



Published in final edited form as:

Behav Genet. 2018 March ; 48(2): 155–167. doi:10.1007/s10519-018-9890-6.

A Brief Critique of the TATES Procedure

Fazil Aliev^{1,2,¶,*}, Jessica E. Salvatore^{1,3,¶,*}, Arpana Agrawal⁴, Laura Almasy⁵, Grace Chan⁶, Howard J. Edenberg⁷, Victor Hesselbrock⁶, Samuel Kuperman⁸, Jacquelyn Meyers⁹, and Danielle M. Dick^{1,10}

¹Department of Psychology, Virginia Commonwealth University, Richmond, Virginia, United States of America

²Department of Actuarial and Risk Management, Karabuk University, Karabuk, Turkey

³Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, United States of America

⁴Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, United States of America

⁵Department of Genetics, Texas Biomedical Research Institute, San Antonio, Texas, United States of America

⁶Department of Psychiatry, University of Connecticut Health Center, Farmington, Connecticut, United States of America

⁷Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, Indiana, United States of America

⁸Division of Child Psychiatry, University of Iowa Hospitals, Iowa City, Iowa, United States of America

⁹Department of Psychiatry, State University of New York Downstate Medical Center, New York, New York, United States of America

¹⁰Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia, United States of America

Abstract

The Trait-based test that uses the Extended Simes procedure (TATES) was developed as a method for conducting multivariate GWAS for correlated phenotypes whose underlying genetic architecture is complex. In this paper, we provide a brief methodological critique of the TATES method using simulated examples and a mathematical proof. Our simulated examples using correlated phenotypes show that more TATES p-values fall outside of the confidence interval

Corresponding authors: faliev@vcu.edu (FA) and jesalvatore@vcu.edu (JES).

¶ Denotes that FA and JES contributed equally as shared first authors.

Compliance with Ethical Standards

Research involving human participants. Not applicable.

Informed consent. Not applicable.

Disclosure of potential conflicts of interest. The authors declare that they have no conflicts of interest.

relative to expectation, and thus the method may result in systematic inflation when used with correlated phenotypes. In a mathematical proof we further demonstrate that the distribution of TATES p-values deviates from expectation in a manner indicative of inflation. Our findings indicate the need for caution when using TATES for multivariate GWAS of correlated phenotypes.

Keywords

multivariate GWAS; TATES

The Trait-Based Association Test that uses the Extended Simes procedure (TATES) was developed in an effort to increase power for multivariate GWAS for phenotypes with complex genetic architectures (van der Sluis et al. 2013). TATES combines p-values across univariate GWAS in order to calculate a single trait-based p-value, while correcting for the correlations among the phenotypes (van der Sluis et al. 2013). As such, it is suggested that TATES provides an efficient and flexible approach for multivariate GWAS for phenotypes whose underlying genetic architecture is either unknown or not likely to conform to a model where genetic variants have a causal effect on the higher-order multivariate phenotype.

Our goal in this paper is to provide a methodological critique of TATES using simulated examples and a mathematical proof. Specifically, we examine the assumption that the TATES method controls for Type I error. For this, we provide simulated examples showing that TATES p-values fall outside of confidence interval more than the expected number of times, thus resulting in inflation of the test results (Type I error), and potentially leading to incorrect conclusions. We further provide a mathematical proof for the simplest two-variable case showing that the distribution of TATES p-values deviates from uniform around 0 when variables are correlated (i.e., if there is no inflation, then around 0 the probability density function of a p-value distribution should be equal to or less than 1). Although there are several critiques regarding limitations of the TATES method relative to other combination methods (e.g., Galesloot et al. 2014; Yang et al. 2016), we note that the observed deviation from the expected “uniform around 0” distribution is novel concern.

Methods

TATES

The TATES test is a modification of the Simes (Simes 1986) and GATES (Li et al. 2011) corrections for multiple testing. The Simes test is a modification of the Bonferroni correction intended to adjust for multiple testing. Assume that p_1, p_2, \dots, p_m are p-values corresponding to test statistics Z_1, Z_2, \dots, Z_m of multiple tests H_1, H_2, \dots, H_m , respectively. It is assumed that the test statistics are continuous. Then, under the null hypothesis the distribution of the p-values is uniform on $[0,1]$. For any given significance level α the test is defined as follows: with $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$ ordered, reject $H_0 = \{H_1, H_2, \dots, H_m\}$ if $p_{(j)} \leq \alpha j/m$ for any $j = 1, 2, \dots, m$ and is based on the inequality:

$$\Pr \left\{ \bigcup_{j=1}^m (p_{(j)} \leq \alpha j/m) \right\} \leq \alpha$$

The resulting Simes p-value is:

$$p_{Simes} = \min_j \left\{ \frac{m}{j} p_{(j)} \right\}.$$

For example, consider three p-values from independent tests for the same null hypothesis: $p_1=0.045$, $p_2=0.046$, $p_3=0.047$. The Bonferroni-corrected p-value needed to reject the null hypothesis for three tests is 0.016 (i.e., $0.05/3$); thus, the null hypothesis is not rejected in this example. In contrast, the Simes test rejects the null hypothesis because

$$p_{(1)} = 0.045, p_{(2)} = 0.046, p_{(3)} = 0.047,$$

$$p_{Simes} = \min \left\{ \frac{3}{1} 0.045, \frac{3}{2} 0.046, \frac{3}{3} 0.047 \right\} = 0.047 < 0.05.$$

