

Application of the Optimal Discovery Procedure to Genetic Case-Control Studies: Comparison with p Values and Asymptotic Bayes Factors

Ioanna Tachmazidou^a Maria De Iorio^b Frank Dudbridge^{a, c, d}

^aMedical Research Council, Biostatistics Unit, Cambridge, ^bDepartment of Epidemiology and Biostatistics, Imperial College London, and ^cDepartment of Non-Communicable Disease Epidemiology and ^dBloomsbury Centre for Genetic Epidemiology and Statistics, London School of Hygiene and Tropical Medicine, London, UK

Key Words

Genome-wide association scan · Hypothesis testing · Multiple testing · Rare variation · Single nucleotide polymorphism · Statistical significance

Abstract

Objectives: The power of genetic association studies is limited by stringent levels of statistical significance. To improve power, Bayes factors (BFs) have been suggested as an alternative measure to the p value, and Storey recently introduced an optimal discovery procedure (ODP) for multiple testing. We aimed to adapt the ODP to genetic case-control studies and to compare its power to p values and asymptotic BFs (ABFs). **Methods:** We propose estimators of the ODP based on prospective and retrospective likelihoods. We performed simulations based on independent common SNPs and on sequence data including rare variants. Effects of causal SNPs were simulated under various distributions of effect size. **Results:** The true ODP is never outperformed, but the estimated ODP has similar power to p values and ABFs. For common SNPs the ODP offers power advantages only in extreme scenarios. However, for rare variants the ODP and ABF detect more associations at low false-positive rates than do p values. **Conclusions:** The ODP can provide higher power than p values for genetic case-control studies of common variants. However, as the ABF has similar power to the ODP and is computed more rapidly, it is our currently preferred method.

Copyright © 2011 S. Karger AG, Basel

Introduction

With recent advances in high-throughput genotyping technology, increasing numbers of genome-wide association studies (GWAS) are being performed to identify single nucleotide polymorphisms (SNPs) and other variants underlying complex traits. Although GWAS have been successful in discovering many validated associations, it is clear that much heritable variation remains to be discovered, owing in part to the limited power of GWAS [1]. Factors that restrict the power include susceptibility variants of low frequency, variants with small to modest effects, incomplete linkage disequilibrium (LD) between markers and susceptibility variants and stringent thresholds of statistical significance when testing thousands of SNPs simultaneously [2]. Given practical constraints on the sample sizes that can be collected, approaches to improve the power of GWAS are desirable. Although the genetic architecture of a complex trait is immutable, power can be improved through study design and improved statistical methods.

The initial analyses of a GWAS consist of single-locus association tests of each marker with the trait using, typically, logistic regression for case-control studies and linear regression for quantitative traits [3, 4]. A summary measure of significance, usually a p value, is obtained for each SNP, and a threshold is applied to determine which SNPs are associated. Replication is considered mandatory before associations are regarded as established [5], and

selection of SNPs for replication can be based on a number of top-ranking SNPs, or the SNPs whose p value is smaller than a threshold, or a pragmatic combination of both criteria. Specification of a p value threshold is guided by control of a particular error rate such as the family-wise type-1 error [6], false discovery rate (FDR) [7, 8] or false-positive reporting probability [9–11]. We are not concerned here with the relative merits of different error measures, which have been discussed elsewhere [12, 13], but note that all are based upon summary statistics, which (at least for FDR) have been sorted into rank order. By far the most commonly used summary is the p value, but recently there has been increased interest in testing associations using Bayes factors (BFs) [4, 13, 14]. Bayesian methods alleviate some of the limitations of p values; for example, the strength of evidence given by BF does not depend on factors that affect power such as sample size or minor allele frequency (MAF). On the other hand, Bayesian analysis requires the specification of a prior distribution for the effect size of truly associated SNPs. See Stephens and Balding [14] for a discussion of Bayesian methods in genetic association studies.

The optimal discovery procedure (ODP) [15] both formalizes a concept of optimal multiple testing and suggests a means to reconcile p values and BFs. Storey [15] developed the ODP in a frequentist setting, and the criterion it optimizes is based on frequentist concepts of true- and false-positive rates over repeated sampling. However, it can be readily interpreted as an empirical BF that draws its prior from the ensemble of tests performed [15]. The advantage of the ODP compared to single significance tests is that when testing each SNP, it borrows strength by using information from all SNPs tested in the study. Specifically the ODP for a particular SNP involves considering the true effect size and the null effect size for the entire set of SNPs in the study. In this way, SNPs with similar effect sizes will reinforce each other, and tests that capture common signal structure in the dataset are enhanced. It is optimal in the sense that it maximizes the expected number of true positives (ETP), given a fixed expected number of false positives (EFP). However, the optimality property only applies when the true effect sizes of all SNPs are known, which is not the case in practice. In applications the ODP must be estimated, which may entail a loss of power compared to the theoretical optimum.

To date, the ODP has only been applied to quantitative gene expression data in relatively small microarray studies [16]. In this paper, we discuss the application of the ODP to large-scale case-control data. We propose estimators based on both prospective and retrospective likeli-

hoods, and compare the power of the estimated ODP to the true ‘oracle’ version. Specifically, we use simulation studies to compare the power of the theoretically optimal ODP and its estimated version to p values from single-SNP logistic regression and the asymptotic BF (ABF) derived by Wakefield [13]. Our main findings are that, whereas the true ODP is never outperformed, the estimated ODP offers advantages over the p value and ABF only in extreme simulation scenarios. In realistic scenarios the estimated ODP yields a similar number of true positives compared to the p value and ABF, and no method exhibits a clear advantage in power when applied to common SNPs. However, when rare variants are present, the ODP and ABF do show a clear increase in power over the p value at low false-positive rates. This may become important as attention turns to studies of whole-sequence data [17, 18]. Taking computational complexity into account, the ABF seems to offer the most attractive operating characteristics among the approaches considered, provided a low false-positive rate is used.

Methods

Logistic Regression

For testing association in case-control data, a common approach is to use a logistic regression model of the form:

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_{0j} + \beta_{1j}x_{ij} + \varepsilon_{ij}, \quad (1)$$

where p_i is the probability that subject i in the study is a case, parameter β_{0j} represents the log-odds of being a case for people with a baseline genotype at SNP j , parameter β_{1j} measures the increase in log-odds due to the genotype of subject i for SNP j coded by the variable x_{ij} , and ε_{ij} are random errors, assumed to be normally distributed about zero. Various codings are possible for x_{ij} corresponding to different genetic models: for example, if subject i carries 0, 1 or 2 copies of the minor allele at SNP j , x_{ij} equals 0, 1, 1, respectively, for a dominant disease model; 0, 0, 1 for a recessive model; or 0, 1, 2 for a log-additive model. We will henceforth assume the log-additive coding, but all expressions can easily be derived for recessive, dominant or other codings.

The likelihood function is

$$L(\beta_{0j}, \beta_{1j}; \mathbf{x}_j, \mathbf{y}) = \prod_{i=1}^N p_i^{y_i} (1-p_i)^{1-y_i}, \quad (2)$$

where $\mathbf{y} = (y_1, \dots, y_N)'$ is the case-control indicator with $y_i = 1$ for cases, 0 for controls.

Maximum likelihood estimates (MLEs) $\hat{\beta}_{0j}$ and $\hat{\beta}_{1j}$ are used to test for association at SNP j comparing the null hypothesis $H_0: \beta_{1j} = 0$ to the alternative $H_0: \beta_{1j} \neq 0$. p values can be derived from the standard methods including the likelihood-ratio and score tests. The Wald statistic $\hat{\beta}_{1j}^2/V_j \sim \chi_1^2$, where $\sqrt{V_j}$ is the (estimated) standard error of $\hat{\beta}_{1j}$, is of particular note for its relation to the ABF described below.

