Evaluating HWE and Association in Genome Wide Association Studies: A Unified Procedure

Stefan Boehringer[1*] and Hajo Holzmann[2]

[1] Institut für Humangenetik, Universitätsklinikum Essen, Essen, Germany,

Hufelandstr. 55, 45122 Essen, Germany

Phone: +49 201 7234533

Email: correspondence@s-boehringer.org

[2] Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, Germany

Hans-Meerweinstr., 35032 Marburg, Germany

Email: holzmann@mathematik.uni-marburg.de

*Corresponding author

Abstract

In current genome wide association studies (GWASs) based on a case-control design, single nucleotide polymorphisms (SNPs) are typically evaluated for association as well as for Hardy-Weinberg equilibrium (HWE) in the control group by performing two separate statistical tests. Since HWE is expected to hold in the control group, SNPs for which the p-value of the HWE test is below some threshold value are excluded from further analysis. For the other SNPs, the value of the HWE statistic is subsequently ignored. This standard approach has several quite obvious drawbacks. First, the threshold value for the HWE test has to be chosen rather arbitrarily. Second, for SNPs which are not excluded the HWE statistic is completely ignored. Thus, the dependency of the two tests is not taken into account in an appropriate manner. Further, we shall argue that such a procedure is not optimal for detecting or avoiding effects of systematic error.

To overcome these drawbacks, we propose a conditional genotype-based test that conditions the Pearson Chi-Square test for association in the 3x2 contingency table on the HWE statistic in the control group. The asymptotic distribution of the conditional test is derived. Here, the main technical tool is a new parametrization for the genotype distribution. The properties of our test, as compared to two competing procedures, are investigated in an extensive simulation study. After comparing power and robustness properties in single locus simulations, our main concern is for the properties of the tests in GWASs simulations. Here, we propose an elaborate simulation setting and study several measures for performance in GWASs, in particular average ranking of alternative SNPs, number of rejected alternative SNPs under a false discovery rate criterion (FDR) and probability of rejecting any alternative SNP (while controlling the FDR). It turns out that the proposed conditional test strongly outperforms its competitors for these performance measures in several distinct GWASs settings.

Finally, we illustrate the new procedure by an application to a data set in an alopecia study. This data set shows mild deviation from HWE in controls and therefore offers interesting insights into behavior under misspecification.

In conclusion, our test makes separate HWE testing superfluous by providing a unified framework and strictly improves on the standard procedure in terms of power and interpretability, thereby making replication more cost effective and improving subsequent fine mapping.

Keywords: Hardy-Weinberg Equilibrium; Genome-Wide Association; Case-Control; False discovery Rate; SNP; Association Testing

1. INTRODUCTION

Genome wide associations studies (GWASs) are currently used in epidemiological studies to identify genetic variants with predictive power for a phenotype [Kruglyak, 2008]. In such a study typically 300,000-500,000 genetic markers are scrutinized, each being tested for significant association with a phenotype. There are many statistical challenges in the analysis of such studies including confounding by population stratification (*e.g.* [Devlin and Roeder, 1999]), marker selection (*e.g.* [Carlson et al., 2004; Consortium, 2007]) and addressing multiple testing (*e.g.* [Benjamini and Hochberg, 1995; Kimmel and Shamir, 2006]). We propose a novel solution for analyzing Hardy-Weinberg equilibrium (HWE), an aspect of case-control studies that has not been sufficiently addressed so far. HWE describes the assumption that the two observed copies (alleles) at a genetic locus are randomly and independently drawn from the previous generation. For a random variable X, describing a locus with two alleles, say $1, 2$, we denote the frequency of one allele with $P(X = 1) = \rho$ and $P(X = 2) = 1 - \rho$. For the joint distribution of both alleles (genotype) we consider the number of copies of, say, allele 1, thereby ignoring parental origin. The HWE assumption can then be summarized by the equation

$$(P(G = 0), P(G = 1), P(G = 2)) = (\rho^2, 2\rho(1 - \rho), (1 - \rho)^2),$$

when G is the random variable for a bi-allelic locus with allele frequency ρ. So called single nucleotide polymorphisms (SNPs) are an important class of bi-allelic loci and we use the terms interchangeably.

In this paper, we focus on case-control studies based on SNPs. Here, HWE is usually assumed to hold in the control group since this was empirically confirmed in many studies (*e.g.* [Yeager et al., 2007]). Assuming HWE in the control group allows to test this assumption by means of a goodness-of-fit test. Testing this assumption might be reasonable, as departures from HWE may be due to systematic error, *e.g.* in the process of observing genotypes (genotyping errors), sample contamination or confounding by population strata (see below). As a conclusion, SNPs departing from HWE are excluded from further analysis(*e.g.* [Gudmundsson et al., 2007; Yeager et al., 2007; Salmela et al., 2008]). However, such a strategy leads to several problems as we will show now. First, the significance level at which to reject a SNP based on HWE is arbitrary and has shifted historically (*e.g.* compare [Salmela et al., 2008; Gudmundsson et al., 2007]). For example, in a study with 500,000 SNPs and a HWE testing procedure at the level $\alpha = .05$, 25,000 SNPs are

expected to be rejected by the test. This situation has led to suggestions of choosing $\alpha = 0.001$ or lower. Second, the conclusion that HWE departures imply systematic error is not warrented and can lead to exclusion of SNPs not influenced by systematic error, thereby ignoring valid information. Third, if a cut-off is used for HWE, the value for the test statistic for SNPs passing the criterion is entirely ignored in subsequent analyses. Intuitively this seems paradoxical as SNPs close to the critical value of the HWE test are treated just like SNPs with p-values for the HWE test close to 1.

We here take the point of view that significant deviations from HWE are the result of finite sample size and random fluctuations from observing a large number of SNPs and therefore the tails of a distribution measuring such deviation will be observed in the absence of systematic error. We will later carefully discuss the relationship with systematic errors. The intuition behind our stance is, that the value of the HWE test statistic is not independent of a subsequent association test, as deviation from HWE in controls tell us that a sample is "atypical". More formally, this leads us to a conditional approach, where a test statistic is considered conditionally on the HWE goodness-of-fit test statistic.

HWE was identified as an important criterion to evaluate quality of association studies [Salanti et al., 2005] and an important source for systematic error [Xu et al., 2002]. The latter reference notes the correlation between false positives and HWE deviations and attributes it mainly to genotyping errors. However, this correlation is naturally implied by our results in the absence of systematic error. Wittke-Thompson et al. [2005] develop a goodness-of-fit test for HWE by separating deviation due to effects of some genetic model and "genuine" HWE deviations. Strong assumptions about the genetic model are needed. Wang and Shete [2008] follow a similar idea by assuming that HWE deviations in cases may be due to a genetic effect. They focus on global testing by adopting a so called tail-strength measure following Taylor and Tibshirani [2006]. Song and Elston [2006] consider a weighted average between the Armitage trend test and a trend statistic comparing HWE in cases and controls. The weighing factor is chosen arbitrarily and is justified by explorative simulations. The most closely related work proposes to ignore HWE in controls altogether [Chen and Chatterjee, 2007], an approach, that is later shown to have unfavorable properties in a GWAS setting. The main difference of our work as compared to the reviewed papers is that we consider HWE in controls to be a nuisance parameter measured by some parameter (η_1) and it is therefore natural to condition the test statistic on a sufficient statistic for η_1.

The paper is organized as follows: In section 2 we introduce a new parametrization of the testing problem and develop the asymptotic theory for a conditional test. In section 3, we scrutinize finite-sample properties of the test by means of simulation studies and compare the test with alternative approaches. Section 4 is concerned with the analysis of a data set on an alopecia phenotype. This analysis gives some insight into behavior under mis-specification as some mild deviation from HWE in controls is observed. In section 5 we conclude with a discussion in which we summarize the properties of our proposed procedure and highlight the benefits for practical data analysis. Technical proofs as well as technical details on the simulation setting are contained in the Appendix.

2. ASYMPTOTIC CONDITIONAL DISTRIBUTION

2.1 Reparametrization of the multinomial distribution

The genotype distribution for N individuals at a SNP can be parametrized by a multinomial distribution $M(N; \pi_1, \pi_2, \pi_3), \pi_1 + \pi_2 + \pi_3 = 1$, with three possible resulting categories and N repetitions. Categories again represent genotypes determined by allele counts of one arbitrarily chosen allele.