TATES is a modification of the Simes test (which requires phenotypes/tests to be independent) for multiple tests corresponding to different phenotypes when the phenotypes are dependent. Assume that given any particular SNP, X_1, \dots, X_m are p-values for m generally dependent phenotypes (Pheno 1, ..., Pheno m). The probability of having at least one true genetic signal among Pheno 1 to Pheno m is estimated with the extended Simes procedure:

$$p_{TATES} = \min_j \{ (m_e / m_{ej}) X_{(j)} \}.$$

Here m_e is the effective number of independent phenotypes, m_{ej} is the number of independent phenotypes among top j phenotypes (after ordering by p-value). To estimate m_e and m_{ej} , the TATES test uses phenotypic information and the argument that p-value correlations and phenotype correlations are related. van der Sluis et al. (2013) used a 6 degree polynomial to approximate the correlations between the p-values and the phenotypes (i.e., the relationship between phenotypic correlation (x) and the p-value correlation (y)), as follows:

$$y = 0.2179x^6 - 0.0219x^5 + 0.1095x^4 + 0.0149x^3 + 0.6226x^2 - 0.0023x - 0.008,$$

$$R^2 = 0.992.$$

When phenotypes are independent, the TATES test is the same as the Simes test. The estimated number of independent phenotypes/p-values among top j phenotypes is defined as:

$$m_{ej} = j - \sum_{i=1}^j (\lambda_i - 1) I(\lambda_i - 1),$$

where I is an indicator function and λ_j is j^{th} eigenvalue of the approximated p-value correlation matrix based on top j phenotypes. We note that this formula corresponds to formula (2) from van der Sluis and et al. (2013) and $m_{em} = m - \sum_{i=1}^m (\lambda_i - 1)I(\lambda_i - 1) = m_e$.

Testing the Distribution of the TATES Statistic

Assuming that p-values come from continuous phenotypes, the method used to calculate TATES p-values should, in theory, produce p-values that are distributed in a way that does not increase Type I error. In the ideal case the distribution will be uniform (Bland 2013; Murdoch et al. 2008). However, even in less than ideal cases, for all “good” statistics, the left side of probability distribution function (pdf) must be ≤ 1 . Otherwise, the results will be inflated, corresponding to how much the pdf > 1 . Since the construction of the TATES test corresponds to a continuous null hypothesis, in this case the p-value distribution should be uniform or at least not exceed 1 around 0. Violation of this assumption (i.e., observing inflation in p-values, as indicated by pdf > 1 around 0) would suggest that the TATES procedure produces an excess of Type I errors, potentially leading to inaccurate conclusions.

To test whether this assumption was met, we conducted simulations in R version 3.1.1 (R Development Core Team 2014) using normally distributed phenotypes. In this paper we provide examples with two and three phenotypes, but the same arguments are true with more than three phenotypes as well. The correlated normal phenotypes were created as linear combinations of independent standard normal distributions. We used two seeds for genotype and phenotype creation and changed coefficients of normal distributions to get different correlations between created phenotypes. The R script for this example, which illustrates inflation in TATES p-values, can be found in Appendix 1. By changing script parameters it is possible to run up to six phenotype examples. Appendix 1 also contains example of three phenotype simulations. For more than six phenotypes, the script can be slightly modified to add more coefficients. For example, to run 8 phenotypes we need to choose `n_pheno=8` and add lines `coeff[7,]=...`, `coeff[8,]=...` after `coeff[6,]=c(0.7,0.9,0.4,0.4,0.1,0.9)` line. Similarly, to add more normal variables we need to change `n_norm` and also number of columns of `coeff[,]` matrix like `coeff[6,]=c(0.7,0.9,0.4,0.4,0.1,0.9,0.4,0.5,0.7)`. The TATES p-values are calculated using the general formulas for any number of phenotypes, as described in van der Sluis et al. (2013).

