SAS MACROs for SNP-phenotype association studies: implementations of the MAX test and MAX-maxT algorithms

Supplementary documentation for project

Dana Aeschliman, Marie-Pierre Dubé
Statistical Genetics Research Group
Montreal Heart Institute

October 18, 2007

Abstract

The MAX test investigates three inheritance models while yielding one p-value. The MAX-maxT test can be useful for producing p-values corrected for multiple testing which take into account the correlation structure of the data. We encode both of these in a suite of SAS MACROs. As well, we have studied the performance of the MAX-maxT test in the presence of missing values and in the presence of population substructure. Last, we compare the performance of the MAX-maxT algorithm to that of SAS PROC CASECONTROL.

1 Introduction

In SNP-phenotype association studies (case-control), the MAX test [3], [12] allows simultaneous investigation of three plausible inheritance models while returning a single p-value. Although the maxT algorithm [10], [4] has been publicly available for some time and a version of it has been used in the analysis of a SNP-phenotype case-control study [9], we are unaware of any generally available implementations of the MAX-maxT algorithm.

This document starts with some initial considerations involved in using the maxT algorithm in the context of the MAX test applied to case-control data. Section 3 shows the results of a series of simulations which investigate the performance of the MAX-maxT algorithm in the presence of missing values and population structure. Following that, Section 4 shows the results of comparing our implementation of the MAX test to the tests in SAS PROC CASECONTROL and is followed by the discussion and conclusions.

2 Initial considerations

We begin by evaluating whether the maxT algorithm's reliance on test statistics for distinct hypotheses having the same or very similar distributions under the null would invalidate its being applied in genetic tests as the three components of the MAX test have different correlations for different SNPs. However, as is evident from the Appendix of [12], under the null hypothesis these correlations depend only on the minimum allele frequency (MAF). By doing a simulation and producing several graphics such as Figures 1 and 2 below, we conclude that the maxT seems to be robust to variation in MAFs.
A desirable property of procedures where many hypotheses are tested simultaneously is that the familywise error rate (the rate at which ANY Type I errors are made for all hypotheses tested in an experiment, the FWER) be maintained at the nominal level regardless of which subset of hypotheses are false. Westfall and Young [10] state that this will be true when subset pivotality holds. We show in the Appendix that subset pivotality at least holds in the ideal case of no population structure or missing values.

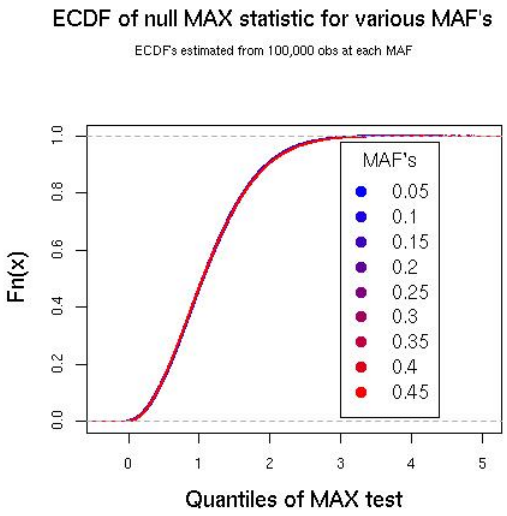

Figure 1: ECDFs of the MAX statistic for various MAFs. The distributions are very comparable.

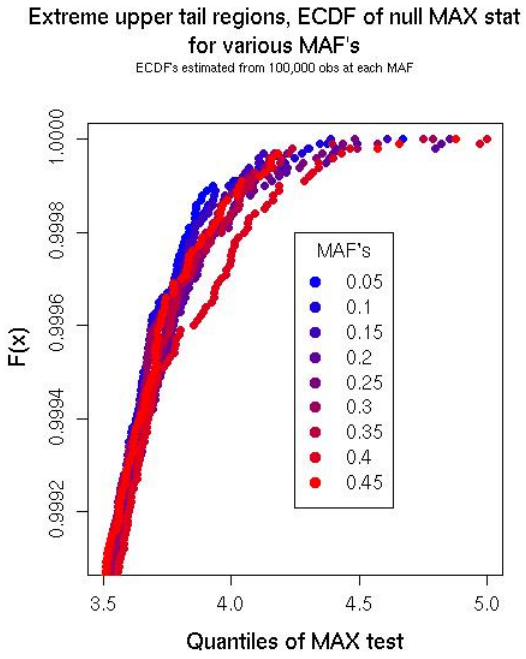

Figure 2: Extreme upper tail regions of the ECDFs of the MAX statistic for various MAFs. The distributions are very comparable.

3 Simulations investigating the performance of the MAX-maxT algorithm under some departures from the ideal

In order for an analysis scheme to be practically useful, it must perform reasonably well under suboptimal conditions. In this section, we perform three simulations: one where some of the data are affected by missingness and two where some of the data are affected by population structure. The two types of population structure that we examine are cryptic relatedness (CR) and population stratification (PS) [11], [13]. The next paragraphs describe the design of these simulations.

In each simulation, we generate 1,050 data sets and permute the case-control status vector at least 1,000 times. All data sets consist of 100 SNPs: three SNPs which are in linkage disequilibrium (LD) with a phenotype-causing SNP, 47 SNPs that are independent, and 25 pairs of SNPs such that the members of a pair are in LD. All data sets contain three SNPs that have (target) correlations of 0.95, 0.65 and 0.2 with the phenotype-causing SNP. The mode of inheritance

in all simulations is additive, i.e. if f_0, f_1, and f_2 are the penetrances given respectively zero, one and two copies of the phenotype-causing allele then $f_1 = (f_0 + f_2)/2$. We choose the MAF for each SNP or for the correlated pairs from a distribution we create after carefully inspecting the table in the supporting online material for [6] at the top of page S12. In generating the phenotype-causing SNPs we follow the algorithm outlined in [13] in the top left corner of p.190. The densities from which we sample relative risks (RRs), minor allele frequencies (MAFs), and baseline penetrances (penetrance assuming that the count of phenotype-causing allele is zero, f_0) are shown in Figure 3. The medians, 90th percentiles and theoretical maximums for f_0, MAF and RR are given in Table 1. The minimum and maximum MAFs are respectively 0.05 and 0.45 for the missingness and CR studies.

