Published in final edited form as:
Genet Epidemiol. 2009 September ; 33(6): 497–507. doi:10.1002/gepi.20402.

Asymptotic Tests of Association with Multiple SNPs in Linkage Disequilibrium

Wei Pan
Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN 55455

Abstract

We consider detecting associations between a trait and multiple SNPs in linkage disequilibrium (LD). To maximize the use of information contained in multiple SNPs while minimizing the cost of large degrees of freedom (DF) in testing multiple parameters, we first theoretically explore the sum test derived under a working assumption of a common association strength between the trait and each SNP, testing on the corresponding parameter with only one DF. Under the scenarios that the association strengths between the trait and the SNPs are close to each other (and in the same direction), as considered by Wang and Elston (2007), we show with simulated data that the sum test was powerful as compared to several existing tests; otherwise, the sum test might have much reduced power. To overcome the limitation of the sum test, based on our theoretical analysis of the sum test, we propose five new tests that are closely related to each other and are shown to consistently perform similarly well across a wide range of scenarios. We point out the close connection of the proposed tests to the Goeman test. Furthermore, we derive the asymptotic distributions of the proposed tests so that p-values can be easily calculated, in contrast to the use of computationally demanding permutations or simulations for the Goeman test. A distinguishing feature of the five new tests is their use of a diagonal working covariance matrix, rather than a full covariance matrix as used in the usual Wald or score test. We recommend the routine use of two of the new tests, along with several other tests, to detect disease associations with multiple linked SNPs.

Keywords

genome-wide association study; logistic regression; multilocus analysis; permutation; single-locus analysis; SNP

INTRODUCTION

Due to the rapid development of large-scale genotyping technologies, it has become feasible to conduct genome-wide association studies (GWASs); in fact, quite a few GWASs have been completed while many others are underway. However, a remaining challenge is how to efficiently analyze data collected from GWASs. This is critical given that most complex diseases or traits are only weakly associated with causal genetic variants. The power of GWASs depends on linkage disequilibrium (LD) between DNA markers (e.g., genotyped SNPs) and causal loci that often are not genotyped: if there exists a causal locus in a DNA region, due to LD, it is likely that one or more SNPs nearby will be associated with the trait too. Hence, it is intuitively reasonable to consider multiple loci in LD in a region; this is the problem we consider here. Roughly speaking, most existing analysis methods fall into one

Correspondence author: Wei Pan, Telephone: (612) 626-2705, Fax: (612) 626-0660, Email: weip@biostat.umn.edu, Address: Division of Biostatistics, MMC 303, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455 0392, U.S.A.

of the two categories: single-locus versus multilocus approaches. The basic idea of the former is to analyze single SNPs one at a time, then combine such individual analyses (Roeder et al 2005 and references therein). For these methods, in addition to possibly failing to utilize information contained in high-order relationships among multiloci, another shortcoming is their reduced power due to the high cost of adjustments for multiple testing that are required to control family-wise error rates or false discovery rates at some nominal levels. On the other hand, to overcome the above two problems with single-locus approaches, multilocus methods jointly model the relationship between a trait and multiloci (e.g., Chapman et al 2003; Qin et al 2002; Schaid et al 2002; Stephens et al 2001; Wei et al 2008; Zhao et al 2003a, b, and references therein). However, there is also an inherent cost of large degrees of freedom (DF) when testing over multiple parameters involved in a joint model.

Wang and Elston (2007) fully recognized the two possibly conflicting goals of capturing information contained in multiloci in LD and reducing the cost of multiple testing or large DF, and thus proposed a weighted score test (WST) to achieve the two goals at the same time: although multiloci are involved, the DF of the WST is only 1. They used simulated data under various scenarios, including mimicking real LD patterns as drawn from HapMap data, to demonstrate improved power of the WST over other existing methods. Here we first study an alternative approach called the sum test in the same context: it is based on the familiar treatment of the trait as the response variable and modeling its mean (e.g. probability of having disease) in terms of multiple SNPs in LD in logistic regression or any generalized linear model (GLM). The basic idea is to test on only one parameter under a working assumption of a common association strength between each of multiple SNPs and the trait. Although in general the association strengths may vary with SNPs, the proposed methods can be more powerful than the WST and several other existing methods while controlling the test size at a specified nominal significance level; similar to the WST, the key is to utilize multiloci while controlling the DF at the minimum 1. Simply put, rather than testing on multiple parameters either individually (followed by a multiple test adjustment as in single-locus approaches) or jointly (with large DF as in most multilocus approaches), the sum test focuses on a scalar function of the multiple parameters with the resulting DF=1.

An original motivation of this work was to explore how the WST works. The numerical results of Chapman and Whittaker (2008) indicate that, if some of the coded SNPs are positively correlated while others are negatively correlated, the WST and the sum test may have low power. Here we exploit theoretical properties of the sum test. In particular, we provide a theoretical explanation for the possible failure of the sum test: if the association strengths between the outcome and the SNPs are not close, especially if some are in opposite directions, the sum test may have low power; this explanation differs from the heuristic argument given by Chapman and Whittaker (2008) that is based on the correlations among the coded SNPs. In addition, the theoretical result motivated the development of four new tests: two of them are asymptotically equivalent to an estimated most powerful test, and are weighted versions of the other two that are closely related to Goeman's test (Goeman et al 2006), which was found to perform well across a wide range of scenarios by Chapman and Whittaker (2008). The proposed tests are all based on a sum of squared marginal regression coefficient estimates or of squared score statistics, possibly weighted inversely by the variances of the coefficient estimates or score statistics. In contrast to the use of computationally intensive permutations or simulations as for Goeman's test, we derive asymptotic distributions of the proposed tests, easily yielding p-values. In spite of their simplicity, our numerical studies showed that the proposed tests performed similarly to Goeman's test. A distinguishing feature of the new tests and Goeman's test is their use of a diagonal working covariance matrix, rather than a full covariance matrix as used in the usual Wald or score test. We offer some intuitive explanations on the possibly improved power of

the new tests over the usual Wald or score test, illustrating the well known fact that there is no uniformly most powerful test in the current context.

METHODS

REGRESSION AND RELATED TESTS

Suppose that we have m independent observations (Y_i, X_i), where subject i has trait (e.g. disease status) value Y_i and genotype $X_i = (X_{i1}, \ldots, X_{ik})$. As in Wang and Elston (2007), we consider the dosage coding of X_{ij} for an additive model: $X_{ij} = 0$, 1 or 2, representing the copy number of one of the two alleles present in SNP j of subject i; other choices include a binary coding of $X_{ij} = 0$ or 1 for a dominant or recessive genetic model, and a more general coding scheme can be implemented, albeit more complicated and will be skipped. To test any possible association between the trait and SNPs, we entertain a generalized linear model (GLM) (McCullagh and Nelder 1983)

$$h[E(Y_i)]=\beta_0+\sum_{j=1}^{k} X_{ij}\beta_j=\beta_0+X_i\beta,$$

where $E(Y_i)$ is the mean of Y_i, $h(\,)$ is a link function and $\beta = (\beta_1, \ldots, \beta_k)'$. Two commonly used models are the linear model with the identity link for continuous or quantitative traits, and the logistic regression model with the Logit link for a binary trait. In this paper, we focus on the logistic model with the dosage coding of X_{ij},

$$\text{Logit}\Pr(Y_i=1)=\beta_0+\sum_{j=1}^{k} X_{ij}\beta_j, \tag{1}$$

where $Y_i = 0$ or 1 indicates whether subject i is a control (i.e. without disease) or a case (i.e. with disease). A global test of any possible association between the trait and SNPs can be formulated as jointly testing on the multiple parameters β_j's with the null hypothesis H_0: $\beta = 0$ by one of the likelihood ratio test (LRT), Wald test and score test in the context of logistic regression (or more generally, of GLM); under H_0, any of the three test statistics has an asymptotically chi-squared distribution with degrees of freedom DF=k. We call the LRT that was used in our numerical examples as the logistic-global (L-G) test. For a large k, the test can be low-powered because of the cost of the large DF.