Bayes Factors

We wish to determine whether the data are generated under the null hypothesis H_0 or the alternative H_1 . The marginal likelihood for hypothesis H_k , $k = 0, 1$, is the probability of observing the data given that the hypothesis is true:

$$p(\mathbf{x}_j, \mathbf{y} | H_i) = \int_{\theta} L(\theta; \mathbf{x}_j, \mathbf{y} | H_i) p(\theta | H_i) d\theta \quad (3)$$

where θ is a vector of parameters, $L(\theta; \mathbf{x}_j, \mathbf{y} | H_k)$ is its likelihood function, and $p(\theta | H_k)$ denotes the prior distribution of the parameters under hypothesis H_i . The BF, defined by $p(\mathbf{x}_j, \mathbf{y} | H_1) / p(\mathbf{x}_j, \mathbf{y} | H_0)$ measures how much the data have increased or decreased the odds of H_1 to H_0 . For values >1 the odds have increased, whereas for values <1 the odds have decreased.

We use the likelihood function for SNP j given by equations 1 and 2, where $\theta_j = (\beta_{0j}, \beta_{1j})'$. To complete the calculation of the BF we require the specification of prior distributions $p(\theta_j | H_1) = p(\beta_{0j}, \beta_{1j} | H_1)$ and $p(\theta_j | H_0) = p(\beta_{0j} | H_0)I(\beta_{1j} = 0)$. For a log-additive effect, the Wellcome Trust Case Control Consortium [4] employ a standardized normal distribution as the prior of β_{0j} under both hypotheses, and an independent $N(0, 0.2^2)$ for β_{1j} under H_1 , which translates to a risk allele odds ratio (OR) with 95% of the distribution in the range -1.5 to 1.5 .

Recently, Wakefield [13] proposed an ABF that approximates the BF for large sample sizes. The ABF is fast and straightforward to calculate, as it only requires the MLE $\hat{\beta}_{1j}$ and standard error $\sqrt{V_j}$ of the log-OR β_{1j} , which can be obtained from the output of a logistic regression. In particular, assuming a normal prior $N(0, W_j)$ on β_{1j} , the ABF in favour of the alternative hypothesis for SNP x_j is given by:

$$\text{ABF}(\mathbf{x}_j) = \left[\sqrt{\frac{V_j + W_j}{V_j}} \exp\left\{-\frac{z_j^2}{2} \frac{W_j}{(V_j + W_j)}\right\} \right]^{-1}, \quad (4)$$

where $z_j^2 = \hat{\beta}_{1j}^2 / V_j$ is the Wald statistic. Wakefield [13] proposed a prior variance W_j that is independent of the MAF of SNP j , and one that varies with the MAF, as it can be argued that smaller MAFs are associated with larger effect sizes. He also derived a prior that produces identical rankings between ABFs and p-values.

To investigate the accuracy of the ABF, Wakefield [10] simulated case-control data with varying numbers of cases and controls, MAFs, and relative risks. The exact BF was calculated using both a rejection algorithm and importance sampling, and the approximation is accurate in all cases [10]. Figure 1B of Wakefield [10] displays the approximate BFs plotted against the exact BFs for a subset of the simulation scenarios, i.e. for a sample size of 250 and 500 cases/controls either under the null model of no association or with a relative risk of 1.3, and MAF 0.05.

Optimal Discovery Procedure

Storey [15] introduced the ODP as a multiple-testing extension of the Neyman-Pearson lemma and proved that it is an optimal thresholding procedure in the sense that for any expected number of false-positive results, the ODP will identify the maximum expected number of true-positive results. This optimality property can also be expressed in terms of FDR [15].

Suppose that markers $1, \dots, M$ are tested for association. Let the data for SNP j have null probability density function f_j and true alternative density g_j , both of which are assumed fixed and known.

Without loss of generality, assume that the null hypothesis is true for tests $m = 1, 2, \dots, M_0$. The ODP statistic can be written as:

$$S_{ODP}(\mathbf{x}_j, \mathbf{y}) = \frac{\sum_{m=1}^M g_m(\mathbf{x}_j, \mathbf{y})}{\sum_{m=1}^{M_0} f_m(\mathbf{x}_j, \mathbf{y})}, \quad (5)$$

The null hypothesis is rejected for all markers for which S_{ODP} exceeds a fixed threshold λ which is chosen to ensure a given EFP. (This threshold must be estimated by a permutation procedure as there is currently no asymptotic theory for it.) However, the optimality of the ODP is a theoretical result that applies when the null and alternative distributions are known. This information is unknown in practice, and therefore the ODP must be estimated. Let \hat{f}_j be the estimator of f_j with all its unknown parameters replaced by their estimates under the constraints of the null hypothesis and \hat{g}_j be the analogous estimate given by the unconstrained parameter estimates. The estimated ODP statistic is given by:

$$\hat{S}_{ODP}(\mathbf{x}_j, \mathbf{y}) = \frac{\sum_{m=1}^M \hat{g}_m(\mathbf{x}_j, \mathbf{y})}{\sum_{m=1}^{M_0} \hat{\omega}_m \hat{f}_m(\mathbf{x}_j, \mathbf{y})}, \quad (6)$$

where $\hat{\omega}$ serves as an estimate of the true status of each hypothesis and is necessary because M_0 is also unknown in practice.

Storey et al. [16] propose ranking the tests by a univariate statistic, and setting $\hat{\omega}_m = 0$ for those appearing to be significant and $\hat{\omega}_m = 1$ otherwise. They use a Kruskal-Wallis test for quantitative data, motivated by speed of computation. For case-control data, we will consider two choices of test for assigning the weights $\hat{\omega}$: the Fisher exact test of allelic association, and the Pearson test of the 2×3 contingency table of genotype counts.

A modified version of $\hat{S}_{ODP}(\mathbf{x}_j)$ is given by including the weights in the numerator as well as denominator:

$$\hat{S}_{ODP}(\mathbf{x}_j, \mathbf{y}) = \frac{\sum_{m=1}^M (1 - \hat{\omega}_m) \hat{g}_m(\mathbf{x}_j, \mathbf{y})}{\sum_{m=1}^M \hat{\omega}_m \hat{f}_m(\mathbf{x}_j, \mathbf{y})}, \quad (7)$$

which could increase power by only considering the likelihood under the alternative hypothesis for SNPs appearing to be significant.

Application to Genetic Association Studies: Prospective Model

The first application of the ODP we consider for case-control data simply uses the likelihood and MLEs from the standard logistic regression model described above. For SNP j the likelihood is averaged over the MLEs for the full set of SNPs:

$$\hat{S}_{ODP}(\mathbf{x}_j, \mathbf{y}) = \frac{\sum_{m=1}^M L(\hat{\beta}_{0m}, \hat{\beta}_{1m}; \mathbf{x}_j, \mathbf{y})}{\sum_{m=1}^M \hat{\omega}_m L(\hat{\beta}'_{0m}, 0; \mathbf{x}_j, \mathbf{y})}, \quad (8)$$

where L is the likelihood from equations 1 and 2, $\hat{\beta}_{0m}$ and $\hat{\beta}_{1m}$ are the MLEs under the alternative hypothesis, and $\hat{\beta}'_{0m}$ is the MLE under the null hypothesis.

When the case-control sampling ratio varies across SNPs (such as when genotypes are missing at random), it is desirable to adjust the intercept parameters appropriately. We describe an approach to this situation in an appendix.

Application to Genetic Association Studies: Retrospective Model

The logistic regression used above is prospective in that it models the probability of being a case, with a given genotype. However, in a case-control study it is the disease status that is fixed by design with the genotypes as the unknown outcome, so that a retrospective model is more appropriate. The widespread use of the prospective model, which can be computationally easier to handle, is supported by the result of Prentice and Pyke [19] according to which, in the absence of missing data, the OR estimators from univariate retrospective and prospective logistic models are equivalent. No such equivalence has been established for the ODP, however, so here we investigate the performance of the ODP when it is constructed from a retrospective model which properly reflects the sampling scheme.

