Using Controls to Test for Hardy-Weinberg Equilibrium: Is It Bona Fide or a Fallacy?

Virginia Ma* Shili Lin†

Abstract

Many statistical methods for analyzing genetic data assume Hardy-Weinberg Equilibrium (HWE), such as those used in genome wide association studies. Therefore, to use such methods, one must check whether the HWE assumption is valid. For a case-control study, researchers have recognized that Hardy Weinberg proportions will be distorted if the marker being tested happens to be associated with the disease. To alleviate this problem, it is typical that HWE tests are carried out only on controls. This raises the following question: do controls as a whole provide a good representation of the population? One may justify this by further assuming that the disease is rare. However, many diseases being studied are in fact common, and the rare disease assumption is often conveniently forgotten. Even if the disease is indeed rare, is it then justifiable to test for HWE in the controls? In this study, we attempt to answer the above questions. Deriving bias when using controls to estimate the population, we demonstrate that the rare disease assumption alone is not sufficient to justify the use of controls as representatives of the population. We additionally investigate whether it is bona fide to perform HWE testing on controls only even if the disease being studied is rare. The results from our study are striking; the Type I error can be severely inflated regardless of whether the disease being investigated is rare or common.

Key Words: Hardy-Weinberg equilibrium, case-control study, rare disease assumption, genome-wide association studies

1. Introduction

Genome-wide Association Studies (GWAS) are frequently performed to look for genetic variants that are associated with an array of common diseases. In such studies, with hundreds of thousands, or even millions, of Single Nucleotide Polymorphisms (SNPs) being generated, the first step of data analysis is often testing for Hardy-Weinberg Equilibrium (HWE) for each of the SNPs (Wittke-Thompson et al. 2005; Yu et al. 2010). This is usually framed as a quality control step to help identify possible errors that could have occurred, including systematic genotyping ones (Gomes et al. 1999; Hosking et al. 2004; Salanti et al. 2005; Moonesighe et al. 2010). Those that do not pass the HWE tests will be eliminated before moving on to the next step, as they are deemed to have violated the HWE assumption and are likely to be caused by errors (McCarthy et al. 2008). To vigorously detect all possible errors, multiple testing procedures are typically not fully implemented, leading to the rejection of a large proportion of the SNPs in some studies (Healey et al. 2000; The Welcome Trust Case Control Consortium 2007; Phillips et al. 2008; Wang and Shete 2010).

In a case-control study setting, it is known that if a locus is indeed associated with the disease of interest, then HWE will be distorted in the cases (Nielsen et al. 1999). Therefore, researchers use controls only, believing that controls are similar to the population in terms of their genetic makeup, which implicitly assumes the disease is rare (Salanti et al. 2005). However, the rare disease assumption itself is one that is often overlooked or merely forgotten and taken for granted.

*Columbus Academy, 4300 Cherry Bottom Road, Columbus, OH 43230

†Department of Statistics, Ohio State University, Columbus, OH 43210

Even if it has been made sure that the disease is indeed rare, relying on this assumption still raises several concerns. The first question: are the controls actually similar to the general population? For studying maternal and imprinting effect, it was argued that the rare disease assumption is necessary, but not sufficient for the premise that controls have similar genetic makeup to the general population (Yang and Lin 2013). This issue was also touched upon in a simulation study devised to investigate HWE testing (Brems 2015). Additionally, what makes the use of controls only under the premise of the rare disease assumption most ironic of all, is the fact that most diseases studied today are in fact common. If the controls are not a good representation of the population, then more variants than necessary may be thrown out since the measure of HWE will be distorted (Wang and Shete 2010). This results in a loss of information that is in fact entirely avoidable if one is made aware of the potential risk using controls only poses when testing for HWE.

In this paper, we address two major issues. Firstly, we investigate if the rare disease assumption itself is sufficient to justify the use of controls only as surrogate for a random sample of the population in the context of HWE testing. We look at the bias incurred when estimating the population genotype frequencies based on the controls. Secondly, we explore whether there is inflated Type I error when using controls only for testing for HWE, regardless of whether the disease of interest is common or rare, and if so, how bad it can be.