In contrast to the global test, another extreme is to conduct individual SNP-by-SNP tests: rather than including all the k SNPs, we include only one SNP in

$$\text{Logit}\Pr(Y_i=1)=\beta_{M,0j}+X_{ij}\beta_{M,j}, \tag{2}$$

where we explicitly distinguish $\beta_M = (\beta_{M,1}, \ldots, \beta_{M,k})'$ in marginal models (2) from β in joint model (1). Then we test H_0: $\beta_{M,j} = 0$ for each $j = 1, \ldots, k$ sequentially. Each test can be done with only one DF, but a multiple test adjustment has to be made, based on either permutation or Bonferroni adjustment. Because the Bonferroni adjustment is known to be conservative, permutation is more widely used, though it is computationally demanding; we

call the univariate test based on permutation U-B. The multiple test adjustment may reduce the power of the test, as shown by Wang and Elston (2007).

Most of the existing methods fall into one of the above two extreme categories. For example, the generalized Hotelling's T^2 test (Fan and Knapp 2003; Xiong et al 2002) and haplotype analyses (e.g., Chapman et al 2003; Schaid et al 2002; Stephens et al 2001; Zhao et al 2003a, b, and references therein) belong to the first category for their joint testing on multiple parameters simultaneously, while other univariate approaches build on SNP-by-SNP tests (Roeder et al 2004 and references therein). The limitations of the two extremes have been well recognized and motivated the development of new methods. A general strategy is to capture full information in all β_j's while reducing the cost of large DF or adjustment for multiple testing, but the key is how to reconcile the two possibly conflicting goals. Wang and Elston (2007) proposed such an approach, called weighted score test (WST): the WST puts higher weights on more important components of Fourier-transformed genotypes and minimizes the DF at one. Below we first study a simple alternative to WST, the sum test; a theoretical analysis clearly indicates a possible limitation of the sum test, which further motivates the development of our new tests.

THE SUM TEST

We first formulate the sum test from a new perspective of striking a compromise between joint modeling and its resulting large DF: while all the SNPs in LD are to be used, we make a key and possibly incorrect working assumption that the SNPs are all equally associated with the trait; that is, rather than unnecessarily aiming to estimate separate β_j's in (1), we use a common β_c in the logistic regression model:

$$\text{Logit}\Pr(Y_i=1)=\beta_{c0}+\sum_{j=1}^{k}X_{ij}\beta_c. \tag{3}$$

To address the question of whether there is any association between the trait Y_i and any SNP, we test H_0: $\beta_c = 0$, which can be easily done by the LRT, Wald or score test in fitting model (3); we used the LRT throughout. Note that fitting model (3) is equivalent to regress Y on a new covariate that is the sum of the genotypes of the multiple SNPs, and hence we call the resulting test the sum test.

Below we explore the advantages and limitations of the sum test. Intuitively, because it tests on only one parameter β_c, there will be no power loss due to large DF (or multiple testing adjustment). Generally, the common association parameter β_c in (3) is an "average" (or more precisely, a function) of the individual $\beta_1, \ldots, \beta_k$; see Appendices A–C for the theory. It is most illustrative to consider the case with the linear model, in which it can be shown (in Appendix A) that

$$\widehat{\beta}_c=\frac{\sum_{j=1}^{k}\sum_{i=1}^{m}X_{ij}^2\widehat{\beta}_{M,j}}{\sum_{i=1}^{m}\left(\sum_{j=1}^{k}X_{ij}\right)^2}, \tag{4}$$

where $\hat{\beta}_c$ is the MLE of β_c in a linear model analogous to (3), and $\hat{\beta}_M$ is the MLE of β_M in the marginal linear models analogous to (2). Note that, because of the consistency of the MLE, with a large sample size in a typical genome-wide association study, equation (4) still

approximately holds if $\hat{\beta}_c$ and $\hat{\beta}_M$ are replaced by their true values β_c and β_M respectively. Note that the weight $\sum_{i=1}^{m} X_{ij}^2$ in (4) is inversely proportional to the variance of $\hat{\beta}_{M,j}$.

Under the global H_0: $\beta_M = 0$, it is trivial to see that H_0: $\beta_c = 0$ also holds, guaranteeing that the sum test will have (asymptotically) correct test size. On the other hand, if the global H_0 is false, then it is likely to have $\beta_c \neq 0$, leading to good power. For this latter purpose, because of the always non-negative weights for $\hat{\beta}_{M,j}$'s, it is best to have $\hat{\beta}_{M,j}$'s (or equivalently, true $\beta_{M,j}$'s for large samples) with the same sign; this is also a requirement for the WST. Hence, to avoid the consequence of that positive and negative components of β_M cancel out, a heuristic approach as discussed by Wang and Elston (2007) is the following: before applying the WST (and the sum test), one needs to adjust the coding of X_{ij}'s so that there are as few negative correlations as possible among the SNPs; for example, if $X_{\cdot l}$ of SNP l is negatively correlated with many other $X_{\cdot j}$'s, we will take new X_{il} as $2 - X_{il}$ for the dosage coding (or switching the values for the two categories in the binary coding). We developed an algorithm as the following: 1) calculate all pairwise correlations; 2) see which SNP s has the largest number, say n_s, of negative correlations; 3) if $n_s > \#SNPs/2$, then flip the coding of SNP s, and repeat the above process; otherwise, stop. Nevertheless, at the end, it is not guaranteed that all SNP pairs will be non-negatively correlated; more importantly, even if they are, it may not help: equation (4) clearly shows that the key is on the signs of the components of $\hat{\beta}_M$; a positively correlated pair of $X_{\cdot j}$ and $X_{\cdot l}$ does not imply that $\text{sign}(\hat{\beta}_{M,i}) = \text{sign}(\hat{\beta}_{M,l})$. The theory clearly illustrates a possible limitation in inferring the performance of the sum test with the use of pairwise SNP correlations, a heuristic suggested by Chapman and Whittaker (2008). Below we propose a class of five new tests that overcome the above issue of the sum test.

