A Generalized Robust Allele-based Genetic Association Test

Lin Zhang[1] and Lei Sun[1,2]

[1]Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto, ON M5G 1Z5, Canada, linzhang@utstat.toronto.edu

[2]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON M5T 3M7, Canada, sun@utstat.toronto.edu

March 3, 2021

Abstract

The allele-based association test, comparing allele frequency difference between case and control groups, is locally most powerful. However, application of the classical allelic test is limited in practice, because the method is sensitive to the Hardy–Weinberg equilibrium (HWE) assumption, not applicable to continuous traits, and not easy to account for covariate effect or sample correlation. To develop a generalized robust allelic test, we propose a new allele-based regression model with individual allele as the response variable. We show that the score test statistic derived from this robust and unifying regression framework contains a correction factor that explicitly adjusts for potential departure from HWE, and encompasses the classical allelic test as a special case. When the trait of interest is continuous, the corresponding allelic test evaluates a weighted difference between individual-level allele frequency estimate and sample estimate where the weight is proportional to an individual's trait value, and the test remains valid under Y-dependent sampling. Finally, the proposed allele-based method can analyze multiple (continuous or binary) phenotypes simultaneously and multi-allelic genetic markers, while accounting for covariate effect, sample correlation and population heterogeneity. To support our analytical findings, we provide empirical evidence from both simulation and application studies.

Keywords: Allele-based association analysis; Correlation; Dependent sample; Hardy–Weinberg equilibrium; Multiple phenotypes; Multiple populations; Robustness.

1 Introduction

An association study aims to identify genetic markers that influence a heritable trait or phenotype of interest, while accounting for environmental effect. To formulate the problem more precisely, let single nucleotide polymorphisms (SNPs) be the genetic markers available for analysis. For each bi-allelic SNP, let a and A be the two possible alleles, and as in convention, let A denote the minor allele with population frequency $p \leq 0.5$. The SNP genotype consists of a paired (but unordered) alleles, aa, Aa or AA. For a case-control association study of a binary trait (Table 1), intuitively one can compare the estimates of p between the case and control groups. Indeed, the resulting allelic test is locally most powerful (Sasieni, 1997). However, the validity of the test hinges on the assumption of Hardy–Weinberg equilibrium (HWE), which assumes that genotype frequencies depend only on allele frequencies. That is, $p_{aa} = (1-p)^2$, $p_{Aa} = 2p(1-p)$ and $p_{AA} = p^2$.

Table 1: **Notations for genotype and allele counts for a case-control study**. The HLA-DQ3 example is from Sasieni (1997), studying women with cervical intraepithelial neoplasia 3. The sample estimates of allele frequency of allele A in the case, control and combined groups are denoted, respectively, as $\hat{p}_r = (2r_2 + r_1)/2r = r_A/2r$, $\hat{p}_s = (2s_2 + s_1)/2s = s_A/2s$ and $\hat{p} = (2n_2 + n_1)/2n = n_A/2n$.

	Genotype Counts				Allele Counts		
	aa	Aa	AA	Total	a	A	Total
Case	r_0	r_1	r_2	r	$2r_0 + r_1$	$r_1 + 2r_2$	$2r$
Control	s_0	s_1	s_2	s	$2s_0 + s_1$	$s_1 + 2s_2$	$2s$
	n_{aa}	n_{Aa}	n_{AA}		n_a	n_A	
Total	n_0	n_1	n_2	n	$2n_0 + n_1$	$n_1 + 2n_2$	$2n$
The HLA-DQ3 example from Sasieni (1997)							
Case	40	45	28	113	125	101	226
Control	273	100	43	416	646	186	832
Total	313	145	71	529	771	287	1058

To evaluate the HWE assumption in an independent sample, one typically applies the Pearson

goodness-of-fit χ^2 test, and under the H_0 of HWE,

$$T_{\text{HWE, Pearson}} = \frac{(n_0 - n(1-\hat{p})^2)^2}{n(1-\hat{p})^2} + \frac{(n_1 - n2\hat{p}(1-\hat{p}))^2}{n2\hat{p}(1-\hat{p})} + \frac{(n_2 - n\hat{p}^2)^2}{n\hat{p}^2} \overset{H_0}{\sim} \chi^2_1. \tag{1}$$

Using the HLA-DQ3 data in Table 1 as an illustration, application of $T_{\text{HWE, Pearson}}$ yields a test statistic of 49.7623 and a p-value of 1.74×10^{-12}, suggesting departure from HWE.

Prior to conducting a case-control genome-wide association study (GWAS), testing HWE in the control sample and excluding SNPs in Hardy–Weinberg disequilibrium (HWD) is a routine part of data quality control; severe HWD is an indication of genotyping error. However, the need for a robust allelic association test remains. First, the p-value threshold used for HWE-based screening in practice is subjective and depends on sample size, often ranging between 10^{-8} and 10^{-12}. For example, Bycroft et al. (2018) recommended 10^{-12} for each batch of $\sim$4,000 samples. With this stringent criterion, it is possible that a SNP that passes the quality control step is truly in HWD at the population level, but not attributed to genotyping error. The corresponding allelic test is then biased, and we provide an application example in Section 5.2. Second, developing a robust and generalized allelic test is analytically important, so that this locally most powerful test (Sasieni, 1997) can be applied to, for example, a continuous trait or a sample of related individuals.

To robustify the classical allelic test against HWD, early work focused on improving variance estimation (Schaid and Jacobsen, 1999), but application is still limited to case-control studies and in simple settings (e.g. independent samples). Thus, most genetic association studies rely on the Y-on-G regression approach, where G is coded additively as 0, 1 and 2 for aa, Aa and AA. The resulting genotype-based additive test is known to be robust to HWD, but the exact HWD-correction mechanism is not clear. Further, data collection in practice typically first samples individuals based on Y, which can be random or Y-dependent sampling (Derkach et al., 2015), and then collects G for the sampled individuals. Thus, it can be argued that G-on-Y is a more fitting statistical framework.

This 'reverse' regression approach can also readily analyze multiple phenotypes simultane-

ously, which was the motivation for MultiPhen (O'Reilly et al., 2012). To deal with the three genotype groups, O'Reilly et al. (2012) used an ordinal logistic regression and stated that the proposed likelihood ratio test does not assume HWE. However, the statistical insight is lacking and analyzing pedigree data remains a challenge. Zhang et al. (2014) provided more background on using 'reverse' regression to simultaneously analyze multiple traits, and also proposed and compared a series of genotype-based GEE association tests.

This work generalizes the locally most powerful allelic association test to more complex settings by developing a novel *allele-based* 'reverse' regression model. Section 2 revisits the classical allelic association test and provides insight on the need for a more flexible formulation of the test. Section 3 develops the new allele-based 'reverse' regression model by first appropriately partitioning the two alleles of a genotype and then specifying the individual allele as the response variable. In addition to the association parameter, the proposed regression includes a parameter that models the dependency between the two alleles of a genotype, explicitly accounting for potential departure from HWE. Section 4 considers more complex settings including related individuals from pedigree data, genetic markers with more than two alleles, and multiple phenotypes or populations. Given the theoretical results presented, simulation experiments in Section 5 are relatively brief with additional empirical evidence from two applications. Section 6 concludes with remarks and discussion.