A convenient approach is to define parameters for the genotype frequencies in controls, using a generalized logistic model:

$$p(X_{ij} = g | Y_i = 0) = \frac{\exp(\alpha_{gj})}{\sum_{g=0,1,2} \exp(\alpha_{gj})} \quad (9)$$

where $g = 0$ is the reference genotype class with $\alpha_{0j} = 0$ and

$$\alpha_{gj} = \log \left[\frac{\Pr(X_{ij} = g | Y_i = 0)}{\Pr(X_{ij} = 0 | Y_i = 0)} \right]$$

for $g = 1, 2$. The genotype frequencies in cases can be written as [20, 21]

$$p(X_{ij} = g | Y_i = 1) = \frac{\exp(\alpha_{gj} + \beta_{jg})}{\sum_{g=0,1,2} \exp(\alpha_{gj} + \beta_{jg})} \quad (10)$$

where β_{jg} is the log OR. The retrospective likelihood is then

$$L_R(\boldsymbol{\alpha}_j, \boldsymbol{\beta}_j; \mathbf{x}_j, \mathbf{y}) = \prod_{i=1}^N \left[\prod_{g=0,1,2} \left[\Pr(X_{ij} = g | Y_i = 0) \right]^{I(x_{ij}=g, y_i=0)} \left[\Pr(X_{ij} = g | Y_i = 1) \right]^{I(x_{ij}=g, y_i=1)} \right] \quad (11)$$

where I is the indicator function taking the value 1 when the operand is true, 0 otherwise.

If we assume Hardy-Weinberg equilibrium in the controls, which is reasonable for a rare disease, we could instead use an allele frequency model with one free parameter, say α'_{ij} . We then define $\alpha_{1j} \equiv \alpha'_{1j} + \log(2)$ and $\alpha_{2j} \equiv 2\alpha'_{1j}$.

The ODP estimate for SNP j based on the retrospective likelihood can be written as

$$\hat{S}_{ODP}^R(\mathbf{x}_j, \mathbf{y}) = \frac{\sum_{m=1}^M L_R(\hat{\boldsymbol{\alpha}}_m, \hat{\boldsymbol{\beta}}_{1m}; \mathbf{x}_j, \mathbf{y})}{\sum_{m=1}^M \hat{\omega}_m L_R(\hat{\boldsymbol{\alpha}}_m^0, 0; \mathbf{x}_j, \mathbf{y})}, \quad (12)$$

where $\hat{\boldsymbol{\beta}}_{1m}$ are the MLEs under the alternative hypothesis, $\hat{\boldsymbol{\alpha}}_m$ are estimated from the control subjects and $\hat{\boldsymbol{\alpha}}_m^0$ are estimated from the cases and controls combined. As the retrospective model conditions on the number of cases and controls observed, there is no need to adjust nuisance parameters $\boldsymbol{\alpha}_j$ to reflect a varying sampling ratio, as is done for the prospective model in the appendix.

True ODP

We will use simulations to examine how much accuracy is lost when estimating the ODP compared to the true ODP using known

model parameters. These parameters can easily be expressed in terms of those used in generating simulated data. The log ORs $\boldsymbol{\beta}_1$ are specified directly as are the weights $\boldsymbol{\omega}$. Suppose we are given the genotype frequencies in the general population, denoted $(\boldsymbol{\pi}_0, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$, and the population prevalence K . Assuming a prospective logistic model for disease with the given OR, its intercept for SNP j is found by the solution γ_j of

$$K = \sum_{g=0,1,2} \frac{\exp(\gamma_j + g\beta_{1j})}{1 + \exp(\gamma_j + g\beta_{1j})} \pi_{gj} \quad (13)$$

The genotype frequencies in controls are

$$\begin{aligned} p(X_{ij} = g | Y_i = 0) &= \frac{p(Y_i = 0 | X_{ij} = g) p(X_{ij} = g)}{p(Y_i = 0)} \\ &= \frac{\pi_{gj}}{(1 + \exp(\gamma_j + g\beta_{1j}))(1 - K)} \end{aligned} \quad (14)$$

The retrospective model parameters $\boldsymbol{\alpha}_j$ can then be fitted to these frequencies and the intercept for the prospective model is given by

$$\begin{aligned} \beta_{0j} &= \log \left[\frac{p(Y_i = 1 | x_{ij} = 0)}{p(Y_i = 0 | x_{ij} = 0)} \right] \\ &= \log \left[\frac{p(x_{ij} = 0 | Y_i = 1) p(Y_i = 1)}{p(x_{ij} = 0 | Y_i = 0) p(Y_i = 0)} \right] \\ &= \log \left[\frac{1 + \exp(\alpha_{1j}) + \exp(\alpha_{2j})}{1 + \exp(\alpha_{1j} + \beta_{1j}) + \exp(\alpha_{2j} + 2\beta_{1j})} \frac{r_j}{1 - r_j} \right] \end{aligned} \quad (15)$$

where r_j is the case-control sampling ratio for SNP j .

Simplification is possible under a rare disease assumption, since then $p(X_{ij} = g | Y_i = 0) \approx \pi_{gj}$ and the retrospective model parameters can be fitted directly to the population frequencies.

Simulations

We performed simulation studies to compare the power of the p value, ABF and ODP. We considered a range of simulation scenarios by varying the MAF of the SNPs considered as well as the risk allele ORs of the causal variants.

We first consider five simulation scenarios in which the SNPs are independent, that is in linkage equilibrium. In each of them, the MAF has a uniform distribution in the range of 0.01 and 0.5, similar to that seen on current genotyping chips, unless otherwise indicated. In the first scenario, the log ORs of the causal SNPs are from a Normal distribution centred around zero with a standard deviation of 0.2. This is the prior that is used by the SNPTEST software for calculating BFs under a log-additive disease model [22], and it reflects prior belief that variants contributing to complex diseases have ORs lying in the range -1.5 to 1.5 with 95% probability. In the second scenario, the log ORs are from an Exponential distribution with rate $1/\log(1.2)$, which implies that the OR has a mean value of about 1.2 with 90% probability < 1.5 . This represents a plausible model, inspired by a comment by Marchini et al. [22], which nevertheless differs from the prior used by the SNPTEST software. In the third scenario, we assume a constant OR of the causal SNPs of 1.2, and the fourth scenario additionally fixes the MAF of the causal SNPs at 0.15. Finally, we consider the situation in which the effect size depends on MAF. It has been

suggested that disease susceptibility loci may have larger effects at smaller MAF [2]. To reflect this dependence, we assume the log ORs of the causal SNPs are from a normal distribution $N(0, W)$, where the variance W follows the formula suggested by Wakefield [13]: $W(\text{MAF}) = \delta_0 \exp(-\delta_1 \text{MAF})$. To choose parameters $\delta_0, \delta_1 > 0$, we assume that for rare M_{lo} and non-rare M_{hi} MAFs, there is an upper bound $\text{OR}_u^{lo} > \text{OR}_u^{hi}$ on the prior for OR, above each of which the OR lies with probability q . The variances at the rare and non-rare SNPs are then given by $W_{lo} = [\log(\text{OR}_u^{lo} / \Phi^{-1}(1 - q))]^2$ and $W_{hi} = [\log(\text{OR}_u^{hi} / \Phi^{-1}(1 - q))]^2$. For our simulations, we use the following parameter values: $q = 0.05$, $\text{OR}_u^{lo} = 2$, $\text{OR}_u^{hi} = 1.5$, $M_{lo} = 0.01$ and $M_{hi} = 0.3$. The MAF of half of the causal variants is assumed to be uniformly distributed $U[0.01, 0.5]$, whereas the MAF of the other half is fixed to 0.01.

For each of the different MAF and OR combinations, we simulate 1,000 datasets containing $N = 2,000$ cases/controls genotyped at 1,000 independent SNPs, 50 of which are assumed to be causal with independent genetic effects.