2. Methods

2.1 Genotype Distribution of the General Population

Consider a SNP with alleles a and A, whose frequencies are p_a and $1 - p_a$, respectively, with a being the minor allele and thus p_a is referred to as the Minor Allele Frequency (MAF). Assuming that the SNP is from a population under HWE, then the probabilities of the three genotypes, AA, Aa, and aa, in the population are defined as,

$$\begin{aligned} p_{AA} &= (1 - p_a)^2, \\ p_{Aa} &= 2p_a(1 - p_a), \\ p_{aa} &= p_a^2. \end{aligned} \tag{1}$$

2.2 Genotype Distribution of Cases

Suppose that the SNP is associated with a disease being studied and that having one or two copies of the minor allele will lead to increased risks of being affected by the disease (i.e. becoming a case). These increased risks are above the baseline level when there are no copies of the minor allele, and can be formally defined in terms of the following relative risks:

$$RR_1 = \frac{P(case \mid Aa)}{P(case \mid AA)} \quad \text{and} \quad RR_2 = \frac{P(case \mid aa)}{P(case \mid AA)}.$$

Then, using the Bayes Rule, the distribution of the three genotypes in the population of cases can be derived in terms of the genotype distribution in the general population, p_{AA}, p_{Aa}, p_{aa}, and the two relative risks, RR_1, RR_2. That is,

$$
\begin{aligned}
p_{AA|case} \equiv P(AA \mid case) &= \frac{p_{AA}}{p_{AA} + p_{Aa} \times RR_1 + p_{aa} \times RR_2}, \\
p_{Aa|case} \equiv P(Aa \mid case) &= \frac{p_{Aa} \times RR_1}{p_{AA} + p_{Aa} \times RR_1 + p_{aa} \times RR_2}, \\
p_{aa|case} \equiv P(aa \mid case) &= \frac{p_{aa} \times RR_2}{p_{AA} + p_{Aa} \times RR_1 + p_{aa} \times RR_2}.
\end{aligned} \tag{2}
$$

These probabilities have been derived many times previously in the literature, including the reproduction in Brems (2015). It is easily seen that the genotype distribution for the cases as defined in the set of equations in (2) is different from that of the general population as defined in (1). We let $C = p_{AA} + p_{Aa} \times RR_1 + p_{aa} \times RR_2$. In order for the three corresponding probabilities to be similar, we need to have $C \approx 1, RR_1/C \approx 1$, and $RR_2/C \approx 1$. These requirements lead to the conclusion that both $RR_1 \approx 1$ and $RR_2 \approx 1$. As such, for a SNP that is associated with the disease, these requirements cannot be satisfied. In other words, the cases are not similar to the general population in their genetic makeup. Hence, it has been correctly concluded by researchers that using cases for testing HWE will lead to inflated Type I error.

2.3 Genotype Distribution of Controls

Similarly, one can derive the distribution of the three genotypes within the population of controls, noting that the population prevalence of the disease is also needed to completely specify them:

$$
\begin{aligned}
p_{AA|cntr} \equiv P(AA \mid control) &= \frac{p_{AA} - p_{AA|case} \times \kappa}{1 - \kappa}, \\
p_{Aa|cntr} \equiv P(Aa \mid control) &= \frac{p_{Aa} - p_{Aa|case} \times \kappa}{1 - \kappa}, \\
p_{aa|cntr} \equiv P(aa \mid control) &= \frac{p_{aa} - p_{aa|case} \times \kappa}{1 - \kappa},
\end{aligned} \tag{3}
$$

where $\kappa \equiv P(case)$ is the population disease prevalence, that is, the probability that a randomly chosen individual from the general population is affected by the disease.

One can easily see that the three control probabilities in (3) also differ from their corresponding population probabilities in (1). In order for the two distributions to be approximately the same, we would require that

$$
\begin{aligned}
p_{AA} - p_{AA|case} \times \kappa &\approx (1 - \kappa) \times p_{AA}, \\
p_{Aa} - p_{Aa|case} \times \kappa &\approx (1 - \kappa) \times p_{Aa}, \\
p_{aa} - p_{aa|case} \times \kappa &\approx (1 - \kappa) \times p_{aa}.
\end{aligned}
$$

Simple algebraic manipulations show that these conditions are equivalent to requiring that the genotype distribution for the cases be approximately the same as the genotype distribution for the general population. As such, regardless of whether the disease is rare or common, the genotype distribution of the controls will not be similar to the distribution of the general population as long as the SNP is associated with the disease with an appreciable effect size.

2.4 The Controls Being Representative of the Population Fallacy

To see that the commonly-held notion that controls can be used as representation of the general population is in fact a fallacy, we rewrite the genotype distribution of the controls using a different representation:

$$P(AA \mid control) = [p(control \mid AA)p_{AA}]/(1-\kappa),$$
$$P(Aa \mid control) = [p(control \mid Aa)p_{Aa}]/(1-\kappa),$$
$$P(aa \mid control) = [p(control \mid aa)p_{aa}]/(1-\kappa).$$

In order for these probabilities to be close to their population counterparts, it is required that

$$P(control \mid Genotype) \approx 1-\kappa,$$

where $Genotype$ can be AA, Aa, or aa. If a disease is rare, then $1-\kappa \approx 1$, which implies that $P(control \mid Genotype)$ would all have to be close to 1. Clearly, this is not possible if the disease is Mendelian. For example, for a recessive disorder with complete penetrance, $P(control \mid aa) = 0$, which is not close to 1. In general, as long as the SNP is associated with the disease with an appreciable effect size, the requirements stated above cannot be satisfied. Thus, the assumption of rare disease has no bearing on whether the distribution of the control and that of the general population are similar.

