	.76	1.3e-1	3.3e-1	3.4e-5
<i>tsc1687605</i>	10	115.265	517	126	.75	2.6e-3	2.3e-1	3.8e-5

NOTE.—SNPs with an associated $P < 4.0 \times 10^{-5}$ for at least one of the three tests are reported. The chromosome, the name of the marker, the position of the marker on the chromosome, the number of genotypes available in cases (N_{ca}) and controls (N_{co}), and the major-allele frequency in the case-control sample as a whole (p) are displayed. An asterisk (*) denotes genomewide significance after Bonferroni correction.

give different results, and, indeed, this was the case with the M_{QLS} , W_{QLS} , and $W_{\chi^2_{corr}}$ identifying 4, 0, and 1 significant markers, respectively, where the marker identified by $W_{\chi^2_{corr}}$ was not among those identified by M_{QLS} . (In this case, the Bonferroni correction took into account the three different tests as well as the number of SNPs tested.) In our simulations, a diagnostic result of >1 in absolute value in either cases or controls corresponded to a noticeable power difference between the tests, but, with larger sample sizes, a smaller diagnostic result might still correspond to a substantial power difference.

We have developed a modified version of the CC-QLS software program³ that outputs the results of our new M_{QLS} test for each SNP, as well as the results of the previously proposed³ $W_{\chi^2_{corr}}$ and W_{QLS} tests. The source code will be available (see M.S.M.'s Web page).

In the simulations and data analysis, we focus on inclusion of small-to-moderate-size outbred families in case-control association studies. However, it is important to note that the M_{QLS} is equally applicable to case-control association testing in founder populations, provided that the genealogy is known. Founder populations—for example, the Tasmanian population¹⁷ and the Hutterites¹⁸—are of interest for the mapping of complex traits for various reasons, including (1) avoidance of the problems of unknown population substructure and (2) the expectation that there would be fewer risk alleles involved in complex disorders in founder populations than in diverse continental populations. The M_{QLS} is computationally feasible, even in a founder population as complicated as the Hutterites, among whom many of the individuals are related through multiple lines of descent and exact likelihood calculation is not feasible.¹⁸

We have examined the question of whether to use prior or posterior kinship coefficients in calculating the weights for W_{QLS} and M_{QLS} and the variances for $W_{\chi^2_{corr}}$, W_{QLS} , and M_{QLS} . We recommend that prior kinship coefficients always be used in calculating the weights; otherwise, the theoretical justification for the statistics might not hold, and they could be badly behaved. In calculating the variance, we found no difference in the results obtained for small-to-moderate-size pedigrees with the two different types of kinship coefficients. Therefore, we recommend prior kinship coefficients for that calculation also, because they are much faster and simpler to compute. For large pedigrees, posterior kinship coefficients are unfeasible to obtain exactly, so it is somewhat academic to debate which is better. Nevertheless, on the basis of our simulations, we can say that, for a design consisting of a small number of large pedigrees, if one is willing to detect a signal that is driven by a combination of linkage and association, then one should obtain higher power with prior kinship coefficients, whereas, if one wants to correct for a known linkage signal to obtain a pure association test, then better type I error properties would be obtained with posterior kinship coefficients. (It is reasonable that these differences should disappear with multiple small-to-moderate-size pedigrees, because, if there is linkage but no association, then different alleles would tend to be associated with the trait in different pedigrees.)

Use of the M_{QLS} requires specification of a constant k in the test statistic. We emphasize that the test is valid for any value of k . To optimize power, we recommend that k be set to the best available estimate of the population prevalence of the trait. Our simulation studies suggest that the power of the test is very robust to the choice of k . When k was misspecified within a factor of 3 or 4 of the

true prevalence, there was little or no loss of power in our simulations.