We repeated each of 10 simulations 100,000 times, where we had 1,000 individuals and 1050 SNPs with minor allele frequency (MAF)=0.5. We ran linear regressions between phenotypes and genotypes and then calculated TATES p-values. These simulations check both the false positive rate and calculate the proportion of times the p-value estimate for the TATES statistic exceeded the expected confidence interval. In 100,000 simulations, we expect that at most 5% of the count of the p-values < 0.05 among 1050 independent SNPs will be > 64 .

Mathematical Proof

In addition to the simulated example, we also include a mathematical proof (Appendix 2) that provides further evidence that when the univariate GWAS p-values are correlated that

the combined TATES p-values violate the distribution assumption around 0. This proof is detailed below under the Results.

Results

We used simulation-based methods to test the critical assumption that the pdf of the TATES p-value distribution is ≤ 1 around 0. In our first simulation example, the TATES statistic was based on normal phenotypes. The number of effective phenotypes ranged between 1.28 and 1.98, and was estimated in R using the formula defined in van der Sluis et al. (2013).

The simulation results are summarized in Table 1. The “false positive ratio column” is calculated based on $1050 \times 100,000$ simulations (i.e. based on all SNPs created in all iterations). This column shows the proportion of TATES p-values ≤ 0.05 among all $1050 \times 100,000$ simulations (TATES p-values). The columns “coefficients of normal variables for phenotype 1(2)” show the linear coefficients of the normal variables involved in creating phenotypes. For example, if we denote normal variables (created in each of 100,000 iterations) z_1, z_2, z_3, z_4 then column value (0.6, 0.7, 0.1, 0.4) for phenotype 1 means that phenotype 1 is defined as $0.6z_1 + 0.7z_2 + 0.1z_3 + 0.4z_4$ for all individuals. The simulated phenotypes and genotypes for the iterations are independent; accordingly, any significant TATES p-value (i.e. TATES p-value ≤ 0.05) is considered as a false positive.

For the confidence interval check, we created 1050 SNPs in each of the iterations and count the number of times the TATES false positive p-values fell outside of the corresponding 95% confidence interval, which in the case of 1050 SNPs is

$1050 \cdot 0.05 + 1.645 \cdot \sqrt{1050 \cdot 0.05 \cdot 0.95} = 64.12$. The last column “out of CI ratio” is the proportion of counts of p-values ≤ 0.05 among 1050 exceeding 64.12 (≥ 65). As the values in this column show, the false positive rate exceeded the expected level of 5%, and increased with increasing correlations between the two phenotypes. This table shows that with highly correlated phenotypes, TATES p-values fall out of the 95% CI up to 18% of the time, which is much more than expected 5%. With small correlations this percent drops to 6–7%. The reason for this is that if the correlation is small then the TATES p-value structure for two phenotypes is

$$\min\{(1 + \text{correlation}) \min(p_1, p_2), \max(p_1, p_2)\}$$

meaning that only one of the p-values (the smallest one) could multiply by $(1 + \text{correlation})$ which is close to one when the correlation is small.

In Table 2 we provide false positive ratios for a three phenotype example. We used the same script as for two phenotype example, changing the number of phenotypes in the script and the corresponding coefficients of three random normal variables. Three columns of the table showing coefficients defined the same way as in two phenotype case. For example (0.6, 0.7, 0.1) for phenotype 1 means that phenotype 1 is defined as $0.6z_1 + 0.7z_2 + 0.1z_3$. Note that in three phenotype case we used linear combinations of three normal variables.

The concerns raised in the simulated example are further evidenced in the mathematical proof where we calculate the exact distribution of the TATES statistic for a two variable example (Appendix 2). In this example we first defined some number d between 0 and 1, then created two variables with uniform distributions. Then we used the first one directly as the p-value for the first phenotype. For the second phenotype's p-value, we used a random combination of two initially created uniform variables (note the first uniform variable directly defined the p-value for phenotype 1). We defined the p-value for phenotype 2 by assigning the first uniform variable with probability d , and the second one with probability $1-d$. In Appendix 2 we prove that the second p-value also has a uniform distribution and, interestingly, the correlation between the two p-value variables is d . The exact pdf of the TATES variable based on p-values of two phenotypes is calculated as

$$f_{X_T}(t) = \frac{2-d^2}{2-d} - \frac{2d(1-d)}{2-d}t.$$

In order to have a uniform distribution, the coefficient of t , i.e., $-2d(1-d)/(2-d)$ of $f_{X_T}(t)$ must be zero. Furthermore, for non-inflated values, we expect that values of the pdf $f_{X_T}(t)$ for t will not exceed 1. However, as t approaches 0, we get $f_{X_T}(0) = (2-d^2)/(2-d) > 1$ ($d > 0$). Again, for the Simes test, which corresponds to the case $d = 0$ (i.e., no correlation between phenotypes), the distribution is correct. This proof shows that the maximum inflation point is $d = 2 - \sqrt{2}$ and $f_{X_T}(0) = 1.1716$, which corresponds to an inflation of approximately 17% for this two variable case. Thus, we see 17% more p-values than expected around 0. The results from this proof thus provide an additional demonstration that when the univariate GWAS phenotypes/p-values are correlated, the combined TATES p-value violates the approximate uniform distribution assumption around zero.