TWO NEW TESTS BASED ON $\hat{\beta}_M$

The problem with the sum test is that, as shown in equation (4), because the estimated common effect $\hat{\beta}_c$ is a linear combination of the components of $\hat{\beta}_M$ with always *positive* coefficients, the test may have reduced power with a small $\hat{\beta}_c$ when the components of $\hat{\beta}_M$ have different signs. Hence, to eliminate the sign problem, we replace the components of $\hat{\beta}_M$ in the linear combination by their squares. Specifically, we propose a test statistic

$$SumSqB = \widehat{\beta}_M' \widehat{\beta}_M = \sum_{j=1}^{k} \widehat{\beta}_{M,j}^2,$$

or, its weighted version with a weight assigned to each component of $\hat{\beta}_M$ based on its variance:

$$SumSqBw = \widehat{\beta}_M' \text{Diag}(V_M)^{-1} \widehat{\beta}_M = \sum_{j=1}^{k} \widehat{\beta}_{M,j}^2 / v_{M,j},$$

where $\text{Diag}(V_M) = diag(v_{M,1}, \ldots, v_{M,k})$ is a diagonal matrix with the elements $v_{M,j}$ as the (estimated) variance of $\hat{\beta}_{M,j}$ from marginal model (2). Note that, albeit related to the Wald test, the above tests differ from the usual Wald test, $\widehat{\beta}_M' V_M^{-1} \widehat{\beta}_M$, in that $V_M = Cov(\hat{\beta}_M)$ is replaced by a diagonal matrix, either an identity matrix I or $\text{Diag}(V_M)$.

Below we describe the asymptotic distributions of the above two test statistics. Based on the theory of generalized estimating equations as shown in Appendix D, under H_0: ($\beta_M = 0$, $\hat{\beta}_M$

asymptotically has a Normal distribution with mean 0, and its covariance matrix can be consistently estimated by a sandwich (or robust) estimator V_M. Hence, asymptotically, each of the two test statistics has a quadratic form of Normal variates, $Q=\widehat{\beta}_M' W^{-1}\widehat{\beta}_M$, with $W = I$ and $W = \text{Diag}(V_M)$ respectively. It is well known (e.g. Johnson and Kotz, 1970, p.150) that the distribution of Q is a weighted sum of k independent chi-squared variates with DF=1, $\sum_{j=1}^{k} c_j\chi_1^2$, where c_j's are the eigen values of $V_M W^{-1}$. Furthermore, by the results of Zhang (2005), $\sum_{j=1}^{k} c_j\chi_1^2$ can be well approximated by $a\chi_d^2+b$ with

$$a=\frac{\sum_{j=1}^{k} c_j^3}{\sum_{j=1}^{k} c_j^2},\quad b=\sum_{j=1}^{k} c_j-\frac{\left(\sum_{j=1}^{k} c_j^2\right)^2}{\sum_{j=1}^{k} c_j^3},\quad d=\frac{\left(\sum_{j=1}^{k} c_j^2\right)^3}{\left(\sum_{j=1}^{k} c_j^3\right)^2}.$$

To calculate a p-value, for example for the SumSqB test,

$$Pr(SumSqB>s|H_0)\approx Pr(\chi_d^2>(s-b)/a).$$

Note that in the above two new tests, we used $\hat{\beta}_M$ from marginal models (2), not $\hat{\beta}$ from joint model (1). Intuitively, because of collinearity, the components of $\hat{\beta}$ have much larger variances than their counter parts of $\hat{\beta}_M$. Our numerical results (not shown) confirmed the much worse performance of the resulting SumSqB or SumSqBw if $\hat{\beta}_M$ was substituted by $\hat{\beta}$. In contrast, the Wald test based on $\hat{\beta}_M$ and that based on $\hat{\beta}$ are asymptotically equivalent because their corresponding score tests are the same, as to be shown later.

TWO NEW TESTS BASED ON MARGINAL SCORE STATISTICS

By the asymptotic connection between the MLE and the score statistic (Cox and Hinkley 1974, p.315), mimicking SumSqB and SumSqBw, we can construct a modified score test based on the marginal model (2). It is easy to verify that the marginal score statistic for $\beta_{M,j}$ under H_0: $\beta_{M,j} = 0$ in (2) is

$$U_{M,j}=\sum_{i=1}^{m} X_{ij}(Y_i-\overline{Y})=X_{\cdot j}'(Y-\overline{Y}),$$

where $\overline{Y}=\sum_{i=1}^{m} Y_i/m$ and 1 is a column vector with elements all 1's. Denote $U_M = (U_{M,1}, \ldots, U_{M,k})'$. We propose a new test based on the sum of the squares of the marginal score statistics,

$$SumSqU=U_M' U_M=(Y-\overline{Y})'XX'(Y-\overline{Y}),$$

where Y and X are the response vector and the design matrix respectively. A weighted form is

$$SumSqUw = U_M' \text{Diag}(I_f)^{-1} U_M,$$

where $I_f = Cov(U_M) = \bar{Y}(1 - \bar{Y})(X - \bar{X})'(X - \bar{X})$ is the expected Fisher information matrix, and $\bar{X}$ is the sample mean of each SNP.

As for SumSqB and SumSqBw, the null distributions of the two test statistics have quadratic forms and can be approximated by $a\chi^2_d + b$. Furthermore, as for SumSqB and SumSqBw, the above two test statistics differ from the usual score test statistic $U_M' I_f^{-1} U_M$ with $I_f = Cov(U_M)$ replaced by a diagonal matrix. Because of the asymptotic equivalence between the Wald test and score test (see equation (5) below for a rigorous proof of their asymptotic equivalence), we expect SumSqB and SumSqU (or, SumSqBw and SumSqUw) to perform similarly, as to be verified numerically next.

Note that the SumSqU test itself can be motivated as a modification to the sum test. It is easy to derive that the score statistic of the sum test for H_0: $\beta_c = 0$ in model (3) is $U_c = \sum_{j=1}^{k} U_{M,j}$, which may be sensitive to varying signs of the components of U_M (and thus of $\hat{\beta}_M$), leading to reduced power; as the SumSqB test, $SumSqU = \sum_{j=1}^{k} U_{M,j}^2$ is invariant to varying signs of $U_{M,j}$'s.

A NEW TEST AS AN ESTIMATED MOST POWERFUL TEST: A UNIFICATION

Below we propose a new test by estimating the most powerful test, which is to be proved to be asymptotically equivalent to the SumSqBw and SumSqUw tests, serving to unify the five new tests proposed here.

It is well known that there is no uniformly most powerful unbiased (UMPU) test on multiple parameters (Cox and Hinkley 1974). On the other hand, under local alternatives $\beta = 0 + \delta b/\sqrt{m}$ with b as a fixed vector and δ a scalar, the most powerful test is based on the test statistic $T_{MP} = b'U$ (Cox and Hinkley 1974, p.319). Because the true β, and hence b are unknown, we aim to estimate T_{MP} by replacing b by an estimate of β:

$$T_{EMP} = \widehat{\beta}_M' U_M.$$

By a standard Taylor expansion of the marginal score equation $U_{M,j}(\hat{\beta}_{M,j}) = 0$, it is easy to verify that

$$\widehat{\beta}_M = I_{M,d}^{-1} U_M + O_P(m^{-1}), \tag{5}$$

where $I_{M,d}$ is a diagonal matrix with $-\partial U_{M,j}(\beta_{M,j})/\partial \beta_{M,j}|_{\beta_{M,j}=0}$ as the *jth* diagonal element, and $U_{M,j}(\beta_{M,j})$ is the score function of marginal model (2). Hence, we have

$$T_{EMP} \approx U_M' I_{M,d} U_M.$$

As before we approximate the asymptotic distribution of T_{EMP} using a scaled and shifted chi-squared distribution.

By the relationship between an observed and an expected information matrix, we have $I_{M,d} \approx \text{Diag}(I_f)$, and thus

$$U_M' I_{M,d} U_M \approx U_M' \text{Diag}(I_f)^{-1} U_M = SumSqUw,$$

which suggests that the SumSqUw is asymptotically equivalent to the estimated most powerful (EMP) test. Similarly, using (5), it is easy to verify that SumSqBw, SumSqUw and EMP are all asymptotically equivalent.