2.5 Bias of Genotype Probability Estimators

We can ascertain the expected degree of deviation when using controls to estimate the genotype frequencies of a population in HWE by calculating the biases and relative biases. For example, the bias for estimating the frequency of aa is,

$$Bias_{aa} = E\left[\frac{n_{aa|cntr}}{n_{cntr}}\right] - p_{aa} = p_{aa|cntr} - p_{aa},$$

where n_{cntr} is the number of controls in the sample while $n_{aa|cntr}$ is the number of controls having the aa genotype. Clearly, the bias is not zero based on the formulas (1) and (3) presented above for the population and control frequencies, respectively, and therefore the estimator is biased. Further, the degree of bias is not dependent on the sample size; thus, the frequency estimator is not consistent. Since p_{aa} can be quite small, especially for variants with a small MAF, we will consider the absolute relative bias as defined in the following:

$$RAbias_{aa} = \frac{|p_{aa|cntr} - p_{aa}|}{p_{aa}}.$$

Calculating the biases and absolute relative biases for the other two genotypes can be carried out similarly.

2.6 Chi-square tests for HWE

We performed multiple Chi-square tests for HWE to ascertain the potential inflation of type I error. The underlying general population is assumed to be under HWE; therefore, a rejection of the null hypothesis H_0 (the SNP marker is in HWE) is treated as committing a type I error. Our first two tests (T1 and T2) are for data from the cases. As well documented in the literature, it is expected that a HWE test based on the cases will lead to an inflation of the type I error rate. As such, these two tests are used merely as a confirmatory investigation for completeness. The last two tests (T3 and T4) are for data from the controls. This setting is the focus of our study, as we are interested in addressing the question that we have raised: is it justifiable to test for HWE using the controls?

T1. Data are generated under the genotype distribution for cases. Chi-square test will be performed assuming that the expected counts for the three genotypes are those from the general population.

T2. Data are generated under the genotype distribution for cases. Chi-square test will be performed assuming that the expected counts for the three genotypes are estimated from the data under H_0.

T3. Data are generated under the genotype distribution for controls. Chi-square test will be performed assuming that the expected counts for the three genotypes are those from the general population.

T4. Data are generated under the genotype distribution for controls. Chi-square test will be performed assuming that the expected counts for the three genotypes are estimated from the data under H_0.

When the population MAF of a SNP marker is known, then T3 will be appropriate. In many real data studies, however, population frequencies are usually unknown and will be estimated from the data. In such cases, T4 is the typical one performed in the first step of a GWAS. We also note that the asymptotic Chi-square distribution for T1 and T3 has 2 degrees of freedom since the known probabilities are used, whereas that for T2 and T4 has 1 degree of freedom as the probabilities are estimated based on the sample being tested.

3. Results

3.1 Assessment of Bias

To illustrate that controls may not be representative of the general population in terms of the genetic makeup, we compute the relative bias of using the genotype probabilities computed from the controls as estimates of the corresponding population frequencies. Plotted in Figure 1 is the absolute relative bias when using an estimator based on the controls only, $n_{aa|cntr}/n_{cntr}$ for p_{aa} in the general population, where n_{cntr} and $n_{aa|cntr}$ are as defined in 2.5. We consider a recessive disease model in which only those with two copies of the minor allele have an elevated risk over the baseline; that is, $RR_1 = 1$ and $RR_2 > 1$. The relative bias is computed over a range of minor allele frequencies, MAF = 0.1, 0.2, 0.3, 0.4, and 0.5. We study two values for RR_2, which are set to be 2 and 3. Furthermore, we entertain two disease prevalence, 0.15 and 0.05, with the latter representing a rare disease scenario. As we can see from the figure, regardless of the combination of the parameters, using controls to estimate the general population will result in bias. The degree of relative bias does depend on the underlying setting. First, for a disease with a larger effect size ($RR_2 = 3$), the relative bias is larger than the corresponding counterpart with a smaller effect size ($RR_2 = 2$). Second, a more common disease ($\kappa = 0.15$) will also lead to greater relative bias compared to a rare disease ($\kappa = 0.05$). Finally, as the minor allele frequency gets larger toward a more common SNP, the relative bias decreases.

