X Chromosome Association Testing in Genome Wide Association Studies

Peter F. Hickey1* and Melanie Bahlo1,2

1Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
2Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, Australia

Genome wide association studies (GWAS) have revealed many fascinating insights into complex diseases even from simple, single-marker statistical tests. Most of these tests are designed for testing of associations between a phenotype and an autosomal genotype and are therefore not applicable to X chromosome data. Testing for association on the X chromosome raises unique challenges that have motivated the development of X-specific statistical tests in the literature. However, to date there has been no study of these methods under a wide range of realistic study designs, allele frequencies and disease models to assess the size and power of each test. To address this, we have performed an extensive simulation study to investigate the effects of the sex ratios in the case and control cohorts, as well as the allele frequencies, on the size and power of eight test statistics under three different disease models that each account for X-inactivation. We show that existing, but under-used, methods that make use of both male and female data are uniformly more powerful than popular methods that make use of only female data. In particular, we show that Clayton’s one degree of freedom statistic [Clayton, 2008] is robust and powerful across a wide range of realistic simulation parameters. Our results provide guidance on selecting the most appropriate test statistic to analyse X chromosome data from GWAS and show that much power can be gained by a more careful analysis of X chromosome GWAS data. Genet. Epidemiol. 35:664–670, 2011. © 2011 Wiley Periodicals, Inc.

Key words: X-linked markers; genetic association studies; association test statistics

INTRODUCTION

Genome wide association studies (GWAS) have yielded many statistically significant, replicable association results, even from simple statistical analyses using single marker tests that focus on the autosomes. The problem of testing for associations between phenotype and genotypes on the X chromosome in mixed-sex samples is one problem that has received surprisingly little attention [Clayton, 2008], yet many complex diseases show clear sex biases in disease frequency highlighting the potential involvement of the sex chromosomes in these diseases [Ober et al. 2008]. Poor analysis of the X chromosome data, which is routinely obtained in GWAS, represents a considerable loss in potential association signals given that the X chromosome makes up approximately 5% (resp. 2.5%) of the female (resp. male) genome.

Association tests on the X chromosome should incorporate into their model that a female has two X chromosomes while a male has one X chromosome and one Y chromosome. The standard autosomal association tests, such as the Cochran-Armitage trend test [Cochran, 1954; Armitage, 1955], are not immediately applicable to X chromosome data since males only have a single X chromosome. It should be noted that there are also two small pseudo-autosomal regions (PARs), comprising approximately 1.9% of the X chromosome, where loci are inherited like autosomal loci. Loci in the PARs can be analysed using standard autosomal methods and so we restrict our study to non-PAR X chromosome loci.

A further complication to the analysis of X chromosome data is due to the process of X-inactivation whereby most female X chromosome loci express only one allele from each pair of alleles. Early in development one of the pair of X chromosomes is chosen to be silenced at random and this silencing is then stably inherited in subsequent somatic cell divisions [Chow et al., 2005]. Thus, for non-PAR loci, a male will only carry one copy while a female will have approximately half her cells with one random copy active with the remainder of her cells having the other copy activated [Amos-Landgraf et al., 2006]. The copy number difference between sexes for the X chromosome raises the problem of testing suitable genetic models as the alternate hypothesis of association between genotype and phenotype. Males are hemizygous for the X chromosome and so have only two distinct genotypes, generically called $A-$ and $B-$. Females have three possible genotypes for X loci: $AA$, $AB$ and $BB$. Thus, simply, association testing in males is akin to allele based testing.

© 2011 Wiley Periodicals, Inc.
whereas association testing in females generally requires a genetic model to be specified.

In GWA studies the autosomal inheritance models usually tested for are the additive, dominant and recessive models. Statistical tests based on the additive model are perhaps the most commonly used in the analysis of data from GWA studies, having good power across a range of genetic models [though exceptions do exist, see Lettre et al., 2007].

Clayton [2008] developed X chromosome specific versions of common autosomal tests that explicitly account for X-inactivation; however, no comparison of these methods to existing methods was included in the paper. Zheng et al. [2007] have also published X-specific association tests and included a simulation study of their proposed methods. The included simulation study reflects a small GWAS with 200 cases and 200 controls. Their simulation is somewhat limited as it only considers an equal number of male and female samples and two different allele frequencies. They do however simulate data both under Hardy-Weinberg equilibrium (HWE) and with various departures from HWE to compare the performance of their six proposed association test statistics (Clayton’s tests were published subsequently to Zheng et al.’s study).