Parameter	Quantile		
	Median	90th percentile	Maximum
$f0$	0.022	0.056	0.1
MAF	0.222	0.401	0.450
RR	2.322	3.558	6.000

Table 1: Median, 90th quantile and theoretical maximum for the generating distributions for f_0, MAF and RR.

In each simulation, we include some pairs of SNPs that have within-pair correlation chosen uniformly on (0.15, 0.85) and common MAF p chosen as outlined above. The way that we generate realizations of these correlated Bernoulli random variables is as follows. Let I_{1i} and I_{2i} represent two loci on a chromosome of individual i. The entries in the vectors I_1 and I_2 will have correlation F when the haplotype is chosen according to Table 2 [1], [11].

Haplotype	(0,0)	(0,1)	(1,0)	(1,1)
Probability	$(1-p)^2 + Fp(1-p)$	$p(1-p)(1-F)$	$p(1-p)(1-F)$	$p^2 + Fp(1-p)$

Table 2: Haplotypes and corresponding multinomial probabilities for generating vectors of Bernoulli variables I_1 and I_2 where each entry has success probability p and correlation between I_{1i} and I_{2i} is F.

3.1 Performance of the MAX-maxT algorithm in the presence of missing values

To investigate the performance of the MAX-maxT algorithm in the presence of missing values we perform a simulation where there are one of four possible

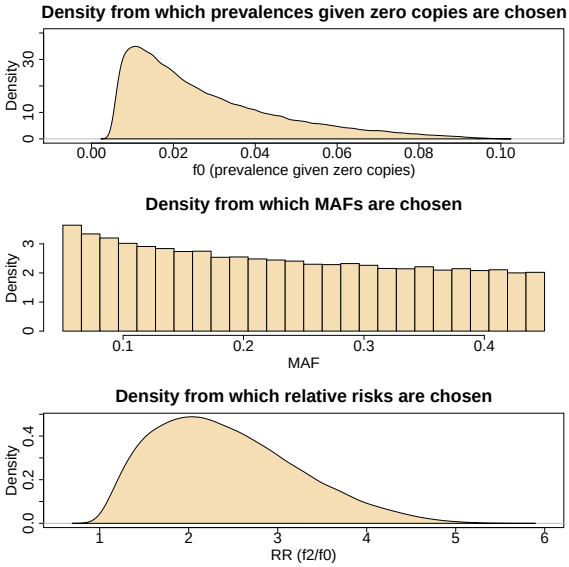

Figure 3: Densities from which baseline prevalences, minimum allele frequencies (MAFs) and relative risks are chosen for all simulations in Section 3.

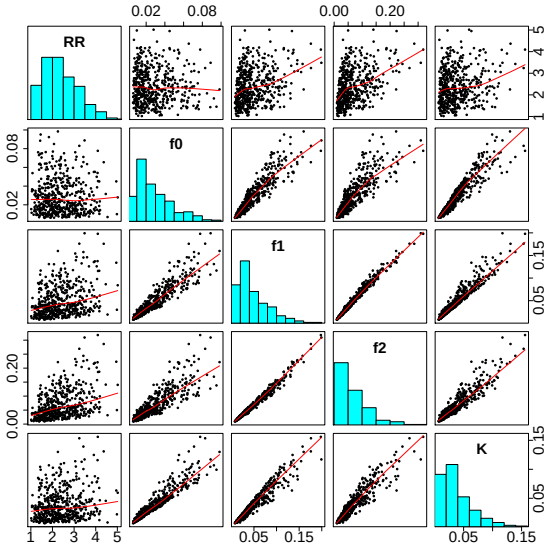

Figure 4: Pairs plot showing 500 realizations from the distribution from which relative risks (RR=f_2/f_0), baseline prevalence (prevalence given zero copies of phenotype-causing allele, f_0), f_1, f_2 and overall population prevalence (K) are generated.

levels of missingness in each data set. At each level of missingness, we generate a total of 1,050 data sets in ten groups of 105. Some data sets have no missing values, others have 2%, 4% or 10% missing values. By "missing" we mean that the state of both alleles is missing for a subject at that SNP. All data sets have 500 cases ($R = 500$) and 500 controls ($S = 500$) for a total of 1,000 subjects ($N = 1000$). Although, in looking at Figure 5, one suspects that the variance of the FWER is greater for the 2% and 4% missing data sets than for the complete data set, neither of these ratios differ significantly from unity.

One troubling aspect here is that the power seems to be independent of the degree of correlation with the phenotype-causing SNP. This could be due to several factors, one of which is our algorithm for generating correlated Bernoulli variables. In the algorithm for the missingness and CR simulations, given the MAF of the phenotype-causing SNP and a target F, we allow the new vector of alleles to be an approximate solution. Prior to running our last simulation (to investigate PS), Subsection 3.4, we rewrote our code for generating correlated Bernoulli variables, forcing the MAF at the new locus to be exactly the same as the MAF at the phenotype-causing SNP. We discuss details of the final algorithm in Subsection 3.4.

A last comment here is that we perform 2,000 permutations for all complete data sets and for 420 of the 1,050 data sets at each level of missingness. The case-control vector of all other data sets is permuted 1,000 times.

3.2 Performance of the MAX-maxT algorithm in the presence of population structure

Two types of population structure that may affect case-control association studies are population stratification (PS) and cryptic relatedness (CR). In this subsection, we create subpopulations having allelic correlations or having different MAFs then we allocate subpopulation members to cases or controls. We investigate the effect that this has on the FWER and the power.

3.3 Cryptic relatedness

Whittemore defines CR as occurring when a population is composed of at least two subpopulations where all marker alleles of individuals in the same subpopulation have correlation F, marker alleles of individuals in distinct subpopulations are independent and all subpopulations have the same mean MAFs [11].