Acknowledgments

We thank Dr. Mark Abney, for discussion and critical comments; Wataru Yoshimura from Affymetrix Application Support, for assistance with the Affymetrix (NetAffx) Web site; and two anon-

ymous reviewers, for helpful comments. This study was supported in part by a David and Lucile Packard Fellowship (to T.T.) and by National Institutes of Health grants HG001645 and HL084715. Data were provided by the COGA (U10AA008401). We gratefully acknowledge COGA and thank Dr. Ray Crowe of the University of Iowa and Dr. Jean MacCluer of the Southwest Foundation for Biomedical Research for their help in obtaining permission to analyze the COGA data.

Appendix A

Extension of M_{QLS} to Multiallelic Case

We extend the M_{QLS} procedure to test for association between a trait and a multiallelic marker. Extensions of the $W_{x^2_{corr}}$ and W_{QLS} tests to multiallelic markers have been given elsewhere.³ Suppose there are a allelic types at the marker, and let $Y = (Y_1, \dots, Y_{a-1})^T$ be an $[(a-1)n]$ vector, where $Y_i = (Y_{i1}, \dots, Y_{in})$ and $Y_{ij} = \frac{1}{2} \times$ (the number of alleles of type i that individual j has). Let $p = (p_1, \dots, p_{a-1})^T$ denote the allele-frequency distribution at the marker in the general population, where $p_i > 0$ is the frequency of allelic type i , and $\mathbf{1}^T p < 1$. Define $r = (r_1, \dots, r_{a-1})^T$ to be the $(a-1)$ vector of expected changes in allele frequencies for a case randomly sampled from the population. Then the mean model for the M_{QLS} in the multiallelic case is $EY = \mu = p \otimes \mathbf{1} + r \otimes (\Phi_{N, N \cup M} A)$, where \otimes is the Kronecker product (see, e.g., the work of Schott^{7(p253)}), and $\mathbf{1}$ is a vector of 1s of length n ; that is

$$EY_{ij} = \mu_{ij} = p_i + r_i \sum_{k=1}^{n+m} 2\phi_{j,k} A_k,$$

where we constrain

$$0 < p_i + r_i \sum_{k=1}^{n+m} 2\phi_{j,k} A_k < 1$$

for all $1 \leq i \leq a-1$, $1 \leq j \leq n$. Under the null hypothesis of no association between the marker and the trait, we have $r = \mathbf{0}$, where $\mathbf{0}$ is a zero vector of length $(a-1)$. Let \mathbf{F} denote an $(a-1)(a-1)$ matrix with (i,j) th entry $\mathbf{F}_{ij} = \frac{1}{2} p_i (1 - p_j)$ if $i = j$ and $\mathbf{F}_{ij} = -\frac{1}{2} p_i p_j$ if $i \neq j$. Note that, under the null hypothesis of no association and no linkage, and when the pedigree founders are drawn from a population in HWE under the null hypothesis, $\text{Var}(Y) = \mathbf{F} \otimes \Phi$. The M_{QLS} test statistic for the multiallelic case is

$$M_{QLS} = \sum_{k=1}^{a-1} \sum_{i=1}^{a-1} (\hat{\mathbf{F}}^{-1})_{ik} (Y_k - \hat{\mu}_{0k})^T \alpha \Gamma^{-1} \alpha^T (Y_i - \hat{\mu}_{0i}),$$

where α and Γ are as defined for equations (5) and (6) and $(\hat{\mathbf{F}}^{-1})_{ik}$ is the (i,k) entry of \mathbf{F}^{-1} evaluated at $p = \hat{p}_0 = (\hat{p}_{01}, \dots, \hat{p}_{0(a-1)})^T$, where \hat{p}_0 is the maximum quasi-likelihood estimate of p when $r = \mathbf{0}$, or equivalently, the BLUE of p based on everyone, which previous work⁸ has shown to be $\hat{p}_{0i} = (\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Phi^{-1} Y_i$ for each i . Under the null hypothesis, the M_{QLS} statistic follows a χ^2 distribution with $a-1$ df (asymptotically, under regularity conditions).