Zheng et al. [2007] show that the optimal choice of test statistic depends on whether HWE holds at the locus and whether males and females have the same risk allele at the locus. They suggest using six different association tests at each X chromosome single nucleotide polymorphism (SNP); SNPs are then ranked based on the minimum $P$-value of the six test statistics, noting of course that the $P$-values can no longer be interpreted as a usual $P$-value, needing to be adjusted according to the correlation between the test statistics.

Searching recent published GWAS literature reveals that neither Clayton’s nor Zheng et al.’s methods appear to have gained widespread use in GWAS analysis. The popular GWAS analysis software PLINK [Purcell et al., 2007] performs analyses on the X chromosome in two ways: (1) ignoring males entirely, or (2) regression analysis of all samples stratified by sex. Clearly it would be desirable to avoid (1) as we are not making use of all the data available, and (2) is undesirable when allele frequencies do not differ between sexes [Clayton, 2008, 2009]. There is currently no study comparing Clayton’s, Zheng et al.’s and PLINK’s methods simultaneously.

To address this we have performed extensive simulations to compare eight of the proposed statistics from Clayton, Zheng et al., and PLINK to assess the power and size (type 1 error rate) of each test under a range of realistic genetic models, allele frequencies and study designs. Our results show that much can be gained by more careful treatment of the X chromosome for association mapping.

**METHODS**

**DISEASE MODELS**

Due to the process of X-inactivation, the hemizygous $B-$ (resp. $A-$) males have the same active allele as the homozygous $BB$ (resp. $AA$) females. We impose this as a constraint on the penetrance parameters in the disease models we consider. Mathematically, let $P \in [0, 1]$ stand for a sample’s control/case status, then the constraint is that

\[
\Pr(P|AA \text{ female}) = \Pr(P|A - \text{ male}) \quad \text{and} \quad \Pr(P|BB \text{ female}) = \Pr(P|B - \text{ male}).
\]

This is similar to Clayton [2008, 2009] who argues that a consequence of X-inactivation is that, from the analysis viewpoint, the hemizygous males should be equivalent to the respective homozygous females at non-PAR X chromosome loci in the absence of interactions with other loci or environmental factors.

Many X-inactivation models have been proposed for females. Due to the random X-inactivation that takes place in females, an $AB$, or heterozygous, female is equally likely to have the $A$ or $B$ allele active in each cell. Since we are not able to ascertain which allele is randomly chosen we need to consider “average” effects of these alleles over all cells.

This averaging can take place in several ways. Model 1 considers the heterozygous females to have the same risk as the hemizygous male non-carriers 50% of the time and the same risk as the hemizygous male carriers 50% of the time. At a population level this means:

\[
\Pr(P|AB \text{ female}) = \frac{1}{2}[\Pr(P|A - \text{ male}) + \Pr(P|B - \text{ male})] = \frac{1}{2}[\Pr(P|AA \text{ female}) + \Pr(P|BB \text{ female})]
\]

(“additive constraint”)

where the last line follows from Equation (1). We call this the “additive constraint” as, mathematically, this constraint on the heterozygous female is identical to the additive model for autosomal genotypes.

Models 2 and 3 of X-inactivation consider the risk allele, $B$, to have a dominant or recessive action in females:

\[
\Pr(P|AB \text{ female}) = \Pr(P|BB \text{ female}) = \Pr(P|B - \text{ male})
\]

(“dominant constraint”)

\[
\Pr(P|AB \text{ female}) = \Pr(P|AA \text{ female}) = \Pr(P|A - \text{ male})
\]

(“recessive constraint”)

Thus disease models are synonymous with X-inactivation models in females. We perform simulations under all three disease/X-inactivation models for females.

**SIMULATION METHODS**

The simulation methodology and notation is adapted from the autosomal simulation methodology of Slager and Schaid [2001]. We use the $f$ superscript to denote female variables and the $m$ superscript to denote male variables.

Consider an X chromosome SNP with two possible alleles (generically labelled $A$ and $B$) under the assumption of HWE; for the X chromosome, HWE means that (i) allele frequencies in males and females are equal in one generation and (ii) HWE holds in females in that generation in the traditional sense [Zheng et al., 2007]. We define allele $B$ to be the risk allele in both males and females, i.e. $b^f = b^m = b$, ($0 < b < 1$) where $b^f$ (resp. $b^m$) is the frequency of the $B$ allele in the female (resp. male) population.