3.3.1 Simulation details

Our simulation breaks with the traditional model by allowing different degrees of correlation within different subpopulations. We perform our simulation by allocating half of our participants to a subpopulation having allelic correlations equal to that exhibited by the inhabitants of the Saguenay/Lac St. Jean region of Quebec, 0.0055 [7], and the other half to a subpopulation having allelic correlations equal to that exhibited by "a random sample of Caucasians drawn

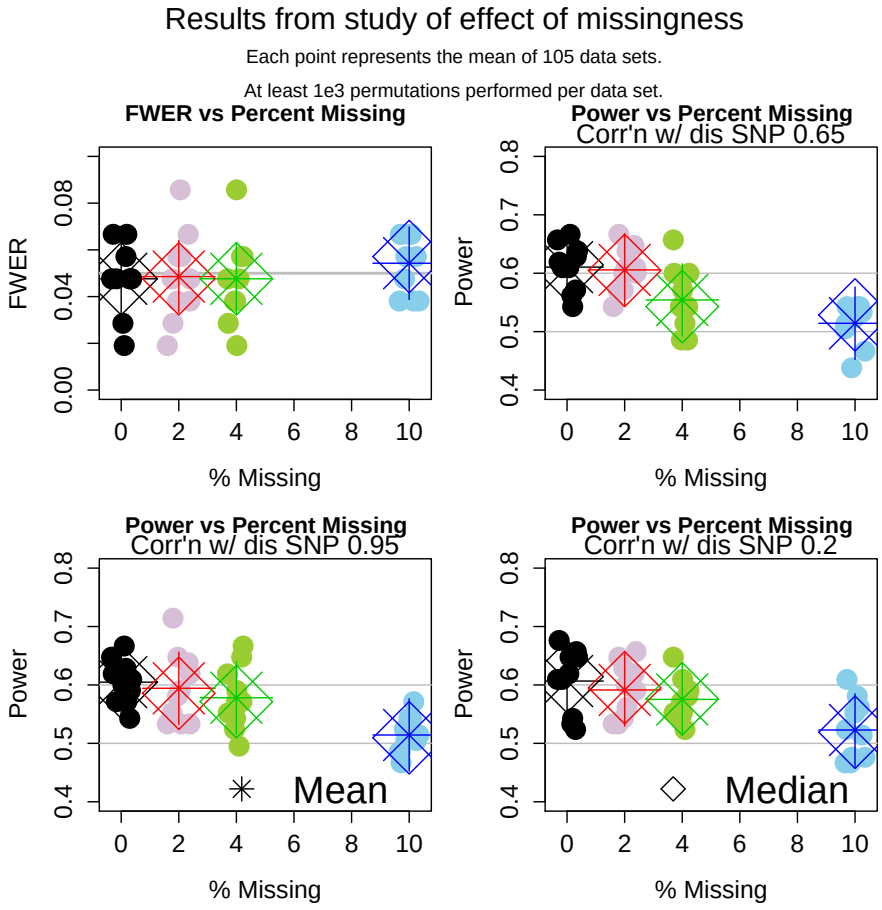

Figure 5: Familywise error rate (FWER) and power in the presence of missing data. Note that the mean FWER stays close to the nominal value even at 10% missing data. For each data set, $R = S = 500$.

from Europe", 0.0006 [1]. For particular realizations of data sets, we follow the algorithm outlined in [13] on the bottom left of p.190 (with the modification previously mentioned). We draw 1,000 study participants for each data set, evenly divided into cases and controls and into the subpopulations.

3.3.2 Results

As we see in Figure 6 below, the results with and without CR are quite comparable. Although the asymptotic FWER seems to equal the nominal rate even in the presence of CR, some small evidence exists that the variance of the FWER in the presence of CR is greater than the variance in the absence of CR (parametric F test p-value 0.054, bootstrap p-value 0.060).

3.4 Population stratification

In the PS model, subpopulations have different MAFs but all alleles are independent.

3.4.1 Simulation details

We use the data from the HapMap project as packaged in the R package SNPassoc to investigate how the MAX-maxT algorithm performs in an instance of PS. In this simulation, the sample size is 256 participants evenly divided between cases and controls. For a number of our SNPs, we randomly allocate 1/2 of our sample to be from the Yoruban subpopulation (YRI) and 1/2 to be from the European (CEU) subpopulation. Thus, as in the CR study above, for each simulated data set, the number in each subpopulation is the same as the number of cases and the number of controls. We draw bootstrap samples from the SNPassoc data set HapMap to use as representative MAFs.
In this simulation we simplify our algorithm for producing SNPs that are correlated with the phenotype-causing SNP. Assume that I_1 is a vector of indicators such that $I_{1i} \in \{0, 1\}$ and that, given $\sum_{i=1}^{N} I_{1i}$, we will generate I_2 such that $\sum_{i=1}^{N} I_{2i} = \sum_{i=1}^{N} I_{1i}$. Note that $\sum_{i=1}^{N} I_{1i} I_{2i} \geq \max(2 * \sum_{i=1}^{N} I_{1i} - N, 0)$. Given the above, our observed measure of correlation, $\hat{F}$ is an increasing function of $\sum_{i=1}^{N} I_{1i} I_{2i}$ and

$$\hat{F} \geq \frac{N(\max(2*\sum_{i=1}^{N} I_{1i} - N, 0)) - (\sum_{i=1}^{N} I_{1i})^2}{(\sum_{i=1}^{N} I_{1i})(N - \sum_{i=1}^{N} I_{1i})}.$$

Here is the algorithm that we use. Given I_1 and a target F, we first check that F is above the minimal value. If not, we replace F by the minimal value. We then calculate the required sum $\sum_{i=1}^{N} I_{1i} I_{2i}$ as the closest integer to

$$\frac{F(\sum_{i=1}^{N} I_{1i})(N - \sum_{i=1}^{N} I_{1i}) + (\sum_{i=1}^{N} I_{1i})^2}{N}.$$

As Table 3 shows, all margins and one entry being fixed, the rest of Table 3 is now known.

