genetic association studies. However, one needs to be cautious when using this statistic. On the basis of simulation results, we found that the analytical distributions of the *TS* and *TSM* statistics are influenced both by the MAF and by genetic models used in association tests. We suggest using the empirical p value, rather than the exact p value, in real situations. A more generalized statistic that does not depend on HWE-test significance in cases should be developed for the incorporation of HWE information and improvement of the power of genetic association studies.

Shizhong Han,[1,4] Li Ma,[2,4] Dawei Li,[1]
and Bao-Zhu Yang[1,3,*]
[1]Division of Human Genetics, Department of Psychiatry, School of Medicine, Yale University, New Haven, CT 06511, USA; [2]Department of Animal Science, University of Minnesota, St. Paul, MN 55108, USA; [3]Veteran's Administration Connecticut Healthcare Center, West Haven, CT 06516, USA
[4]These two authors contributed equally
*Correspondence: bao-zhu.yang@yale.edu

Supplemental Data

Supplemental Data include one figure and can be found with this article online at http://www.ajhg.org/.

Acknowledgments

This work was supported by the U.S. National Institutes of Health (career development award K01 DA024758 to B.Z.Y.). We are grateful for Joel Gelernter's valuable discussion and comments on this letter.

References

1. Feder, J.N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D.A., Basava, A., Dormishian, F., Domingo, R. Jr., Ellis, M.C., Fullan, A., et al. (1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat. Genet. *13*, 399–408.
2. Nielsen, D.M., Ehm, M.G., and Weir, B.S. (1998). Detecting marker-disease association by testing for HW disequilibrium at a marker locus. Am. J. Hum. Genet. *63*, 1531–1540.
3. Czika, W., and Weir, B.S. (2004). Properties of the multiallelic trend test. Biometrics *60*, 69–74.
4. Wittke-Thompson, J.K., Pluzhnikov, A., and Cox, N.J. (2005). Rational inferences about departures from HW equilibrium. Am. J. Hum. Genet. *76*, 967–986.
5. Wang, J., and Shete, S. (2008). A test for genetic association that incorporates information about deviation from Hardy-Weinberg proportions in cases. Am. J. Hum. Genet. *83*, 53–63.
6. Taylor, J., and Tibshirani, R. (2006). A tail strength measure for assessing the overall univariate significance in a dataset. Biostatistics *7*, 167–181.
7. Rohlfs, R.V., and Weir, B.S. (2008). Distributions of Hardy-Weinberg Equilibrium Test Statistics. Genetics *180*, 1609–1616.
8. Hung, H.M., O'Neill, R.T., Bauer, P., and Köhne, K. (1997). The behavior of the P-value when the alternative hypothesis is true. Biometrics *53*, 11–22.

DOI 10.1016/j.ajhg.2009.01.014.

Response to Zang et al. and Han et al.

To the Editor: In July 2008, we proposed a powerful test for the study of genetic association that incorporates information about deviation from Hardy-Weinberg proportions (HWP) in cases.[1] Two approaches were proposed: the mean-based tail-strength (*TS*) measure and the median-based tail-strength (*TSM*) measure. These measures combined p values from the likelihood ratio test (LRT) for association and the exact test for HWP. For both measures, we derived exact formulas to compute p values, and we also provided an approach for obtaining empirical p values with the use of a resampling procedure. The results showed a significant increase in power when using the proposed approaches. The type I errors were also well controlled with the additive model.

In their letter, Zang et al. report that when the underlying genetic model is not additive (recessive or dominant), there is a significant correlation between p values obtained from the LRT and the HWP test. Furthermore, they show that this correlation could lead to excessive false-positive probabilities if one uses the asymptotic formulas provided in our paper.

We agree that under certain situations the correlation between the two p values might not be ignored. However, in our original paper, we discussed limitations of the asymptotic null distributions of *TS* and *TSM*. We stated that "although the exact p values of *TS* and *TSM* are simple and straightforward to compute and interpret, the deviations of underlying assumptions might make the exact p values based on explicit formulas too conservative or too liberal." We therefore proposed an alternative approach for estimating empirical p values of *TS* and *TSM* with the use of a permutation procedure. For this permutation procedure, we resampled the SNP values by using the genotype frequencies calculated from the allele frequencies for both cases and controls. When the permutation procedure is applied, even if the assumptions underlying derivation of asymptotic null distribution are violated, one can still obtain accurate p values.

Tables 1 and 2 in Zang et al.'s letter show that the type I errors of the *TS* and *TSM* measures were inflated for the

Table 1. Estimated Type I Error Probability at 0.05 and 0.01 Significance Levels for Recessive and Dominant Models with Exact Analytical Formulas Versus Resampling Approach