Equation (5) also directly suggests an estimator for $Cov(\hat{\beta}_M)$ as used in deriving the asymptotic distributions of SumSqB and SumSqBw:

$$Cov(\widehat{\beta}_M) = I_{M,d}^{-1} Cov(U_M) I_{M,d}^{-1} = I_{M,d}^{-1} I_f I_{M,d}^{-1},$$

which is the sandwich (also called robust or empirical) covariance estimator used in the GEE for $\hat{\beta}_M$ (Appendix D).

Interestingly, because the score statistic of the sum test for H_0: $\beta_c = 0$ in model (3) is $U_c = \sum_{j=1}^{k} U_{M,j} = 1' U_M$, the sum test can be viewed as an estimated most powerful test by estimating β as $c1$, where c is a constant scalar and 1 is a vector with elements all 1's. This is in agreement with the intuition that the sum test is most powerful when the components of the true β are all equal (and thus with the same sign).

Finally, we note that the usual joint score test, and thus its asymptotically equivalent LRT and Wald test, can be also regarded as estimated most powerful tests. If we use $\hat{\beta}$, instead of $\hat{\beta}_M$ of T_{EMP}, to estimate T_{MP}, we have $T_{EMP,J} = \widehat{\beta} U_M = \widehat{\beta} U \approx U I_f^{-1} U = U_M I_f^{-1} U_M$, where the last two expressions correspond to the score test statistics for joint model (1) and marginal models (2) respectively; the proof is based on two simple facts that, similar to (5), we have $\widehat{\beta} \approx I_f^{-1} U$, and $U_M = U$ as shown next.

GOEMAN'S TEST AND ITS CONNECTION TO THE SUMSQU TEST

Goeman et al (2006) proposed a general empirical Bayes method to test on a large number of parameters, as applicable to the β_j's in the joint logistic regression model (1). A key idea is to treat β_j's as random, rather than as fixed as considered in joint logistic regression. Specifically, $\beta = (\beta_1, \ldots, \beta_k)'$ is assumed a priori from a distribution with mean $E(\beta) = 0$ and covariance $Cov(\beta) = \tau^2 I$. Thus, to test on the original H_0: $\beta = 0$, one can instead test on a new H_0: $\tau^2 = 0$; with the logistic regression model (1), the test statistic turns out to be (Chapman and Whittaker 2008)

$$T_{Go} = \tfrac{1}{2}(U'U - \text{Trace}(I_f)) = \tfrac{1}{2}((Y - \overline{Y})' X X' (Y - \overline{Y}) - \overline{Y}(1 - \overline{Y}) \text{Trace}((X - \overline{X})'(X - \overline{X}))),$$

where U is the score statistic for β under H_0: $\beta = 0$ in joint model (1), and I_f is the corresponding Fisher information matrix: $U = X'(Y - \bar{Y}) = U_M$, and $I_f = Cov(U) = \bar{Y}(1 - \bar{Y})(X - \bar{X})'(X - \bar{X})$.

The null distribution of T_{GO} is unknown and has to be estimated by permutation or simulation, which is computationally intensive.

As noted by Chapman and Whittaker (2008), given Y (as under permutation), the second term of T_{GO} is non-random while the first term has a quadratic form of asymptotic Normal variates. Noting that $U = U_M$, the Goeman test is equivalent to the SumSqU test if permutations are used in the former; more generally, the two tests are different.

The Goeman test was originally proposed for a *large* number of parameters with a relatively small sample size, which is not true in the current context: we have a sample size of hundreds to thousands while the number of parameters is less than 100, e.g. only around 20 in our examples. Hence, the good performance of the Goeman test in the current context as obtained by Chapman and Whittaker (2008) and confirmed here later is somewhat puzzling. Here we offer an explanation. By the specified prior distribution of β_j's as iid with mean 0 and variance τ^2, we have $\widehat{\tau^2} = \sum_{j=1}^{k} \widehat{\beta}_{M,j}^2 / k = SumSqB/k$ as an empirical estimate of τ^2, thus *SumSqB* can be naturally used to test H_0: $\tau^2 = 0$; on the other hand, the asymptotic equivalence between SumSqB and SumSqU, and their close connection to the EMP test, provide a partial justification for SumSqU, and thus for the Goeman test.

RESULTS

SIMULATION SET-UPS

We performed simulation studies by largely following the set-ups given in Wang and Elston (2007) with $k = 4$, 10 and 20 marker SNPs, and sample size $n = 100$, 200 and 500. The disease-causing SNP was assumed to be in the center of the marker SNPs, but was removed from the data. First, we generated a latent vector from a multivariate normal distribution with one of three covariance structures: a compound symmetry (CS) with an equal pairwise correlation $\rho = 0.4$, an AR-1 with the correlation $\rho_{ij} = 0.8^{|i-j|}$ between components i and j, and a correlation matrix with elements ρ_{ij} randomly between 0.3 and 0.7. Second, the latent vector was dichotomized to yield a haplotype with allele frequencies randomly between 0.2 and 0.8 while the minor allele frequency (MAF) for the disease-causing SNP was fixed at 0.2, 0.3 or 0.4. Third, we combined two haplotypes and obtained marker genotype data $X_i = (X_{i1}, \ldots, X_{ik})'$ (and X_{0i} for disease-causing SNP) for subject i. Fourth, the disease status Y_i of subject i was generated from a logistic regression model:

$$\text{Logit}\Pr(Y_i = 1) = \beta_0 + \log(OR) X_{i0}, \tag{6}$$

where we chose $\beta_0 = -\log 4$ to give a background (i.e. not caused by the SNP) disease probability of 0.2, and the odds ratio (OR) ranged from 1 (i.e. no association) to 2. Finally, following the case-control design, we sampled n cases (with $Y_i = 1$) and n controls (with $Y_i = 0$). We excluded the disease-causing SNP, supplying $\{(Y_i, X_i): i = 1,2, \ldots, 2n\}$ as a dataset to various statistical tests. For each set-up, we simulated 1000 datasets, from which we obtained an empirical size or power for each test as its proportion of correctly or incorrectly rejecting its H_0; in particular, the Monte Carlo standard error of an empirical size/power $\hat{p}$ is $\sqrt{\hat{p}(1-\hat{p})/1000} \le 0.016$.

In addition to the sum test and the weighted score test, we also considered a few other ones: the first (denoted as L-G) was a global LRT on H_0: $\beta = 0$ based on the joint model (1); the second was the generalized Hotelling's T^2 test; third, we included the univariate test and Goeman's test based on permutations, U-P and Go-P respectively; finally, we had our five asymptotic tests based on the sum of (weighted) squares of the components of $\hat{\beta}_M$ or of the score statistics.

SIMULATION RESULTS

We considered $k = 4$, 10 and 20 marker SNPs, sample size $n = 100$, 200 and 500, MAF=0.2, 0.3 and 0.4 for the disease-causing SNP, and used nominal significance levels $\alpha = 0.05$ and $\alpha = 0.01$; because the conclusions were similar, to save space, we only gave results for $k = 10$, $n = 500$, MAF=0.2 and $\alpha = 0.05$.