We compare the performances of the p value, ABF (equation 4), prospective estimated ODP (equation 8) and retrospective estimated ODP (equation 12) to the true prospective and retrospective ODP. For the prior of the ABF, we assume that the log OR of the causal SNPs has a $N(0, 0.2^2)$ distribution. Following the optimality criterion of Storey [15], the goal is to maximize the ETP at each fixed level of EFP. To investigate the performance of the methods in terms of optimality, we define a range of thresholds, λ_S for each ODP version, λ_{ABF} for ABF and λ_{pV} for p values, and in each simulated dataset we count how many truly null SNPs have $\hat{S}_{ODP} \geq \lambda_S$, $\text{ABF} \geq \lambda_{BF}$ and p value $\leq \lambda_{pV}$. We can then obtain the EFP for each threshold and the expected false-positive rate (EFPR) after dividing by the number of truly null SNPs. The thresholds for the ODP may be different in each simulation scenario, so they are calculated separately each time. Using the thresholds λ_S, λ_{BF} and λ_{pV} that result in an EFPR of 5%, we count the number of truly causal SNPs that have $\hat{S}_{ODP} \geq \lambda_S$, $\text{ABF} \geq \lambda_{BF}$ and p value $\leq \lambda_{pV}$. This yields the ETP and the corresponding expected true-positive rate (ETPR) for a fixed EFPR of 5%.

We also compare the performance of the methods on sequence data in which the SNPs are correlated and which includes rare variants. We use 5 Mb of sequence data simulated by FREGENE [23] in a population of 20,000 haplotypes using the modelling assumptions of Schaffner et al. [24] with demographic and evolutionary parameters approximating the history of Europeans. Details of the simulated population are given in Hoggart et al. [25]. The population has 176,652 polymorphic sites. From the population we sample 100 datasets of 1,000 cases and 1,000 controls. Because there are many rare variants in a population of European descent, many sites become non-polymorphic. In our analyses, we use all the polymorphic sites, of which there are 91,420 on average.

We select ten variants to be causal and assume an additive disease model between and within the genotypes at the causal sites with a disease prevalence of 1%. We consider three simulation scenarios: (a) the MAF of the causal SNPs and their log OR are sampled from a uniform $U[0.01, 0.05]$ and a Normal distribution $N(0, 0.2^2)$, respectively; (b) the MAF of the causal SNPs and their OR are sampled from uniform distributions $U[0.01, 0.05]$ and $U[1.3, 1.7]$, respectively, and (c) the MAF of the causal SNPs and their OR are sampled from uniform distributions $U[0.01, 0.05]$ and $U[2.2, 2.5]$, respectively.

Results

Independent SNPs

Results from the five simulation scenarios are shown in tables 1–3. We show the ETPR given EFPR of 5% averaged over 1,000 replicates. We show results from the ODP when all weights in the denominator are set to 1 (table 1), when weights are based on Fisher's exact test setting the weight to 0 for the 5% most significant SNPs and 1 otherwise (table 2), and when the same weights are introduced in the numerator too (table 3). Note that the ABF uses the correct prior only in the first scenario. From tables 1–3, we observe that indeed the true ODP is never outperformed, and the true retrospective ODP performs better than the prospective version. However, there is little difference between the estimated versions, with a noticeable advantage for the retrospective version only in the scenario with OR depending on MAF. Overall, the ODP with weights in the numerator as well as the denominator does not perform as well as the other versions. Furthermore, the estimated ODP offers advantages over p values and ABFs only in the extreme scenario in which both MAF and OR are fixed. In the more realistic scenarios, the estimated ODP has similar power to p values and ABFs. Also, it is interesting to note that the ABF is robust to the choice of prior, as it is powerful in all situations considered here and not only when its prior distribution matches the distribution of the log OR. Moreover, we observe that, although the ABF is more interpretable than the p value, both are equally powerful for the situations considered here. Note that power is estimated here at a fixed EFPR of 5%. We also examine the performance of the methods as the EFPR varies between 1 and 20%, with the results shown in figure 1. We observe similar characteristics across the range of EFPR, notably that the true ODP is the most powerful for any EFPR, whereas the estimated ODP has similar power to the p values and ABFs.

We also examine the rankings of SNPs given by the different approaches. In online supplementary table 1, we give Spearman's rank correlation coefficient between the methods averaged over the 1,000 replicates simulated under each scenario (for all online suppl. material, see www.karger.com/doi/10.1159/000323518). Results from the ODP are with all weights in the denominator set to 1; results including weights as above were very similar (data not shown). The correlation is calculated for all SNPs, for the top 50-ranking SNPs by each method, and for the remaining 950 bottom-ranking SNPs. In general, the correlation coefficients calculated for the top-rank-

Table 1. ETPR (%) (and standard error) for 5% EFPR over 1,000 replicates simulated under five different scenarios

	MAF: $U[0.01, 0.5]$ logOR(Aa): $N(0, 0.2^2)$	$U[0.01, 0.5]$ $E(1/\log(1.2))$	0.15 log(1.2)	$U[0.01, 0.5]$ log(1.2)	$U[0.01, 0.5]$ \propto MAF
PV	53.61 (0.14)	51.79 (0.12)	85.18 (0.17)	87.61 (0.12)	19.81 (0.14)
BF	53.38 (0.14)	51.59 (0.12)	84.95 (0.17)	87.99 (0.13)	19.92 (0.14)
rS	52.97 (0.15)	51.43 (0.13)	87.93 (0.14)	87.01 (0.13)	19.41 (0.15)
pS	53.40 (0.14)	51.53 (0.12)	86.15 (0.16)	87.11 (0.11)	15.36 (0.11)
rTS	56.22 (0.16)	55.19 (0.15)	96.72 (0.08)	89.59 (0.14)	25.63 (0.12)
pTS	53.76 (0.14)	55.51 (0.13)	90.71 (0.14)	92.44 (0.10)	19.95 (0.14)

Results from the ODP are with all weights in the denominator set to 1. rS = ODP estimated using the retrospective likelihood; pS = ODP estimated using the prospective likelihood; rTS = true ODP calculated using the retrospective likelihood; pTS = true ODP calculated using the prospective likelihood.

Table 2. ETPR (%) (and standard error) for 5% EFPR over 1,000 replicates simulated under five different scenarios

	MAF: $U[0.01, 0.5]$ logOR(Aa): $N(0, 0.2^2)$	$U[0.01, 0.5]$ $E(1/\log(1.2))$	0.15 log(1.2)	$U[0.01, 0.5]$ log(1.2)	$U[0.01, 0.5]$ \propto MAF
PV	53.82 (0.14)	52.52 (0.13)	84.42 (0.17)	87.56 (0.12)	19.34 (0.14)
BF	53.80 (0.14)	52.16 (0.12)	84.26 (0.17)	87.84 (0.13)	19.28 (0.14)
rS	53.40 (0.14)	51.90 (0.13)	89.92 (0.17)	87.04 (0.13)	19.08 (0.15)
pS	53.42 (0.14)	52.12 (0.12)	86.14 (0.16)	87.11 (0.11)	17.36 (0.14)
rTS	56.78 (0.16)	54.88 (0.12)	96.66 (0.08)	89.54 (0.14)	25.26 (0.12)
pTS	54.03 (0.14)	55.01 (0.13)	91.81 (0.17)	91.54 (0.10)	20.86 (0.14)

Results from the ODP are with denominator weights based on the top 5%-ranking SNPs by Fisher's exact test. rS = ODP estimated using the retrospective likelihood; pS = ODP estimated using the prospective likelihood; rTS = true ODP calculated using the retrospective likelihood; pTS = true ODP calculated using the prospective likelihood.

