Biostatistics (2017) **18**, 3, *pp.* 477–494
doi:10.1093/biostatistics/kxw060
Advance Access publication on 6 February 2017

A powerful and efficient two-stage method for detecting gene-to-gene interactions in GWAS

JAKUB PECANKA*

Leiden University Medical Center, Department of Medical Statistics and Bioinformatics, Leiden, The Netherlands and VU University, Department of Mathematics, Amsterdam, the Netherlands

j.pecanka@lumc.nl

MARIANNE A. JONKER

VU University Medical Center, Department of Epidemiology and Biostatistics, Amsterdam, The Netherlands and Radboud University medical center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands

INTERNATIONAL PARKINSON'S DISEASE GENOMICS CONSORTIUM (IPDGC)

A full list of authors and affiliations is provided in Appendix B of the Supplementary Materials

ZOLTAN BOCHDANOVITS

VU University Medical Center, Department of Clinical Genetics, Amsterdam, The Netherlands

AAD W. VAN DER VAART

Leiden University, Mathematical Institute, Leiden, The Netherlands

SUMMARY

For over a decade functional gene-to-gene interaction (epistasis) has been suspected to be a determinant in the "missing heritability" of complex traits. However, searching for epistasis on the genome-wide scale has been challenging due to the prohibitively large number of tests which result in a serious loss of statistical power as well as computational challenges. In this article, we propose a two-stage method applicable to existing case-control data sets, which aims to lessen both of these problems by pre-assessing whether a candidate pair of genetic loci is involved in epistasis before it is actually tested for interaction with respect to a complex phenotype. The pre-assessment is based on a two-locus genotype independence test performed in the sample of cases. Only the pairs of loci that exhibit non-equilibrium frequencies are analyzed via a logistic regression score test, thereby reducing the multiple testing burden. Since only the computationally simple independence tests are performed for all pairs of loci while the more demanding score tests are restricted to the most promising pairs, genome-wide association study (GWAS) for epistasis becomes feasible. By design our method provides strong control of the type I error. Its favourable power properties especially under the practically relevant misspecification of the interaction model are illustrated. Ready-to-use software is available. Using the method we analyzed Parkinson's disease in four cohorts and identified possible interactions within several SNP pairs in multiple cohorts.

Keywords: Genetic interactions; Epistasis; Two-stage testing; Case-control design; Parkinson's disease.

*To whom correspondence should be addressed.

1. INTRODUCTION

Functional gene-to-gene interaction, also known as *epistasis*, is a biologically plausible explanation for, at least part of, the "missing heritability" of complex traits (Manolio *and others*, 2009; Hemani *and others*, 2013). Although this possibility is widely acknowledged, a straightforward scan for epistatic effects using conventional statistical techniques has not been very successful at finding epistasis in a genome-wide setting (Niel *and others*, 2015). The conventional approach to searching for interaction within pairs of genetic loci is to model and test the presence of two-way interaction within all pairs available for the analysis using a logistic regression model (LRM) (Cordell, 2002; Marchini *and others*, 2005; Park and Hastie, 2008), where interaction amounts to departure from additive effects of genetic variants at different loci with regard to their global contribution to the phenotype. Such approach requires the estimation of the interaction effect for each pair of loci. Given that the number of pairs grows rapidly with the number of loci, both the computational burden of such estimation and the loss of statistical power due to the need to account for the multiplicity of testing make the conventional methods not very suitable for a truly genome-wide search (Steen, 2012; Niel *and others*, 2015). Moreover, despite the novel methodology developed in the last decade (see Section 2), the need for a scalable and powerful strategy to detect epistasis still remains (Niel *and others*, 2015).

A possible way to avoid or at least lessen the two problems is to first identify pairs of genetic loci that are a priori more likely to be involved in epistasis, and subsequently focus only on these pairs when evaluating interaction. Not only is such two-stage design able to reduce the problem of multiplicity of testing by eliminating a large portion of the tests via the pre-assessment, it is also able to lessen the overall computational burden provided that the pre-assessment technique is computationally simpler than the actual test for interaction. In this article, we propose a formal testing procedure for detecting epistasis in precisely such two-stage manner. The initial stage of our method uses the computationally very simple test to assess whether a pair of loci exhibits two-locus genotype dependence in the population of *cases*. For the pairs that fail the independence test the method proceeds to model interaction via the LRM. The two stages of testing are designed to be independent, which means that a correction for the multiplicity of testing is necessary only in the second stage. In addition to the computational advantages, the non-parametric nature of the independence test provides a crucial robustness of the two-stage method with respect to misspecification of the interaction model, that is, the particular functional form of departure from additivity of the effects. Given that the precise form of interaction is generally unknown in practice, it is a nuisance that the classical parametric approach via the LRM requires its specification. And while its misspecification does not affect the type I error of the interaction test, it can influence the power. Since our two-stage procedure relies on the LRM in a comparatively limited way, the power of our method is much less affected by such misspecification. In order to facilitate the use of the method in practice we developed an open-source software package called EpiDetector, which provides an efficient multi-threaded implementation of our two-stage method, making it readily deployable in a truly genome-wide search for two-way interactions of genetic loci using existing data sets.

The outline of this paper is as follows. In Section 2, we give a motivation and context for our approach and a brief overview of the available methods for detecting epistasis. In Section 3, we give the definition of our two-stage method, a proof of type I error control and a strategy for selecting tuning parameters. In Section 4, we present a simulation study which verifies both the type I error control and illustrates the power performance of the method under several practically relevant scenarios. Finally, in Section 5, we apply the new method to Parkinson's disease data.

2. BACKGROUND

The concept of epistasis has been studied in the literature from several different perspectives (Wade *and others*, 2001). Although a consensus definition of the phenomenon is yet to be reached, functional

interactions between disease causing genetic loci have been shown to be closely related to their *irreducible dependencies* (Wade *and others*, 2001). More specifically, interactions can lead to altered dependence structure in the *haplotypes* (i.e. genetic sequences of alleles that progeny inherits from one parent) of the disease causing loci in the population of *cases*, which translates into deviations of two-locus haplotype frequencies from the products of the corresponding single locus allele frequencies in the population of cases. Such deviations are referred to as *linkage disequilibrium* (LD) (Reich *and others*, 2001). Crucially, if two non-interacting loci are located far apart in the genome (e.g. on different chromosomes), they should be in linkage equilibrium (LE). Therefore, LD between them might be due to interaction (Wade *and others*, 2001). Some authors even go as far as defining interaction of two loci as the presence of LD between them in the subpopulation of cases Wu *and others* (2008). In any case, the link between epistasis and LD in the case population is the motivation behind the independence test in the initial stage (S1) of our two-stage procedure. Instead of focusing on the haplotypes, however, we test dependencies within two-locus genotypes, which is motivated primarily by the fact that the resolution of haplotype frequencies is often non-trivial due to the missing phase information (Li *and others*, 2003), and further justified by the fact that dependence of haplotypes leads to dependence also at the genotype level (Wade *and others*, 2001; Wu *and others*, 2008).

In the second stage (S2) of our procedure we "verify" the results of S1 via a score test in the LRM, where interaction is equivalent to non-nullity of the coefficient at a non-linear term. In S2, we correct for multiplicity by the number of tests in S2 via the Bonferroni method, which guarantees strong control of the *family-wise error rate* (FWER) under arbitrary dependence among the parallel tests in S2 provided we make the test statistics in S2 independent of those in S1. Below we present two ways to achieve such independence. For both we use a partial sample of cases in S1 and we either base the S2 tests only on the remaining (disjoint) portion of the data, which leads to independence simply by the (presumed) unrelatedness of the individuals, or we use all of the available data in S2 while adjusting the S2 statistic in a way that removes the correlation between the two stages and thereby leads to asymptotic independence.