2 The classical allelic test revisited

For a given SNP and a binary phenotype of interest, let p_r and p_s denote the allele frequencies, respectively, for the case and control populations (Table 1). The classical allele-based association test, testing H_0: $p_r = p_s$, is based on

$$T_{\text{allelic}} = \frac{(\hat{p}_r - \hat{p}_s)^2}{(\frac{1}{2r} + \frac{1}{2s})\hat{p}(1-\hat{p})} \overset{\text{HWE}, H_0}{\sim} \chi^2_1. \tag{2}$$

The validity of T_{allelic}, however, requires the Hardy–Weinberg equilibrium assumption, because only under HWE, $\widehat{\text{var}}(\hat{p}_r - \hat{p}_s) = (\frac{1}{2r} + \frac{1}{2s})\hat{p}(1-\hat{p})$. Using the HLA-DQ3 data in Table 1 as an example, because the HWE test is significant (Section 1), a direct application of the classical allelic association test in this case is not appropriate.

Indeed, Sasieni (1997) has pointed out that T_{allelic} is valid and locally most powerful if and only if the HWE assumption holds and the genetic effect is additive. In the presence of Hardy–Weinberg disequilibrium, T_{allelic} is known to be anti-conservative. But, we emphasize that this is true only if there is an excess of homozygotes AA, i.e. $\delta = p_{AA} - p^2 > 0$, where δ is a measure of HWD (Weir, 1996). If $\delta < 0$, T_{allelic} is conservative as shown in Figure 1.

To robustify T_{allelic} against HWD, Schaid and Jacobsen (1999) used a multinomial distribution for the genotype counts to improve variance estimate. The resulting test is

$$T_{\text{allelic, Schaid}} = \frac{(\hat{p}_r - \hat{p}_s)^2}{(\frac{1}{2r} + \frac{1}{2s})(\hat{p}(1-\hat{p}) + \hat{\delta})} \overset{H_0}{\sim} \chi^2_1, \tag{3}$$

where $\hat{\delta} = (\hat{p}_{AA} - \hat{p}^2)$ is the sample estimate of the HWD measure. Section 3 provides analytical insight about how $T_{\text{HWE, Pearson}}$ in (1) is related to δ. Here, the effect of $\delta \neq 0$ on the accuracy of T_{allelic} is clear. In the HLA-DQ3 example, $\hat{\delta} = 0.061$ and is highly significant, so the classical allelic test will be too optimistic, $\{T_{\text{allelic}} = 44.8470\} > \{T_{\text{allelic, Schaid}} = 34.3207\}$.

$T_{\text{allelic, Schaid}}$ is effective in making T_{allelic}, the classical allele-based association test, robust to HWD, but its direct application is limited to the simplest setting of case-control studies using independent observations without covariates. It is also not clear how to adapt this comparing-two-proportions analytical framework to, for example, analyzing a continuous trait. Thus, a new formulation of allele-based association test is needed.

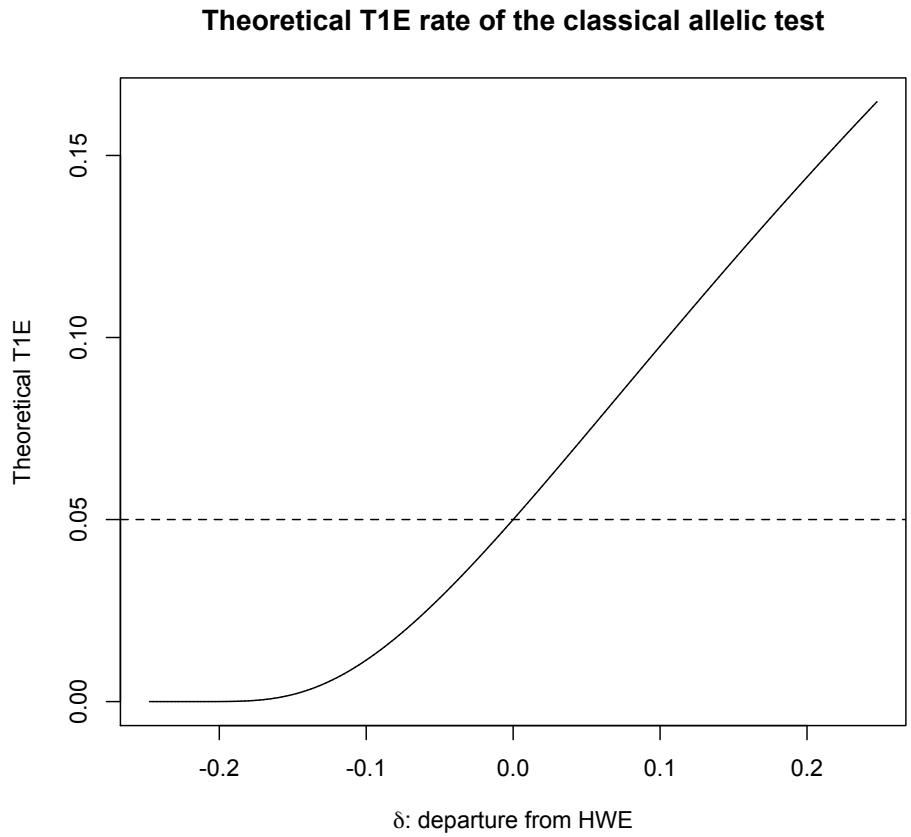

Figure 1: **The theoretical type 1 error rate of the classical allelic test, $T_{\textbf{allelic}}$.** The nominal level is $\alpha = 0.05$, departure from HWE is measured by $\delta = p_{AA} - p^2$, where p is the frequency of the minor allele A and $-p^2 \leq \delta \leq p(1-p)$. When $p = 0.5$, $-0.25 \leq \delta \leq 0.25$.

3 A Generalized Robust Allele-based (RA) Association Test

3.1 Decoupling the two alleles in a genotype

Consider a bi-allelic SNP with genotype $G \in \{aa, Aa, AA\}$, and for the moment assume that there are n independent observations, $G_i, i = 1, \ldots, n$. We partition each G_i as follows,

$$(G_{i1}, G_{i2}) = \begin{cases} (0,0) & \text{if the genotype is } aa \\ (0,1) & \text{if the genotype is } Aa \text{ and } c_i = 0 \\ (1,0) & \text{if the genotype is } Aa \text{ and } c_i = 1 \\ (1,1) & \text{if the genotype is } AA\ , \end{cases} \tag{4}$$

where $c_i \overset{iid}{\sim} \text{Bernoulli}(1/2)$ if $G_i = Aa$. The partition of a genotype is straightforward for homozygotes aa or AA, but not so for a heterozygote Aa. Previous work attempted to split the n_{Aa} observations to $(0,1)$ and $(1,0)$ equally (Schaid et al., 2012; Bourgain et al., 2003), which reduces the variation inherent in a randomly selected allele (Web Appendix A). The use of a fair coin in our proposed approach ensures that $\sum_i G_{i1} \sim \text{Binomial}(n, p_{AA} + p_{Aa}/2)$ and similarly for $\sum_i G_{i2}$, which is critical to valid HWE test as well as association inference.