First, with the CS correlation structure suggesting all the marker SNPs contained the same amount of information about the disease-causing SNP, as shown in Table 1, it is obvious that the sum test and the WST were the winners, closely followed by the Goeman test and our proposed five tests. Second, for the AR-1 correlation structure (Table 2), the sum test, the Goeman test and our proposed five tests performed similarly and better than other methods. Third, with random pairwise SNP correlations (Table 3), the conclusion was similar to that for the CS: the sum test and WST performed best, having a slight edge over the Goeman test and our proposed five tests. Fourth, across all the scenarios, our proposed five tests performed similarly to each other and to the Goeman test, which however depends on the use of computationally intensive permutations or simulations. Note that for the AR-1 or the random correlation structure, pairwise correlations ρ_{ij} were not constant, representing a situation in which the working assumption of "a common association strength" underlying the sum test is violated; nevertheless, the key for the good performance of the sum test (and WST) is that the associations between the response and SNPs were (mostly) in the same direction.

HAPMAP DATA FOR GENE CHI3L2

As in Wang and Elston, we conducted a simulation study based on real LD patterns within the CHI3L2 gene as observed in HapMap data. We downloaded the SNPs of the CHI3L2 gene for the 90 CEU (Utah residents with ancestry from northern and western Europe) individuals from the HapMap site in June 2008. As in Wang and Elston (2007), first, we excluded SNPs with MAF ≤ 0.2, leaving 23 SNPs. Second, we did a single imputation for each of the missing genotypes by randomly drawing an observed genotype of the same SNP. Third, we used the dosage coding for the SNPs and tried to minimize the number of negative correlations among them. Fourth, we deleted redundant SNPs that were perfectly correlated with other SNPs. There was still substantial LD among the remaining 17 SNPs as indicated by the distribution of their pairwise Pearson's correlation coefficients, which ranged from −0.388 to 0.989 with the three quartiles Q_1, Q_2 and Q_3 as 0.364, 0.544 and 0.738. Fifth, we repeatedly sampled (with replacement) subjects from the 90 CEU individuals. Finally, as Wang and Elston, we chose the SNP rs2182114 as disease-causing and generated disease indicators from the logistic regression model (6) with the same β_0 and four possible values of OR. The results for two sample sizes are shown in Table 2. It is clear that the sum test, Goeman's test and our proposed five tests performed similarly and were the winners; in particular, they could have an extra 3% power over that of WTS or the univariate test. The global LRT and the Hotelling's T^2 test had low power, and even possibly inflated test sizes.

For the CHI3L2 gene, we also considered including all the SNPs with MAF> 0.05, rather than MAF> 0.2, and followed exactly the same steps, including choosing SNP rs2182114 as disease-causing. In this way, we ended up with 23 SNPs. Among the 253 SNP pairs, we had

42 pairs with negative correlation coefficients. The distribution of the correlation coefficients for all the pairs was summarized below: min= −0.388, the first quartile Q_1 = 0.142, Q_2 = 0.375, Q_3 = 0.602, and max= 0.989. Hence, compared to that with only 17 SNPs, now some SNPs had weaker correlations, possibly leading to the power loss for the sum test (Table 3). The Goeman test and our proposed tests seemed to be the winners, slightly more powerful than the univariate test, then closely followed by the WST. The other two tests again had much lower power. It seems that the weighted version SumSqBw had a slight edge over SumSqB, albeit not necessarily significant; in contrast, the power difference between SumSqUw and SumSqU was hardly visible.

HAPMAP DATA FOR GENE IL21R

As in Chapman and Whittaker (2008), we also considered the region of gene IL21R, in which LD was low. We followed exactly the same steps as for gene CHI3L2 except that as in Chapman and Whittaker (2008), the disease-causing SNP was selected *randomly* and then excluded from the data in each simulation run. At the end, we had 28 SNPs. There were 54 SNP pairs with negative correlation coefficients. A summary of the correlation coefficients among all the SNP pairs was the following: min= −0.564, Q_1 = 0.097, Q_2 = 0.212, Q_3 = 0.459, and max= 0.991. Hence, the LD pattern in this region was more extreme with some negative or low pairwise correlations, which might explain why the power of the sum test (and WST) was much lower than Goeman's test, our proposed five tests and the univariate test (Table 4). It is interesting to note that the univariate test appeared to be the winner, more powerful than Goeman's test and our proposed five tests in this case.

POWER ANALYSIS UNDER A SIMPLE SCENARIO

We have two immediate observations based on the above numerical studies. First, there is no uniform winner, though often Goeman's test and our proposed five tests have high power and sometimes, the sum test is most powerful. In fact, when testing on multiple parameters, it is well known that no (asymptotically) uniformly most powerful unbiased (UMPU) test exists; if and only if the true value of the parameter vector is known, then an asymptotically UMPU test can be constructed (Cox and Hinkley 1974, p.319). In practice, of course, we do not know the true values of the parameters and hence have no UMPU test, suggesting that, without much prior knowledge, we perhaps should try more than one test. Second, it may be surprising that our proposed five tests can be more powerful than the traditional Wald, score or likelihood ratio test. To be concrete, we contrast the SumSqB test and the standard Wald test; the former uses an identity matrix, rather than a covariance estimate as used in the latter. Intuitively it seems that using a good covariance matrix as in the Wald test should be *always* more productive, but our numerical results suggest otherwise. We first thought that it might be due to bad covariance estimates, which however was not likely given that we had a large sample size and relatively a much smaller number of parameters. Below we offer an explanation under a simplified scenario, which in addition vividly illustrates the above point that the power of a test depends on the true parameter values.

We consider $k = 2$ parameters, $\beta = (\beta_1, \beta_2)'$, and its estimate $\hat{\beta} \sim N(\beta, V)$ with V known: $Var(\hat{\beta}_1) = Var(\hat{\beta}_2) = 1/500$ and $corr(\hat{\beta}_1, \hat{\beta}_2) = \rho$. To test H_0: $\beta = 0$, we can compare the following four tests: the Wald test with a test statistic $\hat{\beta}' V^{-1} \hat{\beta}$, the SumSqB test with $SumSqB = \hat{\beta}'\hat{\beta}$, a univariate test with $Max = \max(|\hat{\beta}_1|, |\hat{\beta}_2|)$, and a simplified sum test with $Sum = \hat{\beta}_1 + \hat{\beta}_2$. The Max test is similar to the U-P test except that the multiple testing on the components of β is analytically adjusted, rather than by permutation. The previous sum test (4) is simplified to the current special case if each covariate is standardized to have the same sample variance 1. For any of the above four test with test statistic $T(\beta)$, the rejection region of the test is simply $R(\beta) = \{\beta: |T(\beta)| > c\}$. The constant c can be obtained either analytically

or numerically (Conneely and Boehnke 2007) to guarantee that $\int_{R(\beta)} f_0(\beta) d\beta_1 d\beta_2 = 1-\alpha=0.95$, where $f_0(\beta)$ is the density function of $\hat{\beta}$ under H_0, i.e., that of a bivariate $N(0, V)$.