Discussion

TATES was developed as a tool to summarize GWAS results across multiple phenotypes in order to obtain a single p-value, while also accounting for the correlations among the phenotypes (van der Sluis et al. 2013). Notable proposed strengths of the TATES method are that it does not assume that a specific genetic model underlies the multiple phenotypes, and it can identify genetic effects that are either phenotype-specific or common among multiple phenotypes.

To control type I error for continuous phenotypes, a statistic must have a pdf ≤ 1 around values close to 0. We accordingly expected the TATES p-values to have a uniform distribution around 0. However, our examples using simulated data showed that it is possible to get inflated results when calculating TATES combined p-values. This concern was further evidenced in a mathematical proof for the simplest two phenotype scenario. These results call into question the use of TATES to test for association across correlated phenotypes, since the TATES test does not satisfy the theoretical assumption that the statistic must be distributed such that the pdf ≤ 1 around 0. The implication of this finding is that TATES may not successfully summarize GWAS results across correlated phenotypes because it can produce results that are inflated (increasing the risk of erroneously rejecting the null hypothesis; Type I error).

Summary and Conclusions

To summarize, TATES was developed as a tool to accommodate complex genetic architectures when conducting multivariate GWAS for correlated phenotypes. However, we note that in many--and likely most--cases TATES p-values are not uniformly distributed around 0, which violates the assumption of a “good” statistic and indicates that TATES p-values are prone to systematic inflation. Our analyses suggest that caution is warranted when using the TATES method to combine p-values across correlated phenotypes.

Acknowledgments

The Collaborative Study on the Genetics of Alcoholism (COGA), Principal Investigators B. Porjesz, V. Hesselbrock, H. Edenberg, L. Bierut, includes ten different centers: University of Connecticut (V. Hesselbrock); Indiana University (H.J. Edenberg, J. Nurnberger Jr., T. Foroud); University of Iowa (S. Kuperman, J. Kramer); SUNY Downstate (B. Porjesz); Washington University in St. Louis (L. Bierut, A. Goate, J. Rice, K. Bucholz); University of California at San Diego (M. Schuckit); Rutgers University (J. Tischfield); Texas Biomedical Research Institute (L. Almasy), Howard University (R. Taylor) and Virginia Commonwealth University (D. Dick). Other COGA collaborators include: L. Bauer (University of Connecticut); D. Koller, S. O'Connor, L. Wetherill, X. Xuei (Indiana University); Grace Chan (University of Connecticut); S. Kang, N. Manz, (SUNY Downstate); J-C Wang (Washington University in St. Louis); A. Brooks (Rutgers University); and F. Aliev (Virginia Commonwealth University). A. Parsian and M. Reilly are the NIAAA Staff Collaborators.

We continue to be inspired by our memories of Henri Begleiter and Theodore Reich, founding PI and Co-PI of COGA, and also owe a debt of gratitude to other past organizers of COGA, including Ting-Kai Li, currently a consultant with COGA, P. Michael Conneally, Raymond Crowe, and Wendy Reich, for their critical contributions. This national collaborative study is supported by NIH Grant U10AA008401 from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and the National Institute on Drug Abuse (NIDA). We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Funding support for GWAS genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research, was provided by the National Institute on Alcohol Abuse and Alcoholism, the NIH GEI (U01HG004438), and the NIH contract "High throughput genotyping for studying the genetic contributions to human disease" (HHSN268200782096C). This work was also supported by F32AA022269 and K01AA024152 (Salvatore); TUBITAK, Turkey, Grant #114C117 (Aliev); K02AA018755 (Dick); and DA32573 (Agrawal). The authors thank Kim Doheny and Elizabeth Pugh from CIDR and Justin Paschall from the NCBI dbGaP staff for valuable assistance with genotyping and quality control in developing the dataset available at dbGaP. This publication is solely the responsibility of the authors and does not necessarily represent the official view of the funders.

References

- Bland M. Do baseline P-values follow a uniform distribution in randomised trials? *PLOS One*. 2013; 8(10):1–5.
- Galesloot TE, van Steen K, Kiemeny LA, Janss LL, Vermeulen SH. A comparison of multivariate genome-wide association methods. *PLoS One*. 2014; 9(4):e95923. [PubMed: 24763738]
- Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. *Am J Hum Genet*. 2011; 88:283–293. [PubMed: 21397060]
- Murdoch DJ, Tsai Y-L, Adcock J. P-Values are random variables. *Am Stat*. 2008; 62(3):242–245.
- R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: 2014.
- Simes RJ. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*. 1986; 73:751–754.
- van der Sluis S, Posthuma D, Dolan CV. TATES: Efficient multivariate genotype-phenotype analysis for genome-wide association studies. *PLoS Genet*. 2013; 9(1):1–9.