In order to motivate our new parametrization, we recall the most simple model of a neutral coalescent process for a fixed population size [Hudson, 1990]. In this model each allele is drawn randomly and independently from the previous generation with replacement. The grouping of two alleles into a genotype (an individual) is not needed in the model as it seeks to describe allele behaviors in terms of frequency changes and time (*e.g.* number of generations) that has passed since two alleles were drawn from the same ancestor. In this setting, the allele frequency ρ suffices to describe the distribution per generation. Following this notion, we interpret the genotype distribution as the result of a two stage process. First, alleles are selected at random for transmission to the current generation (gamete formation), a process that only depends on the allele frequency ρ in the previous generation. Second, alleles are paired to form genotypes (individuals) in the current generation. We introduce a second parameter η that measures deviation of genotype formation from a HWE expectation. More formally we define $(\rho, \eta) = T(\pi_1, \pi_2)$, where

$$\begin{aligned}\rho &= \pi_1 + \pi_2/2 \\ \eta &= \operatorname{sgn}\big(\pi_2 - f_2(\rho)\big)\Big(\sum_{j=1}^{3}\frac{(\pi_j - f_j(\rho))^2}{f_j(\rho)}\Big)^{1/2},\end{aligned}$$

and $\big(f_1(\rho), f_2(\rho), f_3(\rho)\big) = \big(\rho^2, 2\rho(1-\rho), (1-\rho)^2\big)$. One can recover the original parameters of the multinomial by letting $\pi_1 = \rho - \pi_2/2$ and $\pi_2 = f_2(\rho)(1+\eta)$ thereby retaining the full information on the genotype distribution. The parameter η measures an excess or deficit of heterozygous genotypes (allele count of 1 in the genotype) relative to HWE. Not only does this parametrization allow to formulate our problem more conveniently, it also has some merits in reformulating existing statistical procedures concerning HWE. For example, we will later discuss an alternative likelihood ratio (LR) test which has a very simple null hypothesis in our parametrization [Chen and Chatterjee, 2007]. We also note, that graphical representations of genotype frequencies might be more easily interpreted as compared to De Finetti diagrams [Edwards, 2000].

2.2 Maximum likelihood estimates and Fisher information of the reparametrized likelihood

Consider again the multinomial distribution $M(N; \pi_1, \pi_2, \pi_3)$ as above. Let $X = (n_1, n_2, n_3) \sim M(N; \pi_1, \pi_2, \pi_3)$. The maximum likelihood estimates for π_i, $i = 1, 2, 3$, are given by $\hat{\pi}_i = n_i/N$. We have

$$\sqrt{N}\big((\hat{\pi}_1, \hat{\pi}_2)^T - (\pi_1, \pi_2)^T\big) \xrightarrow{d} N(0, \mathcal{J}_0^{-1}),$$

where

$$\mathcal{J}_0^{-1} = \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) \end{pmatrix}, \qquad \mathcal{J}_0 = \begin{pmatrix} \frac{1}{\pi_1} + \frac{1}{\pi_3} & \frac{1}{\pi_3} \\ \frac{1}{\pi_3} & \frac{1}{\pi_2} + \frac{1}{\pi_3} \end{pmatrix},$$

$\mathcal{J}_0$ being the Fisher information of the multinomial.

The ML estimates of (ρ, η) are then given by $(\hat{\rho}, \hat{\eta}) = T(\hat{\pi}_1, \hat{\pi}_2)$. Their asymptotic distribution is easily derived from the $\delta-$method. Indeed,

$$\sqrt{N}\big((\hat{\rho}, \hat{\eta})^T - (\rho, \eta)^T\big) \xrightarrow{d} N(0, \mathcal{I}_0^{-1}),$$

where $\mathcal{I}_0^{-1} = DT\,\mathcal{J}_0^{-1}\,DT^T$. Straightforward algebraic transformations show that

$$\mathcal{I}_0^{-1} = \frac{1}{2}\begin{pmatrix} (1-\eta)(1-\rho)\rho & \eta(1+\eta)(1-2\rho) \\ \eta(1+\eta)(1-2\rho) & -\frac{(\eta+1)\big((\eta-2\eta\rho)^2+2(\rho-1)\rho+\eta(6(\rho-1)\rho+2)\big)}{(1-\rho)\rho} \end{pmatrix}$$

Observe that in HWE equilibrium, $\eta = 0$ and thus $\mathcal{I}_0^{-1}$ is a diagonal matrix and the estimators $(\hat{\rho}, \hat{\eta})$ are asymptotically independent.

2.3 Conditional tests

Consider two multinomial distributions $M(N_i; \pi_{i,1}, \pi_{i,2}, \pi_{i,3})$, $i = 1, 2$ corresponding to control and case groups, with three possible resulting categories and N_i repetitions. Let $X_i = (n_{i,1}, n_{i,2}, n_{i,3}) \sim M(N_i; \pi_{i,1}, \pi_{i,2}, \pi_{i,3})$. Consider the hypothesis

$$H : \pi_{1,j} = \pi_{2,j} =: \pi_j, \qquad j = 1, 2, 3.$$

We set $n_{\cdot j} = n_{1,j} + n_{2,j}$ and $N = N_1 + N_2$. The unrestricted log-likelihood is given by

$$\mathcal{L}(\pi_{1,1}, \pi_{1,2}, \pi_{2,1}, \pi_{2,2}) = \sum_{i=1,2}\sum_{j=1,2} n_{i,j} \log \pi_{i,j} + \sum_{i=1,2} n_{i,3} \log \pi_{i,3},$$

where $\pi_{i,3} = 1 - \pi_{i,1} - \pi_{i,2}$, and the restricted log-likelihood under H

$$\mathcal{L}(\pi_1, \pi_2) = \sum_{j=1,2} n_{\cdot j} \log \pi_j + n_{\cdot 3} \log \pi_3.$$

The unrestricted ML estimates are $\hat{\pi}_{i,j} = n_{i,j}/N_i$, $i = 1, 2$, $j = 1, 2$, and in the equivalent parametrization, $(\hat{\rho}_i, \hat{\eta}_i) = T(\hat{\pi}_{i,1}, \hat{\pi}_{i,2})$. Further, the restricted ML estimates under H are given by $\hat{\pi}_j = n_{\cdot j}/N$. Let $\mathcal{D}(Z)$ denote the distribution of the random variable Z, and let d be a metric on the probabilities on $\mathbb{R}$ which metrizes weak convergence.

Theorem 1. *Suppose that H holds and that further the control group is in HWE, i.e. $\eta_1 = 0$. If $c = N_1/N$ is bounded away from 0 and 1, then for the asymptotic conditional distribution of the LRT statistic, given $\hat{\eta}_1$, we have as $N \to \infty$ that*

$$d\Big(\mathcal{D}\Big(2\big(\mathcal{L}(\hat{\pi}_{1,1}, \hat{\pi}_{1,2}, \hat{\pi}_{2,1}, \hat{\pi}_{2,2}) - \mathcal{L}(\hat{\pi}_1, \hat{\pi}_2)\big)|\hat{\eta}_1\Big), \mathcal{D}\big((Z_1 + Z_2)|\hat{\eta}_1\Big)\Big) = o_P(1), \tag{1}$$

where $\mathcal{D}\Big(\,\cdot\,|\hat{\eta}_1\Big)$ *is the conditional distribution given* $\hat{\eta}_1$, Z_1 *and* $(Z_2, \hat{\eta}_1)$ *are independent,* $Z_1 \sim \chi^2_1$ *and* $Z_2|\hat{\eta}_1 \sim c\chi^2_1\big(N\hat{\eta}_1^2(1-c)/2\big)$.

Theorem 1 is proved in Appendix A.1.

Next we draw some conclusions and comment on the conditional asymptotic distribution.

Remark 1. The asymptotic conditional distribution only depends on $\hat{\eta}_1^2$, the HWE χ^2-statistic, and not on the signed version $\hat{\eta}_1$. Thus, (1) remains true if we condition the LRT statistic on $\hat{\eta}_1^2$ or $Nc\hat{\eta}_1^2$, the latter being the exact form of the HWE χ^2-statistic.

Remark 2. Since the LRT statistic and χ^2 statistic are asymptotically equivalent, (1) remains true if we replace the LRT statistic by the χ^2 statistic for H. More specifically, let

$$\mathcal{X}^2 = \sum_{i=1}^{2}\sum_{j=1}^{3} \frac{\big(n_{i,j} - N_i\hat{\pi}_j\big)^2}{N_i\hat{\pi}_j}.$$

With Remark 1 this yields

Corollary 1. *Under the assumptions of Theorem 1, we have that*

$$d\Big(\mathcal{D}\Big(\mathcal{X}^2|\hat{\eta}_1^2\Big), \mathcal{D}\Big((Z_1+Z_2)|\hat{\eta}_1^2\Big)\Big) = o_P(1), \tag{2}$$

where Z_1 *and* Z_2 *are distributed as in the theorem.*

Remark 3. For evaluating the asymptotic (conditional) P-value of the test statistics in (1) and (2), we have to evaluate the distribution function of the asymptotic conditional distribution. Note that this is the convolution of the χ^2_1-variable and a rescaled non-central χ^2_1 variable. Now, the distribution function $F_1(x;\lambda)$ of $\chi^2_1(\lambda)$ (χ^2_1 with non-centrality parameter (NCP) λ) can be written as (*cf.* [Johnson et al., 1994])

$$F_1(x;\lambda) = \sum_{j=0}^{\infty}\Big(\frac{(\lambda/2)^j}{j!}e^{-\lambda/2}\Big)F_{1+2j}(x),$$

where $F_\nu(x)$ is the distribution function of the central χ^2_ν distribution. Therefore, it would be sufficient to compute the convolutions of χ^2_1 and the rescaled $c\chi^2_{1+2j}$, $j \geq 0$. However, this amounts

to computing the convolution of two gamma variables with distinct scale parameters (*cf.* [Johnson et al., 1994]). Although methods for this problem do exist, we found it preferable to simply compute the convolution of the densities of χ_1^2 and $c\chi_1^2(\lambda)$ numerically, as outlined in the simulation section.

