Published in final edited form as:

Ann Hum Genet. 2008 November ; 72(Pt 6): 812–820. doi:10.1111/j.1469-1809.2008.00467.x.

Probability that a two-stage genome-wide association study will detect a disease-associated SNP and implications for multistage designs

M. H. GAIL^{1,*}, R. M. PFEIFFER¹, W. WHEELER², and D. PEE²

¹ Division of Cancer Epidemiology and Genetics, National Cancer Institute, EPS 8032, Bethesda, MD, 20892-7244, US

² Information Management Services Incorporated, 6110 Executive Boulevard, Suite 310, Rockville, MD, 20852, US

SUMMARY

Large two-stage genome-wide association studies (GWASs) have been shown to reduce required genotyping with little loss of power, compared to a one-stage design, provided a substantial fraction of cases and controls, π_{sample} , is included in stage 1. However, a number of recent GWASs have used $\pi_{\text{sample}} < 0.2$. Moreover, standard power calculations are not applicable because SNPs are selected in stage 1 by ranking their p-values, rather than comparing each SNP's statistic to a fixed critical value. We define the detection probability (DP) of a two-stage design as the probability that a given disease-associated SNP will have among the lowest ranks of p-values at stage 1, and, among those SNPs selected at stage 1, at stage 2. For 8000 cases and 8000 controls available for study and for odds ratios per allele in the range 1.1–1.3, we show that DP is substantially reduced for designs with $\pi_{\text{sample}} \leq 0.25$, and that DP cannot be appreciably increased by analyzing the stage 1 and stage 2 data jointly. These results suggest that multistage designs with small first stages (e.g. $\pi_{\text{sample}} \leq 0.25$) should be avoided, and that additional genotyping in earlier studies with small first stages will yield previously unselected disease-associated SNPs.

Keywords

case-control association study; detection probability; power; proportion positive; ranking and selection; SNP chip

INTRODUCTION

Two-stage genome-wide association studies (GWASs) have been advocated because they can reduce required genotyping with little loss in power (Satagopan & Elston, 2003, Satagopan et al., 2004, Skol et al., 2006). In stage 1 of these designs, all available SNPs are analyzed among a proportion, π_{sample} , of the available cases and controls; only the SNPs judged to be promising in stage 1 are studied in the remaining cases and controls (stage 2). If the cost per genotype is lower in stage 1, larger π_{sample} is required to minimize total cost (Skol, 2007, Wang et al., 2006). Likewise, accounting for the cost of obtaining the cases and controls leads to larger values of π_{sample} (Müller et al., 2007). Optimal designs typically require π_{sample} to be 0.5 or more. Figure 2 in (Skol et al., 2006) shows that power diminishes appreciably if $\pi_{\text{sample}} \leq 0.20$. Nonetheless, some investigators use $\pi_{\text{sample}} < 0.2$ and try to compensate by selecting a large number of SNPs in stage 1 for further study in later stages.

*Correspondence to Mitchell H. Gail, Fax: 1-301-402-0081, E-mail: gailm@mail.nih.gov.

The most studied type of two-stage design is based on an hypothesis testing paradigm: a SNP is said to be disease-associated if its test statistic exceeds a fixed critical value c_1 at stage 1 and if its test statistic at stage 2 (usually based on the combined stage 1 and stage 2 data) exceeds a fixed critical value c_2 (Müller et al., 2007, Satagopan & Elston, 2003, Skol, 2007, Skol et al., 2006, Wang et al., 2006). This hypothesis testing approach not only identifies promising SNPs in the first stage, but also provides an overall p-value for assessing association with disease.

However, many studies are not conducted this way. Instead, a SNP is selected for further study in stage 1 if its p-value ranks among the T_1 smallest p-values. Thus, selection of promising SNPs is not based on a fixed critical value but on ranking the SNPs against each other. This strategy was used in the Cancer Genetic Markers of Susceptibility Study (CGEMS) for prostate cancer (http://cgems.cancer.gov/about/replication_strategy.asp) and in studies of various other diseases (Broderick et al., 2007, Buch et al., 2007, van Es et al., 2007). In the CGEMS prostate cancer study, about $n_1 = 1200$ cases and controls were studied in stage 1 (i.e. $\pi_{\text{sample}} = 0.14$), and roughly $T_1 = 28,000$ SNPs with the smallest p-values were selected for study in 7200 additional cases and controls in subsequent stages.