Yang JJ, Li J, Williams LK, Buu A. An efficient genome-wide association test for multivariate phenotypes based on the Fisher combination function. *BMC Bioinformatics*. 2016; 17(1):19. [PubMed: 26729364]

Appendix 1

R scripts showing inflation in TATES.

```
### 1. Two phenotypes four normal variables case
coef1=c(0.8,0.1,0.2,0.1) ## coefficients to create normal phenotype
coef2=c(0.8,0.6,0.1,0.1)
polynom<-function(x) # polynom from TATES paper
{ if (x==1) {1} else {-0.0008-0.0023*x+0.6226*x^2+0.0149*x^3+
                    0.1095*x^4-0.0219*x^5+0.2179*x^6}
}
### setting initial values for simulation
set.seed(236792) ## seed to keep same genotypes
n_iter=100000 ## number of iterations
n_ind=1000 ## number of individuals
n_snps=1050 ## number of SNPs
maf=0.5 ## MAF
prob=0.05 ## tested probability
## one sided CI limit
confidone=n_snps*prob+1.645*sqrt(n_snps*prob*(1-prob))
## Variable Geno is the table to keep genotypes
Geno=matrix(nrow=n_ind,ncol=n_snps)
for (snp in 1:n_snps) ##filling genotype values
{ Geno[,snp]=sample(c(2,1,0),size=n_ind,replace=T,
                  prob=c(maf^2,2*maf*(1-maf),(1-maf)^2))
}
# In each iteration we create two phenotypes y1,y2
# Then run association and keep both p-values
false_positives1=NULL # to keep tates false positives
false_positives2=NULL # to keep p-value based false positives
set.seed(311456711) # put new seed for phenotypes
conf_sum1=0
for (iter in 1:n_iter)
{ tt=rnorm(4*n_ind) # normal variable size; 4 times n_ind
  z1=tt[1:n_ind] # first normal variable
  z2=tt[(n_ind+1):(2*n_ind)] # second normal variable
  z3=tt[(2*n_ind+1):(3*n_ind)] # third normal variable
  z4=tt[(3*n_ind+1):(4*n_ind)] # fourth normal variable
  # creating phenotypes y1,y2 as linear comb. of z's
  y1=coef1[1]*z1+coef1[2]*z2+coef1[3]*z3+coef1[4]*z4
  y2=coef2[1]*z1+coef2[2]*z2+coef2[3]*z3+coef2[4]*z4
```



```

# calculate Tates based on phenotypes
# defining correlation matrixes using polynom from TATES paper
summand_tates=2-abs(polynom(cor(y1,y2)))
false_positives1[iter]=0
false_positives2[iter]=0
x1=NULL; x2=NULL
for (snp in 1:n_snps)
{ # run association of y's with each snp and keep p-value
  # x1, x2 are variables to keep association p-values
  genn=Geno[,snp]
  x1[snp]=summary(lm(y1~genn))$coef[2,4] #p-value of phenotype 1
  x2[snp]=summary(lm(y2~genn))$coef[2,4] #p-value of phenotype 2
  tates_pheno_cor=min(summand_tates*min(x1[snp],x2[snp]),
    max(x1[snp],x2[snp]))
  false_positives1[iter]=false_positives1[iter]+
    (tates_pheno_cor<=prob)
}
conf_sum1=conf_sum1+(false_positives1[iter]>confidone)
print(paste(iter,cor_plp2,cor(y1,y2),
  sum(false_positives1[1:iter])/(iter*n_snps), conf_sum1/iter))
flush.console() ##to print to the screen
}
cor(y1,y2)
cor_plp2
### 2. Three phenotypes three normal variables case
n_pheno=3
n_norm=3
set.seed(231456799) ## seed to create genotypes
n_iter=100000      ## number of iterations
n_ind= 1000       ## number of individuals
n_snp=1 # or 1050 ## number of snps
maf=0.5          ## MAF
prob=0.05        ## tested probability
## matrix of random coefficients (must have n_pheno rows, and
## at least n_norm columns)
coeff=matrix(0,nrow=max(n_pheno,6),ncol=max(n_norm,6))
## coefficients of dependent normal variables to create
## phenotypes as a linear combinations
coeff[1,]=c(0.2,0.4,0.2,0.0,0.0,0.0)
coeff[2,]=c(0.2,0.1,0.6,0.0,0.0,0.0)
coeff[3,]=c(0.2,0.7,0.6,0.0,0.0,0.0)
coeff[4,]=c(0.2,0.5,0.6,0.1,0.2,0.3)
coeff[5,]=c(0.1,0.4,0.4,0.4,0.1,0.9)
coeff[6,]=c(0.7,0.9,0.4,0.4,0.1,0.9)