		I_2		
		0	1	
I_1	0	$N - 2 * \sum_{i=1}^{N} I_{1i} + \sum_{i=1}^{N} I_{1i}I_{2i}$	$\sum_{i=1}^{N} I_{1i} - \sum_{i=1}^{N} I_{1i}I_{2i}$	$N - \sum_{i=1}^{N} I_{1i}$
	1	$\sum_{i=1}^{N} I_{1i} - \sum_{i=1}^{N} I_{1i}I_{2i}$	$\sum_{i=1}^{N} I_{1i}I_{2i}$	$\sum_{i=1}^{N} I_{1i}$
		$N - \sum_{i=1}^{N} I_{1i}$	$\sum_{i=1}^{N} I_{1i}$	N

Table 3: 2x2 table for two correlated Bernoulli variables

Another difference between the PS simulation and the previous two is that the MAF here varies on (0.1, 0.5) rather than on (0.05, 0.45).
The next subsection explains how we use the data in the R package SNPassoc [5] to arrive at having 4,587 pairs of MAFs for the subsamples from which we draw the bootstrap samples. The part of this report dealing with PS ends with a subsection detailing our simulation results.

3.4.2 Excluding SNPs which would probably not pass HWE

In this subsection we explain how we exclude some of the SNPs from the HapMap data set.
The HapMap data set contains information for 9,305 SNPs for 60 subjects in each of two subpopulations, the Yoruban (YRI) and European ascertained in Utah (CEU), but includes no genotyping information for 464 of the SNPs for the YRI sample. We also exclude 3,977 SNPs for whom the average of the MAFs for the two samples is less than 0.10. We anticipate that, were a population sampled that consists of two subpopulations randomly assigned to cases and controls, a screen might be performed using the controls to exclude SNPs which appear to be out of HWE. We exclude 277 SNPs which we calculate as having too high a probability of eventually being declared to not be in HWE. This process is explained in the next paragraph.
For each SNP, we have p_1 and p_2, the MAFs in the YRI and CEU samples. We assume that each data set will contain 128 members of each subpopulation. The number of individuals from the YRI subpopulation allocated to the cases is Binomial(128, 1/2). From this setting for each SNP, we select a matrix A where each row of A is a realization of $\mathbf{S} = (s_0, s_1, s_2)$, the row representing the controls in the ubiquitous 2x3 table of case-control status by genotype [8]. For each row s of A, we perform the exact test of HWE [2] which is known to maintain a Type I error rate at or below the nominal level. Those members of A having an exact p-value of less than 0.05 are assigned to our set B. We also compute $\mathrm{P}(\mathbf{S} = \mathbf{s} \mid p_1, p_2, G_1 = G_2 = 128)$ using the fact that, conditional on C_1, the number of subpopulation 1 members in the controls, the probability

of a particular 2x3 table showing the genotype counts for the two subpopulations within the controls is the product of two multinomial probabilities and thus

$$\mathrm{P}(\mathbf{S} = \mathbf{s} \mid p_1, p_2, G_1 = G_2 = 128) = \sum_c \mathrm{P}(\mathbf{S} = \mathbf{s} \mid p_1, p_2, G_1 = G_2 = 128, C_1 = c)\mathrm{P}(C_1 = c \mid p_1, p_2, G_1 = G_2 = 128).$$

We assume that sets A and B are independent and estimate P(HWE will be rejected $\mid p_1, p_2, G_1 = G_2 = 128$) as $\frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}$. Of course, given p_1 and p_2, the probability of a realization $\mathbf{s}$ occurring in A is not independent of its belonging to B. For example, in the coordinate plane, for (p_1, p_2) pairs close to $y = x$, realizations with greater probabilities of occurring have smaller probabilities of belonging to B. However, we argue that the assumption of independence should be fairly accurate in the cases where accuracy matters most, for (p_1, p_2) pairs on the boundary between those pairs which would clearly lead to the hypothesis of HWE being rejected for generated realizations and those pairs which would clearly lead to the hypothesis of HWE being accepted. Figures 7 and 8 support this discussion.

3.4.3 Results for PS study

In Figure 9 we show some of the results of our study involving PS. It's difficult to discern from the graphs if the variance is bigger with or without PS so we include Table 4 and note that the statistical evidence for the variance of the FWER being greater in the presence of PS is quite weak (parametric F-test p-value 0.15, bootstrap p-value 0.22). We note that, both in the absence and in the presence of PS, the observed power is slightly greater for the SNPs that are more highly correlated with the phenotype-causing SNP. Oddly, within levels of correlation with the phenotype-causing SNP, the observed power is slightly greater in the presence of PS. We attribute this to sampling variation.

Presence of PS?	FWER	Power, corr=0.95	Power, corr=0.65	Power, corr=0.20
No	0.0002	0.0024	0.0025	0.0016
Yes	0.0004	0.0027	0.0012	0.0014

Table 4: Variances for the FWER and for the power at different levels of correlation with the disease SNP for the PS study.

4 Comparison of the MAX test with SAS PROC CASECONTROL

In this section we compare the performance of the MAX test as implemented in our MACROs to the programs already available in SAS PROC CASECONTROL. Three tests are available in SAS PROC CASECONTROL which are

termed the allele, genetic and trend tests in the SAS documentation. We note that the trend test is one of the components of the MAX test.

4.1 Power

In this subsection we show three different graphs corresponding to three different data generating models: the recessive, additive and dominant models. For each data generating model, we produce power curves for the three tests in SAS PROC CASECONTROL as well as for the MAX test. The parameters MAF, prevalence, sample size, and case-control ratio are the same for all points on a particular graph.

For the purposes of this section, "additive" model just means that the relative risk given one copy of the phenotype-causing allele (γ_1) is between 1 and the relative risk given two copies of the phenotype-causing allele (γ_2). We wanted to smoothly pass through a specific point in parameter space from Table 2 of [12]: the power is 0.8 when the model is additive, the total sample size is 179, the case-control ratio is 1, the MAF is 0.2 and γ_1 and γ_2 are respectively 2 and 3. Thus. we chose f_0 to take equally spaced values on the interval [0.055, 0.099], let $f_1 = -1.569 * f_0 + 0.2548$ and chose f_2 so that the overall prevalence is 0.1. Final parameters for the additive models are shown in Table 5.

We note that, for the recessive case, the MAX and the genotype tests seem to do equally well in the example. For the additive case, the MAX test doesn't do quite as well as the trend test or the allele test. For the dominant model all four tests seem to do equally well on the given example.