Remark 4. The proof shows that one can also test for deviations in the allele frequency with standard χ_1^2 limit distribution when conditioning on the HWE statistic. Indeed, under the assumptions of Theorem 1,

$$N\Big(\frac{c}{a}\big(\hat{\rho}_1 - \hat{\rho}\big)^2 + \frac{1-c}{a}\big(\hat{\rho}_2 - \hat{\rho}\big)^2\Big)|\hat{\eta}_1 \xrightarrow{\mathcal{D}} \chi_1^2,$$

where a is defined in (A.1).

3. SIMULATION STUDY

In this section, the finite sample properties of the proposed testing procedure are scrutinized. First, we describe numerical computation of the distribution function of the test statistic. Second, we show single locus simulations to characterize speed of convergence and make comparisons with two other test statistics. These comparisons will further motivate the application of our proposed procedure in a GWAS setting. Finally, we compare the same statistics in GWAS simulations.

3.1 Computation of the distribution function

Numerical integration is used to compute the distribution function of the convolution $Z := Z_1 + cZ_2(\lambda)$, with NCP $\lambda = N\hat{\eta}_1(1-c)/2$. Denote with $\varphi(x,y) = \varphi(x,y;\lambda,c)$ the joint density of Z_1 and $cZ_2(\lambda)$. Therefore $F_Z(x) = \int_0^x \int_0^z \varphi(t, z-t)dtdz$, if F_Z denotes the distribution function of Z. A first change of variables allows for a rectangular integration area: $\theta : (x,y) \to (x, xy)$, $F_Z(x) = \int_0^x \int_0^1 \varphi \circ \theta(z,t)|D\theta|dtdz$. Note, that $\varphi \circ \theta(z,t)$ has poles at $t = 0$ and $t = 1$ for all values z. The poles behave as $\frac{1}{\sqrt{t}}$ for $t \to 0$ and as $\frac{1}{\sqrt{1-t}}$ for $t \to 1$. For t, we split the integral into subintervals $(0, \frac{1}{4})$, $(\frac{1}{4}, \frac{3}{4})$, $(\frac{3}{4}, 1)$ and apply transformations in the outer intervals in order to eliminate the poles. To this end, we use the identities $\int_0^b f(x) = \int_0^{\sqrt{b}} 2tf(t^2)$ for the pole at 0 and $\int_a^1 f(x) = \int_0^{\sqrt{1-a}} 2tf(1-t^2)$ for the pole at 1.

Therefore, after some algebraic transformations, the distribution function of Z is given by:

$$
\begin{aligned}
& F_Z(x;\lambda,c) \\
= & \int_0^{\frac{1}{4}} \int_0^x \frac{2}{\pi} \exp\left\{-\frac{1}{2}\left(\lambda + s + 4t^2 s(\frac{1}{c} - 1) + \log(c(1-4t^2))\right)\right\} \cosh\left(2t\sqrt{\frac{\lambda s}{c}}\right) ds\ dt \\
+ & \int_{\frac{1}{4}}^{\frac{3}{4}} \int_0^x \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(\lambda + s + st(\frac{1}{c} - 1) + \log(ct(1-t))\right)\right\} \cosh\left(\sqrt{\frac{\lambda st}{c}}\right) ds\ dt \\
+ & \int_{\frac{3}{4}}^{1} \int_0^x \frac{2}{\pi} \exp\left\{-\frac{1}{2}\left(\lambda + \frac{s}{c} + 4t^2 s(1 - \frac{1}{c}) + \log(c(1-4t^2))\right)\right\} \cosh\left(\sqrt{\frac{\lambda s(1-4t^2)}{c}}\right) ds\ dt,
\end{aligned}
$$

which is amenable to numerical integration. We use the software package *adapt* in R version 2.8.0 to compute the integrals [Team, 2008].

3.2 Single locus simulations

Table 1 shows simulations comparing the proposed test statistic with the Pearson test and a test that assumes HWE frequencies in controls [Chen and Chatterjee, 2007]. The latter test is perhaps the most simple way to address the HWE assumption in controls. This test compares genotype frequencies in cases with an HWE-expectation based on an allele frequency estimation in controls, thereby ignoring the parameter η_1 entirely. In (ρ, η) parametrization the test is a LR test with the null hypothesis $\Theta_0 : \rho_1 = \rho_2, \eta_1 = \eta_2 = 0$ and the alternative $\Theta_1 : \eta_1 = 0, (\rho_1 \neq \rho_2 \vee \eta_2 \neq 0)$, where (ρ_1, η_1) are parameters for the control group and (ρ_2, η_2) are parameters for cases. This test completely ignores the HWE distribution in controls and is referred to as the EHWE (expected HWE) test.

Table 1 shows in the first section that all tests maintain the α level faithfully as is expected for the sample size of 10^3 for each group. The second section shows results under the alternative. The most powerful test in cases where $\rho_1 = \rho_2$ and $0 = \eta_1 \neq \eta_2$ is the EHWE test. For example power for EHWE is 99% for $(\rho_1, \rho_2, \eta_1, \eta_2) = (.1, .1, 0, .1)$ whereas it is 77% and 79% for our conditional test and the Pearson test, respectively. In cases where the ηs are equal and ρs differ, our conditional test is the most powerful test. In cases where the Pearson test is most powerful, the difference is not statistically significant such that the Pearson test can be considered as the least powerful test in all scenarios. This result is expected as the Pearson test makes the least assumptions of the compared tests.

Our conditional test and the EHWE test both assume HWE in controls and it appears that they have mutual strengths. However, if we misspecify our test scenario by setting $\eta_1 \neq 0$ some interest-

ing results are apparent from the third section in Table 1. In situations where $\rho_1 = \rho_2, \eta_1 = \eta_2$ but $\eta_1 \neq 0$ the conditional test is conservative and does not exhaust the α-level of 0.05 whereas EHWE is anticonservative and exceeds the α-level by large margins. These qualitative observations have important consequences for a GWAS setting as loci with estimates for $|\eta_1| >> 0$ are expected to occur in the set of SNPs even under the HWE assumptions. In these cases EHWE would be more likely to assign a smaller p-value than our conditional test. Therefore on average, our conditional test adjusts p-values for SNPs with $|\eta_1| >> 0$ upwards whereas EHWE adjusts downwards. Intuitively this should result in an excess of false positives for EHWE as compared to our conditional approach. We address this question by simulations of GWASs.

3.3 GWASs simulations

We conducted simulations of plausible GWAS scenarios in order to assess the impact of our conditional test on the analysis of such studies. Our single locus simulations indicate that our test should have favorable properties as compared to the EHWE and the Pearson test in a GWAS scenario. We constructed an alternative based on a logistic penetrance function. If K loci influence disease status, we have:

$$P(Y = 1|g_1, ..., g_K) \quad = \quad \left(1 + \exp\left\{-\left(\mu + \sum_{i=1}^{K} \beta_i x_i\right)\right\}\right)^{-1} =: Y(\mathbf{g}; \beta). \tag{3}$$

Here, x_i is a score assigned to a genotype, β_i is the effect size of locus i and μ is a baseline penetrance. For genotypes 0 and 2 we assign the respective scores 0 and 1 and for genotype 1 we assign $0, \frac{1}{2}, 1$ to respectively simulate a recessive, additive and dominant mode of inheritance. The prevalence $P(Y = 1)$ of the disease is then given by:

$$\alpha \quad = \quad P(Y = 1) = \sum_{(g_1,..,g_K)\in\mathbf{G}} Y(\mathbf{g}; \beta) P(G = \mathbf{g}) = \sum_{(g_1,..,g_K)\in\mathbf{G}} Y(\mathbf{g}; \beta) \prod_{i=1}^{K} \rho_{i,g_i}, \tag{4}$$

assuming that $\mathbf{g} = (g_1, .., g_K), \beta = (\beta_1, ..., \beta_K)$ and ρ_{i,g_i} denotes the genotype frequency of genotype g_i at locus i. In all simulations we assume all SNPs to be independent for a random sample from the population. In order to pick an alternative hypothesis, we deterministically chose some of the parameters and drew the others randomly. Deterministic parameters were $\alpha, \mu, \beta_1, \rho_1$ and the score assignment. K, $\beta_2, ..., \beta_K, \rho_2, ..., \rho_K$ were drawn randomly as follows. $\beta_i \sim U(1.1, \beta_1), i = 2, ...,$

$\rho_i \sim U(\rho_1/2, \rho_1)$ subject to the constraint $P(Y = 1) = \alpha$. We give details about the procedure in Appendix A.2. In summary, we created alternatives for which we know prevalence, maximal effect size and baseline penetrance with freedom in how the prevalance was distributed over effects of additional loci, mimicking the suspected polygenic nature of most complex genetic diseases. From the allele frequencies expected genotype frequencies under HWE were computed and genotypes were drawn according to these distributions for cases and controls for SNPs under the null. For SNPs under the alternative, genotypes for controls were drawn likewise and genotypes for cases were drawn according to the penetrance model above.