One approach following stage 1 is to continue ranking and selecting the SNPs based on data in stage 2. In particular, one selects a SNP as highly promising at the end of stage 2 if its p-value, which may be based on stage 2 data alone or on some combination of stage 1 and stage 2 data, ranks among the lowest T_2 p-values from the T_1 SNPs studied in stage 2. In this paper we describe the properties of this two-stage selection procedure, and in particular, we calculate the probability that a disease-associated SNP will be selected at the end of stage 2 (the “detection probability”) and the proportion of selected SNPs that are expected to be disease-associated (the “proportion positive”). If a two-stage procedure is designed to have a high detection probability and proportion positive, then one can expect that most disease-associated SNPs will be included in the T_2 selected SNPs and that independent epidemiologic studies will confirm the association with disease in a good proportion of the T_2 selected SNPs.

A second approach is to test for associations with disease among the SNPs selected in stage 1. In a study of gallstone disease (Buch et al., 2007), the authors used the independent stage 2 data alone to produce a p-value. In a study of amyotrophic lateral sclerosis (van Es et al., 2007), the authors ranked SNPs in stage 1 and culled SNPs further by requiring a nominal p-value < 0.1 in stage 2 before assigning a p-value based on independent stage 3 data. To study genetic associations with colorectal cancer, investigators combined data from stages 1 and 2 to produce an overall p-value (Broderick et al., 2007). The statistical properties of these hybrid approaches have not been evaluated, although p-values that depend only on an independent final stage should have nominal significance levels.

In previous work (Gail et al., 2008), we studied ranking procedures for one-stage designs and defined the “detection probability” (DP) as the probability that a disease associated SNP would rank among the T_1 largest test statistics (or T_1 smallest p-values). For a two-stage design, we now define DP as the probability that a disease-associated SNP will rank among the T_1 largest statistics (or T_1 smallest p-values) at stage 1 and among the T_2 largest statistics (or T_2 smallest p-values) at stage 2. As in (Gail et al., 2008), we can also compute the expected proportion of the T_2 selected SNPs that are truly disease-associated, namely the proportion positive (PP), under the assumption that the true number of disease-associated SNPs is known. We provide methods for assessing DP in the two-stage design, not only for a “replication analysis” that is based on the rankings in stage 2 data alone, but also for a final “joint” analysis (Skol et al., 2006) that combines data from stages 1 and 2. We show that for magnitudes of odds ratios commonly found in GWASs, the DP can be much lower for a

two-stage design with $\pi_{sample} \leq 0.25$ than for a one-stage design with the same number of cases and controls, and that DP is hardly increased by an optimal joint analysis of stage 1 and stage 2 data combined when $\pi_{sample} \leq 0.25$. This result is analogous to findings for power in (Skol et al., 2006). Our calculations indicate how much the probability of selecting a disease-associated SNP can be increased by additional genotyping in such studies, and add to arguments in favor of one-stage designs. Finally we discuss implications for the design of GWASs.

METHODS

A two-stage GWAS analyzes T_0 (say 500,000) tagging SNPs in n_1 cases and n_1 controls (stage 1) and selects T_1 promising SNPs for study in stage 2 in independent cases (n_2) and controls (n_2). Following stage 2, T_2 SNPs are selected in the hope that they are associated with disease. The proportion of cases and controls in stage 1 is $\pi_{sample} \equiv n_1/(n_1+n_2)$.

Assuming that the T_0 tagging SNPs are in linkage equilibrium as in (Skol et al., 2006), we proved that SNP genotypes are independent not only in controls, but also in cases for a rare disease (Gail et al., 2008). We used these ideas and asymptotic theory for the Wald and score tests to develop efficient procedures for simulating the chi-square test statistics for the M_0 disease-associated SNPs (“disease SNPs”) and the T_0-M_0 non-disease SNPs in stage 1. For independent stage 2 data, we now extend these methods to simulate independent chi-square tests for T_1 SNPs, of which a random number, M_1 , are disease SNPs that were selected at stage 1, and the remaining T_1-M_1 are non-disease SNPs. We study odds ratios per allele of 1.1, 1.2, 1.3, and 1.5, but we use 1.2 in the following description. We consider two models (Gail et al., 2008) for disease SNPs. In the fixed effects model, the log odds ratio per disease allele is fixed at $\beta = \log(1.2)$ for each disease SNP. Thus the relative odds is 1.44 for a homozygote and 1.2 for a heterozygote. In the random effects model, β is drawn independently for each disease SNP from a normal distribution with mean zero and standard deviation $\tau = (\pi/2)^{1/2} \log(1.2) \approx 1.253 \log(1.2)$. This value of τ yields an expected absolute value of β of $\log(1.2)$. Under both models, $\beta = 0$ for non-disease SNPs. One chi-square test is the squared Wald statistic, $\hat{\beta}^2/\hat{Var}(\hat{\beta})$, for testing $\beta = 0$ in a model for log odds of disease that equals an intercept plus β times the number of minor alleles (Gail et al., 2008). We also studied the corresponding squared score test (Armitage, 1955; Sasieni, 1997). Each of these chi-square tests has the same value whether the major or minor allele confers risk.