While our procedure combines two classical tests in a relatively novel way, historically, the idea of two-stage testing in genetic applications is not new. Over the past decade similar statistical analyses have been successfully utilized for instance in single-locus genome-wide association studies (GWAS) (Thomas *and others*, 2004; Pahl *and others*, 2009), where multi-stage methods were shown to provide substantial power improvements via efficient allocation of budget. Similarly, several of the approaches proposed in the literature to tackle the problem of epistasis rely on multi-stage analyses. The rich landscape of the available methods for searching for epistasis includes, among others, INTERSNP (Herold *and others*, 2009), epiMODE (Tang *and others*, 2008), EpiGPU (Hemani *and others*, 2011), EpiBLASTER (Kam-Thong *and others* (2011), Gini impurity index based supervised learning algorithm (Li *and others*, 2011), iLOCi (Piriyapongsa *and others*, 2012) and Lewinger's pooled-sample two-stage approach (Lewinger *and others*, 2013). The latter method is in fact quite similar to ours, but unlike our method it uses the pooled sample of cases and controls in S1 and a LRM likelihood ratio test (LRT) in S2. As we show in this article, compared to the case-only approach, the pooling of cases and controls in S1 yields an overall less powerful procedure under several relevant scenarios. Particularly in case-control data sets where the fraction of cases is close to the actual prevalence of the phenotype in the general population, the case-only test is considerably more powerful than the pooled-sample test.

3. Method

Consider a sample of n independent individuals from a population $\mathcal{P}$ that can be split into two disjoint subpopulations according to some binary trait (phenotype) Δ. Denote the number of controls (unaffected, $\Delta = 0$) and cases (affected, $\Delta = 1$) in the sample by m_u and m_a, respectively. Consider two biallelic loci ℓ_1 and ℓ_2 with genotypes AA, Aa, aa and BB, Bb, bb, where each individual has one of the nine possible

two-locus genotypes. In other words, for each locus pair the available data consists of triplets (Δ_i, x_i, y_i), $i = 1, \ldots, n$, where x_i and y_i are numerical representations of the genotypes at ℓ_1 and ℓ_2 of the i-th individual. For simplicity, we assume that x_i and y_i count alleles a and b at the two loci, respectively, and take values $0, 1, 2$. Given a desired genome-wide significance level α, our two-stage procedure proceeds as follows. All candidate pairs of loci enter the initial *screening stage* (S1), where genotype independence is tested using a partial subsample of cases at a suitably chosen level α_1. Denote as K_1 the number of pairs of loci passing through S1 by exhibiting sufficient dependence. The K_1 pairs are subsequently entered into the *verification stage* (S2), which tests for interaction in the LRM at level $\alpha_2 = \alpha/K_1$ using statistics independent of those used in S1.

3.1. *Screening stage tests (S1)*

Denote the number of cases simply as m, fix $\delta \in (0, 1)$ and randomly split the m individuals into two parts of sizes m_1 and m_2 such that $m_1 = \lfloor \delta m \rfloor$ and $m_2 = m - m_1$, where we assume that $m = \lfloor \gamma n \rfloor$ for some fixed $\gamma \in (0, 1)$. This makes indexing of statistics by n and m_1 equivalent, which is useful when formulating the asymptotic results in Appendix A of supplementary material available at *Biostatistics* online. Moreover, we assume that the data is sorted so that the first m of n individuals are cases and the first m_1 cases belong to the subsample used in S1. Denote the two-locus genotype probabilities in the population of cases by $p_{kl} = \mathsf{P}(X = k,\ Y = l \mid \Delta = 1)$, where $k, l = 0, 1, 2$ and the random variables X and Y count alleles a and b at ℓ_1 and ℓ_2, respectively. Moreover, put $p_k = \mathsf{P}(X = k \mid \Delta = 1)$ and $q_l = \mathsf{P}(Y = l \mid \Delta = 1)$. Using the m_1 randomly selected cases we estimate these probabilities via maximum likelihood by $\widehat{p}_{kl} = m_{kl}/m_1$, $\widehat{p}_k = m_{k.}/m_1$, $\widehat{q}_l = m_{.l}/m_1$, where $m_{kl} = \sum_{i=1}^{m_1} I_{\{x_i = k, y_i = l\}}$, $m_{k.} = \sum_{i=1}^{m_1} I_{\{x_i = k\}}$, and $m_{.l} = \sum_{i=1}^{m_1} I_{\{y_i = l\}}$. The null hypothesis of independence is $H_0^{\text{ind}}: p_{kl} = p_k q_l$ for all $k, l = 0, 1, 2$, which we test against the omnibus alternative (i.e. $\neg H_0^{\text{ind}}$) using one of the two different statistics. The first choice is the Pearson chisquare statistic

$$T_n^A \;=\; \textstyle\sum_{k,l=0,1,2} d_{kl}^2 / (\widehat{p}_k \widehat{q}_l), \tag{3.1}$$

where $d_{kl} = \sqrt{m_1}(\widehat{p}_{kl} - \widehat{p}_k\, \widehat{q}_l)$. It is a classical result that T_n^A is asymptotically distributed according to the chisquare-four distribution under H_0^{ind}. Moreover, T_n^A is asymptotically equal to the squared Euclidean norm of $\Sigma_A U_n$, where

$$U_n \;=\; (d_{kl})_{k,l=0,1,2}, \tag{3.2}$$

and $\Sigma_A = \operatorname{diag}\{((p_k q_l)^{-1/2})_{k,l=0,1,2}\}$. The elements of both U_n and the diagonal elements of Σ_A are double indexed by k, l and will be ordered into vectors according to the "row-wise" ordering given by $i = 3k + l$. (The particular ordering of the double indexed components into vectors is not essential as long as it is the same for all such vectors.)

Inspired by Lewinger *and others* (2013), we also propose to use a generalized trend test statistic, which tests independence of genotypes while assuming a particular mode of interaction modelled via a non-constant interaction function $z(x, y)$ (such as $z(x, y) = xy$ or $z(x, y) = \min\{1, xy\}$), where we only care about its values on $\{0, 1, 2\} \times \{0, 1, 2\}$. The function $z(x, y)$ is known as *interaction penetrance model* (Marchini *and others*, 2005; Piriyapongsa *and others*, 2012), or simply *interaction model*. The test statistic is defined as

$$T_n^B \;=\; r_n^2 / s_n^2, \tag{3.3}$$

where $r_n = \sum_{k,l=0,1,2} z(u_k, v_l)\, d_{kl}$, s_n^2 is a suitable estimator of $\mathsf{var}\, r_n$, and u_k, v_l are genotype based weights such as $u_k = k$ and $v_l = l$ for $k, l = 0, 1, 2$. Using U_n we write $r_n = (1, \ldots, 1)\Sigma_B U_n$, where

$\Sigma_B = \text{diag}\{(z(u_k, v_l))_{k,l=0,1,2}\}$, which makes T_n^B (asymptotically) equal to a continuous function of U_n. An estimator of the variance of r_n can be based on Lemma A.4 in Appendix A of supplementary material available at *Biostatistics* online. The reason for writing both T_n^A and T_n^B as functions of U_n will be clear from the definition of the adjusted score statistic below.

3.2. *Verification stage tests (S2)*

Given a fixed interaction penetrance function $z(x, y)$, the interaction LRM assumes

$$P(\Delta = 1 \,|\, X, Y) \;=\; \left[1 + \exp(-\beta_0 - \beta_1 X - \beta_2 Y - \beta_3\, z(X, Y))\right]^{-1}. \tag{3.4}$$

Under (3.4) we say that the loci ℓ_1 and ℓ_2 *do not interact* if the hypothesis $H_0^{\text{epi}} : \beta_3 = 0$ is true. H_0^{epi} can be tested using the score statistic (with estimated parameters)

$$C_n \;=\; n^{-1/2} \textstyle\sum_{i=1}^{n} (\Delta_i - \widehat{\Psi}_i)\, \mathsf{z}(x_i, y_i), \tag{3.5}$$

where $\widehat{\Psi}_i \;=\; (1 + \exp(-\widehat{\beta}_0 - \widehat{\beta}_1 x_i - \widehat{\beta}_2 y_i))^{-1}$, $\mathsf{z}(x, y) \;=\; (1, x, y, z(x, y))'$ and $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2$ are the null hypothesis ML estimators of the parameters $\beta_0, \beta_1, \beta_2$. According to classical theory, under H_0^{epi} the score statistic with true values of the parameters is asymptotically normal with zero expectation and the Fisher information matrix (FIM) as its variance. Similarly, by Theorem A.2 of Appendix A of supplementary material available at *Biostatistics* online, C_n is asymptotically normal with zero expectation and a related variance matrix.

Disjoint score statistic Using the $n - m_1$ individuals that were not used in S1 we define the "disjoint" score statistic (with estimated parameters) D_n analogously to (3.5). Due to the independence of individuals, D_n is independent of both T_n^A and T_n^B, which means that we can use D_n in conjunction with either T_n^A or T_n^B to obtain a two-stage procedure (see Section 3.3) which strongly controls FWER (see Section 3.5). We also note that the asymptotic variance of D_n follows directly from Theorem A.2.