3.2 Simulation Study

We considered four sets of genetic models: recessive, dominant, multiplicative, and additive. These four types of models differ in their relative risks with one or two copies of the minor alleles — recessive: $RR_1 = 1, RR_2 > 1$; dominant: $RR_1 = RR_2 > 1$; multiplicative: $RR_2 = RR_1^2$; and additive: $RR_1 = (1 + RR_2)/2$. Under each type of the models, multiple scenarios are considered; full details of the parameter values are given in

Table 1: Simulation models and settings*

Model	Parameter Setting	κ	p_a	RR_1	RR_2
Recessive	R1	0.15	0.3	1	1.4
	R2	0.05	0.1	1	4
	R3	0.15	0.1	1	2
Dominant	D1	0.15	0.3	1.4	1.4
	D2	0.15	0.1	2	2
	D3	0.15	0.3	2	2
Multiplicative	M1	0.15	0.1	1.8	3.2
	M2	0.15	0.1	1.6	2.6
	M3	0.15	0.3	1.8	3.2
Additive	A1	0.15	0.3	1.1	1.2
	A2	0.05	0.1	2	3
	A3	0.15	0.1	2	3

*For each of the four types of models, κ denotes the population prevalence; p_a is the minor allele frequency; RR_1 and RR_2 are the relative risks with one or two copies of the minor alleles, respectively, over the baseline disease risk of having no copy of the minor allele. Note that dominant and multiplicative models are unlikely to be associated with a rare disease and therefore $\kappa = 0.05$ was not investigated for these two models.

Table 1, which include the MAF values as well as the population disease prevalence, κ. In particular, we considered two disease prevalence, $\kappa = 0.15$ and $\kappa = 0.05$ to simulate the situation of a common and rare disease, respectively, for the appropriate models. We also vary the MAF at two levels, 0.1 and 0.3, although not all combinations of κ and MAF are considered. For each setting under a model, we simulated data for 10000 cases and 10000 controls according to the distributions of probabilities provided in equation sets (2) and (3), respectively. Tests T1 and T2 were then applied to the case data, while tests T3 and T4 were applied to the control data. This process was repeated for 10000 simulated data sets.

Figure 2 shows the QQ plots of the quantiles of the p-values (over 10000 simulated data sets) against the quantiles of the uniform(0,1) distribution, in the $-\log_{10}$ scale. Since the data were all simulated under the null hypothesis (H_0 : HWE), the p-values are expected to follow a uniform distribution when tests T3 and T4 are used, if it is indeed justifiable to use the controls to test for HWE. Although understanding the consequences of using the controls is the focus, we also study the results when cases are tested (i.e. T1 and T2 are used) as a comparison. The four plots presented in Figure 2 are the results when the tests are applied to data simulated under the recessive model R1 (see Table 1 for the details of the parameter settings). As expected, the results from T1 and T2 (tests using the cases; Figure 2 (a) and (b)) have severely inflated type I error, as both curves are far above the 95% confidence bands (portrayed by the pair of dashed lines). Although the inflation of type I error is not as severe when testing using the controls (Figure 2 (c) and (d)), the inflation is clearly seen from the curves as well. These results evidently indicate that testing for HWE using controls will also lead to inflation of type I error, casting great doubts on the common practice in the scientific community.

The inflation of type I error when testing using the controls is also clearly seen for the other two recessive models, R2 and R3; their parameter settings are given in Table 1. As shown in the first segments of Figure 3, if the use of controls is justifiable, one would expect only 5% of the p-values to be smaller than 0.05 (or equivalently, greater than 1.3 in the $-\log_{10}$ scale), but this is exceeded by a large margin, regardless of whether the expected are computed based on population genotype frequencies (Figure 3(a)) or estimated from