At stage 1, the T_1 SNPs with the largest chi-square tests (or smallest p-values) are selected. We study two analytical approaches, “replication analysis” and “joint analysis,” similar to (Skol et al., 2006). For a given SNP whose test statistic was in the critical region in stage 1, Skol et al. called an hypothesis test based only on the stage 2 data a “replication analysis.” They used the term “joint analysis” for a final hypothesis test based a linear combination of the test statistics for stage 1 and stage 2 data. Analogously, we use the term “replication analysis” if final selection of the T_2 SNPs from among the T_1 SNPs selected in stage 1 depends only on their rankings derived from the independent stage 2 data. In “joint analysis”, for each of the T_1 SNPs selected in stage 1, we compute $\lambda C_1 + (1-\lambda)C_2$ for $\lambda = 0, 0.05, 0.10, \dots, 1.0$, where C_1 and C_2 are the chi-square statistics observed in stages 1 and 2 respectively. For each λ we estimate DP, and we present the maximal $\hat{DP}$ over the 21 values of λ , together with the corresponding λ_{opt} . If several values maximize DP, we define λ_{opt} as their average.

For the simulations, we assume that 8000 cases and 8000 controls are available to be apportioned between stages 1 and 2 with equal numbers of cases and controls in each stage. We present data for $M_0 = 1$ and 10 and for $\pi_{sample} = 0.125, 0.25, 0.50$, and 1.0, both for fixed effects and random effects models. For each odds ratio, we conducted 12 independent

simulation studies, each with 10,000 replications, to estimate DP for each combination of M_0 , $\pi_{sample} = 0.125, 0.25$ and 0.50 and fixed and random effects models. We allow for realistic random variation in minor allele frequency (MAF). In each simulation and for each of the $T_0 = 500,000$ SNPs, we randomly select a MAF from the distribution in controls in CGEMS (<https://cointegrator.nci.nih.gov/cgems/>), truncated to the interval $[0.05, 0.5]$; the mean MAF is 0.2673 , which is corrected from 0.2763 in (Gail et al., 2008). For each SNP, a chi-square value is generated that depends on the MAF, β , and numbers of cases and controls (Gail et al., 2008). Letting $I(m, ISIM) = 1$ if disease SNP m is among the T_1 SNPs selected after stage 1 and among the T_2 SNPs selected after stage 2 and 0 otherwise, we estimate (Gail et al., 2008) DP as

$$\widehat{DP} = \sum_{m=1}^{M_0} \sum_{ISIM=1}^{10,000} I(m, ISIM) / (10,000 M_0). \quad (1)$$

We note from the exchangeability of the disease SNPs that DP can be interpreted either as the probability that a particular disease SNP will be selected at stage 2 or as the proportion of disease SNPs selected at stage 2 (Gail et al., 2008). The proportion positive, PP, can be estimated (Gail et al., 2008) as

$$\widehat{PP} = (M_0)(\widehat{DP}) / T. \quad (2)$$

We performed these simulations in GAUSS (Aptec Systems, 2005).

RESULTS

The following results are for the chi-square test based on the Wald statistic. Very similar results were found for the chi-square version of the score test (data not shown). We present detailed information for the fixed effects model with odds ratio 1.2 per allele in Table 1. With $M_0 = 1$, $\widehat{DP}$ was 0.882 or higher for the one-stage design ($\pi_{sample} = 1.0$), for various values of T_1 (shown under T_2 in Tables). For $\pi_{sample} = 0.5$, $\widehat{DP}$ was slightly reduced in the replication analysis, but the joint analysis yielded $\widehat{DP}$ closer to that of the one-stage design. The λ_{opt} ranged from 0.30 to 0.45, indicating that the optimal combination usually put more weight on C_2 , even when $n_1 = n_2$. For $\pi_{sample} = 0.25$ or 0.125 , $\widehat{DP}$ was considerably reduced for the replication analysis, and the joint analysis yielded little if any increase in $\widehat{DP}$. For example, for $\pi_{sample} = 0.125$, $T_1 = 25,000$ and $T_2 = 10$, $\widehat{DP}$ was 0.936 for the one-stage analysis, 0.657 for the replication analysis, and 0.658 for the joint analysis. λ_{opt} was smaller for $\pi_{sample} = 0.125$ or 0.25 than for 0.50 . In five of the six scenarios presented in Table 1 for $M_0 = 1$, λ_{opt} was smaller for $T_2 = 100$ than for $T_2 = 1$ or 10 . For $M_0 = 10$, similar results were obtained except for $T_2 = 1$. For $T_2 = 1$, all designs and analyses yielded a $\widehat{DP}$ near 0.100, because the ten disease SNPs competed against each other for selection. For $T_2 = 10$ or 100 , results for $M_0 = 10$ were very close to those for $M_0 = 1$.