Adjusted score statistic In addition to D_n, we define the "adjusted" score statistic as

$$A_n \;=\; C_n - \mathsf{B}(U_n - \sqrt{\delta/(1-\delta)}\, \widetilde{U}_n), \tag{3.6}$$

where $\widetilde{U}_n$ is based on the remaining m_2 cases in an analogous way to U_n of (3.2) and $\mathsf{B} = \mathsf{C}_{SU}\mathsf{V}_U^-$ with V_U^- denoting a pseudoinverse of V_U, which is the asymptotic variance matrix of U_n, and C_{SU} denoting the asymptotic covariance matrix of C_n and U_n. The centering term $\widetilde{U}_n$ in (3.6) ensures that A_n has zero expectation under H_0^{epi}. The question of evaluating A_n is addressed in Section A.3 in Appendix A of supplementary material available at *Biostatistics* online. The matrix B can be calculated using the expressions for V_U and C_{SU} provided by Lemmas A.4 and A.5 and Theorem A.2 (all in Appendix A of supplementary material available at *Biostatistics* online). Theorem A.3 in Appendix A of supplementary material available at *Biostatistics* online gives the asymptotic variance of A_n and shows that under H_0^{epi} the joint asymptotic distribution of A_n and U_n is normal and the asymptotic covariance of A_n and U_n is zero, which makes A_n and U_n asymptotically independent. Consequently, A_n is independent of both T_n^A and T_n^B (or any continuous function of U_n).

In the definition A_n, we regressed C_n onto the full vector U_n in order to make A_n independent of both T_n^A and T_n^B. However, for independence of A_n and T_n^B it is sufficient to regress onto $\Sigma_B U_n$ and doing so might have slight power benefits since $\Sigma_B U_n$ generates a smaller dimensional space compared to U_n as soon as $z(x, y)$ yields zeros for some combination of x and y. In the accompanying software, we implemented the slightly more powerful version when T_n^B is selected.

3.3. *Two-stage testing scheme*

In a search for epistasis, we start with K pairs of loci. Denote the set of all loci pairs as $\mathcal{K} = \{1, \ldots, K\}$. For each pair, we formulate null hypotheses H_{0k}^{ind} (independence of the k-th pair) and H_{0k}^{epi} (no interaction of the k-th pair). For a fixed S1 level $\alpha_1 \in (0, 1)$, we calculate T_{nk} for each $k \in \mathcal{K}$ using the same m_1 cases, where T_{nk} is either T_n^A or T_n^B depending on the user's choice. In S1, we reject H_{0k}^{ind} if $T_{nk} \geq \tau_{\alpha_1}$, where τ_{α_1} is the critical value corresponding to α_1. This yields a set $\mathcal{K}_1 \subset \mathcal{K}$ of $K_1 = |\mathcal{K}_1|$ "flagged" loci pairs. If $K_1 > 0$, then for each pair in $\mathcal{K}_1$ we perform an S2 test at the level of significance $\alpha_2 = \alpha/K_1$ using the statistics R_{nk}, where R_{nk} is either D_n or A_n corresponding to the k-th pair of loci. Given that both D_n and A_n are asymptotically normal, for a two-sided test the standardized absolute values of R_{nk} are compared with the standard normal critical value ξ_{α_2} (i.e. the $(1 - \alpha_2/2)$-quantile) and if ξ_{α_2} is exceeded we reject H_{0k}^{epi} and claim interaction for the k-th pair of loci.

3.4. *Benchmark two-stage testing procedure*

An interesting alternative two-stage procedure was proposed by Lewinger *and others* (2013), who suggest to pool all available cases and controls into a single sample and use it to calculate T_n^D analogously to (3.3). Using the pooled sample, we further define T_n^C analogously to (3.1). The method of Lewinger *and others* (2013) is based on the combination of T_n^D and the likelihood ratio test (LRT). However, the LRT does not seem optimal for a large-scale application such as a genome-wide search for epistasis, because, unlike the score test, it requires parameter estimation both in the null and the full model. Since the two tests are asymptotically equivalent (equally powerful) it seems beneficial to use the score test instead. For the special case of $z(x, y) = xy$ Lewinger *and others* (2013) argue that the statistic T_n^D and the LRT statistic (and therefore also the score statistic C_n) are asymptotically independent. For a general penetrance function $z(x, y)$ a proof of asymptotic independence of both the score test and LRT statistics with both T_n^C and T_n^D can be formulated along the lines of the proof of Theorem A.3 of Appendix A of See supplementary material available at *Biostatistics* online . Consequently, the asymptotic independence permits the two-stage testing scheme of Section 3.3 for T_n^C or T_n^D combined with C_n. We use these two two-stage methods as performance benchmarks for the two-stage procedures based on D_n and A_n in the simulation study of Section 4.

3.5. *Error control in two-stage testing with independent stages*

Due to the (asymptotic) independence, the combinations of either D_n and A_n with either T_n^A or T_n^B lead to valid (i.e. error controlling) two-stage procedures described in Section 3.3, and so do the combinations of C_n with either T_n^C or T_n^D (see Table 1 for the permissible combinations of statistics). A theoretical proof of sufficient error control is formulated in Section A.1 in Appendix A of supplementary material available at *Biostatistics* online. The result is also verified numerically in the simulation study of Section 4.2, where we show sufficient type I error control of the S2 tests regardless of the level α_1 used in S1 for all procedures

Table 1. *An overview of possible combinations of statistics (in each column) that lead to "valid" methods, i.e. methods with independent stages. The statistic T_n^A is defined in (3.1), T_n^B in (3.3), T_n^C and T_n^D are the pooled-sample analogues of T_n^A and T_n^B, respectively, D_n is the disjoint sample score statistic (see Section 3.2), and A_n is defined in (3.6).*

	Pca_4–DS	Pca_4–AS	Pca_1–DS	Pca_1–AS	Ppo_4–CS	Ppo_1–CS
Stage 1	T_n^A	T_n^A	T_n^B	T_n^B	T_n^C	T_n^D
Stage 2	D_n	A_n	D_n	A_n	C_n	C_n

in Table 1. In addition, in Section 4.2, we also illustrate the danger of using a two-stage procedure with dependent tests. We show that the combination of the full sample score statistics C_n with S1 tests based on T_n^A and T_n^B leads to strongly anti-conservative testing procedures.

3.6. *Power within two-stage testing*

In order to provide a fair comparison between the presented methods, we focus on their power performance within a multiple testing context instead of the more usual single test comparison. We refer to such power as MTC-power. If power was considered from the classical power perspective (i.e. single-stage testing with one test), it is easy to see that both A_n and D_n have comparatively lower power than the full sample score statistic C_n. In the case of D_n, this is a straightforward consequence of the smaller sample size (compared to C_n) on which it is based. In the case of the adjusted score A_n the lower power is caused by the inflated variance due to the presence of the regression term. However, the relative power performance changes dramatically if the two statistics A_n and D_n are considered within a two-stage testing scheme applied to a multiple testing scenario with a relatively small number of false hypotheses and unknown (and therefore likely misspecified) interaction model. In Section 4.3, we show that in such setting the S1 tests can (for suitably chosen α_1 and δ) dramatically lower the multiple testing burden faced by the S2 tests and their insensitivity to model misspecification can result in vastly superior procedures Pca_4–DS and Pca_4–AS. For a well-specified interaction model the power advantage of Pca_4–DS and Pca_4–AS is less pronounced or even vanishes. In such case, however, if the case/control ratio in the sample is different from the population prevalence then the pooled-sample based procedures Ppo_4–CS and (especially) Ppo_1–CS perform best among the considered procedures.

3.7. *Choice of S1 level and sample split ratio*

As we showed above, the choice of α_1 is irrelevant with respect to the control of FWER by any two-stage procedure with independent stages. On the other hand, in terms of power it is quite important to choose a suitable value for α_1. In the case of procedures based on D_n or A_n, we must also choose suitable values for the sample split ratio δ, which we define as $\delta = m_1/m$. Therefore, both α_1 and δ can be viewed as tuning parameters of the methods. These parameters can for instance be selected based on a power simulation under some reasonable setting. Alternatively, they can be chosen using a theoretical argument presented in Section A.4 in Appendix A of supplementary material available at *Biostatistics* online, where the optimization requires as input the distribution of genotypes in the general population (via π_{kl}), the values of β_0, β_1, β_2, and finally the value of β_3 towards which the method is tuned.