3.2 Reformulating the test of HWE as an allele-based regression

A critical component of developing a robust allelic association test is the modeling of the Hardy–Weinberg equilibrium assumption. HWE assumes that the two alleles in a genotype are independent of each other, so a natural approach is to use a logistic regression, $\text{logit}(E(G_{i1})) = \log(\frac{p_i}{1-p_i}) = \alpha + \beta G_{i2}$. Indeed, we show in Web Appendix B that the corresponding score test of testing $H_0 : \beta = 0$ approximates $T_{\text{HWE, Pearson}}$; regressing G_{i2} on G_{i1} leads to the same conclusion. However, the differential treatment of G_{i1} and G_{i2} is not ideal. Further, the sample size of

this model is n whereas there are $2n$ alleles from n genotypes. Thus, we consider an alternative regression model that 'doubles' the sample size.

The proposed robust allele-based (RA) regression model for HWE testing is

$$\begin{pmatrix} G_{i1} \\ G_{i2} \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \overset{iid}{\sim} N(0, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}). \tag{5}$$

This RA model has two important features. First, instead of using the location parameter β to represent the dependence between two alleles, we re-parameterize it as the correlation parameter ρ in the covariance matrix to capture HWD. This model reformulation is crucial to methodology development in Section 3.3 where the regression coefficient is reserved for association testing. Second, instead of a logistic regression, we use a Gaussian model even though the outcome is binary. This is because the primary goal here is hypothesis testing not parameter estimation or interpretation. For testing the *location* parameters in regression, Chen (1983) has shown that, under some regularity conditions, the score test statistics for regression models from the exponential family have identical form. In our setting, the Gaussian model enables explicit modeling of HWD through the *correlation* parameter ρ, and the resulting score test encompasses $T_{\text{HWE, Pearson}}$ as a special case which we show below.

Based on the Gaussian model of (5), the score test statistic of testing $H_0 : \rho = 0$ is

$$T_{\text{HWE, RA}} = \frac{(\bar{g}_{12} - \bar{g}^2)^2}{\frac{1}{n}\bar{g}^2(1-\bar{g})^2} = \frac{(\hat{p}_{AA} - \hat{p}^2)^2}{\frac{1}{n}\hat{p}^2(1-\hat{p})^2} = \frac{\hat{\delta}^2}{\frac{1}{n}\hat{p}^2(1-\hat{p})^2} \overset{H_0}{\sim} \chi^2_1, \tag{6}$$

where $\bar{g}_{12} = \sum_i g_{i1}g_{i2}/n = n_{AA}/n = \hat{p}_{AA}$, $\bar{g} = (\sum_i (g_{i1} + g_{i2}))/2n = (2n_{AA} + n_{Aa})/2n = \hat{p}$, and $\hat{\rho} = (\hat{p}_{AA} - \hat{p}^2)/(\hat{p}(1-\hat{p})) = \hat{\delta}/(\hat{p}(1-\hat{p}))$, which is a scaled estimate of HWD, δ.

We first note that the newly developed HWE test statistic is, attractively, proportional to $\hat{\delta} = \hat{p}_{AA} - \hat{p}^2$. Interestingly, after some algebraic manipulations, we show in Web Appendix C that $T_{\text{HWE, RA}}$ in (6) is *identical* to $T_{\text{HWE, Pearson}}$ in (1). This equivalence, however, is under the sim-

plest scenario of an independent sample. For more complex data, several authors have proposed different HWE testing strategies, e.g. Troendle and Yu (1994) developed a method that tests HWE across strata, while Bourgain et al. (2004) proposed a quasi-likelihood method that tests HWE in related individuals. Section 4 shows how the proposed regression (5) can be extended to derive a generalized HWE test. For the moment, we still consider an independent sample but turn our attention to association analysis.

3.3 The generalized robust allele-based (RA) association test via regression

Let G_{i1} and G_{i2} be the two allele-based random variables, as constructed in (4) for a bi-allelic SNP in an independent sample of size n, $i = 1, \ldots, n$. We now also consider Y, a (categorical or continuous) phenotype of interest, and Z, one single covariate for notation simplicity but without loss of generality. The proposed RA regression for association analysis is,

$$\begin{pmatrix} G_{i1} \\ G_{i2} \end{pmatrix} = (\alpha + \beta Y_i + \gamma Z_i) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \overset{iid}{\sim} N(0, \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}). \tag{7}$$

The corresponding score test, testing $H_0 : \beta = 0$, is

$$T_{\mathrm{RA}} = \frac{\{\sum_{i=1}^{n} \sum_{j=1}^{2} (g_{ij} - \hat{p} - \hat{\gamma}(z_i - \bar{z}))y_i\}^2}{2[1 - \frac{\{\sum_i (y_i - \bar{y})(z_i - \bar{z})\}^2}{\sum_i (y_i - \bar{y})^2 \sum_i (z_i - \bar{z})^2}] \sum_i (y_i - \bar{y})^2 (1 + \hat{\rho}) \hat{\sigma}^2} \overset{H_0}{\sim} \chi_1^2, \tag{8}$$

where $\hat{p} = \sum_i (g_{i1} + g_{i2})/2n$, $\bar{y} = \sum_i y_i/n$, $\bar{z} = \sum_i z_i/n$, $\hat{\gamma} = \sum_i (g_{i1} + g_{i2} - 2\hat{p})(z_i - \bar{z})/2\sum_i (z_i - \bar{z})^2$, $\hat{\sigma}^2 = \sum_i \sum_j (g_{ij} - \hat{p} - \hat{\gamma}(z_i - \bar{z}))^2/2n$, and $\hat{\rho} = \sum_i (g_{i1} - \hat{p} - \hat{\gamma}(z_i - \bar{z}))(g_{i2} - \hat{p} - \hat{\gamma}(z_i - \bar{z}))/n\hat{\sigma}^2$. We provide the general form of T_{RA} with multiple Ys and multiple Zs in Web Appendix D.

The proposed T_{RA} unifies previous methods. For example, if Y is binary and there are no

covariates, T_{RA} in (8) is simplified to

$$T_{\text{RA, binary, } \gamma=0} = \frac{(\hat{p}_r - \hat{p}_s)^2}{(\frac{1}{2r} + \frac{1}{2s})(\hat{p}(1-\hat{p}) + \hat{\delta})},$$

which is $T_{\text{allelic, Schaid}}$ in (3). If we further assume HWE, T_{RA} is reduced to

$$T_{\text{RA, binary, } \gamma=0,\ \rho=0} = \frac{(\hat{p}_r - \hat{p}_s)^2}{(\frac{1}{2r} + \frac{1}{2s})\hat{p}(1-\hat{p})},$$

which is the classical allelic test, T_{allelic} in (2).

The proposed RA test also generalizes the allelic association test. To provide analytical insight, consider a *continuous* trait without covariate effect. In that case,

$$T_{\text{RA, } \gamma=0} = \frac{\{\sum_i((g_{i1} + g_{i2})/2 - \hat{p})y_i\}^2}{\frac{1}{2}\sum_i (y_i - \bar{y})^2(\hat{p}(1-\hat{p}) + \hat{\delta})}.$$

Thus, for a continuous trait, the generalized allelic association test evaluates a weighted difference between individual-level allele frequency estimate, $(g_{i1} + g_{i2})/2$, and the whole sample estimate, $\hat{p}$, where the weight is an individual's trait value, y_i. In hindsight, this result may not be surprising. However, the advantages of developing the proposed RA regression model become evident when we analyze more complex data such as data with population heterogeneity and related individuals, which we investigate in the next section.