$\text{OR} \equiv \frac{\text{P(Sick} \mid \text{1 copy)/P(Well} \mid \text{1 copy)}}{\text{P(Sick} \mid \text{0 copies)/P(Well} \mid \text{0 copies)}}$	P(Sick | 0 copies)	Relative risks	
		γ_1	γ_2
1	0.099	1	1.21
1.09	0.096	1.08	1.36
1.19	0.093	1.17	1.52
1.3	0.09	1.27	1.69
1.42	0.087	1.37	1.88
1.55	0.084	1.48	2.08
1.69	0.08	1.6	2.3
1.84	0.077	1.73	2.53
2.01	0.074	1.87	2.78
2.2	0.071	2.02	3.06
2.4	0.068	2.19	3.36
2.62	0.065	2.38	3.7
2.87	0.062	2.57	4.06
3.15	0.058	2.8	4.47
3.46	0.055	3.05	4.91

Table 5: Parameters of various additive models

When comparing the MAX test to PROC CASECONTROL, arguments in terms of power may not be sufficient for choosing one or the other. However, one needs to remember that this discussion occurs in the context of the mode of inheritance being unknown (otherwise one would just choose the component of the MAX test which corresponds to the known mode of inheritance). The next section shows the error rates and FWER at desired $\alpha = 0.05$ for using the MAX test vs. using the (highly questionable but apparently popular) strategy "just take the smallest of the three p-values from PROC CASECONTROL."

4.2 Type I Error Rates and Bonferroni-adjusted FWERs

This subsection shows the results after comparing the MAX test to the results of PROC CASECONTROL and the strategy "just take the smallest of the three p-values." All the work in this subsection was done in R. For PROC CASECONTROL and the questionable strategy, we see that the Type I error rate is about 1.5 times greater than the desired rate and that the FWER is about twice the desired rate. Thus, we predict that using the MAX test can lead to savings which would otherwise be spent further investigating the excess false positives.

5 Discussion and conclusions

We have presented some tests using our implementation of the MAX test and the MAX-maxT algorithm. We have also shown some results implementing these tests in particular cases of missing values and population structure. Although we realize that the number of sample data sets (1,050) that we use in our simulations is relatively small, we state that this sample size should be sufficient for identifying any extreme problems. As well, we consider that we have used rather extreme settings for our simulations (i.e., 10% missing data, MAFs drawn from the CEU and YRI populations). We have seen that the results are quite favorable towards using the software that we provide in the context of SNP-phenotype association studies.

We have not looked at population structure in situations where the different subpopulations have different prevalences except to the degree that this happens inevitably in randomly assigning all study participants to cases and controls. We note, however, that a simulation assigning specific prevalences to different subpopulations could easily be implemented with minor alterations to the existing R code which is available from the authors upon request.

References

[1] Devlin B and Roeder K. (1999) Genomic Control for Association Studies. *Biometrics*, **55**, 997-1004.

[2] Emigh TH. (1980) Comparison of tests for Hardy-Weinberg Equilibrium. *Biometrics*, **36**, 627-642.

[3] Freidlin B, Zheng G., Li Z., Gastwirth JL. (2002) Trend tests for case-control studies of genetic makers: power, sample size and robustness. *Hum. Hered.*, **53**, 146-152.

[4] Ge Y, Dudoit S, Speed T. (2003) Resampling-based multiple testing for microarray data analysis. *TEST*, **12**, 1-77.

[5] González et al. (2007) SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, **23**,644-645.

[6] Hinds DA et al. (2005) Whole genome patterns of common DNA variation in three human populations. *Science*, **307**, 1072-1079.

[7] Pausova et al. (2002) A Genealogical Study of Essential Hypertension with and without Obesity in french Canadians. *Obesity Research*, **10**, 463-470.

[8] Sasieni P. (1997) From Genotypes to Genes: Doubling the Sample Size. *Biometrics*, **53**, 1253-1261.

[9] Sladek R et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. (2007) *Nature*, **445**, 881-885.

[10] Westfall, P.H. and Young, S.S. (1993) *Resampling-Based Multiple Testing*, John Wiley and Sons.

[11] Whittemore A. (2006) Population Structure in Genetic Association Studies. *Proc. J. S. M.*

[12] Zheng G. and Gastwirth JL. (2006) On estimation of the variance in Cochran-Armitage trend tests for genetic association using case-control studies. *Statist. Med.*, **25**, 3150-3159.

[13] Zang Y et al. Robust Genomic Control and Robust Delta Centralization Tests for Case-Control Association Studies. (2007) *Hum. Hered.*, **63**, 187-195.

A Proof of subset pivotality for the MAX-maxT algorithm under ideal conditions

We state that subset pivotality holds for a family of hypotheses resulting from a case-control study of genetic association. Our argument follows the logic of Example 2.1 of Westfall and Young (1993): we specify the distribution of two of the test statistics arising from this situation, showing that this distribution doesn't depend on properties of the other test statistics.

Let j and k be bi-allelic SNPs and M_m be the allele of minimum frequency for $m \in \{j,k\}$. Letting $f_{im} = \Pr(\text{case} \mid i\ M_m \text{ alleles in the genotype})$ we can write the hypotheses to be tested as

$$\mathrm{H}_{0m} : f_{0m} = f_{1m} = f_{2m} = K,\ m \in \{j,k\}$$

where K is the disease prevalence in the population. Assume that alleles on distinct chromosomes are independent and alleles on the same chromosome have correlation F. The joint distribution of the counts Y_{ij} and Y_{ik} of minor alleles at the two loci for individual i is specified. For example, the probability is $(p_1p_2 + F\sqrt{p_1p_2(1-p_1)(1-p_2)})^2$ that the pair of alleles at locus j is (1, 1) while that at locus k is (1, 1). Define some random variables whose values depend on the Ys

$$X(0) = \begin{cases} 0, & Y < 2 \\ 1, & Y = 2 \end{cases}, \quad X(1/2) = \begin{cases} 0, & Y = 0 \\ 1/2, & Y = 1 \\ 1, & Y = 2 \end{cases}, \quad X(1) = \begin{cases} 0, & Y = 0 \\ 1, & Y > 2 \end{cases}.$$