```

```

polynom<-function(x)
{ # polynom from TATES paper
  if (x==1)
  { retpol=1
  } else
  { retpol=-0.0008-0.0023*x+0.6226*x^2+0.0149*x^3+
    0.1095*x^4-0.0219*x^5+0.2179*x^6
  }
  retpol
}
## setting initial values for simulation
## table to keep genotypes
Geno=matrix(nrow=n_ind,ncol=n_snp)
for (snp in 1:n_snp)
{ Geno[,snp]=sample(c(2,1,0),size=n_ind,replace=T,
  prob=c(maf^2,2*maf*(1-maf),(1-maf)^2))
}
## In each iteration we create phenotypes x1,x2,x3(x4,x5,x6),
## run association test and keep all p-values
false_positives=NULL ## to keep tates false positives
s=0
set.seed(31456711) ## 3145671 put new seed for phenotypes
for (iter in 1:n_iter)
{ z=matrix(nrow=n_norm,ncol=n_ind)
  for (i in 1:n_norm)
  { z[i,]=rnorm(n_ind) ## creates st. normal
  }
  # creating "y" phenotypes as linear combination of z's
  y=matrix(nrow=n_pheno,ncol=n_ind)
  yord=matrix(nrow=n_pheno,ncol=n_ind)
  for (i in 1:n_pheno)
  { y[i,]=coeff[i,1]*z[1,]
    for (j in 2:n_norm)
    { y[i,]=y[i,]+coeff[i,j]*z[j,]
    }
  }
  ## running association of y's with each snp and keeping p-value
  ## x1,x2,x3,(x4,x5,x6) are variables to keep assoc. p-values
  x=matrix(nrow=n_pheno,ncol=n_snp)
  false_positives[iter]=0
  for (snp in 1:n_snp)
  { genn=Geno[,snp]
    for (i in 1:n_pheno)
    { x[i,snp]=1

```

```

    x[i,snp]=summary(lm(y[i,]~genn))$coef[2,4] #p of pheno i
  }
  sort1=order(x[,snp])
  xord=x[sort1]      ## sort p-values
  for (i in 1:n_pheno) ## sort phenotypes with sorted p-values
  { yord[i,]=y[sort1[i],]
  }
  ## calculate TATES p-value
  indep=rep(1,n_pheno)
  bind0=yord[1,]
  for (i in 2:n_pheno) ## sort phenotypes with sorted p-values
  { bind0=cbind(bind0,yord[i,])
    cor_mat=apply(cor(bind0),c(1,2),polynom)
    e=eigen(cor_mat)$values
    indep[i]=i      ## find number of indep among top i
    for (m in 1:i)
    { indep[i]=indep[i]-max(e[m]-1,0)
    }
  }
  TATES=xord[1]*indep[n_pheno]
  for (m in 2:n_pheno)
  { TATES=min(TATES,xord[m]*indep[n_pheno]/indep[m])
  }
  false_positives[iter]=false_positives[iter]+(TATES<=prob)
}
if ((iter %% 1000)==0) ##output after every 1000
{ print(paste(iter,sum(false_positives[1:iter])/(iter*n_snp)))
  flush.console()
}
}
print (c(cor(y[1,],y[2,]),cor(y[1,],y[3,]),cor(y[2,],y[3,])))
flush.console()

```

Appendix 2

Mathematical proof for a two variable case demonstrating the inflation of TATES p-values.

In the case of two phenotypes TATES statistics gets the form

$$\begin{aligned}
 X_{TATES} &= \min\{(m_e/m_{e1})\min(X_1, X_2), (m_e/m_{e2})\max(X_1, X_2)\} = \\
 &= \min\{(m_e/1)\min(X_1, X_2), (m_e/m_e)\max(X_1, X_2)\} = \min\{m_e \min(X_1, X_2), \max(X_1, X_2)\}
 \end{aligned}$$

We have the only one coefficient $a = m_e$ in the last formula which is the effective number of p-values among two phenotypes. TATES method is effective only if $a < 2$, otherwise, $a = 2$ is the same as the Simes method, so $X_{TATES} = \min\{a \min(X_1, X_2), \max(X_1, X_2)\}$.

Example (shows that there are uniform variables such that created TATES statistics is not uniform):

Let Z_1, Z_2 be uniform $[0,1]$, fix any d between 0 and 1 and define the binomial variable

$$D = \begin{cases} 1, & d \\ 0, & 1-d \end{cases} \text{ and assume } Z_1, Z_2 \text{ and } D \text{ are independent.}$$

Define $X_1 = Z_1, X_2 = DZ_1 + (1 - D)Z_2$.