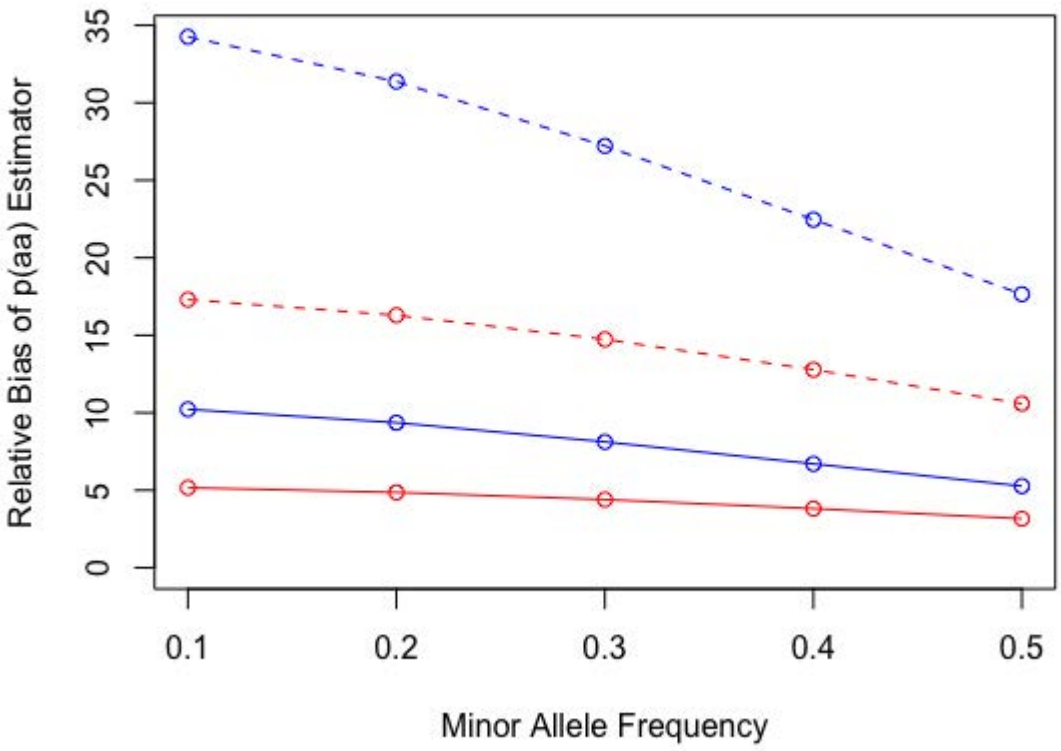
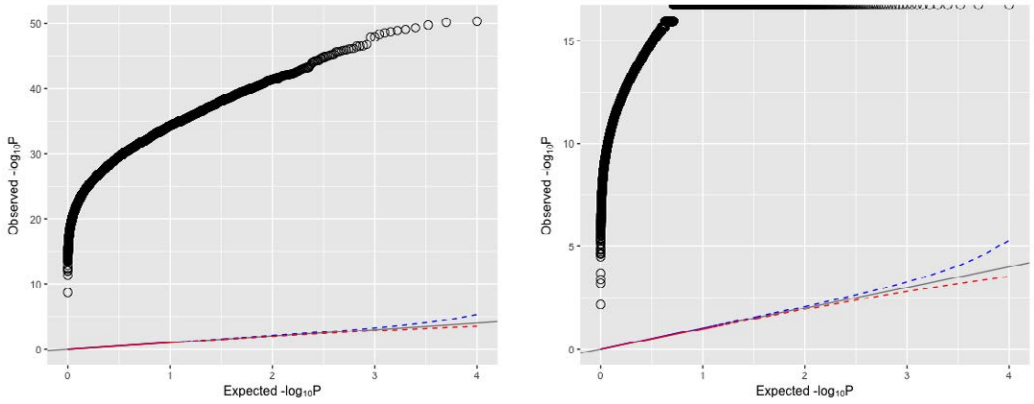


Figure 1: Absolute relative bias when using the genotype frequency from the controls, with $n_{aa|cntr}/n_{cntr}$ as an estimator of the population probability p_{aa}, where n_{cntr} is the sample size of controls and $n_{aa|cntr}$ is the number of controls with the aa genotype. The results shown are from a recessive model; the four lines are distinguished as follows, dashed red: $RR_2 = 2, \kappa = 0.15$; solid red: $RR_2 = 2, \kappa = 0.05$; dashed blue: $RR_2 = 3, \kappa = 0.15$; solid blue: $RR_2 = 3, \kappa = 0.05$.

the controls assuming HWE, the common practice (Figure 3(b)) . These observations also apply to the dominant and multiplicative models (second and third segments of Figure 3, where their specific parameter settings are also given in Table 1.

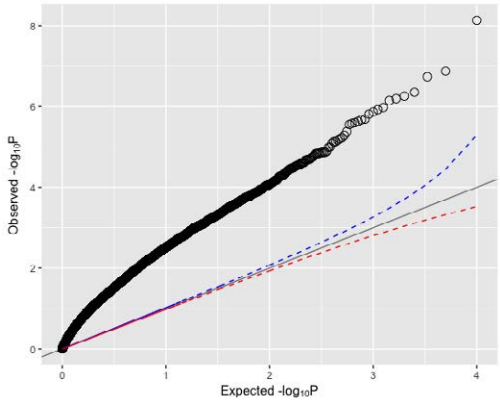
Finally, for additive model A1 (Table 1), the inflation of type I error is still clearly seen when the expected genotype frequencies in the Chi-square test are calculated using the population probabilities, that is, when either T1 or T3 are applied (Figure 4 (a) and (c)). When the expected frequencies are estimated based on the cases, then there is still a hint of type I error inflation (Figure 4 (b)). However, when the expected frequencies are estimated using the controls, then the QQ plot curve is almost entirely within the 95% confidence band. The above observations applied to the other two additive models considered (last segment of Figure 3).

Overall, we observed that the inflation of type I error is more severe when the expected genotype frequencies are calculated from the population parameters than when estimated from the observed data, for all models and parameter settings by comparing between the top and bottom panels of Figure 3.