3.4 Results of GWASs simulation

Table 2 summarizes results of the GWASs simulations which compare our conditional test (indicated by *s), EHWE (indicated by $^\bullet$s) and the Person test (plain notation). We chose several measures to judge the performance of the tests. As ranking is important in practical analysis we show rank criteria in columns Q_x. We pick the SNP at quantile x according to P-value among SNPs under the alternative and average the rank statistic among all SNPs of these selected SNPs. We also evaluated the number of rejected SNPs denoted by m_α for a false discovery rate (FDR) criterion at level α [Benjamini and Hochberg, 1995]. Finally, Φ_α is the power of the FDR procedure, denoting the probability of rejecting any SNP in a GWAS. For all simulations we assumed the prevalence to be 0.1, sample size to be 2×10^3 in both groups and averaged all numbers across 10^3 repetitions after picking an alternative.

In most situations the conditional test is the most powerful. In cases where it is not most powerful the difference is not statistically significant. These cases are seen for the dominant model, where alternative SNPs ranked mostly at the top of the list as is apparent from columns $Q_0^*, Q_0^\bullet, Q_0$ (values between 1.0 and 1.9). In other cases the difference in power can be quite dramatic, *e.g.* for $(OR_1, \rho_1, \mu) = (2.0, .1, 5 \times 10^{-2})$ under the recessive model power is 72% for the conditional test, 5% for EHWE, 0.9% for the Pearson test.

With respect to ranking (columns Q_x), the conditional test ranks SNPs simulated under the alternative with lowest ranks in most situations. This is true for the dominant and additive model. However, under the recessive model EHWE ranks lowest but is has no power to reject the low-ranking SNPs (column $\Phi^\bullet$). Except for the dominant model, all tests rank SNPs from the 0.25-quantile at ranks > 300, highlighting the difficulty to pinpoint relevant SNPs. If an extensive survey is planned that would imply replication of top ranked SNPs in a second sample, the conditional approach is

optimal as it is the only test that can reject SNPs that are ranked lowly.

Columns m show the average number of rejected SNPs per test. It should be noted that these numbers include SNPs drawn under the null hypothesis. These columns therefore need to be interpreted in conjunction with power as low power would indicated rejection of false positives. In all cases, except for non-significant differences, the conditional test rejects most SNPs while having best power, showing that more true positives were rejected on average.

4. DATA ANALYSIS

We applied our testing procedure to a data set previously published [Hillmer et al., 2008] and compared it with the same tests as used in the simulation section. The data set is comprised of 296 males with androgenic alopecia and 383 random controls. Details about the data set are given in the initial publication [Hillmer et al., 2008]. The data set was pruned for SNPs with a minor allele frequency (MAF) < 0.1 (*i.e.* $\min\{\rho, 1-\rho\} < 0.1$) and call rates for SNPs < 0.9.

Figure 4 shows Quantile-Quantile (Q-Q) plots for several tests comparing the empirical distribution to a uniform distribution. The plot for the HWE test in controls (part (A)) shows notable deviation from the expected uniform distribution for expected p-values $< 10^{-2}$. Apart from this deviation in the tail, the HWE statistic shows a close fit with expected values. This deviation most likely excludes systematic error such as population stratification (different mixtures of cases and controls as drawn from several underlying unique distributions) or cryptic relatedness (different correlation structure of samples across groups) [Devlin and Roeder, 1999; Voight et al., 2005] as in principle all SNPs should be affected. A first thing to note is, that an arbitrary cut-off for HWE p-values, say 10^{-3} would not remove all deviating SNPs in the dataset. As our test should not be sensitive to non-systematic deviations from HWE, we did not remove any SNPs from further analysis. It is informative to see how the different tests behave in this situation. We subjected the data to our conditional test, the Pearson test and EHWE.

Part (B), (C) and (D) of figure 4 show Q-Q-plots for the conditional, Pearson and EHWE tests, respectively. All Q-Q-plots show an excess of low p-values as compared to the expected uniform distribution under the null hypothesis. The EHWE shows excessive anti-conservative behavior whereas the conditional and Pearson tests both show close correspondence to a uniform distribution for p-values $> 10^{-3.5}$. The data set contains a strong signal on chromosome 20 such that some tail deviation is expected.

Table 4 shows SNPs ordered according to p-values for EHWE. The corresponding p-values for HWE are located in the range $(10^{-9}, 10^{-1})$ with a majority of SNPs having p-values $< 10^{-2}$. This reflects the sensitivity of EHWE to misspecification as was shown in the simulation section. It is important to note that EHWE has practically no overlap with either the conditional or Pearson test as determined by SNP rankings except for the very strong signal for the top-5 SNPs according to the conditional test. Apart from these SNPs, the top-30 list for EHWE contains highest rankings of 80 and 184 for the conditional and Pearson test, respectively. HWE outliers with p-values $< 10^{-6}$ rank high for EHWE (ranks 2, 5, 9 in the top-10) whereas the conditional and Pearson tests rank them at ranks $> 200,000$ and $> 3,500$, respectively.

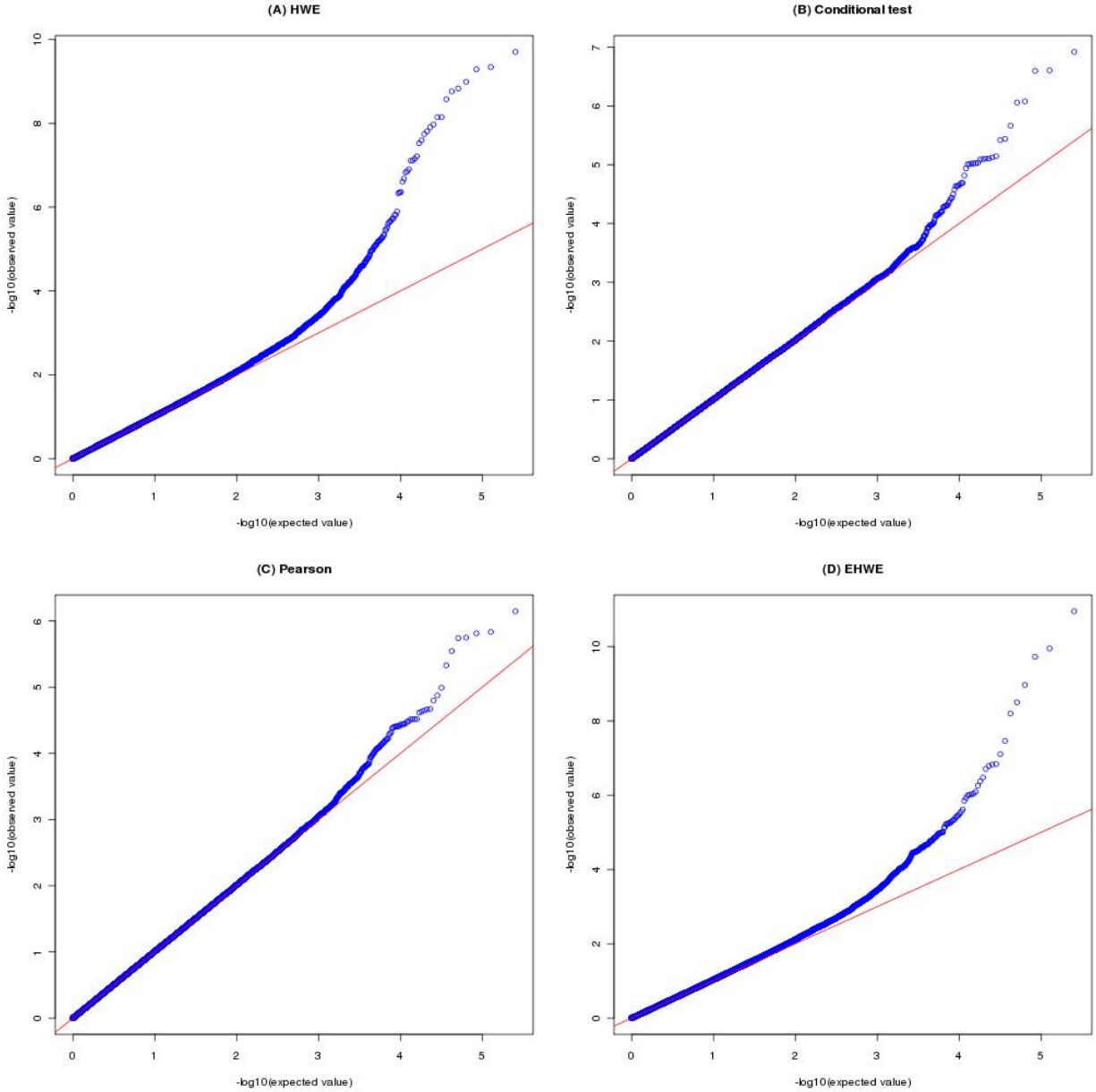
Table 3 shows 30 SNPs ranked according to p-values of the conditional test. Again, overlap with EHWE is weak except for the top-6 SNPs but the Pearson test shows some agreement. Skipping the first six SNPs, the top ranked SNP according to the conditional test has Rank 78 for the Pearson test. From the top 10 list according to Pearson rank 5, 6 and 9 are missing from table 3. These observations can be described more quantitatively by estimating the correlation between p-values of the test statistics as shown in table 5. Pearson and the conditional test show a correlation of 0.95 whereas their correlation with EHWE is < 0.76. The conditional test is the only test showing a negative correlation with HWE (-0.01), EHWE has correlation ~ 0 with HWE (as it ignores the HWE statistic in controls) and Pearson has a correlation of 0.24 with HWE. The negative correlation is not significant in this data set although it is intuitively plausible to assume that HWE should have a strict negative correlation with the conditional test.