Both X_1, X_2 have $U[0,1]$ distribution because

$$\begin{aligned} \Pr\{X_2 < t\} &= \Pr\{X_2 < t, D = 1\} + \Pr\{X_2 < t, D = 0\} = \Pr\{DZ_1 + (1 - D)Z_2 < t, D = 1\} + \Pr\{DZ_1 + (1 - D)Z_2 < t, D = 0\} \\ &= \Pr\{Z_1 < t, D = 1\} + \Pr\{Z_2 < t, D = 0\} = \\ &= \Pr\{Z_1 < t\}\Pr\{D = 1\} + \Pr\{Z_2 < t\}\Pr\{D = 0\} = dt + (1 - d)t = t \end{aligned}$$

Let's calculate correlation between X_1, X_2 . As both variables are uniform $[0,1]$ we get

$$EX_1 = EX_2 = 0.5, \quad \text{var}(X_1) = \text{var}(X_2) = 1/12$$

$$E(X_1X_2) = dE(X_1X_2|D = 1) + (1 - d)E(X_1X_2|D = 0) = d\text{var}(Z_1) + 0.25 = d/12 + 0.25$$

$$\text{cor}(X_1, X_2) = \frac{E(X_1X_2) - EX_1EX_2}{1/12} = \frac{d/12}{1/12} = d$$

Distribution function of TATES statistics $X_{TATES} = \min\{a \min(X_1, X_2), \max(X_1, X_2)\}$ in this case is

$$\begin{aligned}
 F_{X_T}(t) &= \Pr\{X_T < t\} = \Pr\{\min[a \min(X_1, X_2), \max(X_1, X_2)] < t\} = \\
 &= 1 - \Pr\{\min[a \min(X_1, X_2), \max(X_1, X_2)] \geq t\} = 1 - \Pr\{\min(X_1, X_2) \geq (t/a) \cap \max(X_1, X_2) \geq t\} = \\
 &= 1 - \Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a) \cap \max(X_1, X_2) \geq t\} = 1 - [\Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a)\} - \Pr\{X_1 \geq (t/a) \\
 &\cap X_2 \geq (t/a) \cap \max(X_1, X_2) < t\}] = 1 - [\Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a)\} - \Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a) \cap X_1 < t \\
 &\cap X_2 < t\}] = 1 - [\Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a)\} - \Pr\{(t/a) \leq X_1 < t \cap (t/a) \leq X_2 < t\}] = \\
 &= 1 - [\Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a) \cap D = 0\} + \Pr\{X_1 \geq (t/a) \cap X_2 \geq (t/a) \cap D = 1\} - \\
 &- \Pr\{(t/a) \leq X_1 < t \cap (t/a) \leq X_2 < t \cap D = 0\} - \Pr\{(t/a) \leq X_1 < t \cap (t/a) \leq X_2 < t \cap D = 1\}] = \\
 &= 1 - [(1-d)\Pr\{Z_1 \geq (t/a) \cap Z_2 \geq (t/a)\} + d \Pr\{Z_1 \geq (t/a) \cap Z_1 \geq (t/a)\} - \\
 &- (1-d)\Pr\{(t/a) \leq Z_1 < t \cap (t/a) \leq Z_2 < t\} - \Pr\{(t/a) \leq Z_1 < t \cap (t/a) \leq Z_1 < t\}] = \\
 &= 1 - [(1-d)(1-t/a)^2 + d(1-t/a) - (1-d)(t-t/a)^2 - d(t-t/a)] = \\
 &= 1 - d(1-t) - \frac{1-d}{a^2}[(a-t)^2 - (at-t)^2]
 \end{aligned}$$

Derivative of the above is the pdf of Tates statistic

$$f_{X_T}(t) = d + (2/a)(1-d) + 2t(1-d)(1-(2/a)).$$

To have uniform distribution the coefficient of t of $f_{X_T}(t)$ must be zero i.e.,

$$2(1-d)(1-2/a) = 0.$$

It means $d=0$ or $a=2$. In all other cases the test inflates or deflates results. But $d=0$ corresponds to the case $X_2 = Z_2$ which means X_1, X_2 are independent, which is equivalent to the Simes procedure. When $a=2$, this also corresponds to Simes case. Thus, in all other choices of a the statistic inflates or deflates results $f_{X_T}(0) = d + (2/a)(1-d)$ times.