4 Complex data

4.1 Multiple populations, alleles, or phenotypes

The proposed RA regression model of (7) can naturally adjust for population effects by including population indicators, or top principal components, as part of the covariates. Here we emphasize that the potential population effects could include both difference in allele frequency and difference

in HWD between populations. The RA framework models allele frequency heterogeneity through γ and σ^2, while accounting for HWD through ρ.

Without loss of generality, it is instructive to consider the simple case of a case-control study with two populations but without additional covariates. Denote $Z_i = 0$ for population I and $Z_i = 1$ for population II. The corresponding RA regression model is

$$\begin{pmatrix} G_{i1} \\ G_{i2} \end{pmatrix} = (\alpha + \beta Y_i + \gamma Z_i) \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix}, \quad \text{where} \quad \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \end{pmatrix} \sim N(0, \sigma_i^2 \begin{pmatrix} 1 & \rho_i \\ \rho_i & 1 \end{pmatrix}), \qquad (9)$$

$\rho_i = \rho^I$ and $\sigma_i^2 = (\sigma^I)^2$ if $Z_i = 0$; $\rho_i = \rho^{II}$ and $\sigma_i^2 = (\sigma^{II})^2$ if $Z_i = 1$. Using superscripts ^{I} and ^{II} for all the other notations introduced so far, the generalized RA test of $H_0 : \beta = 0$ while accounting for population heterogeneity has the following expression,

$$T_{\text{RA, binary, 2 pop}} = \frac{(w_1(\hat{p}_r^I - \hat{p}_s^I) + w_2(\hat{p}_r^{II} - \hat{p}_s^{II}))^2}{w_1 + w_2},$$

where $w_1 = \frac{2r^I s^I}{n^I} / (\hat{p}^I(1-\hat{p}^I) + \hat{\delta}^I)$, and $\hat{\delta}^I = \hat{p}_{AA}^I - (\hat{p}^I)^2$ captures population-specific HWD in population I; the analytical expressions are the same for w_2 and $\hat{\delta}^{II}$.

If evaluating HWE across multiple populations is the primary objective, we can test $H_0 : \rho^I = \rho^{II} = 0$ and show that $T_{\text{HWE, RA, 2 pop}} = T_{\text{HWE, RA, pop I}} + T_{\text{HWE, RA, pop II}} \sim \chi_2^2$, where the expressions for $T_{\text{HWE, RA, pop I}}$ and $T_{\text{HWE, RA, pop II}}$ are given in (6). We note again the unifying feature of the proposed RA framework. For example, the test of Troendle and Yu (1994) developed specifically for testing HWE across strata has identical form as $T_{\text{HWE, RA, 2 pop}}$.

The RA model of (7) can be also extended to derive a generalized allelic association test for multi-allelic markers with $K > 2$ alleles, with adjustments for covariate effect and Hardy–Weinberg disequilibrium. In that case, we need to introduce two indicator vectors, g_{i1} and g_{i2}, each of length $K-1$, to partition the genotype of individual i; we leave the technical details to Web Appendix E and Web Table 1. Given g_{i1} and g_{i2}, the implementation of the RA regression model of (7)

is straightforward. The RA model can also be used to derive regression-based HWE test of a multi-allelic marker as shown in Web Appendix F.

To analyze multiple J phenotypes simultaneously, we can simply include multiple $Y_j 1$ vectors in the RA model of (7), each representing one phenotype, and then test $H_0 : \beta_j = 0, \forall j \in \{1, 2, \ldots, J\}$. The corresponding score test statistic is χ^2_J distributed under the null of no association. Here we re-iterate that the proposed 'reverse' regression is *allele-based*, conceptually distinct from the *genotype-based* MultiPhen (O'Reilly et al., 2012) which uses an ordinal logistic regression for the three genotype groups.

4.2 Related individuals

We now consider a sample of n correlated individuals with known or accurately estimated pedigree structure (Dimitromanolakis et al., 2019). For notation simplicity but without loss of generality, we present results for analyzing a bi-allelic SNP and one phenotype in a homogeneous population. Let g be a $2n \times 1$ vector of allele indicators for the n genotypes available, where $g = (g'_1, g'_2, \ldots, g'_n)'$ and $g_i = (G_{i1}, G_{i2})'$ for $i \in \{1, \ldots, n\}$, following the allele-partition method in (4), and let $y = (y'_1, y'_2, \ldots, y'_n)'$, $y_i = (Y_i, Y_i)'$, $z = (z'_1, z'_2, \ldots, z'_n)'$, and $z_i = (Z_i, Z_i)'$. The generalized RA regression model for a dependent sample is,

$$g = \alpha 1 + \beta y + \gamma z + \varepsilon, \quad \text{where } \varepsilon \sim N(0, \sigma^2 \Sigma), \tag{10}$$

1 is a $2n \times 1$ vector of 1s, and Σ is a $2n \times 2n$ matrix that captures the genetic correlation between individuals as well as departure from Hardy–Weinberg equilibrium in founders. Founders are individuals who have no ancestors or siblings included in the sample, and their offspring genotypes are in HWE assuming random mating (Web Appendix G).

The specification of Σ is non-trivial, where for any two individuals i and j, $\Sigma_{2(i-1)+l,\, 2(j-1)+l'}$ not only measures the genetic correlation between individual i's lth allele and individual j's l'th

allele, l and $l' \in \{1,2\}$, but also accounts for potential HWD. We note that if $i = j$ and $l = l'$, $\Sigma_{2(i-1)+l,\, 2(j-1)+l'} = 1$. If $i = j$ and $l \neq l'$, $\Sigma_{2(i-1)+l,\, 2(j-1)+l'} = 0$ for a non-founder and $= \rho$ for a founder, where ρ models HWD. Finally, if $i \neq j$, $\Sigma_{2(i-1)+l,\, 2(j-1)+l'} = \phi_{i,j}(1+\rho)$, where $\phi_{i,j}$ is the kinship coefficient between the two individuals (Web Appendix H).

As an illustration, consider a sample of f independent sib-pairs. With a slight abuse of notations, let $\{G_{j11}, G_{j12}, G_{j21}, G_{j22}\}$ denote the the four alleles of the jth sib-pair, $j = 1, \ldots, f$, where $\{G_{j11}, G_{j12}\}$ are for sibling 1 and $\{G_{j21}, G_{j22}\}$ are for sibling 2. In this case, Σ is a block diagonal matrix with

$$\Sigma_j = \begin{pmatrix} 1 & 0 & \phi(1+\rho) & \phi(1+\rho) \\ 0 & 1 & \phi(1+\rho) & \phi(1+\rho) \\ \phi(1+\rho) & \phi(1+\rho) & 1 & 0 \\ \phi(1+\rho) & \phi(1+\rho) & 0 & 1 \end{pmatrix},$$

where $\phi = 0.25$ is the kinship coefficient for a sib-pair.