The proportion positive, PP, can also be estimated for $\beta = \log(1.2)$ from Table 1 and equation (2) assuming M_0 is known. For $M_0 = 1$, $T_1 = 25,000$ and $T_2 = 10$, $\widehat{PP} = 0.065$ for $\pi_{sample} = 0.125$ and 0.093 for the one-stage design. If $M_0 = 10$, the corresponding $\widehat{PP}$ values are 0.654 and 0.898 . For $M_0 = 1$, $T_1 = 25,000$ and $T_2 = 1$, $\widehat{PP} = 0.630$ for $\pi_{sample} = 0.125$ and 0.882 for the one-stage design.

For the random effects model with $\tau = (\pi/2)^{1/2} \log(1.2) \approx 0.2284$ and with $M_0=1$ (Table 2), $\hat{DP}$ ranged from 0.550 to 0.646 for the one-stage design. With $\pi_{sample} = 0.5$, $\hat{DP}$ was modestly reduced for the replication analysis, but only slightly reduced for the joint analysis. λ_{opt} ranged from 0.25 to 0.45. For $\pi_{sample} = 0.25$ or 0.125, $\hat{DP}$ was considerably reduced for the replication analysis, and the joint analysis yielded little if any increase in $\hat{DP}$. For example, for $\pi_{sample} = 0.125$, $T_1 = 25,000$ and $T_2 = 10$, $\hat{DP}$ was 0.599 for the one-stage analysis, 0.465 for the replication analysis, and 0.465 for the joint analysis. λ_{opt} was smaller for $\pi_{sample} = 0.125$ or 0.25 than for 0.50, and λ_{opt} was usually smaller for $T_2 = 100$ than for $T_2 = 1$ or 10. For $M_0 = 10$, results were very close to those for $M_0 = 1$ except when $T_2 = 1$. With $T_2 = 1$, the $M_0 = 10$ disease SNPs competed against each other, reducing $\hat{DP}$ to near 0.100 for all designs and analyses. As for the fixed effects model, application of equation (2) shows that $\hat{PP}$ increases with increasing M_0 .

To examine how much the DP of the two-stage design is reduced compared to the one-stage design for fixed effects models with odds ratios per allele of 1.1, 1.2, 1.3 and 1.5, we plotted $\hat{DP}$ for the two-stage design and joint analysis against the $\hat{DP}$ for the corresponding one-stage design with 8000 cases and 8000 controls for $\pi_{sample} = 0.125$ and $M_0 = 1$ (Figure 1). The bold loci correspond to $T_1 = 25,000$ for $T_2 = 1, 10$ or 100, and the unbolded loci correspond to $T_1 = 1,000$ for $T_2 = 1, 10$ or 100. The points on each locus correspond from left to right to odds ratios 1.1, 1.2, 1.3 and 1.5. The vertical distance from the equiangular line is the decrease in estimated DP from using the two-stage design. For example, for odds ratio 1.2, and for $T_1 = 25,000$ and $T_2 = 100$, the vertical distance is $0.966 - 0.664 = 0.302$; for $T_1 = 1,000$ and $T_2 = 100$, the vertical distance is $0.966 - 0.269 = 0.697$, as can also be seen from Table I. The absolute decrease in $\hat{DP}$ for two-stage designs is substantial for odds ratios 1.1, 1.2 and 1.3, but diminishes for odds ratio 1.5, as both designs attain $\hat{DP}$ approaching 1.0. Except for $T_2 = 1$, very similar graphs are obtained for $M_0 = 10$ (not shown).

For the random effects model with $M_0 = 1$ and $\pi_{sample} = 0.125$, decreases in $\hat{DP}$ for the two-stage design are appreciable for standard deviations of log odds ratios of $(\pi/2)^{1/2}$ times $\log(1.1)$, $\log(1.2)$, $\log(1.3)$, and $\log(1.5)$ (Figure 2). Note that compared to the corresponding fixed effects models, $\hat{DP}$ is smaller for both the one-stage and two-stage designs. Except for $T_2 = 1$, very similar graphs are obtained for $M_0 = 10$ (not shown).

For the fixed effects model with $M_0 = 1$ and $\pi_{sample} = 0.25$, the decreases in $\hat{DP}$ for the two-stage design are smaller than for $\pi_{sample} = 0.125$ (compare Figures 3 and 1). Nonetheless, the decreases in $\hat{DP}$ remain appreciable for odds ratios 1.1 and 1.2 (Figure 3). Except for $T_2 = 1$, very similar graphs are obtained for $M_0 = 10$ (not shown).