Now let's find eigenvalues and exact value of $a = m_e$ for this example

For two variables X_1, X_2 with $corr(X_1, X_2) = d$ the correlation matrix has the form

$$A = \begin{bmatrix} 1 & d \\ d & 1 \end{bmatrix}, \quad \det(A - \lambda I) = 0, \quad \det \begin{bmatrix} 1-\lambda & d \\ d & 1-\lambda \end{bmatrix} = 0,$$

$$(1-\lambda)^2 - d^2 = 0 \text{ and eigenvalues are } \lambda_1 = 1+d, \lambda_2 = 1-d.$$

Coefficient a used in TATES simulated statistic $X_{TATES} = \min\{a \min(X_1, X_2), \max(X_1, X_2)\}$ is

$$a = m_e = m - \sum_{\lambda_i > 1} (\lambda_i - 1) = 2 - (\lambda_1 - 1) = 2 - (1 + d - 1) = 2 - d.$$

This means $f_{X_T}(0) = d + (2/a)(1 - d) = (2 - d^2)/(2 - d)$. Maximum inflation point is $d = 2 - \sqrt{2}$ and gives an inflation of approximately 17% for this two variable case.

Table 1
 Estimation of Percent of TATES p-values Falling out of %95 CI for Two Phenotypes

Simulation #	Coefficients of Normal Variables for Pheno 1	Coefficients of Normal Variables for Pheno 2	Correlation Between Phenotypes	False Positive Ratio	Out of CI Ratio
1	(0.6,0.7,0.1,0.4)	(0.7,0.6,0.1,0.8)	0.9459	0.0551	0.1848
2	(0.1,0.1,0.6,0.4)	(0.1,0.0,0.6,0.8)	0.9343	0.0553	0.1906
3	(0.6,0.1,0.1,0.0)	(0.6,0.3,0.6,0.0)	0.8111	0.0545	0.1624
4	(0.0,0.2,0.1,0.6)	(0.6,0.3,0.4,0.8)	0.8102	0.0545	0.1611
5	(0.8,0.1,0.6,0.4)	(0.2,0.3,0.6,0.8)	0.7566	0.0540	0.1422
6	(0.8,0.1,0.6,0.1)	(0.3,0.6,0.1,0.8)	0.4154	0.0513	0.0708
7	(0.8,0.1,0.6,0.1)	(0.11,0.6,0.1,0.8)	0.2821	0.0506	0.0578
8	(0.8,0.1,0.6,0.1)	(0.8,0.6,0.1,0.8)	0.6475	0.0530	0.1112
9	(0.8,0.1,0.6,0.1)	(0.8,0.6,0.1,1.49)	0.5008	0.0518	0.0823
10	(0.8,0.1,0.2,0.1)	(0.8,0.6,0.1,0.1)	0.8639	0.0550	0.1804

Note. The four numbers in the coefficients columns correspond to coefficients of the standard normal variables used to create phenotypes

Table 2

Estimation of False positive of TATES p-values for Three Phenotypes

Simulation #	Coefficients of Normal Variables for Pheno 1	Coefficients of Normal Variables for Pheno 2	Coefficients of Normal Variables for Pheno 3	Correlations Between Phenos (1,2) (1,3) (2,3)	False Positive Ratio	Out of CI Ratio
1	(0.2,0.4,0.2)	(0.2,0.1,0.6)	(0.2,0.7,0.6)	0.63, 0.94, 0.78	0.0528	0.1074
2	(0.2,0.4,0.2)	(0.2,0.1,0.1)	(0.2,0.7,0.6)	0.83, 0.94, 0.73	0.0532	0.1188
3	(0.2,0.4,0.2)	(0.2,0.1,0.1)	(0.1,0.1,0.6)	0.83, 0.63, 0.69	0.0536	0.1310
4	(0.2,0.8,0.6)	(0.2,0.1,0.1)	(0.2,0.8,0.6)	0.83, 0.95, 0.72	0.0529	0.1102
5	(0.2,0.4,0.2)	(0.2,0.1,0.2)	(0.2,0.8,0.6)	0.81, 0.95, 0.78	0.0540	0.1438
6	(0.2,0.4,0.2)	(0.1,0.8,0.2)	(0.1,0.2,0.6)	0.93, 0.68, 0.53	0.0532	0.1180
7	(0.3,0.3,0.5)	(0.1,0.8,0.2)	(0.1,0.2,0.6)	0.67, 0.92, 0.53	0.0531	0.1159
8	(0.3,0.3,0.5)	(0.3,0.5,0.2)	(0.5,0.2,0.6)	0.83, 0.96, 0.74	0.0528	0.1074
9	(0.3,0.3,0.5)	(0.3,0.5,0.4)	(0.5,0.2,0.1)	0.94, 0.71, 0.75	0.0533	0.1217
10	(0.6,0.6,0.5)	(0.3,0.5,0.4)	(0.5,0.2,0.1)	0.99, 0.84, 0.87	0.0521	0.0893

Note. The three numbers in the coefficients columns correspond to coefficients of three standard normal variables used to create phenotypes