For further illustration, assume all sib-pairs are concordant in phenotype (i.e. r pairs of cases and s pairs of controls) with no covariates, the score statistic of testing $H_0 : \beta = 0$ is

$$T_{\text{RA, sib-pair, binary-concordant, } \gamma=0} = \frac{(\bar{g}_r - \bar{g}_s)^2}{(\frac{1}{4r} + \frac{1}{4s})(1 + 2\phi(1+\hat{\rho}))\bar{g}(1-\bar{g})}, \tag{11}$$

where $\bar{g}_r = \sum_{j=1}^{f} \sum_{k=1}^{2} \sum_{l=1}^{2} y_{jk} g_{jkl}/4r$, $\bar{g}_s = \sum_{j=1}^{f} \sum_{k=1}^{2} \sum_{l=1}^{2} (1-y_{jk}) g_{jkl}/4s$, $\bar{g} = \sum_{j=1}^{f} \sum_{k=1}^{2} \sum_{l=1}^{2} g_{jkl}/4f$, and $\hat{\rho} = \sum_{j=1}^{f} \sum_{l=1}^{2} \{(g_{j11} - \bar{g})(g_{j2l} - \bar{g}) + (g_{j12} - \bar{g})(g_{j2l} - \bar{g})\}/(4f\phi\bar{g}(1-\bar{g})) - 1$. It is compelling that the expression of (11) is similar to that of the classical allelic test in (2). However, the denominator in (11) explicitly adjusts for the inherent genetic correlation *between* (not within unless inbreeding) the sibling alleles through ϕ, as well as any potential HWD through $\hat{\rho}$.

Figure 2: **Empirical type 1 error rate and power of the classical allelic association test and the proposed robust allelic (RA) test.** The nominal level is $\alpha = 0.05$, evaluated based on 10^4 simulation replicates. Note that when $\delta > 0$, the classical allelic test has inflated type 1 error rate as shown in (a) and (b), so the corresponding power in (c) and (d) is not meaningful and shown in a lighter shade. Also note that the HWD measure δ is bounded by the minor allele frequency p, $-p^2 \leq \delta \leq p(1-p)$. Results of type 1 error control when analyzing rare variants with $p = 0.01$ or 0.05 are shown in Web Figure 2, and when using the GWAS significance level of $\alpha = 5 \times 10^{-8}$ (evaluated based on 10^{10} simulation replicates) are shown in Web Table 2.

5 Empirical evidence

5.1 Simulation studies

To numerically demonstrate the robustness of T_{RA} to Hardy–Weinberg disequilibrium as compared with T_{allelic}, we simulated a case-control study with an independent sample of 1,000 cases and 1,000 controls. The minor allele frequency (MAF) was $p = 0.2$ or 0.5 for the minor allele A. The amount of HWD as measured by $\delta = p_{AA} - p^2$ ranged from the minimum of $-p^2$ to the maximum of $p(1-p)$. Then $p_{AA} = p^2 + \delta$ and $p_{Aa} = 2p(1-p) - 2\delta$, and $(n_{aa}, n_{Aa}, n_{AA}) \sim \text{Multinomial}\{n, (1 - p_{Aa} - p_{AA}, p_{Aa}, p_{AA})\}$. For power evaluation at $\alpha = 0.05$, we assumed an additive model with disease penetrances $f_0 = P(Y = 1|G = aa) = 0.08$, $f_1 = P(Y = 1|G = Aa) = 0.10$, and $f_2 = P(Y = 1|G = AA) = 0.12$.

The empirical type 1 error results in Figure 2(a) and 2(b) confirm the theoretical results in Figure 1: the proposed T_{RA} is accurate across the whole range of HWD values, while T_{allelic} is not robust against HWD. Further, the empirical power results in Figures 2(c) and 2(d) highlight the fact that the classical allelic test could have reduced power when the number of homozygotes AA is fewer than what is expected under the HWE assumption (i.e. when $\delta < 0$), which is not well acknowledged in the existing literature.

We also assessed the accuracy of T_{RA} and T_{allelic} when analyzing SNPs with low MAF ($p =$

0.01 and 0.05), and results in Web Figure 2 are similar to that in Figure 2. In addition, we evaluated the performance of the tests using the GWAS significance level of $\alpha = 5 \times 10^{-8}$, with 10^{10} simulation replicates for each parameter setting considered (Web Table 2); results for MAF > 0.1 are not shown, because the performance of T_{RA} is better for larger MAF as expected. Overall, the proposed RA test is accurate or slightly conservative for the whole range of HWD. In contrast, the classical allelic test has grossly inflated type 1 error rate up to 1.35×10^{-5} for $\alpha = 5 \times 10^{-8}$ (Web Table 2).

5.2 Application 1 - revisit the study of Wittke-Thompson et al. (2005)

Studying Hardy–Weinberg disequilibrium not due to genotyping error, Wittke-Thompson et al. (2005) identified 60 bi-allelic SNPs from 41 case-control association studies. Focusing on association analysis of these 60 SNPs, we compared T_{allelic} with the proposed T_{RA} (Figure 3). As expected based on the simulation results in Figure 2, for SNPs with $\hat{\delta} > 0$, T_{allelic} can appear to be more powerful, due to increased type 1 error rate, than the proposed T_{RA}. For example, for the most significant SNP, p-value$_{\mathrm{allelic}} = 1.82 \times 10^{-10}$ and p-value$_{\mathrm{RA}} = 6.60 \times 10^{-9}$. However, for this SNP $\hat{\delta} = 0.052 > 0$ with p-value$_{\mathrm{HWE}} = 3.09 \times 10^{-4}$; the HWE test is significant but not suggesting genotyping error. Thus, T_{allelic} is too optimistic. In contrast, for the third most significant SNP, $\hat{\delta} = -0.031 < 0$ and p-value$_{\mathrm{HWE}} = 0.040$. In that case, T_{allelic} is conservative while the proposed T_{RA} is not only robust but also more powerful, where p-value$_{\mathrm{allelic}} = 5.84 \times 10^{-6}$ and p-value$_{\mathrm{RA}} = 8.86 \times 10^{-7}$.