5. DISCUSSION

In this paper we have proposed a new testing procedure for case-control studies in a GWAS setting. One major advantage of our test is that is removes the need for two separate tests (HWE, Pearson) and thereby eliminates the need for an arbitrary cutoff that is normally introduced to avoid false positives. However, the problem of systematic error needs still to be addressed. For example a Q-Q plot can reveal deviations from the expected χ^2_1 distribution for the HWE test. If a deviation is seen in the tail of the distribution, a very strict cutoff can be used to eliminate these SNPs. This should typically be a p-value of around 10^{-6} thereby establishing a much better correspondence between HWE statistic and systematic error.

Another interesting property of our test is that it is more powerful than competing tests, in partic-

Figure 1: Q-Q-plots for p-values derived from the obesity data set comparing an empirical distribution with a uniform distribution on a double logarithmic scale. Part (A) shows the HWE test in controls, (B) is the conditional test, (C) the Pearson test and (D) is EHWE.



ular in the analysis of GWAs. Intuitively, this result is expected from the distribution of our test statistic. The non-centrality parameter relates linearly to the HWE statistic and therefore gives more mass to the tail for a SNP with a big HWE statistic as compared with a SNP that shows close correspondence with HWE. Therefore, the test has lower critical values for SNPs in HWE as compared to SNPs with deviations from HWE. Loosly speaking, our conditional test is more likely to reject "promising" SNPs in HWE and less likely to reject "atypical" SNPs with departures from HWE.

Interestingly, a more direct approach to address the assumption of HWE in controls, namely EHWE, is not effective in the analysis of GWAs. Instead of modeling HWE, EHWE completely ignores the HWE statistic by constraining the parameter η_1 to zero in the control group (see table 5). While doing so can improve power in single locus simulations, it decreases power in GWASs and is additionally very sensitive to model misspecification. What is the difference between single locus and GWAS simulations? In the single locus situations the tails of the HWE statistic does only contribute to the average according to its probability mass, whereas in a GWAS no averaging takes place over SNPs such that SNPs with HWE statistics in the tail will be individually represented and are even likely to stick out in tests like Pearson or EHWE for only this reason. This explains why EHWE should be avoided in GWASs.

As a further conclusion, we would like to point out that simulations on a single locus basis do not allow to judge performance in a GWAS as our results underline.

We have analyzed a data set that shows some deviation from the expected distribution for the HWE statistic in controls. This deviation is most likely due to population stratification [Voight et al., 2005]. While this deviation could be addressed (*e.g.* [Devlin and Roeder, 1999]), we chose to re-analyze this data set in its published form. While our test is robust against deviation from HWE in the controls *i.e.* $\eta_1 \neq 0$ in controls, obviously, if $\eta_1 \neq \eta_2$ under the null due to population stratification, our test shows anti-conservative behavior. Therefore, population stratification still has to be quantified and addressed, if a biased analysis is to be expected. In this respect the conditional test seems to behave similarly to the Pearson statistic although it seems to be slightly more anti-conservative.

In conclusion, our test offers a unified and more powerful approach to association testing as compared to standard analyses. These two aspects seem to warrant wide adoption and we aim to integrate our procedure into standard software packages. The test statistic is identical to that of

the Pearson test. Only computing P-values requires a numeric integration. For high-throughput analysis P-values can be precomputed on a grid for the non-centrality parameter and values of the test statistic. Actual P-values can be computed utilizing bi-linear interpolation as we did in the GWAS simulations.

With regard to population stratification it is interesting to compute the distribution by relaxing the assumption $\eta_1 = 0$ in controls. Such a framework would allow for joint analysis of population stratification, HWE and association testing.

6. ACKNOWLEDGMENTS

We are grateful to Ruth Pfeiffer for a critical appraisal of the manuscript. Hajo Holzmann gratefully acknowledges financial support from the Claussen-Simon Stiftung and from the Landesstiftung Baden-Württemberg ("Juniorprofessorenprogramm"). The alopecia data set was kindly provided by Felix Brockschmidt, Michael Steffens and Markus Nöthen.

A. APPENDIX

A.1 Proof of Theorem 1

Proof. We set

$$\tilde{\mathcal{L}}(\rho_1, \eta_1, \rho_2, \eta_2) = \mathcal{L}\big(T^{-1}(\rho_1, \eta_1), T^{-1}(\rho_2, \eta_2)\big), \qquad \tilde{\mathcal{L}}(\rho, \eta) = \mathcal{L}\big(T^{-1}(\rho, \eta)\big).$$

We let $(\hat{\rho}_1, \hat{\eta}_1, \hat{\rho}_2, \hat{\eta}_2)$ and $(\hat{\rho}, \hat{\eta})$ be the unrestricted and restricted ML estimators of the transformed parameters, respectively. Denote the true value for the ρ's by $\rho_1 = \rho_2 = \rho_0$ (by assumption, $\eta_1 = \eta_2 = 0$). Set

$$a = (1 - \rho_0)\rho_0, \qquad b = 2, \tag{A.1}$$

so that $\big(\mathcal{I}_0^{-1}\big)_{1,1} = a$, $\big(\mathcal{I}_0^{-1}\big)_{2,2} = b$, in case $\eta = 0$ (note that $(\mathcal{I}_0^{-1})_{1,2} = (\mathcal{I}_0^{-1})_{2,1} = 0$). Now

$$\sqrt{N}\big((\hat{\rho}_1, \hat{\eta}_1, \hat{\rho}_2, \hat{\eta}_2)^T - (\rho_0, 0, \rho_0, 0)^T\big) \xrightarrow{d} N(0, \mathcal{I}_{Joint}^{-1}),$$

where $\mathcal{I}_{Joint}^{-1} = \mathrm{diag}\big(a/c, b/c, a/(1-c), b/(1-c)\big)$. A standard argument in likelihood theory (see *e.g.* [Ferguson, 1996], p. 145, eq (4)) now shows that under our assumptions,

$$\begin{aligned} & 2\big(\mathcal{L}(\hat{\rho}_1, \hat{\eta}_1, \hat{\rho}_2, \hat{\eta}_2) - \mathcal{L}(\hat{\rho}, \hat{\eta})\big) \\ = \;& N\big(\hat{\rho}_1 - \hat{\rho}, \hat{\eta}_1 - \hat{\eta}, \hat{\rho}_2 - \hat{\rho}, \hat{\eta}_2 - \hat{\eta}\big)\mathcal{I}_{Joint}\big(\hat{\rho}_1 - \hat{\rho}, \hat{\eta}_1 - \hat{\eta}, \hat{\rho}_2 - \hat{\rho}, \hat{\eta}_2 - \hat{\eta}\big)^T + o_P(1) \\ = \;& N\Big(\frac{c}{a}\big(\hat{\rho}_1 - \hat{\rho}\big)^2 + \frac{1-c}{a}\big(\hat{\rho}_2 - \hat{\rho}\big)^2\Big) + N\Big(\frac{c}{b}\big(\hat{\eta}_1 - \hat{\eta}\big)^2 + \frac{1-c}{a}\big(\hat{\eta}_2 - \hat{\eta}\big)^2\Big) + o_P(1). \\ = \;& A_{1,n} + A_{2,n} + o_P(1). \end{aligned} \tag{A.2}$$