For the random effects model with $M_0 = 1$ and $\pi_{sample} = 0.25$, the decreases in $\hat{DP}$ for the two-stage model are smaller than for $\pi_{sample} = 0.125$ (compare Figures 4 and 2). The decreases in $\hat{DP}$ are small for $T_1 = 25,000$, but, for $T_1 = 1,000$, remain appreciable for standard deviations of log odds ratios of $(\pi/2)^{1/2}$ times $\log(1.1)$, $\log(1.2)$, $\log(1.3)$, and $\log(1.5)$ (Figure 4). Except for $T_2 = 1$, very similar graphs are obtained for $M_0 = 10$ (not shown).

An executable pre-compiled GAUSS program is available from the first author to estimate DP and PP for two-stage GWASs.

DISCUSSION

We studied the detection probability (DP) of a two-stage GWAS design, that is, the chance that a given disease-associated SNP will have among the lowest ranks of p-values (or highest ranks of chi-square statistics) at stages 1 and 2. Our data for fixed effects models indicate that the DP from a two-stage design with $\pi_{sample} \leq 0.25$ and 8000 cases and controls

can be substantially less than that of the corresponding one-stage design with the same numbers of cases and controls for odds ratios per allele of 1.1, 1.2, and 1.3, which are typical of statistically significant odds ratios found in recent large GWASs. For the range of values $T_1 \leq 25,000$ that we studied, a “joint” analysis cannot appreciably increase the DP of the two-stage design if $\pi_{\text{sample}} \leq 0.25$. Similar results are found for corresponding random effects disease models, which yield somewhat smaller DPs. To achieve an adequate DP, the first stage must have enough cases and controls to assure that a high proportion of disease-associated SNPs have among the T_1 lowest p-values at stage 1. Thus, if 16,000 cases and controls were available for study, choosing $\pi_{\text{sample}} \leq 0.25$ would yield acceptable DP, as seen from calculations for the one-stage design (Gail et al., 2008). Except for settings where enormous numbers of cases and controls are available for study, however, designs with $\pi_{\text{sample}} \leq 0.25$ should be avoided. Software is available from the first author to allow researchers to study other parameter values and sample sizes.

Our data suggest that additional stage 1 genotyping in most previous studies with $\pi_{\text{sample}} \leq 0.25$ will yield additional promising SNPs and that future multistage designs should not use $\pi_{\text{sample}} \leq 0.25$, unless the numbers of available cases and controls are very large. Other considerations also favor using larger values of π_{sample} . As the cost per genotype of chips designed for stage 1 decreases relative to that for specialized platforms for subsequent stages, cost considerations argue for larger values of π_{sample} (Skol, 2007, Wang et al., 2006). The costs of obtaining cases and controls also favor a larger value of π_{sample} (Müller et al., 2007).

The two-stage ranking and selection procedure analyzed in this paper differs from two-stage procedures that apply the same fixed critical values to data from each SNP and are designed to select promising SNPs in stage 1 and provide a final p-value for testing an association following stage 2, as in (Skol et al., 2006). In particular, the two-stage ranking and selection procedure does not attempt to control the overall p-value, but only to obtain a very promising set of T_2 SNPs at the end of stage 2. Despite these different goals and methods, Figure 2 in (Skol et al., 2006) shows that power diminishes appreciably, and cannot be retrieved by joint analysis, if $\pi_{\text{sample}} \leq 0.20$ and $\pi_{\text{marker}} \equiv T_1/T_0 \leq 0.1$, in agreement with our findings for DP.

In some circumstances, power calculations can be used to approximate DP. For a one-stage design with $M_0 = 1$, equation (2.6) in (Gail et al., 2008) shows that DP can be approximated by the power that corresponds to an hypothesis test with size $\alpha = T_1/T_0$. Although power calculations performed in this way and extended to the two-stage design may approximate the DP under certain conditions, the results in (Skol et al., 2006) were not based on such significance levels and critical values. We illustrate these differences using the program provided by (Skol et al., 2006) at <http://csg.sph.umich.edu>. For $\pi_{\text{sample}} = 0.125$, $\pi_{\text{marker}} \equiv T_1/T_0 = 25,000/500,000 = 0.05$, genome-wide false-positive rate 0.05, which corresponds to $\alpha = 0.05/500,000 = 10^{-7}$ for the joint analysis, a single fixed minor allele frequency of 0.2673 for all SNPs, and an assumed disease prevalence of 0.10, this program yields power estimates of 0.83 for the replication analysis and 0.83 for the joint analysis. The corresponding critical value for a normal deviate for stage 1 is 1.96, and for stage 2, the critical values are 4.611 for the replication analysis and 5.189 for the joint analysis. For the ranking procedure with $M_0 = 1$, $T_1 = 25,000$ and $T_2 = 1, 10$, or 100, the DP was 0.63, 0.65, and 0.66 respectively for the replication analysis and 0.64, 0.66 and 0.66 respectively for the joint analysis (Table 1) in the realistic setting in which minor allele frequencies are drawn from the distribution in CGEMS, with mean 0.2673. If instead it was assumed that all SNPs had the same minor allele frequency, 0.2673, as was assumed for the power calculations, the corresponding estimates of DP were 0.74, 0.75 and 0.75 for the joint analysis. These calculations illustrate that there are quantitative differences between assessments based on