Appendix B

Connection between MZ Twins and Invertibility of Φ

We prove that Φ is invertible if and only if the set N does not include both members of any MZ twin pair. Because Σ is a covariance matrix, it must be symmetric and positive semidefinite and is invertible if and only if it is positive definite. Φ inherits these properties. Note that Σ and hence Φ is positive definite if and only if there is no linear combination $c^T Y$, $c \in \mathbb{R}^n \setminus \{0\}$, such that $\Pr(c^T Y = 0) = 1$. This is because $\text{Var}(c^T Y) = c^T \Sigma c$, and $\text{Var}(c^T Y) = 0$ if and only if $\Pr(c^T Y = 0) = 1$. If N includes i and j who are MZ twins, then $\Pr(Y_i - Y_j = 0) = 1$, and Φ is not invertible. Suppose Φ is not invertible. Then there must be some individual i in N such that Y_i can be written as a linear combination of Y_{-i} . $\Pr(Y_i = d^T Y_{-i}) = 1$, $d \in \mathbb{R}^{n-1}$, where Y_{-i} is Y excluding the i th element. Consider the situation in which every individual in $N \setminus \{i\}$ is heterozygous at the binary marker, which has positive probability for $0 < p < 1$. Then, if i is not MZ twin to anyone in $N \setminus \{i\}$, it is possible for i to have any genotype, and so it cannot be true that $\Pr(Y_i = d^T Y_{-i}) = 1$. Thus, $\Pr(Y_i = d^T Y_{-i}) = 1$ implies that i has an MZ twin in $N \setminus \{i\}$.

Appendix C

Derivation of M_{QLS} Mean Model

We show that, under a two-allele disease model, for an arbitrary set of possibly related outbred individuals, the ratio

$$\frac{\mu_i^* - p}{\mu_a^* - p}$$

(given in the “Development and Justification of the M_{QLS} Test” section) tends to

$$\sum_{j=1}^{n+m} 2\phi_{ij}A_j,$$

as the effect size (or differences among penetrance probabilities) tend to zero. Throughout, we condition on the pattern of missing phenotype information. Consider a two-allele disease model with penetrance probabilities k , $k - c_2$, and $k - c_3$ for individuals who have 2, 1, or 0 alleles of type 1, respectively, where $k - 1 \leq c_2 \leq k$ and $k - 1 \leq c_3 \leq k$, with at least one of c_2 and c_3 nonzero. (Note that, as c_2 and c_3 tend to zero, the population prevalence K_p will tend to k .) Under this model, we have

$$\mu_a^* - p = E(Y_i|A_i = 1) - p = \frac{p(1-p)[-c_2(1-2p) + c_3(1-p)]}{k + (1-p)[-2c_2p - c_3(1-p)]}.$$

Furthermore, we can express

$$\begin{aligned} \mu_i^* - p &= E(Y_i|A) - p \\ &= \frac{-pP(A) + \sum_{s \in \{0, .5, 1\}} sP(Y_i = s) \sum_{Y_{-i}} P(Y_{-i}|Y_i = s)P(A|Y)}{P(A)}, \end{aligned} \quad (C1)$$

where A is the entire phenotype vector, Y_{-i} denotes $(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_{n+m})$, Y_{-1} denotes (Y_2, \dots, Y_{n+m}) , and $Y_{-(n+m)}$ denotes (Y_1, \dots, Y_{n+m-1}) , where Y_{n+1}, \dots, Y_{n+m} denote the true (unobserved) genotype values for individuals $n + 1, \dots, n + m$. As c_2 and c_3 approach 0, we have the expansions

$$P(A) = k^{n_c}(1-k)^{n_u} \left\{ 1 - \frac{1}{k} [2c_2p(1-p) + c_3(1-p)^2] \sum_{j=1}^{n+m} A_j \right\} + o(c_2) + o(c_3)$$