3.8. *A possibility of control-based pre-assessment*

In Section 2, we formulated an argument for performing the S1 tests using the cases. However, if two interacting loci are in LE in the general population, any LD among the cases necessarily leads to LD also among the controls. In other words, it is possible to base the S1 tests on the controls instead of the cases. This seems desirable given that many existing data sets are *imbalanced* in favor of the controls, which makes their usage in S1 ideal. Interestingly, our simulation study (results not shown) revealed that in a setting with high population prevalence of cases and highly imbalanced data towards the controls, the two-stage procedures with S1 tests performed using the controls are overall the best performers, regardless of whether the interaction model is well-specified or not. The reason for this is the ever decreasing marginal power gains by the score test in a setting with an increasing number of controls when the number of cases remains fixed (Foppa and Spiegelman, 1997), or vice versa. Here we focus on the case-only based methods because most relevant diseases have prevalence much lower than 50%, which makes the degree of LD

induced among the controls much lower than among the cases, thus making the control-based methods substantially less powerful in such a setting.

4. SIMULATION STUDY

In order to investigate the behavior of the two-stage methods, we performed a simulation study. First, we verified type I error control by the two-stage methods listed in Table 1. Then, in order to illustrate the potential anti-conservativeness of two-stage methods that improperly combine dependent test statistics, we also considered two "invalid" two-stage procedures Pca_4–CS and Pca_1–CS, which use T_n^A and T_n^B in S1 and the full sample score statistic C_n in S2. Second, we performed a power comparison of the methods, where we also focused on the influence of the penetrance model (i.e. $z(x,y)$). We simulated a large number of case-control data sets with interacting loci using several penetrance models (listed in Table 2) and analyzed it using model A, where the interaction term is multiplicative in the minor allele counts. This is a popular choice when modelling epistasis via LRM in practice, however, it is often assumed by default without much justification. As we show below, a severe misspecification of $z(x,y)$ can have a strong influence on the performance of parametric tests that rely on it (e.g. the score test). Crucially, the Pearson-type statistics T_n^A and T_n^C are non-parametric and do not assume any particular shape for $z(x,y)$, which makes the two-stage procedures based on T_n^A and T_n^C much more robust towards misspecification of $z(x,y)$.

4.1. *Software*

The data simulation and analysis in this article were performed using an open-source software package called EpiDetector, which we developed specifically for this purpose. EpiDetector is a command line tool similar in usage to the popular genetic data analysis package PLINK (Purcell *and others*, 2007). The software can be readily deployed to perform a GWAS search for interactions using the methods of this paper. The source code is available upon request.

4.2. *Type I error control*

Using a fixed case count $m_a = 1000$ and three different control counts $m_u = 500$, $m_u = 3000$, and $m_u = 7000$, we simulated 10 case-control data sets, each with 1000 non-interacting loci and a randomly generated case-control status, which we subsequently analyzed using all of the considered two-stage methods, namely Pca_4–DS, Pca_4–AS, Pca_1–DS, Pca_1–AS, Ppo_4–CS, Ppo_1–CS. For the former four methods,

Table 2. *Various choices of the interaction penetrance function $z(x,y)$. The tables represent the interaction models A, C, I, J, O, P, Z (top left to right; bottom left to right; the names were chosen completely arbitrarily) and shows values of $z(x,y)$ for the nine possible combinations of genotypes.*

A	y = 0	1	2
x = 0	0	0	0
1	0	1	2
2	0	2	4

C	y = 0	1	2
x = 0	0	1	0
1	1	0	1
2	0	1	0

I	y = 0	1	2
x = 0	0	0	0
1	0	1	1
2	0	1	1

J	y = 0	1	2
x = 0	0	0	0
1	0	2	0
2	0	0	3

O	y = 0	1	2
x = 0	0	0	0
1	0	1	2
2	0	2	3

P	y = 0	1	2
x = 0	0	0	0
1	0	0	1
2	0	1	1

Z	y = 0	1	2
x = 0	0	0	0
1	0	0	1.2
2	0	1.2	2.5

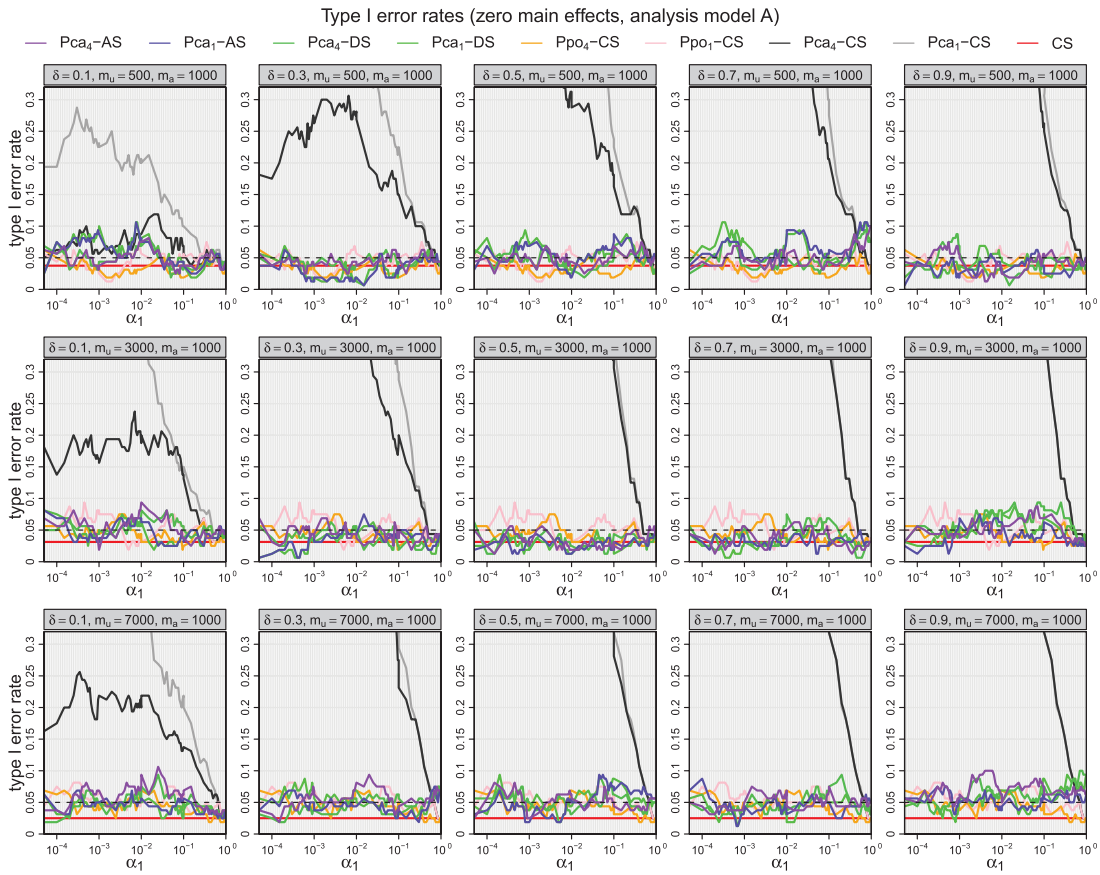

Fig. 1. Type I error rates for various two-stage methods with $z(x, y) = xy$ as functions of α_1 with five different values of δ ranging between 0.1 and 0.9 (left to right) and $m = 500$, $m = 3000$, $m = 7000$ (top to bottom), $m_a = 1000$ and $\beta_1 = \beta_2 = 0$). In each plot the y-axes have the same range from 0 to 0.3.

we considered five different values of δ ranging between 0.1 and 0.9. Figure 1 shows the S2 type I error rates of various two-stage methods as functions of α_1. The error rates were calculated only among the tests for which independence was rejected by the S1 tests at level α_1. The plots confirm our expectations about proper type I error control for all combinations of sample sizes and δ for all six "valid" two-stage methods. Moreover, it is clear that for almost all combinations of δ and sample sizes the observed type I error rates of the "invalid" procedures Pca_1–CS and Pca_4–CS skyrocket as α_1 decreases. In fact, in most plots the error rates very quickly approach one (not visible due to the y-axis upper bound of 0.3) providing a very clear warning about improperly designed two-stage methods with dependent stages.