Figure 1 plots the rejection regions of the four tests with two different values of $\rho = 0.3$ and $\rho = 0.7$. In addition, 1000 random variates drawn from $N(\beta, V)$ are also plotted on each panel. Three non-zero β values were considered: $\beta = (0, 0.05)'$, (0.05, 0.05) or (−0.05, 0.05). Table 5 gives the empirical power of each test under each set-up. It is most illustrative to notice the varying rejection regions of the tests, which determine the power of the tests. For example, along the identity (i.e. 45-degree) line, the rejection region of the Wald test covers that of the other tests while the sum test gives the smallest rejection region, suggesting that among the four tests, the Wald and sum tests are respectively the least and most powerful when $\beta_1 \approx \beta_2 \neq 0$; on the other hand, if $\beta_1 \approx -\beta_2 \neq 0$, the roles of the Wald and sum tests are reversed, as confirmed by Table 5. In particular, it is clear that the use of the (true!) covariance matrix of $\hat{\beta}$ in the Wald test may lose power as compared to the SumSqB test (and sum test) if the components of the true β are indeed close to each other; furthermore, the power difference increases with the correlation between the components.

DISCUSSION

We have studied association mapping with multiple SNPs, possibly in LD. These SNPs may form an LD block, or more generally fall within a sliding window in a GWA scan. All the tests can be equally applied when other covariates are included in a regression model. We have mainly studied two ways of boosting statistical power: one is through dimension reduction as done in the sum test, while the other is to ignore possible correlations among the parameter estimates or score statistics as in the class of the SumSq tests. Interestingly, both the sum test and the class of the SumSq tests, along with the standard joint score test (and its asymptotically equivalent Wald test and LRT), can be all regarded as *estimated* most powerful tests with various estimates of the true association parameters.

A surprising result of this study is the confirmed good performance of the sum test in detecting possible associations between a trait and multiple SNPs in LD under some situations. The basic idea underlying the sum test is to adopt a working assumption on the existence of a common association strength between the trait and each of multiple SNPs. Intuitively, if any of the SNPs is associated with the trait, due to LD, other correlated SNPs should also be associated with the trait; even if their association strengths vary, their "average" (or more generally, some function of them) is likely to differ from 0, thus motivating the sum test with DF=1 on this "average" effect. Under situations with both positive and negative SNP associations with the trait, resulting in that the targeted "average" effect is much weaker than some individual effects, the sum test, as WST, may not be powerful. Largely for this reason, Chapman and Whittaker (2008) did not recommend the use of the sum test. However, because the sum test, as any SumSq test, is so easy and general to use with applicability to either continuous or discrete traits with or without other covariates, and to other study designs as long as a GLM can be employed, we do not dismiss its use. The good performance of the sum test is not limited to simulated data: Zhong and Pan (2008) applied it to a GAW16 Rheumatoid Arthritis (RA) dataset (Plenge et al 2007); for the four LD blocks containing well known RA-associated loci, among the sum test, Goeman's test and univariate test, each test won in one case while all three performed similarly well in the fourth case (with empirical power close 1 due to too strong signals). Nevertheless, as shown here and in Zhong and Pan (2008), the performance of the sum test depends on the coding of SNPs in an unknown way; further studies are needed to identify optimal SNP codings for the sum test for any given dataset.

Our proposed five tests all have a quadratic form as a sum of squares of (weighted) marginal coefficient estimates or score statistics. It is perhaps surprising that, though the covariance matrix of the marginal coefficient estimates or score statistics can be consistently estimated, the new tests use only an identity matrix or only the diagonal elements of the corresponding covariance matrix, deviating from the standard Wald test and score test. We tried with the use of the full covariance matrices, and the results (not shown) were not good; in fact, the resulting Wald or score test was asymptotically equivalent to the LRT (L-G) for the joint model (1), which did not perform well as shown in our numerical examples. We emphasize that the possibly improved (or degraded) performance of our proposed tests over the standard LRT, Wald or score test should generally hold for any regression models, even with the presence of other covariates. Rather than using a sum of squares, we also tried a sum of absolute values of (weighted) marginal coefficient estimates; it performed no better than the sum of squares (results not shown). Importantly, all of the proposed five tests can be conducted by recoursing to their asymptotic null distributions, avoiding the use of time-consuming permutations or simulations as for Goeman's test. We also compared the use of the asymptotic null distributions of the proposed tests to that of their permutational distributions, leading to similar results (not shown). Because of the consistently good and similar performance of Goeman's test and our proposed tests, and computational advantages of our proposed tests, especially with the sum of the squared score statistics, over Goeman's test, we recommend the routine use of SumSqU and SumSqUw, along with the sum test, the global LRT (or Wald or score test) and the univariate test (U-P) for the lack of any uniformly most powerful test.

Acknowledgments

This research was partially supported by NIH grants GM081535 and HL65462. The author is grateful to Drs Tao Wang and Rob Elston for sharing their data and computer program, to Fang Han for help with Figure 1, and to Dr Melanie Wall for helpful discussions that motivated the technical development in Appendices A–C. The author thanks the reviewers for helpful comments.

References

Chapman JM, Whittaker J. Analysis of multiple SNPs in a candidate gene or region. Genetic Epidemiology. 2008; 32:560–566. [PubMed: 18428428]

Chapman JM, Cooper JD, Todd JA, Clayton DG. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum Hered. 2003; 56:18–31. [PubMed: 14614235]

Clayton D, Chapman J, Cooper J. Use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol. 2004; 27:415–428. [PubMed: 15481099]

Conneely KN, Boehnke M. So many correlated tests, so little time! Rapid adjustment of p values for multiple correlated tests. Am J Hum Genet. 2007; 81:1158–1168. [PubMed: 17966093]

Cox, DR.; Hinkley, DV. Theoretical Statistics. Chapman and Hall; London: 1974.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet. 2001; 29:229–232. [PubMed: 11586305]

Fan R, Knapp M. Genome association studies of complex diseases by case-control designs. Am J Hum Genet. 2003; 72:850–868. [PubMed: 12647259]

Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, et al. The structure of haplotype blocks in the human genome. Science. 2002; 296:2225–2229. [PubMed: 12029063]

Goeman JJ, van de Geer S, van Houwelingen HC. Testing against a high dimensional alternative. J Royal Stat Soc B. 2006; 68:477–493.

Huber, PJ. Proceedings if the Fifth Berkeley Symposium in Mathematical Statistics and Probability. Berkeley: University of California Press; 1967. The behavior of maximum likelihood estimates under nonstandard conditions.

Johnson, NL.; Kotz, S. Distributions in Statistics, Continuous Univariate Distributions. 2. Boston: Houghton-Mifflin; 1970.

Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

McCullagh, P.; Nelder, JA. Generalized linear models. Chapman and Hall; London: 1983.

Plenge RM, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N Engl J Med. 2007; 357:1199–209. [PubMed: 17804836]

Qin ZS, Niu T, Liu JS. Partition-ligation EM algorithm for haplotype inference with single nucleotide polymorphisms. Am J Hum Genet. 2002; 71:1242–1247. [PubMed: 12452179]

Roeder K, Bacanu SA, Sonpar V, Zhang X, Devlin B. Analysis of single-locus tests to detect gene/disease associations. Genet Epidemiol. 2005; 28:207–219. [PubMed: 15637715]

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet. 2002; 70:425–434. [PubMed: 11791212]

Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001; 68:978–989. [PubMed: 11254454]

Wang T, Elston RC. Improved power by use of a weighted score test for linkage disequilibrium mapping. Am J Hum Genet. 2007; 80:353–360. [PubMed: 17236140]

Wei Z, Li M, Rebbeck T, Li H. U-statistics-based tests for multiple genes in genetic association studies. Annals of Human Genetics. 2008; 72:821–833. [PubMed: 18691161]

White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982; 50:1–25.