		Type I Error Probability							
		Exact *TS*		Exact *TSM*		Empirical *TS*		Empirical *TSM*	
MAFs	Models	0.05	0.01	0.05	0.01	0.05	0.01	0.05	0.01
0.1	recessive	0.0790	0.0280	0.0780	0.0290	0.0470	0.0090	0.0450	0.0090
	dominant	0.0540	0.0100	0.0510	0.0090	0.0210	0.0030	0.0220	0.0040
0.3	recessive	0.0730	0.0310	0.0730	0.0310	0.0530	0.0120	0.0540	0.0120
	dominant	0.0610	0.0150	0.0600	0.0150	0.0350	0.0060	0.0350	0.0060
0.5	recessive	0.0520	0.0190	0.0530	0.0170	0.0430	0.0100	0.0430	0.0100
	dominant	0.0520	0.0120	0.0490	0.0120	0.0390	0.0050	0.0360	0.0050

recessive and dominant models when the asymptotic null distribution was applied. We would like to point out that even though the type I errors would be inflated with the analytical formulas we provided, the use of simulation-based computation of p values should always maintain good control over the type I error, regardless of the genetic model or minor-allele frequency (MAF) used. As a proof of principle, we performed simulations with the parameters used by Zang et al. in their Tables 1 and 2. We considered recessive and dominant models for the disease simulations. The MAFs were set at 10%, 30%, and 50%. With the simulation procedure proposed in our paper, for each scenario, we generated 1000 replicates under the null hypothesis of no association between the SNP and the disease, each with 500 cases and 500 controls. Table 1 reports the observed type I error probabilities at the defined significances of 0.05 and 0.01 with both exact formulas and the permutation procedure. The *TS* and *TSM* type I errors found with the use of asymptotic null distributions are close to those reported by Zang et al. But all the type I errors of *TS* and *TSM* found with the use of the empirical permutation procedure are well under control when either the dominant or recessive model is assumed.

Also, Zang et al. extended our work to allow for correlation between the p values, denoted by *TSC*, and provided the asymptotic distribution of the new *TSC* measure. The rationale they provide for this is that simulation-based approaches to determine p values have limited applications in genome-wide association studies. Although we agree with this rationale, we would like to point out certain limitations of *TSC*. First, because computation of correlation between the LRT and the exact test for HWP that we used is difficult, they used the trend test instead of the LRT and the chi-square test instead of the exact test for HWP. The chi-square test for HWP may not obtain an accurate p value, even in relatively large samples, especially when the MAF is small.[2] Second, they assumed that the joint distribution of the HWP test and the trend test is a bivariate normal distribution. The normality assumption is widely used for multivariate distributions; however, its validity depends on various underlying regularity conditions. Therefore, these underlying assumptions could affect the performance of the *TSC* measure.

Importantly, the *TSC* measure proposed by Zang et al. is limited to the trend test. They also show (in Appendix B) that the MAX3 test is more powerful than the *TSC* measure in most situations, thereby limiting the utility of *TSC*. It should be noted that our *TS* and *TSM* measures are more flexible. In our paper, we used LRT to perform the association test by using cases and controls and combined the resultant p value with the p value from the HWP test. We noted that other statistical tests could be used in place of LRT in our methodology. Therefore, our approach allows for combining p values from the MAX3 test with those from HWP test in order to develop an even more powerful association test.

In Han et al.'s letter, they also claim that the asymptotic null distributions of *TS* and *TSM* were influenced by the different genetic models and MAFs (causing correlations between two p values) and assessing HWP by using exact test (causing nonuniform p values of the HWP test) and, therefore, suggest using empirical p values in real situations. As mentioned in our response to Zang et al., we agree that the empirical p values are more appropriate than the exact p values in many situations and suggest that investigators use the permutation procedure proposed in our original paper. The skewness property of the p values for the HWP test that Han et al. mention can also be found in Figure 1B in our original paper. We thank them for providing further insight into the coarse nature of the p values from the exact HWP test.

Our aim in the original paper was to show that if significant departure from HWP in cases is observed, that information can be used to obtain higher significance for the genetic association test. Therefore, our simulations were based on replicates having departure from HWP. It is important to note that if the cases do not deviate from HWP, it is not appropriate to perform the test we have proposed. This could be the reason that Han et al. did not replicate our results reported in Table 4 in the original paper. Looking back at the original paper, it seems we did not make this point clear. We are glad to have this opportunity to clarify this point.

In conclusion, we have proposed an approach to combine p values obtained from genetic association tests (e.g., LRT) with p values obtained from the HWP test. We further emphasize that our approach is not limited to LRT, and any test for case-control association (trend, MAX3) could be used in place of LRT. The correlation between p values will depend on several factors, including underlying genetic models, minor-allele frequency, and choice of genetic tests (LRT, trend, MAX3, etc.). Analytical formulas to compute such correlation will be of limited value because of further assumptions involved in their computation and/or the use of inferior tests. Therefore, we recommend using resampling-based methods to assess the significance of our proposed tests. With the advancement of computational power, such resampling approaches are feasible for at least those SNPs found significant in the discovery phase of two-stage genome-wide association studies.

Jian Wang[1] and Sanjay Shete[1,*]
[1]Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA
*Correspondence: sshete@mdanderson.org

References

1. Wang, J., and Shete, S. (2008). A test for genetic association that incorporates information about deviation from Hardy-Weinberg proportions in cases. Am. J. Hum. Genet. *83*, 53–63.
2. Wigginton, J.E., Cutler, D.J., and Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. Am. J. Hum. Genet. *76*, 887–893.

DOI 10.1016/j.ajhg.2009.01.016.