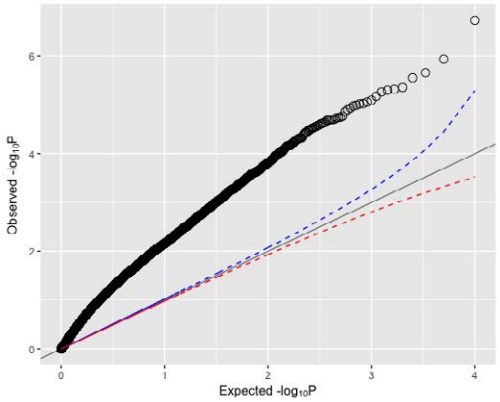


(a) T1

(b) T2

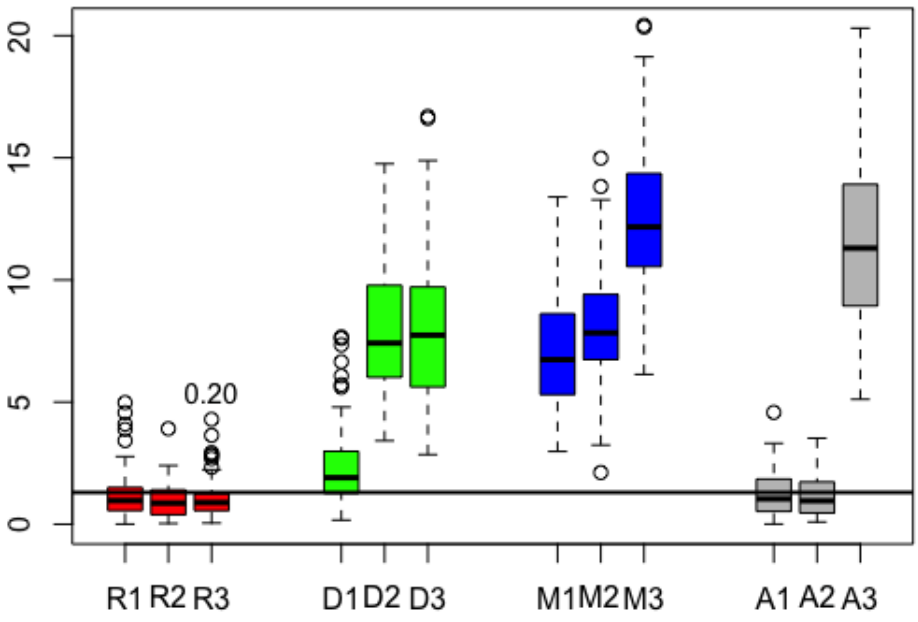


(c) T3

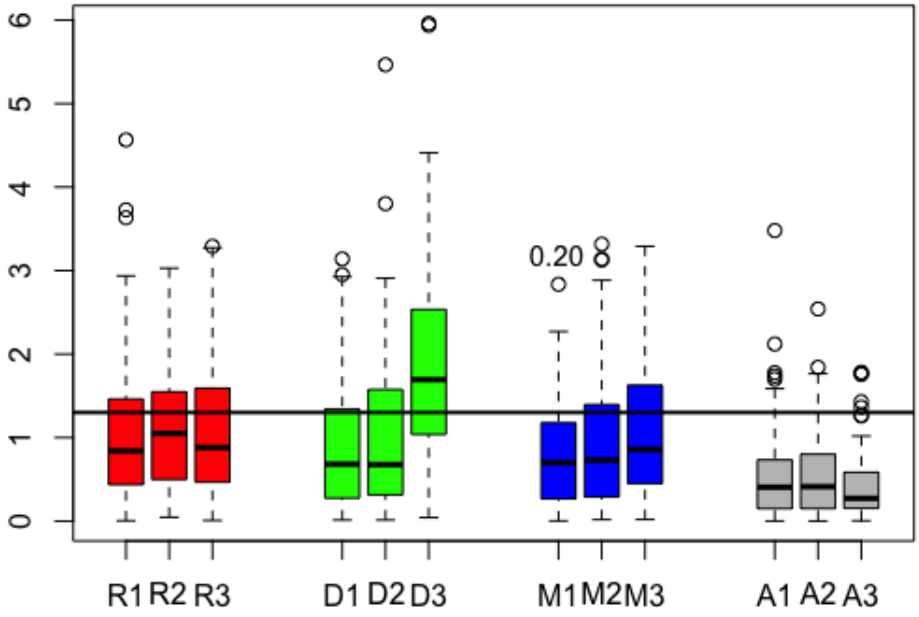


(d) T4

Figure 2: QQ plots of $-\log_{10}$ (p-value) of the observed vs. the expected. The area bounded by the red and blue dashed curves around the grey diagonal line represent a 95% confidence band. A curve above the band indicates inflated Type I error. The results shown are from data simulated under a recessive model, setting R1 in Table 1: $RR_2 = 1.4, \kappa = 0.15, p_a = 0.3$. Note that the peculiar feature in (b) is due to the limits on the number of significant digits.