4.3. *Power performance*

In the power simulation, we considered all seven penetrance models A, C, I, J, O, P, Z of Table 2 with $m_a = 2000$ cases and $m_u = 3000, 7000, 11000$ controls, where we aimed to reflect an often encountered imbalance in favor of controls in real world data sets. Using the LRM we simulated cases and controls from a population with prevalence of cases of around 5% until the desired sample sizes were reached by discarding the excess cases and/or controls (i.e. case-control sampling). During the simulation the main

effects were set to zero (i.e. $\beta_1 = \beta_2 = 0$). For each penetrance model and each combination of sample sizes, we also considered three minor allele frequency (MAF) scenarios combining MAFs 0.35 and 0.1. Using the LRM we simulated 5000 independent loci pairs for each such scenario with interaction effect parametrized via $OR = \log \beta_3$. Each simulated data set was then analyzed using penetrance model A, that is, with $z(x, y) = xy$. Each such test was assumed to be performed in the context of 10^8 tests, thus the single-stage score test p-values were Bonferroni-corrected by 10^8, while the S2 tests were performed at level $\alpha_2 = 0.05/(10^8\alpha_1)$. Such S2 levels were motivated by the fact that $10^8\alpha_1$ is the expected number of tests in S2 under the combined null hypothesis of independence and no interaction. Given that in a large epistasis search strong LD and epistasis are rare provided pairs of loci on the same chromosome are not included in the two-stage analysis, the actual number of tests in S2 should in fact be near $10^8\alpha_1$. Finally, in the simulation study the values of α_1 and δ were optimized to get the best empirical power for each setting. This of course would not be directly possible in a real world data analysis, however, the theoretical method for determining suitable values for the tuning parameters discussed in Section 3.7 provides a way to solve this issue.

Figures 2–4 show the results. The total of 63 power plots provides a substantial insight into how the various methods stack up against each other and against the classical single-stage test for different case-control ratios and MAF combinations under various degrees of misspecified interaction model during the analysis ranging from well-specified to strongly misspecified. Given the high number of considered scenarios it is not straightforward to summarize the observed influences of parameter choices. However, it is very clear that the various considered methods are influenced by these choices to a varying degree. This is to say that in general it is clear that model misspecification alters the relative performance of the methods and that the parametric single-stage CS test. Moreover, the parametric two-stage methods that require specifying $z(x, y)$ in both stages (i.e. Pca_1–AS, Pca_1–DS, Ppo_1–CS) suffer more pronounced decrease of power due to model misspecification compared to the methods that do not require it (i.e. Pca_4–AS, Pca_4–DS, Ppo_4–CS). Additionally, it is extremely noteworthy that under model misspecification (with fixed population prevalence and fixed allele frequencies) not all methods benefit from an overall larger sample size. This is particularly evident for the pooled-sample methods Ppo_4–CS and Ppo_1–CS, where under all scenarios with misspecified model the two methods in fact lose power when the number of controls is increased. Although this might seem somewhat counterintuitive, it is a direct consequence of the pooled-sample tests losing power as the fraction of cases in the sample approaches the prevalence of cases in the population. In the three considered case-control ratios in the simulation, by keeping the number of cases fixed and increasing only the number of controls the sample prevalence of cases approaches the population prevalence (around 5%), which results in a vast loss of overall power of the two-stage procedures based on the pooled-sample test of independence. This effect is most pronounced under the high MAF scenario in Figure 2, but it is present in the other two figures as well. On the other hand, the newly proposed two-stage methods, especially the partially non-parametric Pca_4–DS and Pca_4–AS very desirably benefit from the increased sample size. Moreover, not only are Pca_4–DS and Pca_4–AS positively affected by the increased sample size, they are the top performers overall in many of the model misspecification scenarios and their advantage over the competition increases with increased imbalance towards the controls (e.g., simulation models C, J, P, Z). This is the case for both Pca_4–DS and Pca_4–AS under the high MAF scenario of Figure 2, while for the other two MAF scenarios in Figures 3 and 4 Pca_4–DS takes the upper hand over Pca_4–AS. Given that both Pca_4–DS and Pca_4–AS yield reasonable performance compared to the other methods also under the correctly specified model, it seems safe to say that the two methods, and especially Pca_4–DS, are very strong performers and overall can be seen as superior to the other considered methods.

As for the comparison among Pca_4–DS and Pca_1–DS versus Pca_4–AS and Pca_1–AS it seems fair to say that the disjoint tests (Pca_4–DS and Pca_1–DS) generally outperform the adjusted tests (Pca_4–AS and Pca_1–AS). In terms of power, under many considered scenarios the two methods are equivalent, however, under several scenarios the disjoint test was clearly superior. We also observed that the adjusted score test

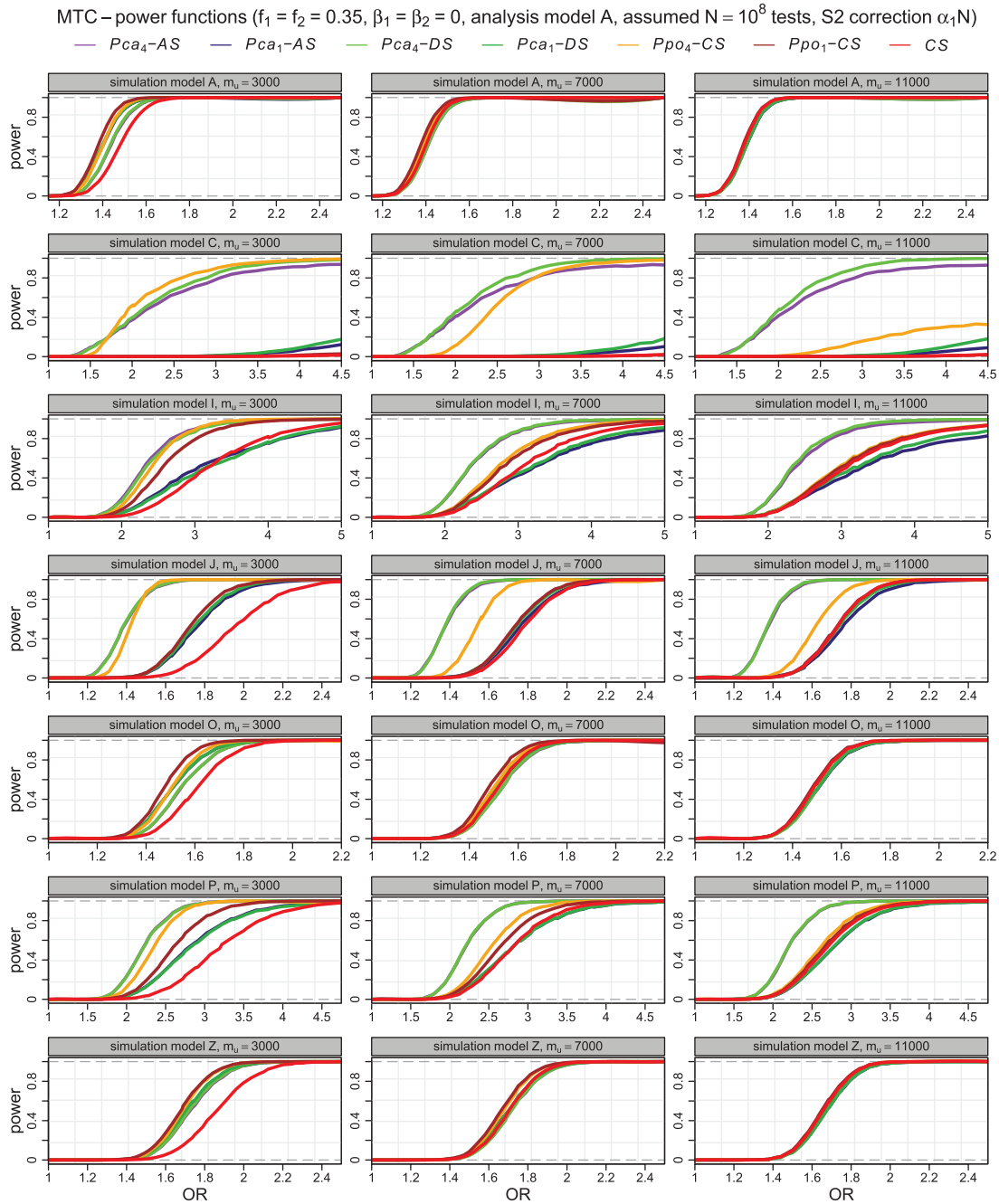

Fig. 2. Empirical MTC-powers at optimal input parameters α_1 and δ for the nine testing methods as functions of interaction odds ratio $OR_3 = \log \beta_3$ for simulation models A,C,I,J,O,P,Z with analysis model A, $m_a = 2000$ with MAF close to 0.35 for both loci in each test, phenotype population prevalence of 5% and $\beta_1 = \beta_2 = 0$.