Now, straightforward computations give the covariance in the following (degenerate) asymptotic covariance matrix,

$$\sqrt{N}\big((\hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{21}, \hat{\pi}_{12}, \hat{\pi}_1, \hat{\pi}_2) - (\pi_1, \pi_2, \pi_1, \pi_2, \pi_1, \pi_2)\big) \to N(0, \Sigma_\pi),$$

where $(\pi_1, \pi_2) = T^{-1}(\rho_0, 0)$ and

$$\Sigma_\pi = \begin{pmatrix} \frac{(1-\pi_1)\pi_1}{c} & -\frac{\pi_1\pi_2}{c} & 0 & 0 & (1-\pi_1)\,\pi_1 & -\pi_1\pi_2 \\ -\frac{\pi_1\pi_2}{c} & \frac{(1-\pi_2)\pi_2}{c} & 0 & 0 & -\pi_1\pi_2 & (1-\pi_2)\,\pi_2 \\ 0 & 0 & \frac{(1-\pi_1)\pi_1}{1-c} & -\frac{\pi_1\pi_2}{1-c} & (1-\pi_1)\,\pi_1 & -\pi_1\pi_2 \\ 0 & 0 & -\frac{\pi_1\pi_2}{1-c} & \frac{(1-\pi_2)\pi_2}{1-c} & -\pi_1\pi_2 & (1-\pi_2)\,\pi_2 \\ (1-\pi_1)\,\pi_1 & -\pi_1\pi_2 & (1-\pi_1)\,\pi_1 & -\pi_1\pi_2 & (1-\pi_1)\,\pi_1 & -\pi_1\pi_2 \\ -\pi_1\pi_2 & (1-\pi_2)\,\pi_2 & -\pi_1\pi_2 & (1-\pi_2)\,\pi_2 & -\pi_1\pi_2 & (1-\pi_2)\,\pi_2 \end{pmatrix}$$

Therefore, from the δ-method,

$$\sqrt{N}\big((\hat{\rho}_1, \hat{\eta}_1, \hat{\rho}_2, \hat{\eta}_2, \hat{\rho}, \hat{\eta})^T - (\rho_0, 0, \rho_0, 0, \rho_0, 0)^T\big) \to N(0, \mathcal{I}_{comp}^{-1}), \tag{A.3}$$

where

$$\mathcal{I}_{comp}^{-1} = \begin{pmatrix} a/c & 0 & 0 & 0 & a & 0 \\ 0 & b/c & 0 & 0 & 0 & b \\ 0 & 0 & a/(1-c) & 0 & a & 0 \\ 0 & 0 & 0 & b/(1-c) & 0 & b \\ a & 0 & a & 0 & a & 0 \\ 0 & b & 0 & b & 0 & b \end{pmatrix}$$

Thus, it follows that $A_{1,n}$ and $A_{2,n}$ in (A.2) are asymptotically independent, and that $A_{1,n}$ is, also conditionally on $\hat{\eta}_1$, asymptotically distributed as χ_1^2. As for $A_{2,n}$, from (A.3) it follows that

$$d\Big(\mathcal{D}\Big(\sqrt{N}\big(\hat{\eta}_2, \hat{\eta}\big)|\hat{\eta}_1\Big), \big(N(\mu_{\hat{\eta}_1}, \Sigma_{cond})|\hat{\eta}_1\big)\Big) = o_P(1),$$

where

$$\mu_{\hat{\eta}_1} = (0, c\sqrt{N}\hat{\eta}_1)^T, \qquad \Sigma_{cond} = \begin{pmatrix} b/(1-c) & b \\ b & (1-c)b \end{pmatrix}.$$

Thus,

$$d\Big(\mathcal{D}\Big(A_{2,n}|\hat{\eta}_1\Big), \mathcal{D}\Big(W|\hat{\eta}_1\Big)\Big) = o_P(1),$$

where

$$W = \frac{c}{b}\big((1-c)^{1/2}X - (1-c)\sqrt{N}\hat{\eta}_1\big)^2 + \frac{1-c}{a}\big((1-c)^{1/2}X + c\sqrt{N}\hat{\eta}_1 - X/(1-c)^{1/2}\big)^2$$

and $X \sim N(0, b)$ is independent of $\hat{\eta}_1$. W reduces in distribution to

$$W|\hat{\eta}_1 \stackrel{d}{=} c\Big(Y - \frac{\sqrt{N}\hat{\eta}_1(1-c)^{1/2}}{\sqrt{b}}\Big)^2|\hat{\eta}_1 \sim c\chi_1^2\big(N\hat{\eta}_1^2(1-c)/b\big),$$

where $Y \sim N(0,1)$ is independent of $\hat{\eta}_1$. The theorem follows.

□

A.2 Simulations

We here give details about the simulation of genotypes for the GWAS simulations. With the penetrance model (3) and the choosing of parameters as described in the simulation section the alternative hypothesis is fully specified. However, computationally it is not straightforward to choose the number of loci under the alternative K and random penetrance parameters $\beta = (\beta_2, ..., \beta_K)$ such that

$$\alpha = P(Y = 1),$$

for a given α. In order to efficiently draw parameters we use a step-wise procedure. We construct a parameter vector $\theta^{(i)} = (\alpha, \mu, \beta_1, \beta_2, ..., \beta_{i+1}, \rho_1, \rho_2, ..., \rho_{i+1})$ by adding parameters $\beta_{i+1} \sim U(1.1, \beta_1)$, $\rho_{i+1} \sim U(\rho_1/2, \rho_1)$ to $\theta^{(i-1)}$. We then estimate prevalence $P(Y = 1)$ by $\hat{\alpha}$ and accept $\theta^{(i)}$ if $\hat{\alpha} < \alpha$. Otherwise we draw new parameters β_{i+1}, ρ_{i+1}. We stop the procedure, if $\hat{\alpha} \in (\alpha - \epsilon, \alpha + \epsilon)$. We use $\epsilon = 10^{-2}$ in all simulations.

As computation of α would require summation over all possible genotype combinations in formula (4), the number of which grows exponentially with the number of loci, we instead estimate α by Monte-Carlo integration. Instead, we compute $\hat{\alpha} = \sum_{(g_1,..,g_K)\in\mathbf{G_0}} Y(\mathbf{g};\beta)G(\mathbf{g};\rho)$, with $\mathbf{G_0} = (g_{1j}, .., g_{ij})_{j=1}^M$, where each g_{lj} is independently drawn from the corresponding genotype distribution of locus l. We use $M = 10^5$ in the simulations.

After specifying the alternative, we independently draw genotypes from the distribution as specified by $\rho = (\rho_1, ..., \rho_K)$ and assign a phenotype according to the penetrance model. As we typically use $\alpha = .1$ in the simulations, the ratio of generated cases and controls is around 1:9, which is a tolerable excess of controls.

Finally we simulate S control loci for which the distribution between cases and controls is identical. We use $S = 3 \times 10^5$ in the simulations. We average results from simulations over 10^3 runs for each alternative that was generated once as outlined above.

REFERENCES

Benjamini, Y., and Hochberg, Y. [1995], "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *J. Royal Statist. Soc. B* **57**, 289–289.

Carlson, C. S., Eberle, M. A., Rieder, M. J., Yi, Q., Kruglyak, L., and Nickerson, D. A. [2004], "Selecting a Maximally Informative Set of Single-Nucleotide Polymorphisms for Association Analyses Using Linkage Disequilibrium," *American J. Human Genetics* **74**, 106–120.

Chen, J., and Chatterjee, N. [2007], "Exploiting Hardy-Weinberg Equilibrium for Efficient Screening of Single SNP Associations from Case-Control Studies," *Hum. Hered.* **63**, 196–204.

Consortium, W. T. C. C. [2007], "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.," *Nature* **447**, 661–678.

Devlin, B., and Roeder, K. [1999], "Genomic Control for Association Studies," *Biometrics* **55**, 997–1004.

Edwards, A. W. F. [2000], *Foundations of Mathematical Genetics* Cambridge University Press.