and

$$\begin{aligned} \sum_{Y_{-i}} P(Y_{-i}|Y_i = s)P(A|Y) &= k^{n_c}(1-k)^{n_u} \\ &\times \left\{ 1 + \sum_{Y_{-i}} [P(Y_{-i}|Y_i = s) \left(\frac{c_2}{1-k} n_{.s,u} - \frac{c_2}{k} n_{.s,c} + \frac{c_3}{1-k} n_{0,u} - \frac{c_3}{k} n_{0,c} \right)] \right\} + o(c_2) + o(c_3), \end{aligned} \quad (C2)$$

where n_c is the number of affected, n_u is the number of unaffected, $n + m - n_c - n_u$ is the number of individuals of unknown phenotype in the study, $n_{v,c}$ is the number of affected individuals having a Y value of v (regardless of whether Y is actually observed or is missing in the study), and, similarly, $n_{v,u}$ is the number of unaffected individuals having a Y value of v . Note that, for example,

$$n_{.s,u} = \sum_{j=1}^{n+m} \mathbf{1}_{Y_j = .s} \mathbf{1}_{A_j = \frac{-k}{1-k}},$$

so

$$\sum_{Y_{-i}} P(Y_{-i}|Y_i = s) n_{.s,u} = \sum_{j=1}^{n+m} \mathbf{1}_{A_j = \frac{-k}{1-k}} \sum_{Y_{-i}} P(Y_{-i}|Y_i = s) \mathbf{1}_{Y_j = .s} = \sum_{j=1}^{n+m} \mathbf{1}_{A_j = \frac{-k}{1-k}} P(Y_j = .s | Y_i = s),$$

where $P(Y_j = .5 | Y_i = s)$ can be calculated in terms of the relationship between the pair of individuals (i, j) , without having to consider multiple individuals jointly. Applying a similar argument to the other terms in equation (C2) and reorganizing terms, we can obtain

$$\sum_{s \in \{0, .5, 1\}} sP(Y_i = s) \sum_{Y_{-i}} P(Y_{-i} | Y_i = s)P(A | Y) = k^{n_c}(1 - k)^{n_u}$$

$$\times \left[E(Y_i) - \frac{c_2}{k} \sum_{j=1}^{n+m} A_j E(Y_i | Y_j = .5)P(Y_j = .5) - \frac{c_3}{k} \sum_{j=1}^{n+m} A_j E(Y_i | Y_j = 0)P(Y_j = 0) \right] + o(c_2) + o(c_3) .$$

Plugging into equation (C1) and noting that, for a pair of outbred individuals (i, j) ,

$$E(Y_i | Y_j = 0)P(Y_j = 0) = p(1 - p)^2(1 - 2\phi_{ij})$$

and

$$E(Y_i | Y_j = .5)P(Y_j = .5) = p(1 - p)[2p + 2\phi_{ij}(1 - 2p)] ,$$

we obtain

$$\mu_a^* - p = E(Y_i | A) - p = \frac{1}{P(A)} \{ k^{n_c-1}(1 - k)^{n_u} 2p(1 - p) \}$$

$$\times [c_3(1 - p) - c_2(1 - 2p)] \sum_{j=1}^{n+m} A_j \phi_{ij} + o(c_2) + o(c_3) .$$

Then, plugging in the derived expressions for $\mu_a^* - p$ and $P(A)$, and letting c_2 and c_3 tend to 0, we get

$$\lim_{c_2, c_3 \rightarrow 0} \frac{\mu_a^* - p}{\mu_a^* - p} = \sum_{j=1}^{n+m} 2\phi_{ij} A_j .$$

If we let r represent the quantity $\mu_a^* - p$, then this leads to our model $\mu_a^* = p + r \sum_{j=1}^{n+m} 2\phi_{ij} A_j$.