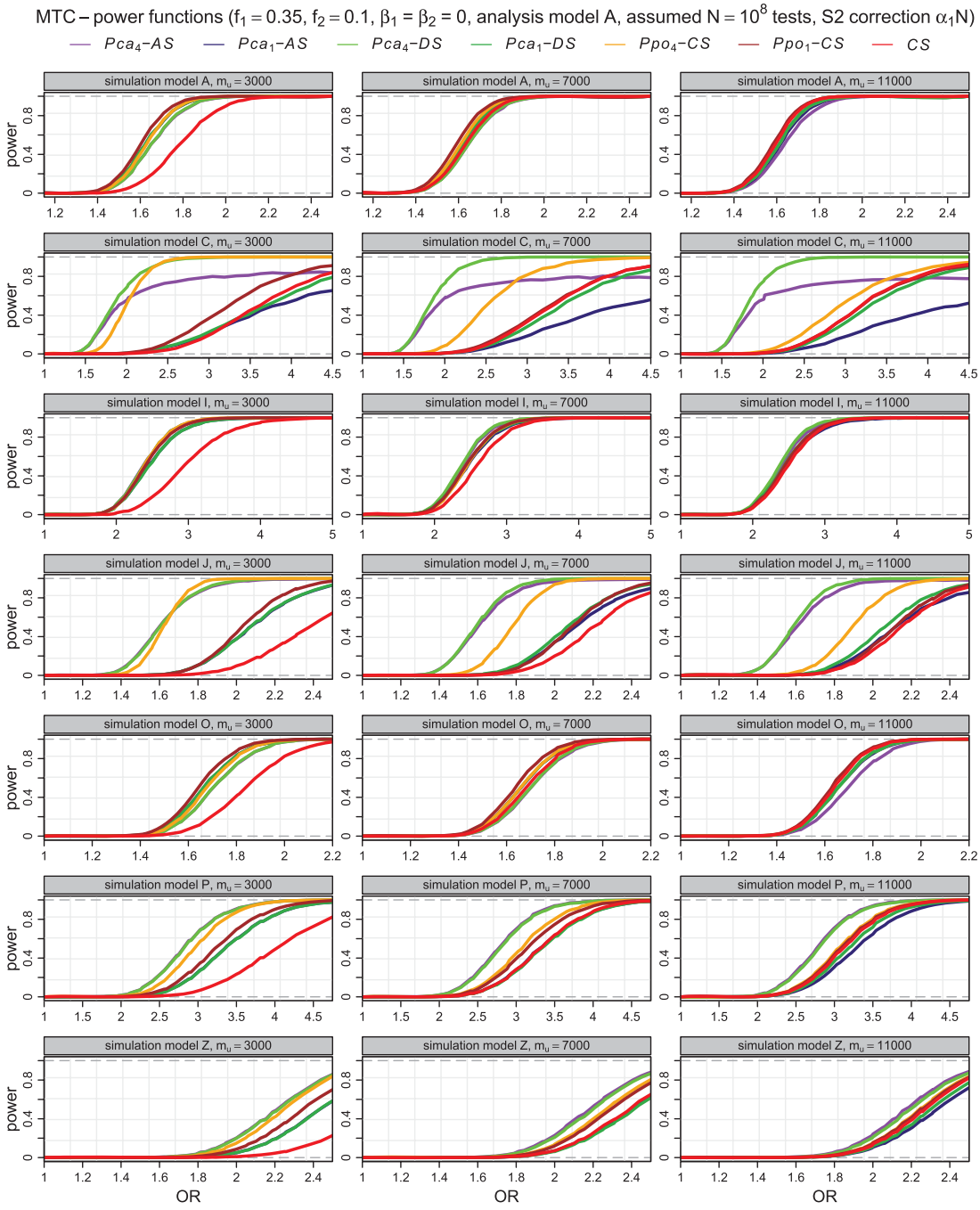

Fig. 3. Empirical MTC-powers at optimal input parameters α_1 and δ for the nine testing methods as functions of interaction odds ratio $OR_3 = \log \beta_3$ for simulation models A,C,I,J,O,P,Z with analysis model A, $m_a = 2000$ with MAF close to 0.35 and 0.1 for the two loci within each test, phenotype prevalence of 5% and $\beta_1 = \beta_2 = 0$.

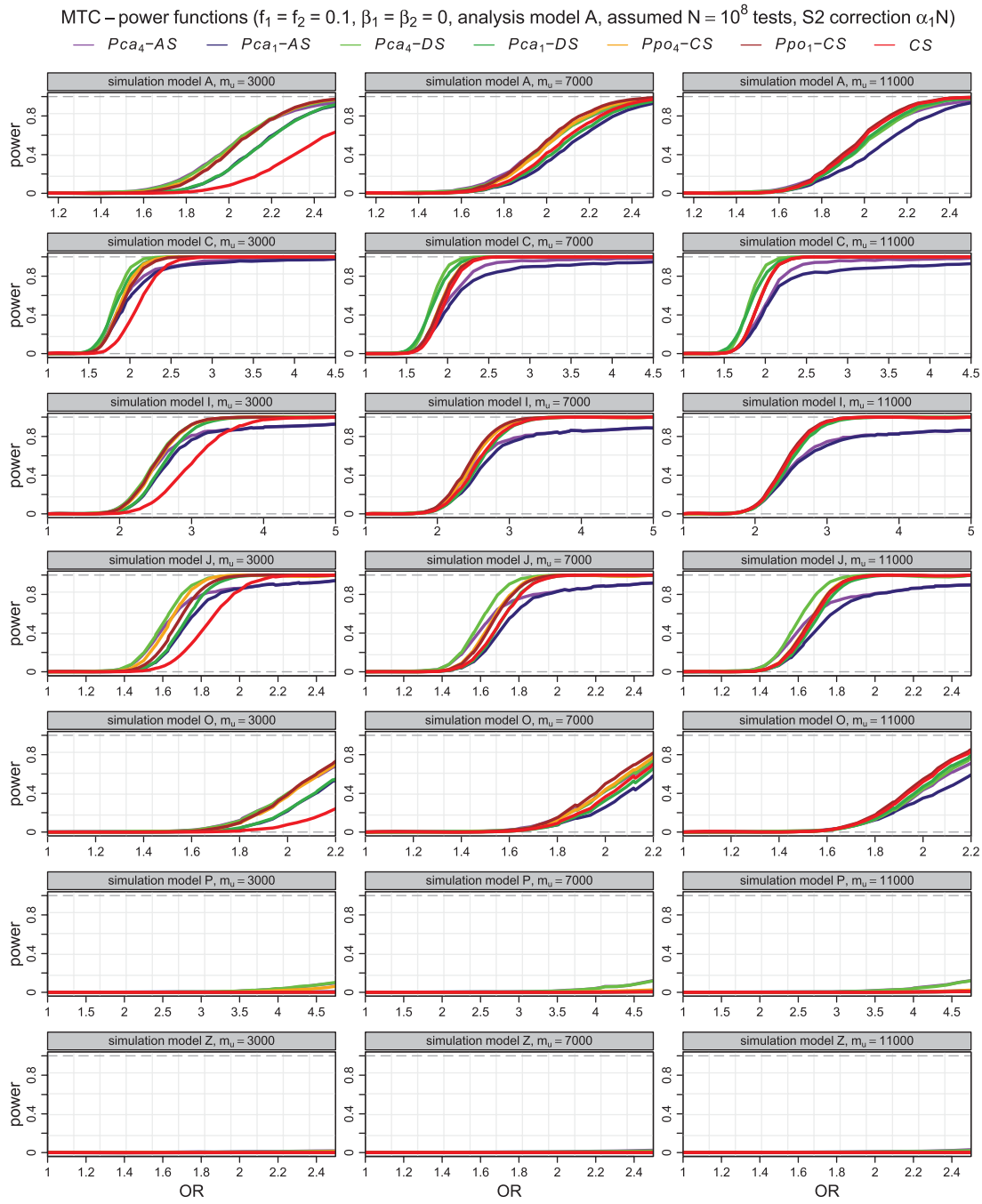

Fig. 4. Empirical MTC-powers at optimal input parameters α_1 and δ for the nine testing methods as functions of interaction odds ratio $OR_3 = \log \beta_3$ for simulation models A,C,I,J,O,P,Z with analysis model A, $m_a = 2000$ with MAF close to 0.1 for both loci within each test, phenotype prevalence of 5% and $\beta_1 = \beta_2 = 0$.

is more susceptible to numerical instability problems especially for very small allele frequencies. This is probably due to the more involved nature of the statistic, which requires an estimation of a higher number of variance and covariance matrices.