Genet. Epidemiol.
Under HWE, the distribution of genotypes in the female population is defined by:
\[
g' = \left( \begin{array}{c} g'_0 \\ g'_1 \\ g'_2 \end{array} \right) = (Pr(AA), Pr(AB), Pr(BB)) = ((1 - b)^2, 2b(1 - b), b^2)
\]
and the distribution of genotypes in the male population by
\[
g'' = \left( \begin{array}{c} g''_0 \\ g''_1 \\ g''_2 \end{array} \right) = (Pr(A - male), Pr(B - male))
\] 
\[= (1 - b, b).
\]
We assume a random sample of \( R \) cases and \( S \) controls giving a total sample size of \( N = R + S \). The trait prevalence, \( K \), is the proportion of people in the population with the trait of interest.

Let \( p_i \) denote the probability that a female case has genotype \( i, i \in \{ AA, AB, BB \} \); \( q_i \) the probability that a female control has genotype \( i \). For each female genotype we define a penetrance parameter
\[
f_i = Pr(\text{Affected|Female, genotype}=i).
\]
The values of these penetrance parameters are not usually known [Slager and Schaid, 2001]; however, they can be estimated from the trait prevalence, \( K \), and the genotypic relative risks (GRRs), \( \gamma_i \), by:
\[
f_0 = \frac{K}{\sum_{i \in \{ AA, AB, BB \}} f_i}, \quad f_1 = f_0 \gamma_1 \quad \text{and} \quad f_2 = f_0 \gamma_2
\]
where the GRRs are pre-defined as \( \gamma_1 = f_1/f_0 \) and \( \gamma_2 = f_2/f_0 \).

Following Equation (1), \( g' = Pr(\text{Affected|A - male}) = f_0 \) and \( g'' = Pr(\text{Affected|B - male}) = f_2 \). This of course implies that \( g'' = g''_{0} = \gamma_2 \).

The values of \( p \) and \( q \) in females are calculated from the equations
\[
p_i = \sum_{j \in \{ AA, AB, BB \}} f_i s_j, \quad q_i = \sum_{j \in \{ AA, AB, BB \}} (1 - f_i) s_j, \quad i \in \{ 0, 1, 2 \}
\]
and in males from
\[
p_j = \sum_{i, j \in \{ AA, AB, BB \}} f_j s_i, \quad q_j = \sum_{i, j \in \{ AA, AB, BB \}} (1 - f_j) s_i, \quad j \in \{ 0, 2 \}.
\]

Genotypes can then be simulated conditional on phenotype using a multinomial distribution; for example, the \( p' \) female cases’ genotypes follow a trinomial distribution with probability vector \( p' = (p'_0, p'_1, p'_2) \).

**X CHROMOSOME ASSOCIATION TESTS**

We briefly explain each statistic here and refer the reader to the original articles for formulae and further details. All these methods are single-marker tests and assume the same risk allele in males and females unless otherwise noted.

1. \( Z^2_A \) [Zheng et al., 2007]: a simple allele counting test that compares the frequency of each allele within cases and controls. Males will only have half the impact on the results of this test as females [Clayton, 2008]. Provided HWE holds, the \( Z^2_A \) test has an approximate \( \chi^2 \) distribution on 1 degree of freedom (df) under the null hypothesis.

2. \( Z^2_{mg} \) [Zheng et al., 2007]: combines an allele-counting test of the male genotype table with a Cochran-Armitage trend test of the female genotype table and follows a \( \chi^2 \) distribution on 1 df under the null hypothesis.

3. \( Z^2_{mg'A} \) [Zheng et al., 2007]: combines an allele-counting test of the male genotype table with a Cochran-Armitage trend test of the female genotype table. The test allows for males and females to have different risk alleles at a locus and follows a \( \chi^2 \) distribution on 2 df under the null hypothesis.

4. \( Z^2_{mg'A} \) [Zheng et al., 2007]: combines an allele-counting test of the male genotype table with an allele-counting test of the female genotype table and follows a \( \chi^2 \) distribution on 1 df under the null hypothesis. As the test involves an allele-counting test of the female genotype table it requires HWE to hold in females for the test to be of the correct size.