Appendix D

Details of M_{QLS} Power Improvement Diagnostic

Recall that the test statistics $W_{\chi^2_{\text{corr}}}$, W_{QLS} , and M_{QLS} each have the form $S^T[\text{Var}_0(S)]^{-1}S$, where $S = V^T Y$, $V = (V_1, \dots, V_i, \dots, V_n)^T$, and V_i can be viewed as the total weight given by the test statistic to individual i , with

$$V_i = \frac{W_{\text{test},i}}{(w_{\text{test}}^T \mathbf{1})} - \frac{W_{\text{null},i}}{(w_{\text{null}}^T \mathbf{1})} .$$

Here, $S = \hat{p}_{\text{test}} - \hat{p}_{\text{null}}$. Note that, by construction, $\bar{V} = \frac{1}{n} \sum_{i=1}^n V_i = 0$ for each test statistic. To measure the difference in weights between a pair of statistics, say $W_{\chi^2_{\text{corr}}}$ and M_{QLS} , for case and control individuals, we propose

$$\psi(W_{\chi^2_{\text{corr}}}, M_{\text{QLS}}, \text{case}) = \sqrt{\frac{1}{n_c - 1} (B_{\chi^2_{\text{corr}}} - B_{M_{\text{QLS}}})^T (B_{\chi^2_{\text{corr}}} - B_{M_{\text{QLS}}})}$$

and

$$\psi(W_{\chi^2_{\text{corr}}}, M_{\text{QLS}}, \text{control}) = \sqrt{\frac{1}{n - n_c - 1} (C_{\chi^2_{\text{corr}}} - C_{M_{\text{QLS}}})^T (C_{\chi^2_{\text{corr}}} - C_{M_{\text{QLS}}})} ,$$

where, for each statistic, the corresponding B is a vector of length n with i th component

$$B_i = \frac{n_c \mathbf{1}_{c,i} V_i}{\mathbf{1}_c^T V},$$

which is V_i normalized by the mean total weight among cases if i is a case and is 0 if i is a control. Similarly, for each statistic, the corresponding C is a vector of length n with i th component

$$\frac{(n - n_c)(1 - \mathbf{1}_{c,i})V_i}{(\mathbf{1} - \mathbf{1}_c)^T V},$$

which is V_i normalized by the mean total weight among controls if i is a control and is 0 if i is a case. The definitions of the ψ s for any other pair of statistics are analogous. Note that $B_{\chi_{\text{corr}}^2} = \mathbf{1}_c$ and $C_{\chi_{\text{corr}}^2} = \mathbf{1} - \mathbf{1}_c$, so that $\psi(W_{\chi_{\text{corr}}^2}, M_{\text{QLS}}, \text{case})$ reduces to the absolute value of the coefficient of variation of $V_{M_{\text{QLS}}}$ among cases, and, similarly, $\psi(W_{\chi_{\text{corr}}^2}, M_{\text{QLS}}, \text{control})$, $\psi(W_{\chi_{\text{corr}}^2}, W_{\text{QLS}}, \text{case})$, and $\psi(W_{\chi_{\text{corr}}^2}, W_{\text{QLS}}, \text{control})$ reduce to the absolute values of the coefficients of variation of $V_{M_{\text{QLS}}}$ among controls, $V_{W_{\text{QLS}}}$ among cases, and $V_{W_{\text{QLS}}}$ among controls, respectively, where the coefficient of variation is the SD divided by the mean. In the situation in which there is a group \mathcal{G} of permutations of the $n + m$ individuals such that (1) every element of \mathcal{G} preserves genotyped/missing status; (2) every element of \mathcal{G} preserves A ; (3) every element of \mathcal{G} preserves $\Phi_{N, N \cup M}$; (4) for every pair of genotyped case individuals i and j , there is a permutation in \mathcal{G} that maps i to j ; and (5) for every pair of genotyped control individuals k and l , there is a permutation in \mathcal{G} that maps k to l , then the coefficients of variation for the total weights of the cases and for the total weights of the controls are both equal to 0 for the M_{QLS} and the W_{QLS} statistics, and these two statistics are equivalent to the $W_{\chi_{\text{corr}}^2}$ test. For example, these conditions hold, at a marker with no missing genotypes, if the cases are affected sib pairs and the controls are unaffected unrelated individuals.