5.3 Application 2 - a cystic fibrosis (CF) gene modifier study

Here we applied T_{RA} to simultaneously analyze two phenotypes using a sample with related individuals from the Canadian cystic fibrosis (CF) gene modifier study. The two phenotypes of interest are lung function, a quantitative trait (Taylor et al., 2011), and meconium ileus (MI), a

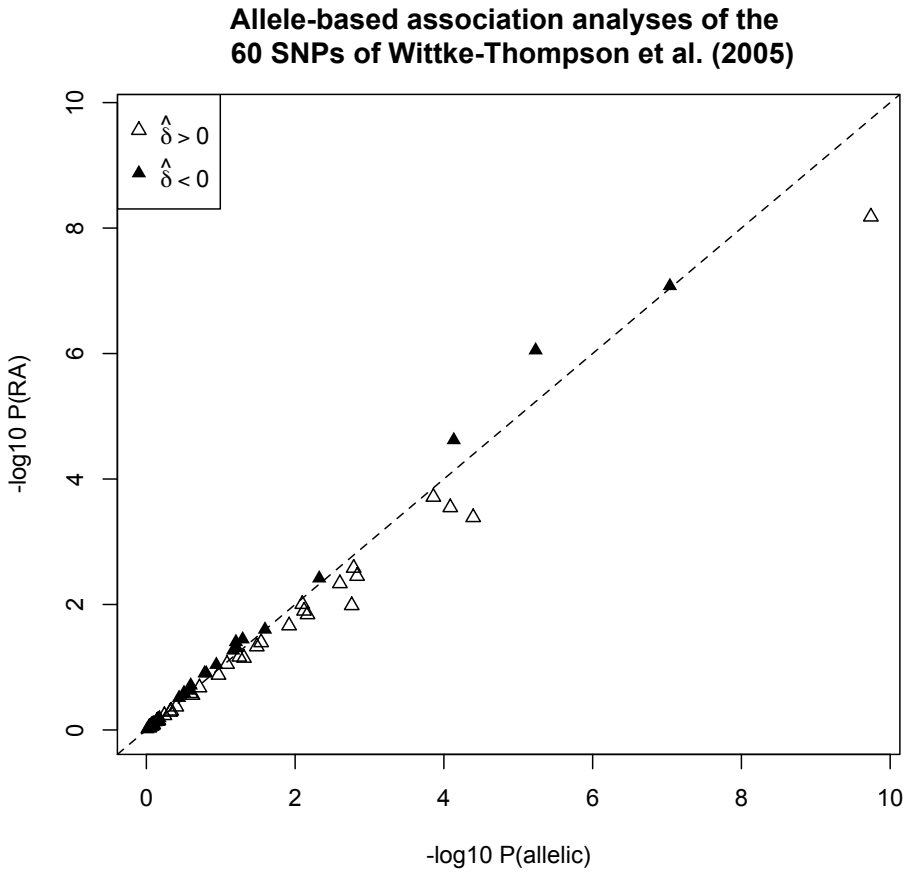

Figure 3: **Results of application 1.** Allele-based association tests of the 60 SNPs identified in Wittke-Thompson et al. (2005), contrasting the proposed RA method, T_{RA}, with the classical allelic test, T_{allelic}. Unfilled triangles are for SNPs with $\hat{\delta} > 0$ (T_{allelic} having inflated type 1 error), and filled triangles are for SNPs with $\hat{\delta} < 0$ (T_{allelic} having deflated type 1 error); see Figure 1 for theoretical results and Figure 2 for simulation results regarding type 1 error control of the two methods.

binary trait (Gong et al., 2019). The sample of $2,540$ CF subjects includes $2,420$ singletons and 60 independent sib-pairs. For completeness, we first analyzed each phenotype individually using the proposed *allele-based* RA framework, and we compared the results with the traditional *genotype-based* method via the standard (generalized) linear mixed models (LMM or GLMM). We then analyzed both phenotypes simultaneously using T_{RA}.

Figures 4(a) and 4(b) show that results of genotype- and allele-based methods are largely con-

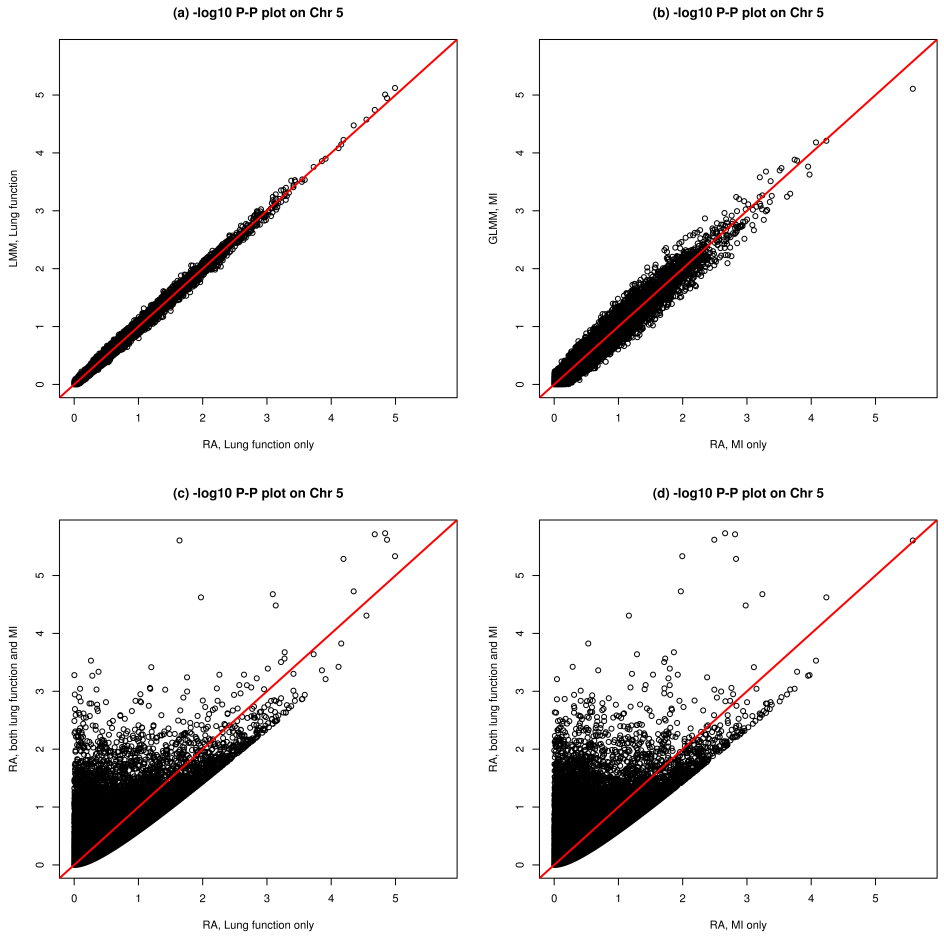

Figure 4: **Results of application 2 - Chromosome 5-wide.** Genetic association studies of lung function and meconium ileus of $34,378$ bi-allelic markers on chromosome 5, using a sample of $2,540$ individuals with cystic fibrosis of which $2,420$ are singletons and 120 are from 60 sib-pairs. LMM and GLMM are *genotype-based* association analyses based on, respectively, the standard linear mixed model for a continuous trait (i.e. lung) and generalized LMM for a binary trait (i.e. MI), and RA is the proposed allele-based association method that can also simultaneously analyze multiple traits using a sample of related individuals. Genome-wide results are shown in Web Figure 1.

sistent; see Section 6 for a discussion. For the most significant SNP associated with MI (Figure 4(b)), the p-value of T_{RA} is 2.62×10^{-6}, slightly smaller than 7.80×10^{-6} of the GLMM method. In addition, the proposed T_{RA} method can simultaneously analyzed both phenotypes and identify suggestive SNPs that have p-values several orders of magnitude smaller than those from

Table 2: **Results of application 2 - Previously reported SNPs.** The p-values in this table differ from those in Sun et al. (2012) and Corvol et al. (2015), because the analyses here only included the Canadian samples and individuals with both phenotypes measured. Other details see legend to Figure 4.