Table 3. ETPR (%) (and standard error) for 5% EFPR over 1,000 replicates simulated under five different scenarios

	MAF: $U[0.01, 0.5]$ logOR(Aa): $N(0, 0.2^2)$	$U[0.01, 0.5]$ $E(1/\log(1.2))$	0.15 log(1.2)	$U[0.01, 0.5]$ log(1.2)	$U[0.01, 0.5]$ \propto MAF
PV	53.82 (0.14)	52.52 (0.13)	84.42 (0.17)	87.56 (0.12)	19.34 (0.14)
BF	53.80 (0.14)	52.16 (0.12)	84.26 (0.17)	87.84 (0.13)	19.28 (0.14)
rS	53.04 (0.15)	51.01 (0.13)	96.04 (0.17)	85.68 (0.13)	18.38 (0.16)
pS	53.31 (0.14)	52.07 (0.13)	88.14 (0.15)	87.43 (0.12)	17.38 (0.13)
rTS	56.78 (0.16)	54.88 (0.12)	96.66 (0.08)	89.54 (0.14)	25.26 (0.12)
pTS	54.03 (0.15)	55.01 (0.13)	91.81 (0.15)	91.54 (0.11)	20.86 (0.13)

Results from the ODP are with weights also introduced into the numerator. rS = ODP estimated using the retrospective likelihood; pS = ODP estimated using the prospective likelihood; rTS = true ODP calculated using the retrospective likelihood; pTS = true ODP calculated using the prospective likelihood.

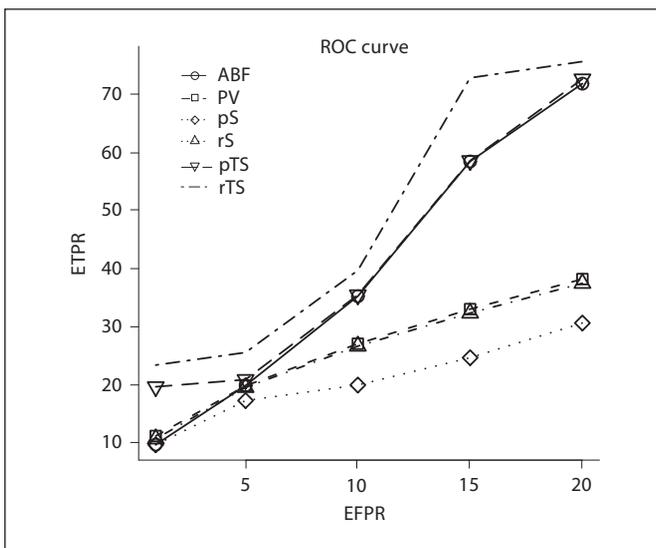
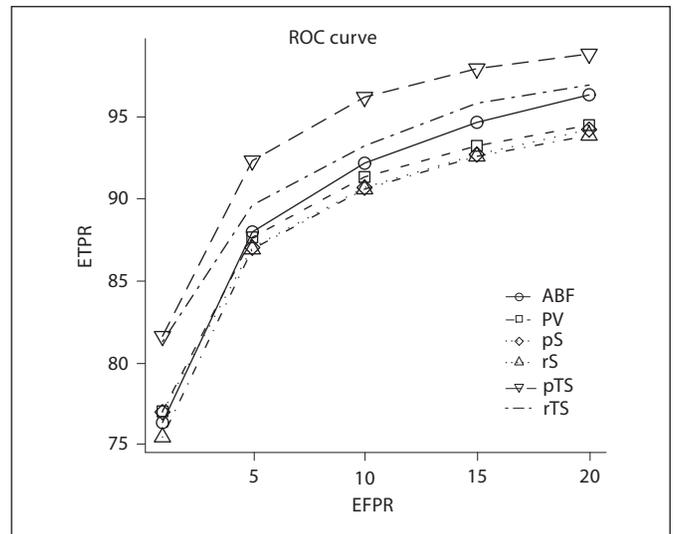
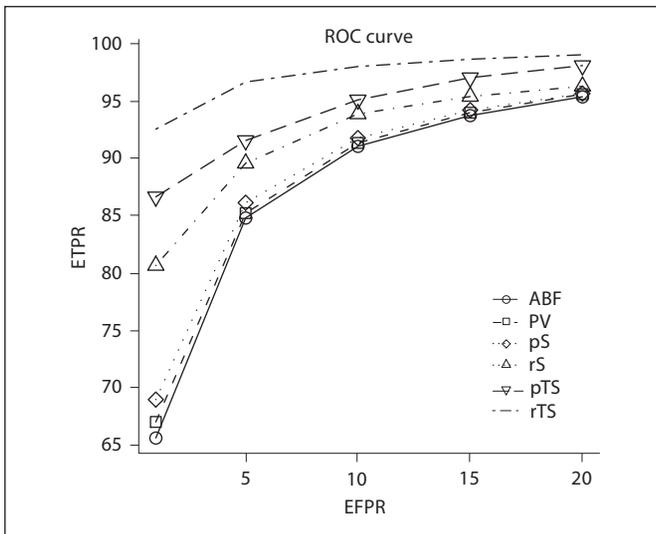
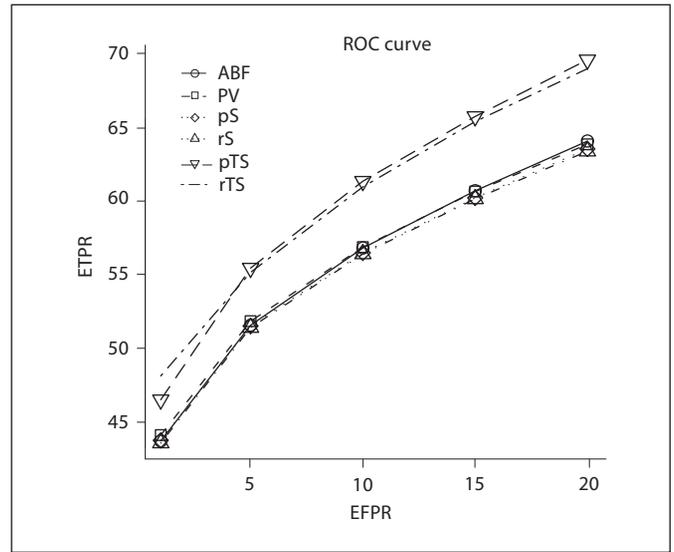
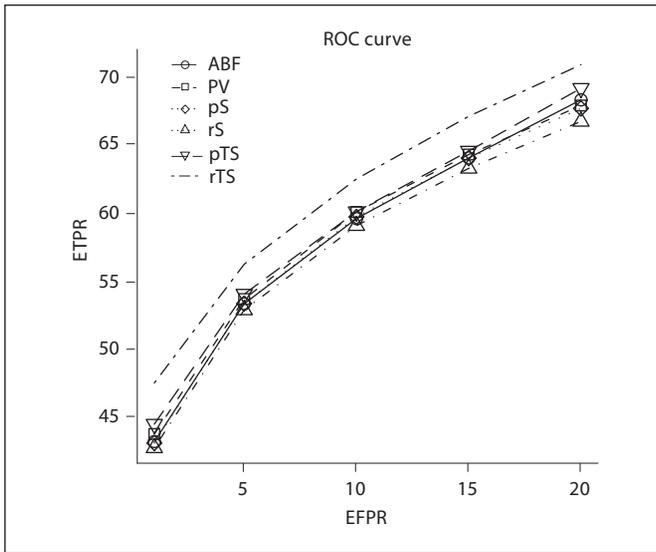


Fig. 1. ETPR (%) for a range of EFPR (%) over 1,000 replicates simulated under five different scenarios. Results from the ODP are with denominator weights based on the top 5%-ranking SNPs by Fisher's exact test. Top left: $MAF \sim U[0.01, 0.5]$, $\log OR(Aa) \sim N(0, 0.2^2)$. Top right: $MAF \sim U[0.01, 0.5]$, $\log OR(Aa) \sim E(1/\log(1.2))$. Middle left: $MAF = 0.15$, $OR(Aa) = 1.2$. Middle right: $MAF \sim U[0.01, 0.5]$, $OR(Aa) = 1.2$. Bottom: $MAF \sim U[0.01, 0.5]$, $\log OR(Aa) \propto MAF$. rS = ODP estimated using the retrospective likelihood; pS = ODP estimated using the prospective likelihood; rTS = true ODP calculated using the retrospective likelihood; pTS = true ODP calculated using the prospective likelihood.