Xiong M, Zhao J, Boerwinkle E. Generalized T^2 test for genome association studies. Am J Hum Genet. 2002; 70:1257–1268. [PubMed: 11923914]

Zhang J-T. Approximate and asymptotic distributions of Chi-squared-type mixtures with applications. Journal of the American Statistical Association. 2005; 100:273–285.

Zhao H, Pfiffer R, Gail MH. Haplotype analysis in population genetics and association studies. Pharmacogenomics. 2003a; 4:171–178. [PubMed: 12605551]

Zhao LP, Li S, Khalid N. Assessing haplotype-based association with multiple SNPs in case-control studies. Am J Hum Genet. 2003b; 72:1231–1250. [PubMed: 12704570]

Zhong, W.; Pan, W. Power comparison of statistical tests of association with multiple SNPs with the GAW16 rheumatoid arthritis data. Research Report 2008–015, Division of Biostatistics, University of Minnesota. 2008. Available at www.biostat.umn.edu/rrs.php

APPENDIX A

RELATIONSHIPS BETWEEN $\hat{\beta}_c$, $\hat{\beta}$ and $\hat{\beta}_M$ IN LINEAR REGRESSION

Denote $Y = (Y_1, \ldots, Y_m)'$, $X = (X_{.1}, \ldots, X_{.k})$, $X_c = \sum_{j=1}^{k} X_{.j}$ and $\beta = (\beta_1, \ldots, \beta_k)'$. Note that $X_c = X1$ with $1 = (1, \ldots, 1)'$. Without loss of generality, assuming that Y and each predictor $X_{.j}$ have been centered at 0 such that the intercept term is 0. Because under the normality assumption, the maximum likelihood estimate (MLE) (and equivalently, the least squares estimate, LSE) of β is $\hat{\beta} = (X'X)^{-1}X'Y$, we have $X'Y = (X'X)\hat{\beta}$. On the other hand, the MLE of β_c is

$$\widehat{\beta}_c=(X_c'X_c)^{-1}X_c'Y=(X_c'X_c)^{-1}1'X'Y=(X_c'X_c)^{-1}1'(X'X)\widehat{\beta},$$

where $(X_c'X_c)^{-1}1'(X'X)$ is a row vector with the sum of its elements being

$$(X_c'X_c)^{-1}1'(X'X)1=(X_c'X_c)^{-1}(X_c'X_c)=1.$$

Thus $\hat{\beta}_c$ is a linear combination of $\hat{\beta}_1, \ldots, \hat{\beta}_k$. Because the sum of the coefficients equals to 1, it can be interpreted that $\hat{\beta}_c$ is a *weighted average* of $\hat{\beta}_1, \ldots, \hat{\beta}_k$. Furthermore, the weights equal to the column sums of $X'X$ divided by $X_c'X_c \geq 0$, hence, the signs of the weights may not be positive; however, if all the $X_{\cdot j}$'s are positively correlated, then the weights are all positive.

With collinearity, it is well known that the components of $\hat{\beta}$ may have large variances, which will not be able to explain the good performance of $\hat{\beta}_c$. In fact, by $Cov(\hat{\beta}) = \sigma^2(X'X)^{-1}$ with $Var(Y_i) = \sigma^2$, we have

$$Var(\widehat{\beta}_c)=(X_c'X_c)^{-1}1'(X'X)Cov(\widehat{\beta})(X'X)1(X_c'X_c)^{-1}=\sigma^2(X_c'X_c)^{-1},$$

which is not directly related to possibly nearly singular $(X'X)^{-1}$. Furthermore, because the MLE from the marginal models (2) is $\widehat{\beta}_M=(X'X)_d^{-1}X'Y$, where $(X'X)_d = Diag(X'X)$, it is easy to see that

$$\widehat{\beta}_c=(X_c'X_c)^{-1}X_c'Y=(X_c'X_c)^{-1}X_c'(X'X)_d\widehat{\beta}_M=\frac{(\sum_{i=1}^m X_{i1}^2,\ldots,\sum_{i=1}^m X_{ik}^2)\widehat{\beta}_M}{\sum_{i=1}^m\left(\sum_{j=1}^k X_{ij}\right)^2}.$$

APPENDIX B

RELATIONSHIPS BETWEEN $\hat{\beta}_c$, $\hat{\beta}$ and $\hat{\beta}_M$ IN GLM

For simplicity, we assume that there is no intercept term; otherwise, it can be handled similarly (but with more complicated notations). In general, there is no closed form solution for MLEs $\hat{\beta}$ and $\hat{\beta}_c$, it is difficult to characterize the relationship between the two. However, with the canonical link function $h(\ )$, e.g., the logit link in logistic regression, the score equations for GLMs analogous to models (1)–(3) satisfy respectively (McCullagh and Nelder 1983)

$$\begin{gathered}\sum_{i=1}^m X_i'\left(Y_i - h^{-1}(X_i\widehat{\beta})\right)=0,\\ \sum_{i=1}^m X_{ij}\left(Y_i - h^{-1}(X_{ij}\widehat{\beta}_{M,j})\right)=0 \quad \text{for } j=1,\ldots,k,\\ \sum_{i=1}^m\sum_{j=1}^k X_{ij}\left(Y_i - h^{-1}(\sum_{j=1}^k X_{ij}\widehat{\beta}_c)\right)=0,\end{gathered}$$

where $h^{-1}(\ )$ is the inverse of the link function and its operation on a vector or matrix is element-wise. Hence,

$$\sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}h^{-1}(\sum_{j=1}^{k} X_{ij}\widehat{\beta}_c)=\sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}Y_i$$
$$=\sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}h^{-1}(X_i\widehat{\beta})$$
$$=\sum_{i=1}^{m}\sum_{j=1}^{k} X_{ij}h^{-1}(X_{ij}\widehat{\beta}_{M,j}),$$

which implicitly determines a relationship between $\hat{\beta}_c$ and $\hat{\beta}$, and $\hat{\beta}_c$ and $\hat{\beta}_M$.

APPENDIX C

CONVERGENCE RESULTS

Now we consider the following question: given that the joint model (1) is the true model, and thus (3) is possibly a misspecified model, what does $\hat{\beta}_c$ converges to as the sample size goes to infinity? First, obviously, by the consistency of MLE under a correct model, $\hat{\beta}$ converges to its true value β^*. Second, based on the property of MLE under a mis-specified model (Huber 1967; White 1982), $\hat{\beta}_c$ converges to β_c^*, which is the value of β_c that minimizes the Kullback-Leibler distance between the true distribution $f(\beta^*)$ of (Y_i, X_i)'s and the distribution $g(\beta_c)$ implied by the misspecified model (3).

Because of the relationship between $\hat{\beta}$ and $\hat{\beta}_c$, there is also an implied relation between β^* and β_c^*, say $\widehat{\beta_c^*}=r(\widehat{\beta^*})$; for example, $r(\ )$ is a linear function in linear regression. Therefore, testing H_0: $\beta_c^*=0$ in model (3) is equivalent to testing H_0: $r(\beta^*) = 0$ on the individual parameters $\beta_1^*, \ldots, \beta_k^*$ in model (1).