5. APPLICATION

We applied the two-stage methods to the analysis of four Parkinson's disease (PD) cohorts from the IPDGC. For a detailed description of the data see Pankratz *and others* (2009), IPDGC (2011); IPDGC and WTCCC2 (2011). For the analysis we extracted those SNPs that could be assigned to protein coding genes, which resulted in 11382 (USA-NIA), 13484 (NL), 13116 (GE) and 10602 (NINDS-CIDR) gene-based SNPs to be tested for gene-to-gene interactions on PD status. Table 3 shows the numerical summary of the data sets. Representing the genotypes numerically by counting the rarer allele at each SNP, we deployed six two-stage methods, namely Pca_4–DS, Pca_1–DS, Pca_4–AS, Pca_1–AS, Ppo_4–CS, Ppo_1–CS. In order to limit the influence of possible population differences among the four data sets, instead of merging the data sets we performed the tests for each SNP-pair separately in each data set where the given pair was present. Using 70% of the cases in S1 (i.e., $\delta = 0.7$) for all four data sets yielded up to four S1 p-values for each SNP-pair. We subsequently combined these p-values using a Fisher-type weighted combination method (see Section A.5 in Appendix A of supplementary material available at *Biostatistics* online) with the weights simply based on case counts underlying each S1 p-value. This yielded a single combined S1 p-value for each SNP-pair, which we compared with a selected level α_1 for each method. Based on the observed optimality of S1 levels in the simulation we selected $\alpha_1 = 10^{-6}$ for Pca_4–DS, Pca_1–DS, Pca_4–AS and Pca_1–AS, while for Ppo_4–CS, Ppo_1–CS we used $\alpha = 10^{-4}$. The S2 tests were also performed separately for each data set (where present) and the resulting p-values were again combined using the same combination method except this time the weights were based on the harmonic means of the case and control counts used in each S2 test, since the harmonic mean gives the effective rate of convergence of the score statistic in the LRM (Foppa and Spiegelman, 1997). Given that the resulting combined p-values in S1 were independent of the combined p-values in S2, it was sufficient to correct the combined S2 p-values only by the number of tests in S2. In addition to the two-stage methods, we also calculated the single-stage score test p-values in each cohort, which we combined using the same p-value combination method. Given that there were 92 250 923 tests in total, the resulting combined p-values were compared with the Bonferroni corrected level $5.4 \cdot 10^{-10}$.

Using the corrected combined S2 p-values we identified several genome-wide significant SNP pairs, while the single-stage tests yielded none. Aiming at maximizing the unambiguity of the results, we took advantage of the availability of multiple cohorts and report only the SNP pairs for which the combined S2 test was significant and nominally supported by more than one cohort. With multiple cohorts it would have been possible to attempt replication directly by reserving one of the data sets, however, given that majority of the replication data set would not have been used at all, attempting replication would have come at a relatively high cost in terms of power. Therefore, we opted for a single analysis, but present

Table 3. *Numerical summaries of the analyzed Parkinson's disease data sets.*

Cohort	SNPs	Cases	Controls	Genotyping platform
USA-NIA	11382	971	3034	Illumina HapMap 550
Dutch (NL)	13484	772	2024	Illumina610-Quad
German (GE)	13116	742	944	Illumina HumanHap550
NINDS-CIDR	10602	876	857	IlluminaCNV370

only the multi-cohort supported findings. While technically this does not amount to a replication of our findings, we suggest that it increases their credibility substantially. Note that a significant S1 test was not required to be supported by multiple cohorts, which is consistent with the rationale for the pre-tests as providing primarily circumstantial evidence that guides the verification stage analyses. Any reasonable a priori suggestion of a genetic interaction would then be formally tested in the S2 stage, hence nominal support from multiple sources was only required for this stage.

Our analysis identified two SNP pairs with multi-cohort evidence of epistasis. For both pairs the multi-cohort evidence for interaction comes from the Dutch (NL) and German (GE) cohorts where both of the incriminated SNPs in each pair were present. Given that all SNPs in the analysis were gene based, our results yield evidence for genome-wide significant gene-to-gene interaction for the risk to develop PD (Table 4). The two identified gene pairs are DUSP12 in combination with DOCK4 and UBE2J1 in combination with GPR107. Interestingly, at least three out of these four genes have biological functions that are closely related to the pathogenesis of PD. DUSP12, also known as hYVH1, is a dual-specificity phosphatase that was shown to physically interact with Hsp70 in order to prevent heat-shock induced cell death (Sharda *and others*, 2009). Hsp70 in turn has been shown to affect PD pathogenesis by affecting aberrant alpha-synuclein aggregation (Zhang and Cheng, 2014; Gao *and others*, 2015). DOCK4 is known to regulate neurite differentiation by activation of Rac1 (Xiao *and others*, 2013). Rac1 in turn was shown to rescue neurite retraction caused by G2019S LRRK2, a well-known pathogenic mutation causing familiar PD (Chan *and others*, 2011). UBE2J1, also known as UBC6, is a member of the Parkin-Ubiquitin Proteasomal System pathway (see NCBI - BioSystems) and directly interacts with Parkin (Mengesdorf *and others*, 2002), another well-known PD gene. Finally, GPR107 is a G-protein coupled receptor. While a different G-protein coupled receptor gene, namely GPR37, is known to be a risk gene with respect to PD, a direct connection between GPR107 and PD is unclear at this point.

Table 4. *Results of the analysis of PD data sets. The table shows both combined and single-cohort p-values and MAFs for two pairs of genes that were identified for interaction using* Pca_1–DS *(first pair) and* Pca_1–AS *(second pair) with* $\delta = 0.7$ *and* $\alpha_1 = 10^{-6}$. *The implicating (i.e. leading to significance) p-values for each pair are shown in bold.*

GENE1	Chr1:RS1	GENE2	Chr2:RS2	T_n^B	A_n	D_n	C_n
UBE2J1	6:rs17798549	GPR107	9:rs4837460	**8.27e-07**	1.34e-02	**5.97e-06**	8.65e-02
				T_n^B (NL)	A_n (NL)	D_n (NL)	C_n (NL)
				9.32e-04	8.59e-02	2.60e-03	4.48e-01
				T_n^B (GE)	A_n (GE)	D_n (GE)	C_n (GE)
				4.60e-05	1.89e-02	4.56e-05	2.44e-02
				f_1 (NL)	f_2 (NL)	f_1 (GE)	f_2 (GE)
				0.043	0.011	0.035	0.023
GENE1	Chr1:RS1	GENE2	Chr2:RS2	T_n^B	A_n	D_n	C_n
DUSP12	1:rs1063179	DOCK4	7:rs12705795	**8.30e-07**	**3.59e-06**	1.30e-01	8.88e-01
				T_n^B (NL)	A_n (NL)	D_n (NL)	C_n (NL)
				8.11e-08	3.45e-02	4.54e-02	9.17e-01
				T_n^B (GE)	A_n (GE)	D_n (GE)	C_n (GE)
				7.75e-01	2.78e-06	9.82e-01	5.79e-01
				f_1 (NL)	f_2 (NL)	f_1 (GE)	f_2 (GE)
				0.04	0.062	0.05	0.064

6. DISCUSSION

In this article, we formulated two variants of a two-stage method for genome-wide search for epistasis. We compared the methods in terms of error control and power performance with the classical single-stage approach and existing similar two-stage approaches. For a good power performance our approach requires two things, which are a low level of background dependence in the general population and suitable choice of input parameters, namely the interaction penetrance function and the S1 level and sample proportion. Although the two-stage methods are valid in terms of type I error regardless of the degree of background dependence among the loci in the data, a low level of background dependence allows the S1 tests to substantially limit the multiple testing burden encountered by the S2 tests. Crucially, in a genome-wide search for epistasis strong background dependence among a large portion of the tested loci pairs is unlikely. Regarding the choices of input parameter, the fact that they influence the overall power performance of the methods can be seen as both a limitation and a virtue in terms of the power potential, since their presence creates a possibility for tailoring the methods towards specific practical scenarios. In any case, in order to make the accompanying software as user-friendly as possible we made a default (but changeable) choice for the interaction model and provided an automated way of choosing the parameters α_1 and δ based on the theoretical procedure described in Section 3.7.