Table 4. ETPR of the ODP estimated using the retrospective likelihood for 5% EFPR over 1,000 replicates

	$k\%: 0$	2	5	7	20
MAF $U[0.01, 0.5]$, logOR $N(0, 0.2^2)$	52.97	53.22	53.40	53.50	53.44
MAF $U[0.01, 0.5]$, logOR $E(1/\log(1.2))$	51.43	52.02	51.90	52.08	51.56
MAF 0.15, OR 1.2	87.93	88.80	89.92	90.26	91.12
MAF $U[0.01, 0.5]$, OR 1.2	87.01	87.06	87.04	87.08	86.92
MAF $U[0.01, 0.5]$, logOR \propto MAF	19.41	18.92	19.08	19.18	19.40

Weights introduced in the denominator of the ODP are based on the top $k\%$ -ranking SNPs by Fisher's exact test.

Table 5. ETPR of the ODP estimated using the retrospective likelihood for 5% EFPR over 1,000 replicates

	$k\%: 0$	2	5	7	20
MAF $U[0.01, 0.5]$, logOR $N(0, 0.2^2)$	52.97	52.95	52.97	53.02	52.91
MAF $U[0.01, 0.5]$, logOR $E(1/\log(1.2))$	51.43	51.45	51.44	51.45	51.38
MAF 0.15, OR 1.2	87.93	89.45	90.66	91.00	91.78
MAF $U[0.01, 0.5]$, OR 1.2	87.01	86.96	87.00	87.02	87.00
MAF $U[0.01, 0.5]$, logOR \propto MAF	19.41	19.36	19.40	19.35	19.10

Weights introduced in the denominator of the ODP are based on the top $k\%$ -ranking SNPs by the genotypic association test.

ing SNPs tend to be higher than the ones calculated using all or the bottom-ranking SNPs. Therefore, as well as having similar power, the rankings returned by the various methods are similar for the causal SNPs, whereas they rank the null- or the lower-signal SNPs differently. In particular, we observe that overall, there is high correlation between the rankings given by p values and ABFs. The rankings between the different estimated ODP methods are also quite similar in general, as well as those of the estimated ODP compared to p values and ABFs. However, the correlation of the rankings from the true ODP and the other approaches tends to be low when the correlation coefficient is calculated for all SNPs or for the bottom-ranking SNPs. These correlation coefficients are much improved when they are calculated for the top 50-ranking SNPs.

Sensitivity to Weight Specification

To estimate the weights $\hat{\omega}$ in the estimated ODP, we assume that $k\%$ of the SNPs have a true association with the trait, so the top $k\%$ SNPs ranked by Fisher's exact test are given $\hat{\omega} = 0$ and the rest have $\hat{\omega} = 1$. To investigate the effect of weights, we vary k between 0–20% and compare

the ETPR of the retrospective estimated ODP at 5% EFPR (table 4). We observe that for reasonable values of k , there are no significant differences in the results. We also estimated the weights by ranking the SNPs according to the genotypic test instead of Fisher's exact test (table 5). The results indicate that there are no significant differences in the true-positive rate of the ODP when the weights are estimated using Fisher's exact test versus the genotypic test. This could be explained by the fact that, although the ranking of SNPs from different univariate tests is not identical, the ranking of the top SNPs tends to be similar with most differences in ranking occurring towards the middle or bottom, where there is less evidence of association among the SNPs.

Sequence Data

Table 6 shows the ETPR corresponding to 5% EFPR for the three simulations based on sequence data. ODP is estimated using the retrospective likelihood with weights based on the top 0.5% SNPs ranked by Fisher's exact test. We observe that there is now a systematic advantage of the ABF and ODP over the p value, and the ABF seems to perform slightly better than the ODP. This advantage

Table 6. ETPR (%) (and standard error) for 5% EFPR over 100 replicates of sequence data simulated with ten causal variants and 1,000 cases/controls

	MAF: $U[0.01, 0.05]$ logOR(Aa): $N(0, 0.2^2)$	$U[0.01, 0.05]$ log($U[1.3, 1.7]$)	$U[0.01, 0.05]$ log($U[2.2, 2.5]$)	$U[0.01, 0.5]$ $N(0, 0.2^2)$
PV	27.50 (1.559)	69.70 (1.514)	98.40 (0.369)	53.60 (1.618)
BF	36.50 (1.617)	77.80 (1.292)	99.10 (0.287)	55.80 (1.671)
rS	33.50 (1.438)	77.90 (1.282)	98.80 (0.327)	59.80 (1.676)

rS = ODP estimated using the retrospective likelihood with denominator weights based on the top 0.5%-ranking SNPs by Fisher's exact test.

seems to arise from the lower frequency of the causal SNPs. More precisely, the advantage is due to the combination of MAF and sample size that leads to low power for individual rare SNPs. We obtained identical results by increasing the MAF and decreasing the sample size proportionally (data not shown), and therefore expect similar results for rarer SNPs tested in larger samples.

We also simulated 100 datasets in which the MAF of the causal SNPs has a uniform distribution in the range of 0.01 and 0.5, and their log ORs are from the Normal distribution $N(0, 0.2^2)$, which is the first scenario considered in the section on independent SNPs. The results (table 6) indicate a small but significant advantage of the ODP over ABF and the p value, with the p value having the worst performance. This is in contrast to our results for independent SNPs (tables 1–3) in which the three methods had similar power. The ODP has succeeded by borrowing strength from the correlation structure in the data, and this is possibly more likely for common causal SNPs than rare ones.

In online supplementary table 2 we report Spearman's rank correlation coefficient between the methods averaged over the 100 replicates. The correlation coefficient is calculated for all SNPs, for the top 50-ranking SNPs and for the remaining SNPs. We observe that the rankings of p values and ABFs are comparable, especially for the top-ranking SNPs. However, those given by the estimated ODP are very dissimilar to those from both p values and ABFs when the correlation coefficient is calculated for all SNPs or for the bottom-ranking SNPs. These correlation coefficients are much improved when they are calculated for the top 50-ranking SNPs. These results are similar to those for independent SNPs, and again indicate high concordance between methods for truly associated SNPs.

We also examine the performance of the methods as the EFPR varies between 1 and 20%. The results are

shown in figure 2 and illustrate that the estimated ODP has higher ETPR than both p values and ABFs for values of EFPR $\geq 10\%$. In particular, it can be observed that the ETPR of the ABF does not increase appreciably for values of the EFPR $> 5\%$. This is due to the effect of many rare variants present in the sequence data, for which the sampling variance V in equation 4 is much larger than the prior variance W , leading to an ABF near 1. This leads to a portion of the null distribution of the ABF over which it is nearly constant, and the ETPR is therefore also nearly constant. In general, the portion of the null distribution showing this behaviour will depend on the genomic structure, but ours was chosen to be typical of Europeans. The ODP avoids this problem, because it takes into account the distribution of MAF over the whole set of SNPs, whereas ABF and p values only consider the MAF of the tested SNP. We also observed similar behaviour for the more accurate BF calculated by the SNPTEST software (not shown).

Discussion

The main motivation for using the ODP is that it explicitly exploits multiplicity by pooling information across SNPs when conducting multiple inferences in order to gain extra power compared to p values and ABFs. The ODP provides a theoretically optimal procedure against which any other method can be compared. However, it requires the specification of the true likelihood functions for each SNP, which are largely unknown in practice. Therefore, the ODP must be estimated, and the main question we addressed is whether the optimality of the ODP is retained by the estimated ODP.