Let $X_{im}(t)$ represent a generic coded count for indivual i at locus m. Note that the joint distribution $(X_{ij}(t), X_{ik}(t'))$ is specified under the assumptions above and in particular $\mathrm{cov}(X_{ij}(t), X_{ik}(t'))$ can be computed. We can write the numerator of a component of the MAX test at locus m as $U_m(t) = S\sum_{i=1}^{R} X_{1im}(t) - R\sum_{i=1}^{S} X_{0im}(t)$ where the first index on $X_{1im}(t)$ denotes case-control status. Noting that, under H_{0m}, $T_m(t) = U_m(t)/\sqrt{\mathrm{var}_{\mathrm{H}_{0m}}(U_m(t))}$ is asymptotically distributed as $\mathrm{N}(0,1)$, it remains only to specify $\mathrm{cov}_{H_{0j}\cup H_{0k}}(T_j(t), T_k(t'))$ in order to have completely specified the joint distribution $(T_j(t), T_k(t'))$. Thus,

$$\begin{aligned}
&\mathrm{cov}_{H_{0j}\cup H_{0k}}(T_j(t), T_k(t')) \\
&\quad = \mathrm{E}_{H_{0j}\cup H_{0k}}(T_j(t), T_k(t')) \\
&\quad = (\mathrm{var}_{\mathrm{H}_{0j}}(U_j(t))\mathrm{var}_{\mathrm{H}_{0k}}(U_k(t'))^{-1/2}\mathrm{E}_{H_{0j}\cup H_{0k}}(U_j(t), U_k(t')) \\
&\quad = (\mathrm{var}_{\mathrm{H}_{0j}}(U_j(t))\mathrm{var}_{\mathrm{H}_{0k}}(U_k(t'))^{-1/2}RSN\mathrm{cov}(X_{1ij}(t), X_{1ik}(t')).
\end{aligned}$$

The distribution of the MAX test statistics at j and k, being a function only of the statistics $T_m(t)$, is now completely specified.

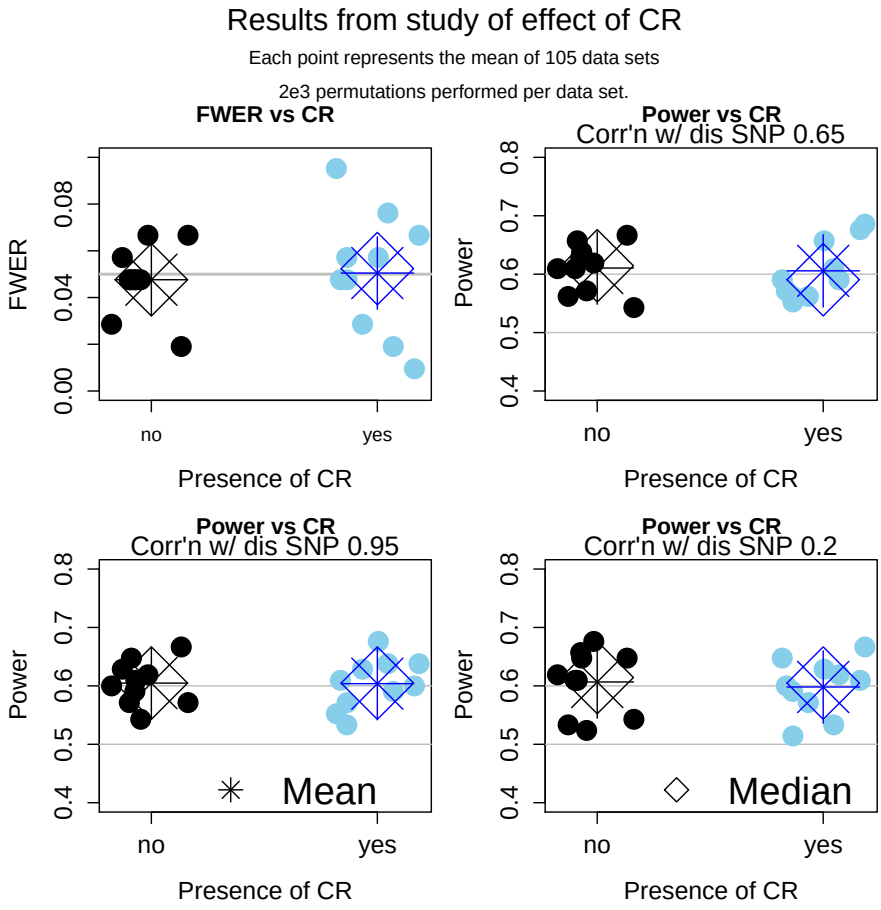

Figure 6: Performance of MAX-maxT algorithm in the presence of cryptic relatedness. Expectation of the FWER seems to be unaffected.