power and those based on detection probability, even though the broad conclusion that π_{sample} should not be too small is supported by both analyses.

It is worthwhile to recount some differences between power and detection probability. Power is the probability that the test statistic for a given SNP will fall into the pre-determined critical region for a one- or two-stage design that is chosen to control a genome-wide significance level, as for example in (Skol et al., 2006). Power thus depends on the chosen significance level; DP depends, instead, on T_0 , T_1 , and T_2 . The power to reject the null hypothesis for a given SNP does not depend on the test statistics for any other SNP; DP depends on the test results for all SNPs. Power does not depend on the number of disease-associated SNPs, M_0 ; DP can be sharply reduced by competition among disease-associated SNPs, especially if T_2 is less than M_0 . Most power calculations assume that all disease-associated SNPs have the same minor allele frequency; DP calculations routinely allow for allele-frequencies to be drawn from a realistic distribution of allele minor frequencies. DP estimates the probability that a given disease-associated SNP will have among the smallest T_2 p-values at the end of a two-stage study; power, as routinely calculated, does not have this interpretation. If disease-associated SNPs are exchangeable, as we assume in the fixed effect and random effect models (see METHODS), DP also has an interpretation as the proportion of disease-associated SNPs that have among the smallest T_2 p-values at the end of a two-stage study. Thus, the estimate of DP can be used to estimate how many more disease-associated SNPs with similar odds ratios to those already found are likely to be discovered by conducting another study with a larger stage1 sample.

Satagopan and colleagues (Satagopan et al., 2004, Satagopan et al., 2002) also studied ranking procedures to identify disease-susceptibility SNPs in two-stage designs, but used different rank-based criteria from DP and also assumed that the disease allele was known. The two-sided versions of a Wald test or a score test that we used have the same value whether one counts major or minor alleles, and hence are particularly appropriate for GWASs (Devlin & Roeder, 1999, Pfeiffer & Gail, 2003).

The ranking and selection methods used in this paper depend on the assumption that tagging SNPs are independent (Gail et al., 2008), an assumption that is also widely used in power calculations, e.g. (Skol et al., 2006). Zaykin and Zhivotovsky (Zaykin & Zhivotovsky, 2005) analyzed different ranking criteria for one-stage designs and found that selection probabilities were little affected by correlations of p-values within linkage disequilibrium blocks or among such blocks. In unreported simulations in which non-disease associated SNPs were paired and each member of the pair assigned the same chi-square value (perfect correlation within pairs), we found almost no effect on the estimates of DP and PP, compared to the situation in which SNP genotypes are independent. Thus it is likely that our estimates of DP are robust to local correlations among tagging SNPs.

In view of the potential losses in DP in multistage designs and trends in costs favoring large values of π_{sample} , the one-stage design becomes increasingly attractive. Another advantage of the one-stage design is that it yields data that can readily be used in meta-analyses. For example, if a preliminary study identifies a particular SNP as associated with disease, data from independent one-stage studies can be used to test the association and provide an unbiased estimate of the corresponding odds ratio. Later stages in a multistage design would provide no information if that SNP had not been tested in the later stages.

Acknowledgments

The Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute supported this work. We thank the reviewers for comments that improved the paper.