Ferguson, T. S. [1996], *A Course in Large Sample Theory* Chapman & Hall/CRC.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L. T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J. T., Agnarsson, B. A., Baker, A., Sigurdsson, A., Benediktsdottir, K. R., Jakobsdottir, M., Xu, J., Blondal, T., Kostic, J., Sun, J., Ghosh, S., Stacey, S. N., Mouy, M., Saemundsdottir, J., Backman, V. M., Kristjansson, K., Tres, A., Partin, A. W., Albers-Akkers, M. T., Marcos, J. G., Walsh, P. C., Swinkels, D. W., Navarrete, S., Isaacs, S. D., Aben, K. K., Graif, T., Cashy, J., Ruiz-Echarri, M., Wiley, K. E., Suarez, B. K., Witjes, J. A., Frigge, M., Ober, C., Jonsson, E., Einarsson, G. V., Mayordomo, J. I., Kiemeney, L. A., Isaacs, W. B., Catalona, W. J., Barkardottir, R. B., Gulcher, J. R., Thorsteinsdottir, U., Kong, A., and Stefansson, K. [2007], "Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24," *Nat. Genet.* **39**, 631–637.

Hillmer, A. M., Brockschmidt, F. F., Hanneken, S., Eigelshoven, S., Steffens, M., Flaquer, A., Herms, S., Becker, T., Kortüm, A., Nyholt, D. R., Zhao, Z. Z., Montgomery, G. W., Martin, N. G., Mühleisen, T. W., Alblas, M. A., Moebus, S., Jöckel, K., Bröcker-Preuss, M., Erbel, R., Reinartz, R., Betz, R. C., Cichon, S., Propping, P., Baur, M. P., Wienker, T. F., Kruse, R., and Nöthen, M. M. [2008], "Susceptibility variants for male-pattern baldness on chromosome 20p11," *Nat. Genet.* **40**, 1279–81.

Hudson, R. R. [1990], "Gene genealogies and the coalescent process," *Oxford Surveys in Evolutionary Biology*, 7, 1–44.

Johnson, N. L., Kotz, S., and Balakrishnan, N. [1994], *Continuous univariate distributions. Vol. 1* Wiley.

Kimmel, G., and Shamir, R. [2006], "A Fast Method for Computing High-Significance Disease Association in Large Population-Based Studies," *American J. Human Genetics* **79**, 481–492.

Kruglyak, L. [2008], "The road to genome-wide association studies.," *Nat. Rev. Genet.* **9**, 314–318.

Salanti, G., Amountza, G., Ntzani, E. E., and Ioannidis, J. P. A. [2005], "Hardy–Weinberg equilibrium in genetic association studies: an empirical evaluation of reporting, deviations, and power," *European J. Human Genetics* **13**, 840–848.

Salmela, E., Lappalainen, T., Fransson, I., Andersen, P. M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M. L., Schreiber, S., and Kere, J. [2008], "Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncovers Population Structure in Northern Europe," *PLoS ONE* **3**.

Song, K., and Elston, R. C. [2006], "A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine-mapping in case-control studies," *Statistics in Medicine* **25**, 105–126.

Taylor, J., and Tibshirani, R. [2006], "A tail strength measure for assessing the overall univariate significance in a dataset," *Biostat.* **7**, 167–181.

Team, R. D. C. [2008], *R: A Language and Environment for Statistical Computing*, Vienna, Austria:.

Voight, B. F., Pritchard, J. K., and Abecasis, G. [2005], "Confounding from cryptic relatedness in case-control association studies," *PLoS Genet.* **1**, e32.

Wang, J., and Shete, S. [2008], "A test for genetic association that incorporates information about deviation from Hardy-Weinberg proportions in cases," *American J. Human Genetics* **83**, 53–63.

Wittke-Thompson, J. K., Pluzhnikov, A., and Cox, N. J. [2005], "Rational Inferences about Departures from Hardy-Weinberg Equilibrium," *American J. Human Genetics* **76**, 967–986.

Xu, J., Turner, A., Little, J., Bleecker, E. R., and Meyers, D. A. [2002], "Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error?," *Human Genetics* **111**, 573–574.

Yeager, M., Orr, N., Hayes, R. B., Jacobs, K. B., Kraft, P., Wacholder, S., Minichiello, M. J., Fearnhead, P., Yu, K., Chatterjee, N., Wang, Z., Welch, R., Staats, B. J., Calle, E. E., Feigelson, H. S., Thun, M. J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F. R., Giovannucci, E., Willett, W. C., Cancel-Tassin, G., Cussenot, O., Valeri, A., Andriole, G. L., Gelmann, E. P., Tucker, M., Gerhard,

D. S., Fraumeni, J. F., Hoover, R., Hunter, D. J., Chanock, S. J., and Thomas, G. [2007], "Genome-wide association study of prostate cancer identifies a second risk locus at 8q24," *Nat. Genet.* **39**, 645–649.

Table 1: Single locus simulations comparing the power for the conditional test (Φ^*), the EHWE test ($\Phi^\bullet$) and the Pearson test (Φ). 10^3 controls and 10^3 cases were drawn from the distributions (ρ_0, η_0) and (ρ_1, η_1), respectively. The significance level was chosen as $\alpha = .05$ and the power was estimated from 10^4 simulations.

ρ_0	η_0	ρ_1	η_1	$\Phi^*_{.05}$	$\Phi^\bullet_{.05}$	$\Phi_{.05}$
0.10	0.00	0.10	0.00	0.048	0.051	0.049
0.50	0.00	0.50	0.00	0.053	0.053	0.053
0.10	0.00	0.10	0.05	0.199	0.320	0.173
0.50	0.00	0.50	0.05	0.179	0.273	0.155
0.10	0.00	0.10	0.10	0.772	0.990	0.791
0.50	0.00	0.50	0.10	0.575	0.820	0.507
0.10	0.00	0.11	0.00	0.149	0.138	0.135
0.50	0.00	0.51	0.00	0.088	0.084	0.082
0.10	0.00	0.12	0.00	0.457	0.426	0.423
0.50	0.00	0.52	0.00	0.209	0.188	0.188
0.10	0.05	0.10	0.05	0.029	0.313	0.049
0.50	0.05	0.50	0.05	0.037	0.275	0.050
0.10	-0.05	0.10	-0.05	0.043	0.255	0.049
0.50	-0.05	0.50	-0.05	0.041	0.273	0.051
0.10	0.05	0.10	0.10	0.195	0.990	0.332
0.50	0.05	0.50	0.10	0.162	0.819	0.154
0.10	-0.05	0.10	0.00	0.048	0.053	0.136
0.50	-0.05	0.50	0.00	0.052	0.052	0.154
0.10	0.05	0.12	0.05	0.277	0.653	0.444
0.50	0.05	0.52	0.05	0.141	0.425	0.195
0.10	-0.05	0.14	-0.05	0.884	0.967	0.940
0.50	-0.05	0.54	-0.05	0.469	0.761	0.590

Table 2: Simulations of GWAs under different scenarios based on 10^3 simulations. $OR_1 = exp(\beta_1)$ is the maximal effect size of all loci, ρ_1 is the corresponding allele frequency, μ is the baseline penetrance and M denotes the mode of inheritance. Sample size was 2×10^3 for controls and cases. Q_x is the average rank of the locus under the alternative with the x-quantile according to p-value among SNPs under the alternative. m_α is the average number of rejected SNPs according to an FDR criterion at level α and Φ_α is the corresponding power (for details see text). A * represents results for the conditional test, $^\bullet$ represents EHWE and a plain notation indicates the Pearons test.

OR_1	ρ_1	μ	M	Q_0^*	$Q_0^\bullet$	Q_0	$Q_{.25}^*$	$Q_{.25}^\bullet$	$Q_{.25}$	$m_{.05}^*$	$m_{.05}^\bullet$	$m_{.05}$	$\Phi_{.05}^*$	$\Phi_{.05}^\bullet$	$\Phi_{.05}$
1.5	0.2	0.050	D	1.3	1.5	1.9	33.2	34.1	71.3	1.4	1.5	1.1	0.253	0.253	0.199
1.5	0.2	0.050	A	41.6	78.0	79.3	1534.6	2322.9	2353.2	0.2	0.1	0.1	0.107	0.039	0.045
1.5	0.2	0.050	R	37.6	21.7	164.9	15138.0	8203.9	21819.8	1.0	0.1	0.0	0.599	0.033	0.010
1.5	0.2	0.005	D	1.0	1.0	1.0	12.1	13.0	17.7	6.2	6.1	4.5	0.570	0.559	0.548
1.5	0.2	0.005	A	6.7	16.2	15.0	2668.7	3990.4	4016.2	0.4	0.1	0.1	0.246	0.065	0.060
1.5	0.2	0.005	R	16.5	5.5	77.7	11075.8	5175.3	16663.3	1.1	0.2	0.0	0.524	0.108	0.013
1.8	0.1	0.050	D	1.0	1.0	1.0	2.3	2.3	2.3	3.8	3.9	3.7	0.475	0.487	0.503
1.8	0.1	0.050	A	11.3	21.7	21.5	540.6	862.9	866.8	0.3	0.2	0.2	0.095	0.062	0.068
1.8	0.1	0.050	R	19.3	6.7	108.7	12700.6	5896.9	19041.4	1.0	0.2	0.0	0.563	0.085	0.007
1.8	0.1	0.005	D	1.0	1.0	1.0	8.0	8.0	8.0	12.6	13.0	11.7	0.373	0.376	0.420
1.8	0.1	0.005	A	1.2	1.9	1.8	326.3	575.5	581.0	1.1	0.7	0.6	0.331	0.238	0.221
1.8	0.1	0.005	R	6.3	1.5	28.8	5735.5	1670.7	9783.6	0.8	0.8	0.0	0.440	0.274	0.016
2.0	0.1	0.050	D	1.0	1.0	1.0	3.0	3.0	3.2	5.2	5.1	4.3	0.251	0.261	0.276
2.0	0.1	0.050	A	8.3	17.1	18.9	1209.8	1887.0	1814.3	0.4	0.2	0.2	0.099	0.042	0.035
2.0	0.1	0.050	R	44.5	23.7	339.1	24396.8	14778.6	33595.1	1.5	0.1	0.0	0.720	0.050	0.009
2.0	0.1	0.005	D	1.0	1.0	1.0	11.8	12.3	14.2	13.2	12.6	10.4	0.497	0.502	0.498
2.0	0.1	0.005	A	1.1	1.6	1.6	486.9	828.4	829.3	1.4	0.7	0.6	0.383	0.250	0.211
2.0	0.1	0.005	R	33.2	15.6	265.5	21262.2	11883.1	30055.9	1.3	0.1	0.0	0.678	0.061	0.004