Web Resources

The URLs for data presented herein are as follows:

Affymetrix, <https://www.affymetrix.com/analysis/netaffx/index.affx> (for NetAffx information on the location of SNPs in table 7 and identification of genes that are in close proximity to the SNPs)

EntrezGene, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=gene> (for information about genes that are close to SNPs in table 7)

M.S.M.'s Web page, <http://www.stat.uchicago.edu/~mcpeek/software/index.html> (for the source code described in the text)

Online Mendelian Inheritance in Man (OMIM), <http://www.ncbi.nlm.nih.gov/Omim/> (for alcohol dependence, *FBXO15*, *VAPA*, *CYP2C18*, and *HTR2A*)

References

1. Spielman R, McGinnis R, Ewens W (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
2. Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I: DNA pooling. *Genome Res* 8: 1273–1288
3. Bourgain C, Hoffjan S, Nicolae R, Newman D, Steiner L, Walker K, Reynolds R, Ober C, McPeck MS (2003) Novel case-control test in a founder population identifies P-selectin as an atopy-susceptibility locus. *Am J Hum Genet* 73:612–626
4. Slager SL, Schaid D (2001) Evaluation of candidate genes in case-control studies: a statistical method to account for related subjects. *Am J Hum Genet* 68:1457–1462
5. Browning S, Briley J, Chandra G, Charnecki J, Ehm M, Johansson K, Jones B, Karter A, Yarnall D, Wagner M (2005) Case-control single marker and haplotypic association analysis of pedigree data. *Genet Epidemiol* 28:110–122
6. Edenberg HJ, Bierut LJ, Boyce P, Cao M, Cawley S, Chiles R, Doheny KF, Hansen M, Hinrichs T, Jones K, et al (2005) Description of the data from the Collaborative Study on the Genetics of Alcoholism (COGA) and single-nucleotide polymorphism genotyping for Genetic Analysis Workshop 14. *BMC Genetics Suppl* 6:S2
7. Schott JR (1996) Matrix analysis for statistics. John Wiley, New York
8. McPeck MS, Wu X, Ober C (2004) A quasi-likelihood method for allele frequency estimation. *Biometrics* 60:359–367
9. Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101
10. Sun L, Cox N, McPeck MS (2002) A statistical method for identification of polymorphisms that explain a linkage result. *Am J Hum Genet* 70:399–411
11. Grant BF, Harford TC, Dawson DA, Chou P, Dufour M, Pickering R (1994) Prevalence of DSM-IV alcohol abuse and dependence: United States 1992. *Alcohol Health Res World* 18: 243–248
12. Hwu HG, Chen CH (2000) Association of 5HT2A receptor gene polymorphism and alcohol abuse with behavior problems. *Am J Med Genet* 96:797–800
13. Zhu X, Cooper R, Kan D, Cao G, Wu X (2005) A genome-wide linkage and association study using COGA data. *BMC Genetics Suppl* 6:S128
14. Chiu YF, Liu SY, Tsai YY (2005) A comparison in association

- and linkage genome-wide scans for alcoholism susceptibility genes using single-nucleotide polymorphisms. *BMC Genetics Suppl* 6:S89
15. Chen L, Liu N, Wang S, Oh C, Carriero NJ, Zhao H (2005) Whole-genome association studies on alcoholism comparing different phenotypes using single-nucleotide polymorphisms and microsatellites. *BMC Genetics Suppl* 6:S130
 16. Rabinowitz D, Laird N (2000) A unified approach to adjusting association tests for population admixture with arbitrary pedigree structure and arbitrary missing marker information. *Hum Hered* 50:211–223
 17. Stankovich J, Bahlo M, Rubio JP, Wilkinson CR, Thomson R, Banks A, Ring M, Foote SJ, Speed TP (2005) Identifying nineteenth century genealogical links from genotypes. *Hum Genet* 117:188–199
 18. Abney M, McPeck MS, Ober C (2005) Estimation of variance components of quantitative traits in inbred populations. *Am J Hum Genet* 66:629–650