References

- Aptec Systems. The Gauss System, Version 6. Maple Valley, WA: 2005.
- Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics*. 1955; 11:375–386.
- Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I, Houlston RS. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. *Nat Genet*. 2007; 39:1315–1317. [PubMed: 17934461]
- Buch S, Schafmayer C, Volzke H, Becker C, Franke A, Von Eller-Eberstein H, Kluck C, Bassmann I, Brosch M, Lammert F, Miquel JF, Nervi F, Wittig M, Roskopf D, Timm B, Holl C, Seeger M, Elsharawy A, Lu T, Egberts J, Fandrich F, Folsch UR, Krawczak M, Schreiber S, Nurnberg P, Tepel J, Hampe J. A genome-wide association scan identifies the hepatic cholesterol transporter ABCG8 as a susceptibility factor for human gallstone disease. *Nat Genet*. 2007; 39:995–999. [PubMed: 17632509]
- Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; 55:997–1004. [PubMed: 11315092]
- Gail MH, Pfeiffer RM, Wheeler W, Pee D. Probability of detecting disease-associated single nucleotide polymorphisms in case-control genome-wide association studies. *Biostatistics*. 2008; 9:201–215. [PubMed: 17873152]
- Müller HH, Pahl R, Schäfer H. Including sampling and phenotyping costs into the optimization of two stage designs for genome wide association studies. *Genet Epidemiol*. 2007; 31:844–852. [PubMed: 17549751]
- Pfeiffer RM, Gail MH. Sample size calculations for population- and family-based case-control association studies on marker genotypes. *Genet Epidemiol*. 2003; 25:136–48. [PubMed: 12916022]
- Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics*. 1997; 53:1253–61. [PubMed: 9423247]
- Satagopan JM, Elston RC. Optimal two-stage genotyping in population-based association studies. *Genet Epidemiol*. 2003; 25:149–157. [PubMed: 12916023]
- Satagopan JM, Venkatraman ES, Begg CB. Two-stage designs for gene-disease association studies with sample size constraints. *Biometrics*. 2004; 60:589–97. [PubMed: 15339280]
- Satagopan JM, Verbel DA, Venkatraman ES, Offit KE, Begg CB. Two-stage designs for gene-disease association studies. *Biometrics*. 2002; 58:163–170. [PubMed: 11890312]
- Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*. 2006; 38:209–13. [PubMed: 16415888]
- Skol AD, Scott LJ, Abacasis GR, Boehnke M. Optimal designs for two-stage genome-wide association studies. *Genet Epidemiol*. 2007; 31:776–788. [PubMed: 17549752]
- Van Es MA, Van Vught PW, Blauw HM, Franke L, Saris CG, Andersen PM, Van Den Bosch L, De Jong SW, Van 't Slot R, Birve A, Lemmens R, De Jong V, Baas F, Schelhaas HJ, Sleegers K, Van Broeckhoven C, Wokke JHJ, Wijmena C, Robberecht W, Veldink JH, Ophoff RA, Van Den Berg LH. ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet Neurology*. 2007; 6:869–877. [PubMed: 17827064]
- Wang HS, Thomas DC, Pe'er I, Stram DO. Optimal two-stage genotyping designs for genome-wide association scans. *Genet Epidemiol*. 2006; 30:356–368. [PubMed: 16607626]
- Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. *Genetics*. 2005; 171:813–823. [PubMed: 16020784]

Figure 1.

Detection probabilities of two-stage designs with joint analysis versus one-stage designs for fixed effects disease models with 12.5% of cases and controls in stage 1 ($\pi_{sample}=0.125$). Other parameters include $T_0=500,000$ SNPs, numbers of SNPs selected at stage 1, $T_1=25,000$ (bold loci) or 1,000 (unbolded loci), number of SNPs selected at stage 2, $T_2=1, 10$, or 100, $M_0=1$ disease-associated SNP, 8,000 total cases, and 8,000 total controls. Points on each locus from left to right correspond to odds ratios per allele of 1.1, 1.2, 1.3, and 1.5.

Figure 2.

Detection probabilities of two-stage designs with joint analysis versus one-stage designs for random effects disease models with 12.5% of cases and controls in stage 1 ($\pi_{\text{sample}}=0.125$). Other parameters include $T_0=500,000$ SNPs, numbers of SNPs selected at stage 1, $T_1=25,000$ (bold loci) or 1,000 (unbolded loci), number of SNPs selected at stage 2, $T_2=1, 10$, or 100, $M_0=1$ disease-associated SNP, 8,000 total cases, and 8,000 total controls. Points on each locus from left to right correspond to standard deviations of the random effects of log odds ratios per allele of $(\pi/2)^{1/2}$ times $\log(1.1)$, $\log(1.2)$, $\log(1.3)$, and $\log(1.5)$.

Figure 3.

Detection probabilities of two-stage designs with joint analysis versus one-stage designs for fixed effects disease models with 25% of cases and controls in stage 1 ($\pi_{sample}=0.25$). Other parameters include $T_0=500,000$ SNPs, numbers of SNPs selected at stage 1, $T_1=25,000$ (bold loci) or 1,000 (unbolded loci), number of SNPs selected at stage 2, $T_2=1, 10$, or 100, $M_0=1$ disease-associated SNP, 8,000 total cases, and 8,000 total controls. Points on each locus from left to right correspond to odds ratios per allele of 1.1, 1.2, 1.3, and 1.5.

Figure 4.