Our simulations indicate that although the true ODP is never outperformed, the estimated ODP has similar

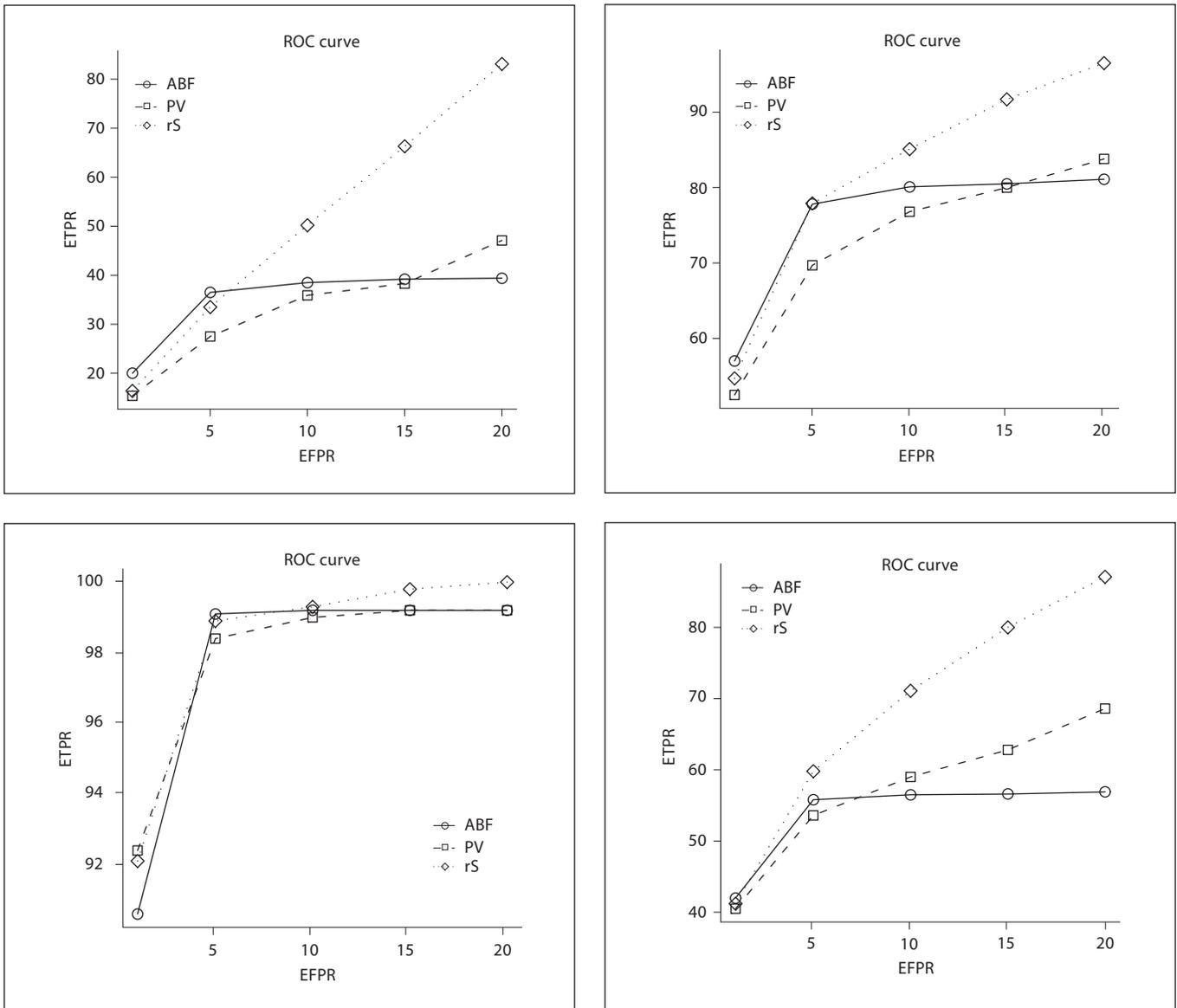


Fig. 2. ETPR (%) for a range of EFPR (%) over 100 replicates of sequence data simulated with ten causal variants and 1,000 cases/controls. Top left: $MAF \sim U[0.01, 0.05]$, $\log OR(Aa) \sim N(0, 0.2^2)$. Top right: $MAF \sim U[0.01, 0.05]$, $OR(Aa) \sim U[1.3, 1.7]$. Bottom left: $MAF \sim U[0.01, 0.05]$, $OR(Aa) \sim U[2.2, 2.5]$. Bottom right: $MAF \sim U[0.01, 0.5]$, $\log OR(Aa) \sim N(0, 0.2^2)$. rS = ODP estimated using the retrospective likelihood with denominator weights based on the top 0.5%-ranking SNPs by Fisher's exact test.

power to p values and ABFs in realistic situations, although it does better when the effects have an unexpected distribution. The ODP shows a small improvement in power over ABFs when SNPs are in LD, and ABFs in turn have slightly higher power than p values. In the analysis of rare SNPs, the ODP exhibits noticeable improvements in power compared to ABF and p values for $EFPR \geq 10\%$,

whereas it performs at least as well as the ABF for $EFPR \leq 5\%$. We observed good operating characteristics of the ABF. It seems to be robust to the choice of prior, as it performs well for all OR distributions considered. Furthermore, although the ABF is an asymptotic result, we found it to be powerful for rare SNPs. In fact, we did not significantly improve power by using more accurate BFs

calculated by the SNPTEST software [4] (results not shown).

In this work, we describe how the ODP can be used for significance testing of SNPs in case-control studies. The procedure we employ is straightforward to implement, as it uses maximum likelihood estimates of parameters of interest, which can be easily obtained by a logistic regression performed separately on each SNP. We make the distinction between ODP based on prospective and retrospective likelihoods. The former is conveniently implemented in standard software, whereas the latter correctly models the sampling scheme of a case-control study. For single SNP analysis, the OR estimators from prospective and retrospective models are the same [19], and we wanted to examine whether this also applies to the ODP. Our results show that although the true retrospective ODP performs better than the prospective version, there is little operational difference between the estimated versions, and the prospective version could be used to little detriment.

The calculation of the ODP is computationally demanding. The ODP of a single SNP involves computation of the null and alternative likelihoods over all the SNPs in the whole dataset, and this can prove impractical for genome-wide studies. Indeed, our simulations considered a much reduced number of SNPs compared to a GWAS, precisely because of the computational issue. Simulation is required to determine an ODP threshold for a given false-positive rate, further increasing the computational load. The significantly higher computational effort required by the ODP, combined with the fact that its power is comparable to that of the ABF for common variants, means that there seems little reason to favour it over the ABF, especially for GWAS. However, when there are a lot of rare SNPs in the region under investigation, the ODP may have more power than the ABF at certain false-positive rates.

Improved performance of the ODP could be obtained by improving its estimation. We have described an implementation of the ODP for case-control association studies, based on maximum likelihood estimation performed separately for each SNP. It is well known that this approach produces more variable estimates for the ensemble of SNPs than simultaneous estimation based on a penalized likelihood, using techniques such as ridge regression or Lasso. Such methods have been used for GWAS [25] and could provide better estimates of the ODP as they analyze all SNPs simultaneously. Cao et al. [26] develop this idea in a Bayesian framework by using a shrinkage prior on variance components, with the posterior means from the Bayesian model used in the ODP. Their

approach is more powerful than the ODP in the gene expression analysis setting, in which the number of replicates is small. Thanks to the conjugacy of the model, the Bayesian ODP is relatively computationally efficient, though still more intensive than the ODP based on univariate maximum likelihood. These approaches are a promising direction for further development of the ODP for large-scale genetic association studies.

Compared to our results, Storey et al. [16] show greater gains in power when using the ODP in microarray studies of gene expression. A possible explanation for this difference in ODP performance could be the different correlation structure of gene expression data compared to SNP data. Another likely explanation is that the sample sizes used by Storey et al. [16] were considerably smaller than those we considered for genetic association studies. In that case, there is more to be gained by borrowing information across tests, as we also found when considering rare variants: our results in that case are more comparable to those of Storey et al. [16], but we also found that the ABF had similarly increased power.

Storey [15] bridges the gap between the frequentist ODP and Bayesian hypothesis testing by comparing the ODP to a Bayes rule, and describes how the ODP can be interpreted as an empirical BF. It is therefore not surprising that its power is similar or slightly worse than that of an analytic BF (asymptotic or not) for common SNPs and when the prior for the BF is well specified. Similarly, as our results bear out, the ODP has greater power than an ABF when the true distribution of effects is very unexpected, or for very rare SNPs.