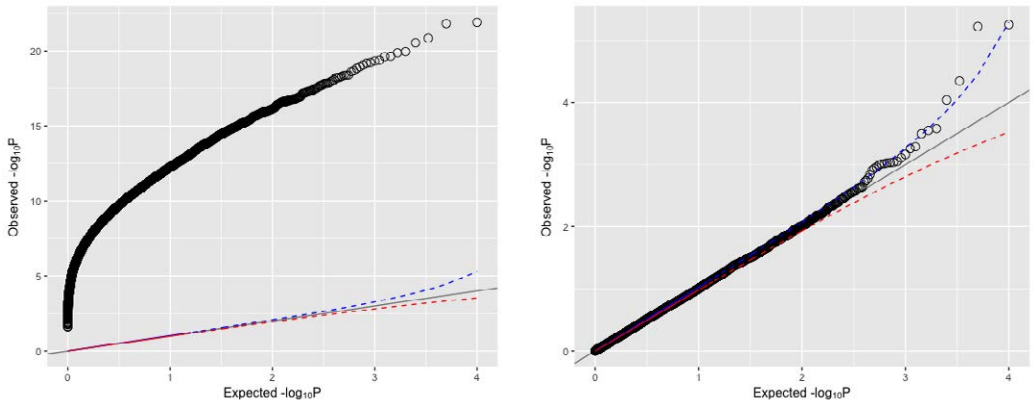


(a) T3



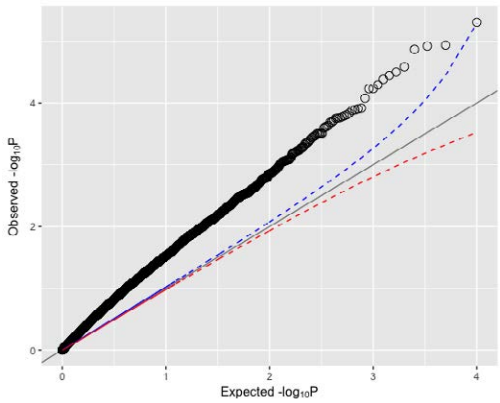
(b) T4

Figure 3: Boxplots of $-\log_{10}$ of the observed p-values over 10000 simulated data sets. The boxplots are arranged in four triplets, with each triplet representing the three specific parameter settings for each of the four types of models described in Table 1. The top plot (a) displays results from the T3 test and the bottom (b) from the T4 test. For the two boxplots (other than the additive tests) where the proportion of $-\log_{10}$ (p-value) > 1.3 (i.e. p-value < 0.05) is less than 25% (solid line above the upper end of the box), the percentages are indicated above the corresponding boxplots. As one can see from these two percentages, they are both at 20%, indicating type I error is also severely inflated. The solid black line across all boxplots marks 1.3 in the $-\log_{10}$ scale.

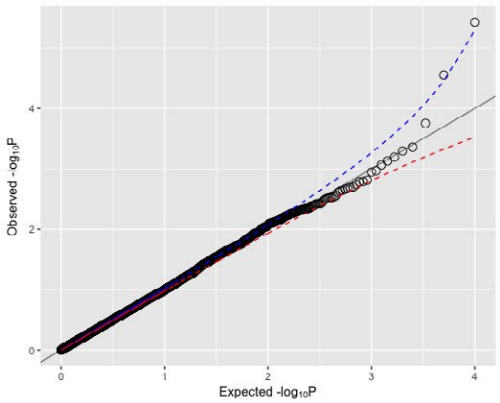


(a) T1

(b) T2



(c) T3



(d) T4

Figure 4: QQ plots of $-\log_{10}$ (p-value) of the observed vs. the expected. The area bounded by the red and blue dashed curves around the grey diagonal line represent a 95% confidence band. A curve above the band indicates inflated Type I error. The results shown are from data simulated under an additive model, setting A1 in Table 1: $RR_1 = 1.1, RR_2 = 1.2, \kappa = 0.15, p_a = 0.3$.

4. Discussion

It has been a long tradition that Hardy-Weinberg Equilibrium is carried out on controls in a GWAS with a case-control design as a quality control step to detect genotyping and other errors. However, whether this practice is a bona fide one has not been sufficiently investigated. Specifically, the reason for its appropriateness is rarely discussed, and if any, the assumption of rare disease is typically cited (Wang and Shete 2010). As such, its validity has not been seriously challenged in general. Although Brems (2015) discussed potential problems of HWE testing when using controls, the study therein was limited in scope as he only considered the setting in which the expected frequencies were taken to be from the general population, which deviates from usual practical usage. Further, bias was never discussed, thus he did not weigh in on the notion that controls as a whole are not a good representation of the population.