Detection probabilities of two-stage designs with joint analysis versus one-stage designs for random effects disease models with 25% of cases and controls in stage 1 ($\pi_{sample}=0.25$). Other parameters include $T_0=500,000$ SNPs, numbers of SNPs selected at stage 1, $T_1=25,000$ (bold loci) or 1,000 (unbolded loci), number of SNPs selected at stage 2, $T_2=1, 10$, or 100, $M_0=1$ disease-associated SNP, 8,000 total cases, and 8,000 total controls. Points on each locus from left to right correspond to standard deviations of the random effects of log odds ratios per allele of $(\pi/2)^{1/2}$ times $\log(1.1)$, $\log(1.2)$, $\log(1.3)$, and $\log(1.5)$.

Table 1

Probability of detecting a disease SNP (DP) and optimal stage1 weight for joint analysis, λ_{opt} , for the fixed effects model with log odds ratio per allele $\beta = \log(1.2)$

π_{sample}	Analysis	Number of disease SNPs, $M_0=1$						Number of disease SNPs, $M_0=10$					
		$T_1=1000$			$T_1=25,000$			$T_1=1000$			$T_1=25,000$		
		$T_2=1$	$T_2=10$	$T_2=100$	$T_2=1$	$T_2=10$	$T_2=100$	$T_2=1$	$T_2=10$	$T_2=100$	$T_2=1$	$T_2=10$	$T_2=100$
0.125	Replicate	.266	.268	.269	.630	.654	.664	.096	.276	.277	.100	.652	.669
	Joint	.267	.268	.269	.635	.656	.664	.096	.278	.277	.100	.654	.669
	λ_{opt}	.25	.17	.27	.35	.35	.25	.03	.00	.00	.27	.30	.20
0.25	Replicate	.612	.624	.626	.803	.854	.881	.100	.616	.625	.100	.829	.881
	Joint	.613	.624	.626	.825	.865	.885	.100	.616	.625	.100	.843	.884
	λ_{opt}	.32	.30	.15	.40	.35	.25	.15	.00	.00	.33	.40	.30
0.50	Replicate	.797	.869	.897	.690	.832	.912	.100	.819	.894	.100	.767	.911
	Joint	.843	.884	.900	.858	.917	.953	.100	.838	.895	.100	.881	.953
	λ_{opt}	.40	.35	.20	.45	.45	.40	.35	.30	.05	.50	.45	.40
1.00	One-stage ^a	.882	.933	.966	.882	.933	.966	.100	.898	.966	.100	.898	.966

^aThe DP estimate for the one-stage analysis is from the three simulations with common M_0 and with $\pi_{sample} = 0.125, 0.25$, and 0.50 .

The number selected in a one-stage design, T_1 , is listed under the column labeled T_2

Table 2

Probability of detecting a disease SNP (DP) and optimal stage1 weight for joint analysis, λ_{opt} , for the random effects model with standard deviation of log odds ratio per allele $\tau = (\pi/2)^{1/2} \log(1.2)$

π_{sample}	Analysis	Number of disease SNPs, $M_0=1$						Number of disease SNPs, $M_0=10$					
		$T_1=1000$			$T_1=25,000$			$T_1=1000$			$T_1=25,000$		
		$T_2=1$	$T_2=10$	$T_2=100$	$T_2=1$	$T_2=10$	$T_2=100$	$T_2=1$	$T_2=10$	$T_2=100$	$T_2=1$	$T_2=10$	$T_2=100$
0.125	Replicate	.311	.314	.316	.449	.471	.487	.096	.307	.315	.100	.474	.493
	Joint	.311	.314	.316	.452	.472	.487	.096	.307	.315	.100	.475	.494
	λ_{opt}	.10	.22	.08	.35	.32	.30	.00	.00	.00	.40	.25	.25
0.25	Replicate	.448	.459	.464	.512	.549	.582	.099	.433	.453	.100	.545	.584
	Joint	.451	.460	.464	.524	.557	.586	.099	.433	.453	.100	.553	.589
	λ_{opt}	.30	.15	.25	.40	.35	.35	.00	.00	.00	.30	.35	.35
0.50	Replicate	.512	.558	.584	.471	.530	.596	.099	.496	.557	.100	.518	.594
	Joint	.535	.568	.589	.540	.584	.628	.099	.498	.557	.100	.572	.629
	λ_{opt}	.40	.30	.20	.50	.40	.40	.05	.10	.00	.20	.40	.40
1.00	One-stage ^a	.554	.601	.647	.554	.601	.647	.100	.589	.651	.100	.589	.651

^aThe DP estimate for the one-stage analysis is from the three simulations with common M_0 and with $\pi_{sample} = 0.125, 0.25, \text{ and } 0.50$.
The number selected in a one-stage design, T_1 , is listed under the column labeled T_2