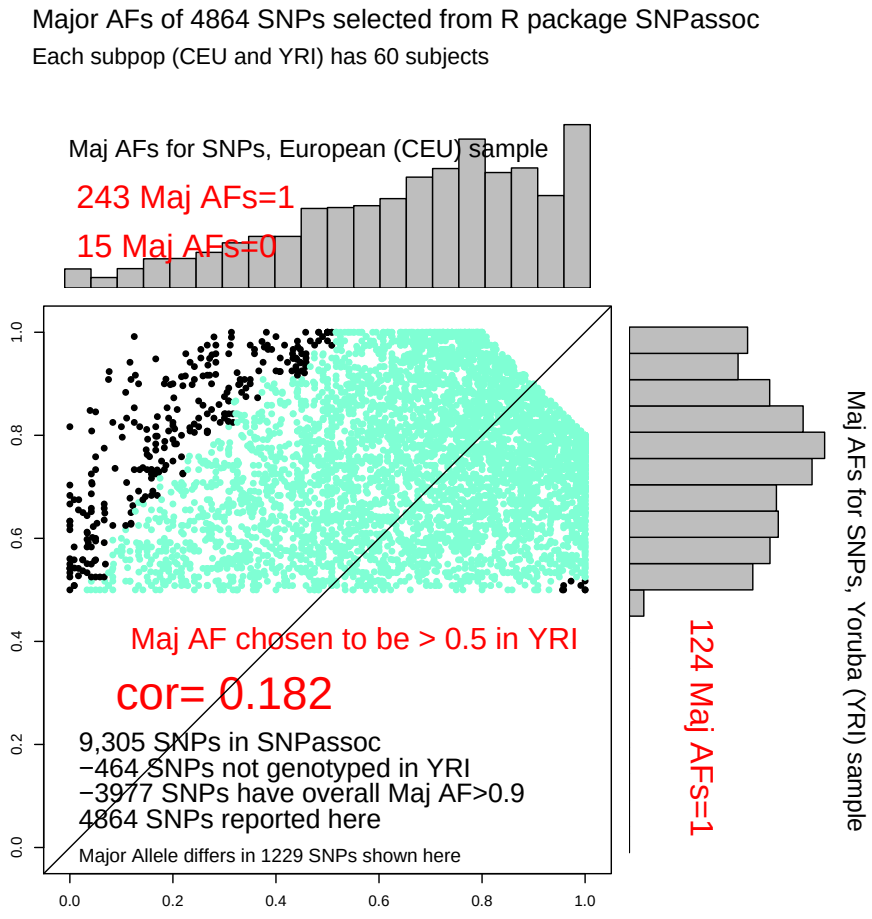

Figure 7: Major Allele Frequencies in YRI sample vs those in the CEU sample from the R package SNPassoc. **Black** dots are those (p_1, p_2) pairs for which estimated P(HWE will be rejected $\mid p_1, p_2, G_1 = G_2 = 128) \geq 0.95$.

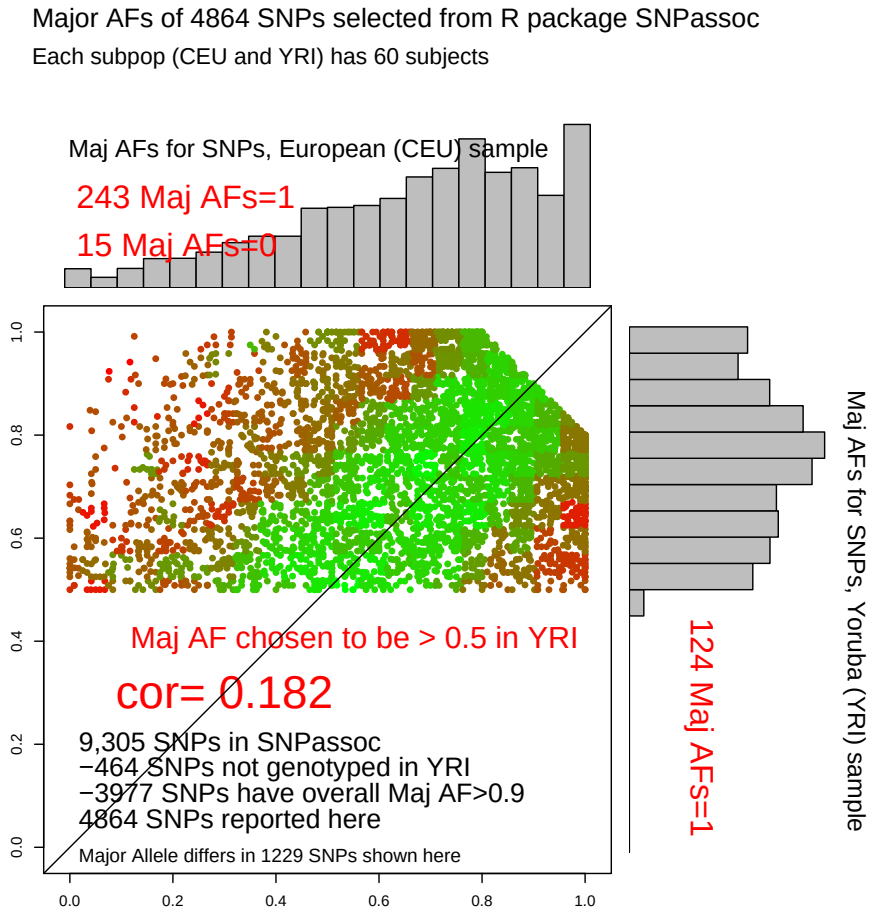

Figure 8: Major Allele Frequencies in YRI sample vs those in the CEU sample from the R package SNPassoc. Red dots are those (p_1, p_2) pairs which lead to realizations n for whom the p-value of the HWE exact test have negative Spearman correlations with the probability of occurrence.

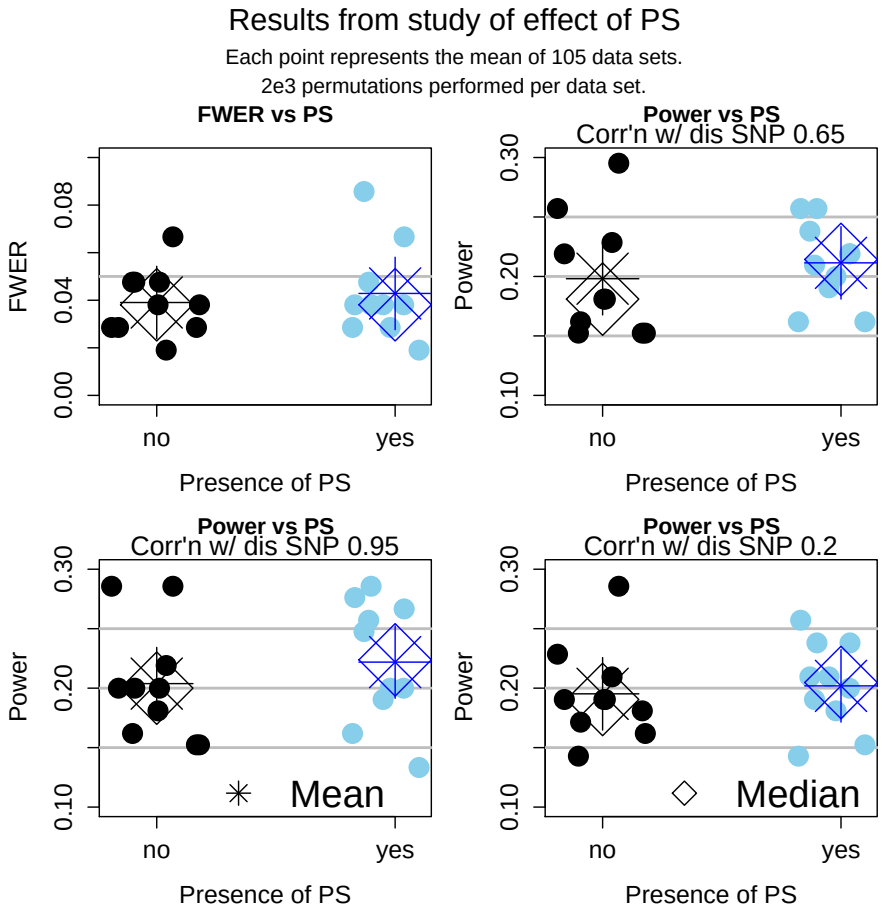

Figure 9: Performance of MAX-maxT algorithm in the presence of severe population stratification. Expectation of the FWER seems to be unaffected.