In this article, a much more comprehensive study was carried out. We used theoretical analysis on the bias when using genotype estimators based on controls for a general population in HWE to show that the assertion of controls being "similar" to the general population in terms of genetic makeup is in fact a myth rather than the truth, even if the disease is indeed rare. In fact, the bias and relative bias are only dependent on the population allele frequencies and the relative risks, not directly on the disease prevalence. Although the relative bias is seen to have a larger magnitude for a common disease than that for a rare disease, the bias is obviously apparent for a rare disease as well.

Armed with the evidence that controls may not truly represent the general population, we carried out a simulation study to investigate the degree of inflated type I error when controls are used to test for HWE. Using a variety of disease models and parameter settings, we show that, other than the additive model, there is typically severe inflation of type I error, regardless of whether known population frequencies or estimated frequencies based on controls are used as the expected in the Chi-square test. It is seen that the inflation of type I error may be smaller for a rare disease than a common disease, but for the former, even HWE testing in the disease-free controls can still lead to severe inflation of type I error.

Acknowledgements

This work was carried out as a summer project by the first author under the mentorship of the second author, whose work was supported in part by the National Science Foundation grant DMS-1208968. The first author would also like to acknowledge that this work was partly motivated by the research in the Masters Thesis of Mr. Matt Brems.

REFERENCES

Brems, M. (2015), "The Rare Disease Assumption: The Good, The Bad, and The Ugly," Master's Thesis. The Ohio State University.

Gomes, I., Collins, A., Lonjou, C., Thomas, N. S., Wilkinson, J., Watson, M., Morton, N. (1999), "Hardy-Weinberg Quality Control," *Annals of Human Genetics*, 63, 535–538.

Healey, C. S., Dunning, A. M., Teare, M. D., Chase, D., Parker, L., Burn, J., Chang-Claude, J., Mannermaa, A., Kataja, V., Huntsman, D. G., Pharaoh, P. D. P., Luben, R. N., Easton, D. F., Ponder, B. A. J. (2000), "A Common Variant in *BRCA2* Is Associated With Both Breast Cancer Risk and Prenatal Viability," *Nature Genetics*, 26, 362–364.

Hosking, L., Lumsden, S., Lewis, K., Yeo, A., McCarthy, L., Bansal, A., Riley, J., Purvis, I., Xu, C. F. (2004), "Detection of Genotyping Errors by Hardy-Weinberg Equilibrium Testing," *European Journal of Human Genetics*, 12, 395–399.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., Hirschhorn, J. N. (2008), "Genome-Wide Association Studies for Complex Traits: Consensus, Uncertainty and Challenges," *Nature Review Genetics*, 9, 356–369.

Moonesignhe, R., Yesupriya, A., Chang, M., Dowling, N. F., Khoury, M. J., Scott, A. J. (2010), "A Hardy-Weinberg Equilibrium Test for Analyzing Population Genetic Surveys With Complex Sample Designs," *American Journal of Epidemiology*, 171, 932–941.

Nielsen, D., Ehm, M. G., Weir, B. S. (1999), "Detecting Marker-Disease Association by Testing for Hardy-Weinberg Disequilibrium at a Marker Locus," *American Journal of Human Genetics*, 63, 1531–1540.

Phillips, J. A., Poling J. S., Phillips, C. A., Stanton, K. C., Austin, E. D., Cogan, J. D., Wheeler, L., Yu, C., Newman, J. H., Dietz, H. C., Loyd, J. E. (2008), "Synergistic Heterozygosity for *TGFβ1* SNPs and *BMPR2* Mutations Modulates the Age at Diagnosis and Penetrance of Familial Pulmonary Arterial Hypertension," *Genetics in Medicine*, 10, 359–365.

Salanti, G., Amountza, G., Ntzani, E. E., Ioannidis, J. P. A. (2005), "Hardy-Weinberg Equilibrium in Genetic Association Studies: an Empirical Evaluation of Reporting, Deviations, and Power," *European Jounral of Human Genetics*, 13, 840–848.

The Wellcome Trust Case Control Consortium. (2007), "Genome-Wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls," *Nature*, 446, 661–678.

Wang, J., Shete, S. (2010), "Using Both Cases and Controls for Testing HArdy-Weinberg Proportions in a Genetic Association Study," *Human Heredity*, 69, 212–218.

Wittke-Thompson, J. K., Pluzhnikov, A., Cox, N. J. (2005), "Rational Inferences about Departures from Hardy-Weinberg Equilibrium," *American Journal of Human Genetics*, 76, 967–986.

Yu, C., Zhang, S., Zhou, C., Sile, S. (2010), "Likelihood Ratio Test of Population Hardy Weinberg Equilibrium for Case-Control Studies," *Genetic Epidemiology*, 33, 275–280.

Yang, J., Lin, S. (2013), "Robust Partial Likelihood Approach for Detecting Imprinting and Maternal Effects Using Case-Control Families," *The Annals of Applied Statistics*, 7, 249–268.