Top CF lung function associated SNPs from Corvol et al. (2015)				
Chr	SNP	Lung function only		Lung function & MI simultaneously
		$-\log_{10} p_{\text{LMM}}$	$-\log_{10} p_{\text{RA}}$	$-\log_{10} p_{\text{RA, lung \& MI}}$
3	rs2246901	3.21	**3.25**	2.63
5	rs3749615	3.25	**3.27**	3.57
6	rs2395185	6.65	**6.77**	6.08
11	rs10466455	5.84	**5.86**	4.84
Top CF meconium ileus associated SNPs from Sun et al. (2012)				
Chr	SNP	MI only		Lung function & MI simultaneously
		$-\log_{10} p_{\text{GLMM}}$	$-\log_{10} p_{\text{RA}}$	$-\log_{10} p_{\text{RA, lung \& MI}}$
1	rs4077468	5.34	**5.47**	4.87
1	rs7512462	4.54	**4.82**	4.56
1	rs7419153	3.68	**4.07**	3.35
1	rs12047830	3.10	**3.20**	2.63

studying one phenotype at a time (Figures 4(c) and 4(d)). However, these results are not genome-wide significant and establishing true association requires follow-up studies.

Table 2 summarizes the association results for previously reported and replicated SNPs associated with CF lung function (Corvol et al., 2015) and MI association (Sun et al., 2012). For all the SNPs in Table 2, the proposed RA test yields slightly larger $-\log_{10}$(p-values) than LMM or GLMM, suggesting that the allele-based method has the potential to be more powerful than the traditional genotype-based approach. The RA simultaneous analysis of the two phenotypes, however, did not lead to more significant results; this is not surprising because these SNPs were selected based on the two previous single-phenotype analyses.

6 Discussion

To generalize the concept of comparing allele frequencies between case and control groups to more complex data settings, here we developed a novel robust allele-based (RA) regression framework that regresses individual alleles on phenotypes. Motivated by the earlier work of Chen (1983) for testing regression coefficients in independent samples, the proposed method utilizes the Gaussian model of (10) that (i) leads to a valid allelic association test through testing the regression coefficient β, (ii) adjusts for covariate effect through additional regression coefficient γ, (iii) explicitly models potential departure from Hardy–Weinberg equilibrium through ρ in the covariance matrix Σ, (iv) accounts for sample correlation through kinship coefficient ϕ in Σ, and (v) analyzes either a binary or a continuous phenotype, or multiple phenotypes simultaneously, where the phenotype data can be subjected to Y-dependent sampling.

The generalized allelic association test also unifies existing allelic tests. Under the HWE assumption and for a case-control study using an independent sample without covariates, the score test of $H_0 : \beta = 0$ based on the simplified RA model (7) is identical to the classical allelic test in (2), $T_{\text{RA, binary, } \gamma=0, \ \rho=0} = T_{\text{allelic}}$. In the presence of HWD, $T_{\text{RA, binary, } \gamma=0} = T_{\text{allelic, Schaid}}$.

A crucial stage of this work is creating the two allele-based random variables, G_{i1} and G_{i2}, and leveraging the 'power' of regression in new settings. The idea of reformulating an existing test statistic as a regression to facilitate method extension is not new. In their Reader Reaction to the generalized non-parametric Kurskal-Wallis test of Acar and Sun (2013) for handling group uncertainty, Wu and Guan (2015) presented "*a rank linear regression model and derived the proposed GKW statistic as a score test statistic*". More recently, Soave and Sun (2017) showed that by first reformulating the original Levene's test, testing for variance heterogeneity between k groups in an independent sample without group uncertainty, as a two-stage regression, the extension to more complex data is more straightforward.

The concept of 'reverse' regression has also been explored before, focusing on regressing *geno-*

type on phenotype, notably by O'Reilly et al. (2012) for simultaneously analyses of multiple phenotypes. The corresponding MultiPhen method uses an ordinal logistic regression for the three genotype groups and then applies a likelihood ratio test in independent samples. Zhang et al. (2014) then examined GEE-based association tests for related individuals, but fundamentally these early tests are genotype-based.

Another stream of genotype-based 'reverse' or retrospective approach started with the quasi-likelihood method of Thornton and McPeek (2007) for case-control association testing with related individuals. The method first defines $X_i = G_i/2 \in \{0, 1/2, 1\}$, then links the mean of X_i with Y_i via a logit transformation and uses the kinship coefficient matrix as the covariance matrix of X_i, and finally obtains a quasi-likelihood score test. Subsequently, Feng (2014) and Feng et al. (2011) extended the method of Thornton and McPeek (2007) to a quasi-likelihood regression model that can incorporate multiple phenotypes. Although $E(X_i) = E(G_i/2)$ can be interpreted as allele frequency, the quasi-likelihood score test is fundamentally a genotype-based association method. Further, the use of the kinship matrix alone as the covariance matrix requires the assumption of HWE.

Most existing family-based association studies rely on the Y-on-G prospective regression framework via LMM or GLMM (Eu-Ahsunthornwattana et al., 2014). For the application study in Section 5.3, we applied both the proposed RA method, and LMM (for the continuous CF lung function) and GLMM (for the binary meconium ileus status). Although there are differences in the (single-phenotype) analyses (Figures 4(a) and 4(b)), results are remarkably consistent. Interestingly, in the simplest case of an independent sample with no covariates, the corresponding RA test statistic has identical form as that derived from the genotype-based prospective regression model, as well as that from the non-parametric trend test (Web Appendix I). This inference similarity indirectly confirms the validity of the proposed approach ('reverse', allele-based and Gaussian model applied to a binary variable), but does not take away the contribution of this work. In particular, unlike LMM or GLMM, the proposed RA method can analyze multiple phenotype simultaneously as shown in Figures 4(c) and 4(d).

One of the challenges related to the proposed framework is the interpretation of parameter estimate for β even though its corresponding hypothesis testing is valid. Thus, we emphasize that the method developed here is tailored for rapid association testing not effect size estimation. We direct the readers to Krutchkoff (1967); Halperin (1970) for parameter estimation in the context of 'reverse' regression. The proposed RA framework provides a statistically efficient and computationally fast way for genome-wide association scans. For example, the analysis of Application 2 took 23.16 hours using one CPU core of an Intel Xeon Processor (Skylake, IBRS) at 2.40 GHz; the computation time is reduced to less than 2 hours when analyzing only the unrelated individuals in the CF sample.

Another difficulty present in any 'reverse' regression approach is the modeling and interpretation of gene-gene or gene-environment interactions. It is also not clear how to perform an allelic association test for an X-chromosomal variant; see Chen et al. (2018) for genotype-based association methods. Finally, although the proposed RA method has good type 1 error control when analyzing rare variants (Web Figure 2 and Web Table 2), power of the RA method in its current implementation, analyzing one SNP at a time, is likely to be low for a rare variant. Extending the RA methodology to simultaneously analyze multiple rare variants (Derkach et al., 2014) is of future interest but beyond the scope of this work.