APPENDIX D

DISTRIBUTIONAL THEORY OF β_M

We formulate fitting separate marginal models as joint modeling in the context of generalized estimating equations (GEE) to derive the asymptotic distribution of $\hat{\beta}_M$. In GEE, for "cluster" i, denote y_i, = $(y_{i1}, \ldots, y_{ik})' = Y_i 1_k$ with 1_k as a k-dimensional vector containing all ones, and $z_{ij} = j$. Fitting the k marginal models (2) is equivalent to fitting a GEE model:

$$\text{Logit}\,\Pr(y_{ij}=1)=\sum_{j=1}^{k} I(z_{ij}=j)\beta_{M,0j}+\sum_{j=1}^{k} I(z_{ij}=j)X_{ij}\beta_{M,j},$$

with a working independence correlation structure. By the GEE theory (Liang and Zeger 1986), $\hat{\beta}_M$ has an asymptotic Normal distribution with mean β_M and a covariance matrix that can be consistently estimated by a sandwich (or robust) estimator, which is available from commonly used statistical packages such as R and SAS.

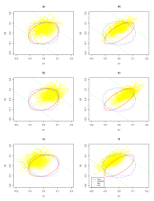

Figure 1.
Rejection regions of the four tests at significance level $\alpha = 0.05$ for bivariate Normal data, along with 1000 points generated from the true distribution for each set-up; see Table 5 for more details on the set-ups.

Table 1

Empirical sizes and powers of various tests with nominal significance level $\alpha = 0.05$ for simulated data with three correlation structures (Corr).

								SumSq				
Corr	**OR**	**Sum**	**WST**	**L-G**	**T^2**	**U-P**	**Go-P**	**Bw**	**B**	**Uw**	**U**	**EMP**
CS	1.0	.051	.053	.047	.049	.046	.047	.044	.046	.044	.043	.045
	1.2	.098	.096	.059	.062	.072	.084	.076	.076	.077	.080	.080
	1.4	.235	.226	.089	.093	.153	.206	.198	.199	.199	.193	.199
	1.6	.395	.399	.145	.153	.239	.366	.357	.363	.358	.356	.360
	1.8	.578	.578	.255	.262	.379	.530	.518	.506	.518	.519	.520
	2.0	.711	.713	.357	.366	.480	.670	.661	.657	.661	.662	.666
AR-1	1.0	.055	.048	.053	.054	.037	.049	.047	.047	.048	.048	.049
	1.2	.132	.115	.078	.080	.107	.131	.123	.123	.124	.125	.127
	1.4	.350	.315	.192	.194	.289	.354	.354	.353	.354	.352	.357
	1.6	.599	.549	.361	.370	.504	.583	.584	.583	.585	.577	.589
	1.8	.798	.743	.549	.560	.704	.796	.782	.779	.783	.785	.785
	2.0	.895	.868	.726	.727	.845	.907	.897	.891	.896	.901	.898
Rand	1.0	.044	.043	.048	.051	.050	.048	.044	.046	.044	.046	.046
	1.2	.134	.130	.078	.079	.087	.121	.116	.113	.116	.114	.117
	1.4	.320	.318	.148	.153	.200	.290	.279	.280	.281	.284	.281
	1.6	.546	.550	.243	.246	.360	.523	.505	.510	.505	.500	.506
	1.8	.753	.748	.383	.391	.537	.729	.716	.717	.718	.721	.720
	2.0	.863	.864	.530	.540	.688	.848	.837	.835	.837	.836	.840

Table 2

Empirical sizes and powers of various tests with nominal significance level $\alpha = 0.05$ for simulated data based on the real LD pattern of gene CHI3L2 with MAF> 0.2 (#SNP=16).

								SumSq				
n	**OR**	**Sum**	**WST**	**L-G**	T^2	**U-P**	**Go-P**	**Bw**	**B**	**Uw**	**U**	**EMP**
200	1.0	.050	.041	.094	.036	.053	.052	.051	.049	.052	.053	.055
200	1.2	.181	.160	.142	.058	.169	.182	.177	.181	.177	.179	.180
200	1.4	.521	.480	.292	.173	.483	.516	.512	.513	.512	.513	.518
200	1.6	.803	.774	.521	.375	.764	.818	.814	.816	.813	.811	.818
500	1.0	.051	.043	.074	.032	.054	.057	.056	.056	.056	.054	.057
500	1.2	.387	.356	.188	.113	.333	.381	.370	.376	.370	.370	.371
500	1.4	.886	.867	.606	.483	.886	.899	.901	.901	.901	.896	.901
500	1.6	.994	.992	.927	.879	.997	.995	.995	.997	.995	.994	.995

Genet Epidemiol. Author manuscript; available in PMC 2010 September 1.

Table 3

Empirical sizes and powers of various tests with nominal significance level $\alpha = 0.05$ for simulated data based on the real LD pattern of gene CHI3L2 with MAF> 0.05 (#SNP=22).

								SumSq				
n	**OR**	**Sum**	**WST**	**L-G**	T^2	**U-P**	**Go-P**	**Bw**	**B**	**Uw**	**U**	**EMP**
200	1.0	.047	.040	.103	.034	.048	.058	.047	.055	.049	.053	.052
200	1.2	.174	.151	.148	.068	.148	.177	.162	.161	.168	.174	.173
200	1.4	.481	.438	.302	.164	.424	.520	.498	.472	.507	.512	.512
200	1.6	.743	.719	.500	.353	.735	.808	.789	.766	.791	.801	.796
500	1.0	.051	.043	.072	.034	.050	.056	.049	.049	.049	.050	.050
500	1.2	.350	.315	.173	.113	.290	.372	.357	.342	.358	.363	.360
500	1.4	.849	.819	.527	.413	.863	.894	.894	.877	.894	.891	.895
500	1.6	.991	.985	.894	.836	.991	.995	.996	.992	.996	.995	.996

Table 4

Empirical sizes and powers of various tests with nominal significance level $\alpha = 0.05$ for simulated data based on the real LD pattern of gene IL21R with MAF>0.2 (#SNP=27).

								SumSq				
n	**OR**	**Sum**	**WST**	**L-G**	T^2	**U-P**	**Go-P**	**Bw**	**B**	**Uw**	**U**	**EMP**
200	1.0	.046	.050	.098	.063	.057	.052	.046	.047	.047	.047	.048
200	1.2	.078	.078	.107	.078	.087	.087	.078	.078	.079	.084	.082
200	1.4	.204	.215	.200	.148	.256	.265	.260	.264	.265	.261	.267
200	1.6	.351	.366	.344	.275	.500	.474	.451	.457	.457	.464	.470
500	1.0	.050	.049	.054	.031	.055	.047	.042	.045	.044	.042	.045
500	1.2	.165	.174	.142	.107	.183	.204	.207	.202	.208	.202	.211
500	1.4	.432	.444	.408	.333	.652	.600	.587	.582	.589	.594	.594
500	1.6	.607	.611	.717	.667	.908	.831	.833	.836	.836	.828	.839

Table 5

Empirical powers of the four tests with nominal significance level $\alpha = 0.05$ for simulated bivariate Normal data.

Set-up	ρ	β	Wald	SumSqB	Max	Sum
a	0.3	$(0, 0.05)'$	0.164	0.143	0.158	0.121
b	0.3	$(0.05, 0.05)'$	0.226	0.258	0.242	0.312
c	0.3	$(-0.05, 0.05)'$	0.373	0.239	0.274	0.059
d	0.7	$(0, 0.05)'$	0.263	0.102	0.158	0.133
e	0.7	$(0.05, 0.05)'$	0.180	0.224	0.222	0.296
f	0.7	$(-0.05, 0.05)'$	0.725	0.171	0.292	0.082