Table 3: Analysis of an alopecia data set. Results are ordered according to P-values of the conditional test and the first 30 SNPs are shown. Columns in parentheses show order statistics. C denotes the conditional test, P denotes the Pearson test and E denotes EHWE. λ denotes the NCP of the conditional test.

SNP	C	P	(P)	EHWE	(E)	HWE	λ
rs1998076	1.21e-07	7.15e-07	1	4.23e-07	14	5.9e-01	0.24
rs6075852	2.50e-07	1.47e-06	2	9.65e-07	18	6.6e-01	0.17
rs2180439	2.53e-07	1.54e-06	3	1.01e-06	20	6.6e-01	0.17
rs201571	8.37e-07	4.68e-06	7	2.44e-06	23	5.8e-01	0.26
rs6113491	8.75e-07	1.78e-06	4	3.13e-06	25	1.2e-01	2.01
rs6137444	2.17e-06	1.02e-05	8	3.63e-06	27	6.5e-01	0.18
rs10992241	3.64e-06	2.00e-04	78	1.08e-05	45	9.3e-01	0.01
rs6137473	3.83e-06	1.60e-05	10	1.38e-05	50	4.9e-01	0.41
rs6047768	7.16e-06	2.43e-05	15	2.12e-05	63	4.0e-01	0.59
rs2207878	7.51e-06	2.25e-05	13	2.05e-05	60	3.2e-01	0.84
rs6113424	7.87e-06	3.06e-05	19	2.59e-05	75	5.1e-01	0.37
rs201543	7.88e-06	3.06e-05	17	2.59e-05	73	5.1e-01	0.37
rs6035995	8.04e-06	3.04e-05	16	2.56e-05	71	4.7e-01	0.44
rs2024885	8.08e-06	3.06e-05	18	2.59e-05	74	5.1e-01	0.37
rs1884592	9.28e-06	3.89e-05	26	3.27e-05	85	5.9e-01	0.25
rs6137476	9.49e-06	3.89e-05	28	3.27e-05	87	5.9e-01	0.25
rs1555264	9.53e-06	3.64e-05	25	3.14e-05	82	5.1e-01	0.37
rs6047769	9.57e-06	3.62e-05	23	3.09e-05	81	5.0e-01	0.38
rs2328683	9.82e-06	4.08e-05	31	3.54e-05	93	6.5e-01	0.18
rs6047731	9.84e-06	3.89e-05	27	3.27e-05	86	5.9e-01	0.25
rs4805229	1.15e-05	4.82e-05	33	3.60e-05	94	9.8e-01	0.00
rs655683	1.53e-05	2.13e-05	11	5.93e-06	36	1.8e-01	1.49
rs927059	2.05e-05	7.34e-05	43	6.47e-05	109	5.3e-01	0.33
rs6106434	2.05e-05	8.50e-05	49	7.08e-05	113	5.5e-01	0.30
rs1009840	2.17e-05	5.23e-05	35	4.90e-05	100	2.2e-01	1.22
rs4896028	2.26e-05	6.82e-05	41	6.69e-05	110	3.1e-01	0.86
rs500629	2.31e-05	8.08e-05	46	8.71e-05	124	5.6e-01	0.29
rs6137547	2.32e-05	6.40e-05	39	5.77e-06	34	5.3e-01	0.38
rs9300398	2.65e-05	5.04e-05	34	8.83e-05	125	2.5e-01	1.15
rs4771987	3.16e-05	1.05e-04	56	9.49e-05	135	5.6e-01	0.29

Table 4: Analysis of an alopecia data set. Results are ordered according to P-values of EHWE and the first 30 SNPs are shown. Columns in parentheses show order statistics. C denotes the conditional test, P denotes the Pearson test and E denotes EHWE. λ denotes the NCP of the conditional test.

SNP	EHWE	C	(C)	P	(P)	HWE	λ
rs1506694	1.13e-11	8.17e-03	2195	3.43e-03	900	7.9e-03	4.25
rs12357377	1.13e-10	9.68e-01	245387	2.23e-01	57416	4.5e-07	16.24
rs2189935	1.88e-10	2.47e-01	63594	4.02e-02	10418	2.2e-04	6.93
rs3760877	1.07e-09	1.51e-01	39079	1.44e-02	3768	6.2e-03	6.79
rs396999	3.17e-09	1.00e+00	253779	6.01e-01	153007	7.2e-09	23.04
rs1873921	6.30e-09	1.06e-01	27530	2.40e-03	640	3.1e-02	6.12
rs13362504	3.45e-08	1.73e-02	4630	4.57e-03	1208	8.4e-03	4.91
rs11232869	7.84e-08	7.72e-01	195223	2.85e-01	73175	3.3e-05	7.20
rs1491485	1.45e-07	9.71e-01	246040	6.68e-02	17202	1.7e-09	22.81
rs12282	1.50e-07	2.36e-04	80	6.35e-04	184	9.4e-01	0.00
rs632547	1.63e-07	9.88e-01	250757	4.88e-01	124485	2.1e-06	12.32
rs7857803	1.96e-07	5.35e-01	135381	8.97e-02	23171	3.5e-03	7.71
rs4307321	3.35e-07	8.29e-01	209871	2.02e-01	52160	4.7e-04	10.17
rs1998076	4.23e-07	1.21e-07	1	7.15e-07	1	5.9e-01	0.24
rs647731	5.53e-07	5.18e-02	13690	3.01e-02	7797	6.2e-02	2.49
rs2687860	8.08e-07	1.00e+00	253697	3.76e-01	96292	1.5e-07	24.61
rs295117	9.08e-07	9.85e-01	249783	4.01e-01	102635	8.9e-05	13.95
rs6075852	9.65e-07	2.50e-07	2	1.47e-06	2	6.6e-01	0.17
rs9285864	9.70e-07	8.32e-01	210441	1.49e-02	3907	1.2e-07	21.28
rs2180439	1.01e-06	2.53e-07	3	1.54e-06	3	6.6e-01	0.17
rs2975520	1.19e-06	1.00e+00	253844	8.32e-01	211079	1.0e-09	25.35
rs875001	1.42e-06	1.49e-01	38498	6.47e-02	16671	3.4e-02	3.19
rs201571	2.44e-06	8.37e-07	4	4.68e-06	7	5.8e-01	0.26
rs2425628	2.78e-06	9.71e-01	246163	4.55e-01	115990	2.7e-04	10.94
rs6113491	3.13e-06	8.75e-07	5	1.78e-06	4	1.2e-01	2.01
rs970952	3.51e-06	6.56e-01	165884	2.47e-01	63608	1.9e-03	6.09
rs6137444	3.63e-06	2.17e-06	6	1.02e-05	8	6.5e-01	0.18
rs10824842	3.93e-06	1.00e+00	253840	3.42e-01	87705	8.3e-04	32.92
rs1555257	4.53e-06	3.37e-03	965	3.01e-03	803	2.5e-01	1.03
rs7744253	4.67e-06	8.75e-01	221379	3.55e-01	90907	9.4e-04	7.74

Table 5: Correlation structure between the conditional, EHWE, Pearson, HWE tests for the obesity data set.

	Conditional	EHWE	Pearson	HWE
Conditional	1.00	0.76	0.95	-0.01
EHWE	0.76	1.00	0.72	0.01
Pearson	0.95	0.72	1.00	0.24
HWE	-0.01	0.01	0.24	1.00