The proposed framework, however, is flexible and promising in other directions. For example, the inclusion of parameter ρ in the RA model (7) is advantageous for both method comparison and further development. In the absence of Y and Z and sample correlation, we have shown that the score test of $H_0 : \rho = 0$ is identical to the classical Pearson's χ^2 test of HWE in (1), $T_{\text{HWE, RA}} = T_{\text{HWE, Pearson}}$. For more complex data, instead of developing individual remedies addressing specific challenges, the proposed method provides a principled approach for extensions. For example, we have shown in Section 4.1 that by introducing a population indicator we can derive a HWE test across populations that, in the simple setting of an independent sample, has identical form with that of Troendle and Yu (1994).

In terms of association testing, the value of explicitly modeling HWD via ρ in the regression is two fold. First, if there is a strong prior evidence for HWE, we can restrict ρ to be zero and establish a locally most powerful score test. Second, for the special case of a case-control study, Song and Elston (2006) and Wang and Shete (2010) have argued that departure from HWE in the case group provides additional evidence for association. However, the existing methods are ad-hoc. For example, the method of Song and Elston (2006) first conducts genotype-based association test and Pearson χ^2 test of HWE separately, then aggregates the two (dependent) tests by a weighted sum, and finally evaluates the statistical significance via simulations. The proposed RA regression framework conceptually offers a new approach that could directly incorporate group-specific ρ into association inference, which we will explore as future work.

7 Acknowledgment

The authors thank Dr. Lisa Strug and her lab for providing the cystic fibrosis application data. This research is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC, RGPIN-04934 and RGPAS-522594), and the Canadian Institutes of Health Research (CIHR, MOP-310732) to LS. LZ is a trainee of the CANSSI-ONTARIO STAGE training program at the University of Toronto.

References

Acar, E. F., and L. Sun, 2013: A generalized kruskal–wallis test incorporating group uncertainty with application to genetic association studies. *Biometrics*, **69 (2)**, 427–435.

Bourgain, C., M. Abney, D. Schneider, C. Ober, and M. S. McPeek, 2004: Testing for hardy–weinberg equilibrium in samples with related individuals. *Genetics*, **168 (4)**, 2349–2361.

Bourgain, C., and Coauthors, 2003: Novel case-control test in a founder population identifies

p-selectin as an atopy-susceptibility locus. *The American Journal of Human Genetics*, **73 (3)**, 612–626.

Bycroft, C., and Coauthors, 2018: The uk biobank resource with deep phenotyping and genomic data. *Nature*, **562 (7726)**, 203–209.

Chen, B., R. V. Craiu, L. J. Strug, and L. Sun, 2018: The x factor: A robust and powerful approach to x-chromosome-inclusive whole-genome association studies. *arXiv preprint arXiv:1811.00964*.

Chen, C.-F., 1983: Score tests for regression models. *Journal of the American Statistical Association*, **78 (381)**, 158–161.

Corvol, H., and Coauthors, 2015: Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nature Communications*, **6**, 8382.

Derkach, A., J. F. Lawless, and L. Sun, 2015: Score tests for association under response-dependent sampling designs for expensive covariates. *Biometrika*, **102 (4)**, 988–994.

Derkach, A., J. F. Lawless, L. Sun, and Coauthors, 2014: Pooled association tests for rare genetic variants: a review and some new results. *Statistical Science*, **29 (2)**, 302–321.

Dimitromanolakis, A., A. D. Paterson, and L. Sun, 2019: Fast and accurate shared segment detection and relatedness estimation in un-phased genetic data via truffle. *The American Journal of Human Genetics*, **105 (1)**, 78–88.

Eu-Ahsunthornwattana, J., E. N. Miller, M. Fakiola, S. M. Jeronimo, J. M. Blackwell, H. J. Cordell, W. T. C. C. C. 2, and Coauthors, 2014: Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genetics*, **10 (7)**, e1004 445.

Feng, Z., 2014: A generalized quasi-likelihood scoring approach for simultaneously testing the genetic association of multiple traits. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **63 (3)**, 483–498.

Feng, Z., W. W. Wong, X. Gao, and F. Schenkel, 2011: Generalized genetic association study with samples of related individuals. *The Annals of Applied Statistics*, 2109–2130.

Gong, J., and Coauthors, 2019: Genetic association and transcriptome integration identify contributing genes and tissues at cystic fibrosis modifier loci. *PLoS Genetics*, **15 (2)**, e1008 007.

Halperin, M., 1970: On inverse estimation in linear regression. *Technometrics*, **12 (4)**, 727–736.

Krutchkoff, R., 1967: Classical and inverse regression methods of calibration. *Technometrics*, **9 (3)**, 425–439.

O'Reilly, P. F., C. J. Hoggart, Y. Pomyen, F. C. Calboli, P. Elliott, M.-R. Jarvelin, and L. J. Coin, 2012: Multiphen: joint model of multiple phenotypes can increase discovery in gwas. *PloS One*, **7 (5)**, e34 861.

Sasieni, P. D., 1997: From genotypes to genes: doubling the sample size. *Biometrics*, 1253–1261.

Schaid, D. J., and S. J. Jacobsen, 1999: Biased tests of association: comparisons of allele frequencies when departing from hardy–weinberg proportions. *American Journal of Epidemiology*, **149 (8)**, 706–711.

Schaid, D. J., J. P. Sinnwell, and G. D. Jenkins, 2012: Regression modeling of allele frequencies and testing hardy weinberg equilibrium. *Human Heredity*, **74 (2)**, 71–82.

Soave, D., and L. Sun, 2017: A generalized levene's scale test for variance heterogeneity in the presence of sample correlation and group uncertainty. *Biometrics*, **73 (3)**, 960–971.

Song, K., and R. C. Elston, 2006: A powerful method of combining measures of association and hardy–weinberg disequilibrium for fine-mapping in case-control studies. *Statistics in Medicine*, **25 (1)**, 105–126.

Sun, L., and Coauthors, 2012: Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. *Nature Genetics*, **44 (5)**, 562.

Taylor, C., and Coauthors, 2011: A novel lung disease phenotype adjusted for mortality attrition for cystic fibrosis genetic modifier studies. *Pediatric Pulmonology*, **46 (9)**, 857–869.

Thornton, T., and M. S. McPeek, 2007: Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *The American Journal of Human Genetics*, **81 (2)**, 321–337.

Troendle, J., and K. Yu, 1994: A note on testing the hardy–weinberg law across strata. *Annals of Human Genetics*, **58 (4)**, 397–402.

Wang, J., and S. Shete, 2010: Using both cases and controls for testing hardy–weinberg proportions in a genetic association study. *Human Heredity*, **69 (3)**, 212–218.

Weir, B., 1996: Genetic analysis ii. *Sinauer: Sunderland, MA*.

Wittke-Thompson, J. K., A. Pluzhnikov, and N. J. Cox, 2005: Rational inferences about departures from hardy–weinberg equilibrium. *The American Journal of Human Genetics*, **76 (6)**, 967–986.

Wu, B., and W. Guan, 2015: Reader reaction on the generalized kruskal–wallis test for genetic association studies incorporating group uncertainty. *Biometrics*, **71 (2)**, 556–557.

Zhang, Y., Z. Xu, X. Shen, W. Pan, A. D. N. Initiative, and Coauthors, 2014: Testing for association with multiple traits in generalized estimation equations, with application to neuroimaging data. *NeuroImage*, **96**, 309–325.