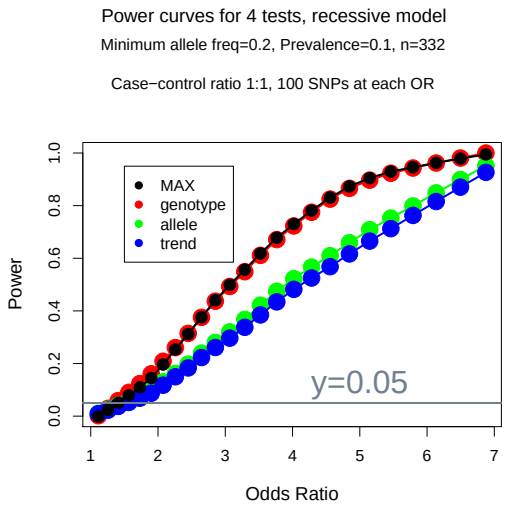

Figure 10: Power of four analysis schemes in the case of a recessive trait.

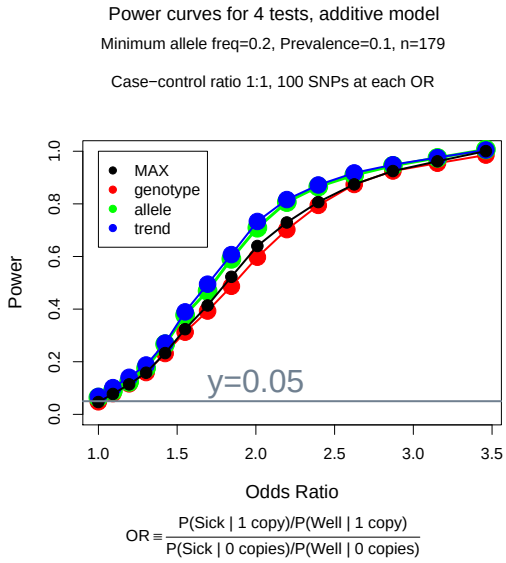

Figure 11: Power of four analysis schemes in the case of an additive trait.

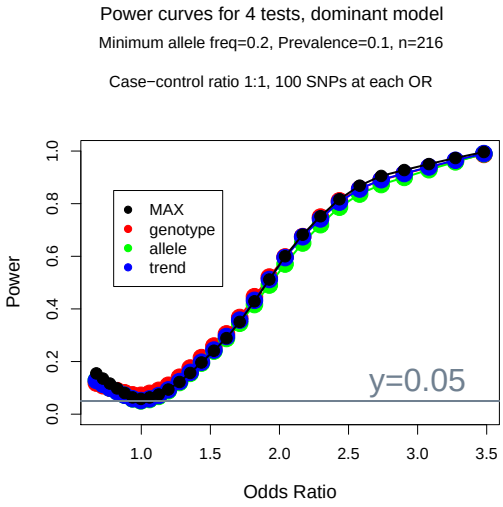

Figure 12: Power of four analysis schemes in the case of a dominant trait.

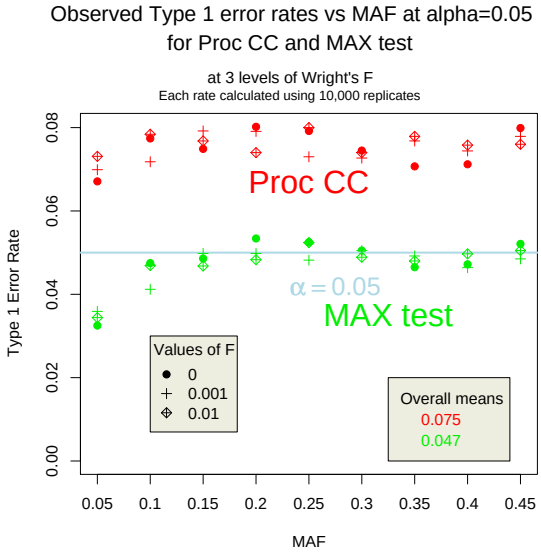

Figure 13: Type I error rates for the MAX test and for the "just use the smallest of the 3 p-values" approach to applying SAS PROC CASECONTROL.

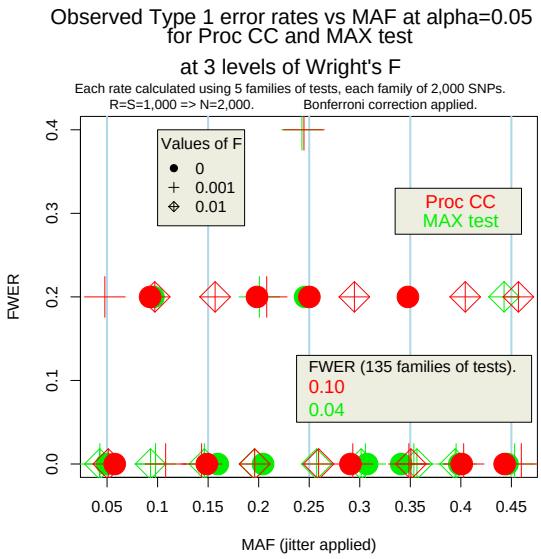

Figure 14: Bonferroni-adjusted FWER for the MAX test and for the "just use the smallest of the three p-values" approach to applying the tests in SAS PROC CASECONTROL.