Guindani et al. [27] propose a Bayesian discovery procedure as a generalization of the Bayes rule and the ODP. Compared to the ODP, their approach amounts to a more sophisticated way of choosing weights. Marginal improvement in inference was shown for microarray expression data, and an application of this approach to case-control data should also yield modest benefits. However, the Bayesian discovery procedure is yet more computationally intensive than the ODP.

The selection of SNPs to follow up can be based on ranking as well as on statistical significance. The optimality of the ODP is established as a thresholding procedure, but not necessarily for ranking. We found the rankings of the methods considered to be very highly correlated, with similar numbers of true positives within the top-ranking SNPs. While there are other methods with potentially better performance for ranking [28], the fact that ODP, ABF and p values may be used for both ranking and thresholding is an attractive property of these methods.

We have shown that for association studies of common SNPs, there is little difference in power whether the SNPs are tested by p value, ABF or ODP. Furthermore, in our simulations, there were no differences in the variance of the number of true positives either, so that overall, there is little to choose between the methods in terms of power. As the ABF can be computed more rapidly, it is our currently preferred method for common variants.

Nevertheless, the higher power of the ODP when all parameter values are known suggests that improved multivariate estimators of the ODP could yet yield more powerful approaches for significance testing of SNPs in association studies. Moreover, for rare SNPs, there are gains in power to be obtained from ODP compared to the ABF and p value. This should prove important with the currently increasing interest in testing very rare SNPs observed in whole-sequence data.

Appendix

Prospective ODP with Variable Sampling Ratio

The intercepts β_{0m} in equation 8 are a function of the sampling ratios, population prevalence, genotype frequencies and ORs, but the sampling ratio for SNP m is not expected to be informative for the OR of SNP j . Therefore, using the naive values of $\hat{\beta}_{0m}$ and $\hat{\beta}'_{0m}$ in the ODP for SNP j is likely to mis-specify the distributions used in equation 8 and therefore reduce its power.

References

- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. *Nature* 2009;461:747–753.
- Wang WYS, Barratt BJ, Clayton DG, Todd JA: Genome-wide association studies: theoretical and practical concerns. *Nat Genet* 2005;6:109–118.
- Balding D: A tutorial on statistical methods for population association studies. *Nat Genet* 2006;7:781–791.
- The Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661–678.
- Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, Hirschhorn JN, Abecasis G, Altshuler D, Bailey-Wilson JE, Brooks LD, Cardon LR, Daly M, Donnelly P, Fraumeni JF, Freimer NB, Gerhard DS, Gunter C, Guttmacher AE, Guyer MS, Harris EL, Hoh J, Hoover R, Kong CA, Merikangas KR, Morton CC, Palmer LF, Phimister EG, Rice JP, Robert J, Rotimi C, Tucker MA, Vogan KJ, Wacholder S, Wijsman EM, Winn DM, Collins FS, NCI-NHGRI Working Group on Replication in Association Studies: Replicating genotype-phenotype associations. *Nature* 2007;447:655–660.
- Dudbridge F, Gusnanto A: Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 2008;33:406–418.
- Storey J, Tibshirani R: Statistical significance for genome-wide studies. *Proc Natl Acad Sci* 2003;100:9440–9445.
- Sabatti C: Avoiding false discoveries in association studies. *Methods Mol Biol* 2007;376:195–211.
- Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N: Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434–442.
- Wakefield J: A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* 2007;81:208–227.
- Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, Knig IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H, WTCCC, Cardiogenics Consortium: Genomewide association analysis of coronary artery disease. *N Engl J Med* 2007;357:443–453.

Instead, we suggest adjusting $\hat{\beta}_{0m}$ using the OR and genotype frequency of SNP m but the sampling ratio of SNP j . The adjusted intercept, denoted $\hat{\beta}_{0m}^*$, is obtained as the solution to

$$r_j = \frac{\exp(\hat{\beta}_{0m}^*)}{1 + \exp(\hat{\beta}_{0m}^*)} \hat{p}_{0m} + \frac{\exp(\hat{\beta}_{0m}^* + \hat{\beta}_{1m})}{1 + \exp(\hat{\beta}_{0m}^* + \hat{\beta}_{1m})} \hat{p}_{1m} + \frac{\exp(\hat{\beta}_{0m}^* + 2\hat{\beta}_{1m})}{1 + \exp(\hat{\beta}_{0m}^* + 2\hat{\beta}_{1m})} \hat{p}_{2m} \quad (16)$$

where r_j is the sampling ratio for SNP j and $(\hat{p}_{0m}, \hat{p}_{1m}, \hat{p}_{2m})$ are the estimated marginal genotype frequencies in the ascertained sample.

Using $\hat{\beta}_{0m}^*$ in equation 8 yields the adjusted prospective ODP, given by

$$\hat{\sigma}_{ODP}^{\text{adj}}(\mathbf{x}_j) = \frac{\sum_{m=1}^M L(\hat{\beta}_{0m}^*, \hat{\beta}_{1m}; \mathbf{x}_j)}{\sum_{m=1}^M \hat{\omega}_m L(\hat{\beta}_{0m}^*, 0; \mathbf{x}_j)} \quad (17)$$

Acknowledgements

This work was supported by the Medical Research Council (WBSU.1052.00.012.00001.01). We thank Clive Hoggart for providing the FREGENE population of sequence data.

- 12 Manly KF, Nettleton D, Hwang JT: Genomics, prior probability, and statistical tests of multiple hypotheses. *Genome Res* 2004;14:997–1001.
- 13 Wakefield J: Bayes factors for genome-wide association studies: comparison with p values. *Genet Epidemiol* 2008;33:79–86.
- 14 Stephens M, Balding DJ: Bayesian statistical methods for genetic association studies. *Nat Rev Genet* 2009;10:681–690.
- 15 Storey J: The optimal discovery procedure: a new approach to simultaneous significance testing. *J R Statistic Soc B* 2007;69:347–368.
- 16 Storey J, Dai J, Leek J: The optimal discovery procedure for large-scale significance testing, with applications to comparative microarray experiments. *Biostatistics* 2007;8:414–432.
- 17 Shaffer C: Next-generation sequencing outpaces expectations. *Nat Biotechnol* 2007;25:149.
- 18 Schuster S: Next-generation sequencing transforms today's biology. *Nat Methods* 2008;5:16–18.
- 19 Prentice RL, Pyke R: Logistic disease incidence models and case-control studies. *Biometrika* 1979;66:403–411.
- 20 Satten GA, Carroll RJ: Conditional and unconditional categorical regression models with missing covariates. *Biometrics* 2000;56:384–388.
- 21 Epstein MP, Satten GA: Inference on haplotype effects in case-control studies using unphased genotype data. *Am J Hum Genet* 2003;73:1316–1329.
- 22 Marchini J, Howie B, Myers S, McVean G, Donnelly P: A new multipoint method for genome-wide association studies via imputation of genotypes. *Nat Genet* 2007;39:906–913.
- 23 Hoggart CJ, Chadeau-Hyam M, Clark TG, Lampariello R, Whittaker J, De Iorio M, Balding D: Sequence-level population simulations over large genomic regions. *Genetics* 2007;177:1725–1731.
- 24 Schaffner SF, Foo C, Gabriel S, Reich D, Daly MJ, Altshuler D: Calibrating a coalescent simulation of human genome sequence variation. *Genome Research* 2005;15:1576–1583.
- 25 Hoggart CJ, Clark T, De Iorio M, Whittaker J, Balding D: Genome-wide significance for dense SNP and resequencing data. *Genet Epidemiol* 2008;32:179–185.
- 26 Cao J, Xie XJ, Zhang S, Whitehurst A, White MA: Bayesian optimal discovery procedure for simultaneous significance testing. *BMC Bioinformatics* 2009;10:1471–2105.
- 27 Guindani M, Müller P, Zhang S: A Bayesian discovery procedure. *J R Stat Soc Series B Stat Methodol* 2009;71:905–925.
- 28 Louis TA, Ruczinski I: Efficient evaluation of ranking procedures when the number of units is large, with application to SNP identification. *Biometrical J* 2010;52:34–49.