We investigated the relative performance of the methods under various scenarios, which combined different sample sizes, allele frequencies and interaction models. We showed that under many of those scenarios our two-stage approach has the potential to significantly outperform both the classical single-stage test and the competing two-stage procedures. Especially for misspecified interaction models our method can achieve superior power performance. Such robustness with respect to the interaction model is crucial in real world applications. The results of analysis of PD data sets further support this claim. We believe that our results could move the problem of the "missing heritability" a step closer towards unraveling.

ACKNOWLEDGMENTS

We would also like to thank all of the subjects who donated their time and biological samples to be a part of the Parkinson's disease study by IPDGC. *Conflict of Interest*: None declared.

FUNDING

This project was funded as part of Project 3.7.3 (Statistical Genetics) of the Center for Medical Systems Biology of the Netherlands Genomics Initiative.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

REFERENCES

CHAN, D., CITRO, A., CORDY, J. M., SHEN, G. C. AND WOLOZIN, B. (2011). Rac1 protein rescues neurite retraction caused by G2019S leucine-rich repeat kinase 2 (LRRK2). *The Journal of Biological Chemistry* **286**, 16140–9.

CORDELL, H. J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics* **11**, 2463–2468.

FOPPA, I. AND SPIEGELMAN, D. (1997). Power and sample size calculations for case-control studies of gene-environment interactions with a polytomous exposure variable. *American Journal of Epidemiology* **146**, 596–604.

Gao, X., Carroni, M., Nussbaum-Krammer, C., Mogk, A., Nillegoda, N. B., Szlachcic, A., Guilbride, D. L., Saibil, H. R., Mayer, M. and Bukau, B. (2015). Human Hsp70 disaggregase reverses parkinson's-linked α-synuclein amyloid fibrils. *Molecular Cell* **59**, 781–793.

Hemani, G., Knott, S. and Haley, C. (2013). An evolutionary perspective on epistasis and the missing heritability. *PLoS Genetics* **9**, e1003295.

Hemani, G., Theocharidis, A., Wei, W. and Haley, C. (2011). EpiGPU: exhaustive pairwise epistasis scans parallelized on consumer level graphics cards. *Bioinformatics* **27**, 1462–1465.

Herold, C., Steffens, M., Brockschmidt, F. F., Baur, M. P. and Becker, T. (2009). INTERSNP: genome-wide interaction analysis guided by a priori information. *Bioinformatics* **25**, 3275–3281.

International Parkinson Disease Genomics Consortium (IPDGC). (2011). Imputation of sequence variants for identification of genetic risks for parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649.

International Parkinson Disease Genomics Consortium (IPDGC) and Wellcome Trust Case Control Consortium 2 (WTCCC2). (2011). A two-stage meta-analysis identifies several new loci for parkinson's disease. *PLoS Genetics* **7**, e1002142.

Kam-Thong, T., Czamara, D., Tsuda, K., Borgwardt, K., Lewis, C. M., Erhardt-Lehmann, A., Hemmer, B., Rieckmann, P., Daake, M., Weber, F., Wolf, C., Ziegler, A., Pütz, B., Holsboer, F., Schölkopf, B. *and others*. (2011). EPIBLASTER – fast exhaustive two-locus epistasis detection strategy using graphical processing units. *European Journal of Human Genetics* **19**, 465–471.

Lewinger, J. P., Morrison, J. L., Thomas, D. C., Murcray, C. E., Conti, D. V., Li, D. and Gauderman, W. J. (2013). Efficient two-step testing of gene-gene interactions in genomewide association studies. *Genetic Epidemiology* **37**, 440–451.

Li, J., Horstman, B. and Chen, Y. (2011). Detecting epistatic effects in association studies at a genomic level based on an ensemble approach. *Bioinformatics* **27**, i222–i229.

Li, S.S., Khalid, N., Carlson, C. and Zhao, L. P. (2003). Estimating haplotype frequencies and standard errors for multiple single nucleotide polymorphisms. *Biostatistics* **4**, 513–522.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F., McCarroll, S. A. *and others*. (2009). Finding the missing heritability of complex diseases. *Nature* **461**, 747–753.

Marchini, J., Donnelly, P. and Cardon, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics* **37**, 413–417.

Mengesdorf, T., Jensen, P. H., Mies, G., Aufenberg, C. and Paschen, W. (2002). Down-regulation of parkin protein in transient focal cerebral ischemia: a link between stroke and degenerative disease? *PNAS* **99**, 15042–7.

Niel, C., Sinoquet, C., Dina, C. and Rocheleau, G. (2015). A survey about methods dedicated to epistasis detection. *Frontiers in Genetics* **6**, 1–19.

Pahl, R., Schäfer, H. and Müller, H. (2009). Optimal multistage designs – a general framework for efficient genome-wide association studies. *Biostatistics* **10**, 297–309.

Pankratz, N., Wilk, J. B., Latourelle, J. C., DeStefano, A. L., Halter, C., Pugh, E. W., Doheny, K. F., Gusella, J. F., Nichols, W. C., Foroud, T. *and others*. (2009). Genomewide association study for susceptibility genes contributing to familial parkinson disease. *Human Genetics* **124**, 593–605.

Park, M. Y. and Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. *Biostatistics* **9**, 30–50.

PIRIYAPONGSA, J., NGAMPHIW, C., INTARAPANICH, A., KULAWONGANUNCHAI, S., ASSAWAMAKIN, A., BOOTCHAI, C., SHAW, P. J. AND TONGSIMA, S. (2012). iLOCi: a SNP interaction prioritization technique for detecting epistasis in genome-wide association studies. *BMC Genomics* **13**, 1–15.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A. R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. W., DALY, M. J. *and others*. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* **81**, 559–575.

REICH, D. E., CARGILL, M., BOLK, S., IRELAND, J., SABETI, P. C., RICHTER, D. J., LAVERY, T., KOUYOUMJIAN, R., FARHADIAN, S. F., WARD, R. *and others*. (2001). Linkage disequilibrium in the human genome. *Nature* **411**, 199–204.

SHARDA, P. R., BONHAM, C. A., MUCAKI, E. J., BUTT, Z. AND VACRATSIS, P. O. (2009). The dual-specificity phosphatase hYVH1 interacts with Hsp70 and prevents heat-shock-induced cell death. *Biochemical Journal* **418**, 391–401.

STEEN, K. VAN. (2012). Travelling the world of gene–gene interactions. *Briefings in Bioinformatics* **13**, 1–19.

TANG, W., WU, X., JIANG, R. AND LI, Y. (2008). Epistatic module detection for case-control studies: A bayesian model with a gibbs sampling strategy. *PLoS Genetics* **5**, e1000464.

THOMAS, D., XIE, R. AND GEBREGZIABHER, M. (2004). Two-stage sampling designs for gene association studies. *Genetic Epidemiology* **27**, 401–414.

WADE, M. J., WINTHER, R. G., AGRAWAL, A. F. AND GOODNIGHT, C. J. (2001). Alternative definitions of epistasis: dependence and interaction. *Trends in Ecology & Evolution* **16**, 498–504.

WU, X., JIN, L. AND XIONG, M. (2008). Composite measure of linkage disequilibrium for testing interaction between unlinked loci. *European Journal of Human Genetics* **16**, 644–651.

XIAO, Y., PENG, Y., WAN, J., TANG, G., CHEN, Y., TANG, J., YE, W. C., IP, N. Y. AND SHI, L. (2013). The atypical guanine nucleotide exchange factor Dock4 regulates neurite differentiation through modulation of Rac1 GTPase and actin dynamics. *The Journal of Biological Chemistry* **288**, 20034–45.

ZHANG, Z. AND CHENG, Y. (2014). miR-16-1 promotes the aberrant α-synuclein accumulation in parkinson disease via targeting heat shockprotein 70. *Scientific World Journal* **938348**, 1–8.

[*Received June* 13, 2016; *revised June* 13, 2016; *accepted for publication November* 05, 2016]