5. \( S_1 \) [Clayton, 2008]: analogous to a Cochran-Armitage trend test of a combined male and female genotype table and follows a \( \chi^2 \) distribution on 1 df under the null hypothesis.

6. \( S_2 \) [Clayton, 2008]: analogous to a Pearson’s \( \chi^2 \) test on 2 df of a combined male and female genotype table and follows a \( \chi^2 \) distribution on 2 df under the null hypothesis.

7. \( CAf \): a Cochran-Armitage trend test of only the female genotype data.

8. \( \chi^2 \): Pearson’s \( \chi^2 \) test on 2 df of only the female genotype data.

Tests 1–6 use both male and female data while tests 7–8 are female-only tests and are the default tests when a naive analysis of X chromosome data is performed using the –model option in the popular GWAS analysis software PLINK (v1.07) [Purcell et al., 2007]. The \( S_1 \) and \( S_2 \) statistics are available in the R package \texttt{snpMatrix} [Clayton and Leung, 2010]. Zheng et al. [2007] also developed two statistics, \( Z^2_{mg} \) and \( Z^2_{mg'A} \), to specifically test for sex-linked allelic heterogeneity at a locus (i.e. different risk alleles between the sexes at a locus). The asymptotic null distributions of these statistics are somewhat more difficult to compute and the authors do not provide software to compute the relevant quantiles. Furthermore, such risk loci are outside the scope of our study so we do not include these two tests.

**SIMULATION PARAMETERS**

We fix the number of cases at \( R = 2,000 \) and the number of controls at \( S = 2,000 \), chosen to reflect contemporary GWAS sample sizes. We selected disease prevalences of \( K = 1/10, 1/100 \) and 1/1000, typical of complex diseases such as asthma, Type 1 diabetes and multiple sclerosis, respectively. Throughout we consider males and females to have the same prevalence of the disease.

We vary the disease model, the sex ratios within the case and control cohorts and the the allele frequency of the risk allele in the population. Since \( \gamma_2 = \gamma_2 \), the disease model is completely specified by the GRR vector \( \gamma = (\gamma_1, \gamma_2) \). The null model corresponds to \( \gamma^{null} = (1, 1) \), the additive constraint to \( \gamma^{add} = (1 + t)/2, 1) \), the dominant constraint to \( \gamma^{dom} = (t, 1) \), and the recessive constraint to \( \gamma^{rec} = (1, t) \), where \( t > 1 \), with larger \( t \) corresponding to a locus conferring greater risk.
The percentages of the case and control cohorts that are female are given by $\%_{\text{case}}$ where $\%_{\text{case}} = \text{the percentage of cases that are female},$ and $\%_{\text{control}} = \text{the percentage of controls that are female}.$ For example, $\%_{\text{case}} = 75\%$ means 75% of cases are female and 25% are male. The parameter $\%_{\text{control}}$ is chosen to reflect either a GWAS where cases and controls are matched by sex (i.e. $\%_{\text{case}} = \%_{\text{control}}$) or a GWAS where the control cohort is 50% female ($\%_{\text{control}} = 50\%$). The latter is to simulate a GWAS where the control cohort is obtained from a genotype database such as the Illumina iControlDB [http://www.solexa.com/science/icontroldb.ilmn]. When such a database is used the control cohort is unlikely to be matched by sex to the case cohort and, assuming random sampling, approximately 50% of the control cohort will be female.

We consider the following parameter values for the simulations:

- $b \in \{0.01, 0.02, \ldots, 0.05, 0.1, \ldots, 0.95, 0.96, \ldots, 0.99\}$
- $\gamma \in \{\gamma_{0}^{\text{null}}, \gamma_{0}^{\text{add}}, \gamma_{0}^{\text{dom}}, \gamma_{0}^{\text{rec}}\}$
- $\%_{\text{case}} \in \{\%_{\text{null}}, \%_{\text{add}}, \%_{\text{dom}}, \%_{\text{rec}}\}$

This leads to 1,350 combinations of parameters for each level of the disease prevalence, $K,$ we refer to a combination of parameters as a simulation design.

The simulations and implementation of the eight tests statistics were written in the statistical programming language R version 2.10.0 [R Development Core Team, 2009].

We simulate 100,000 replicates for each simulation design and calculate each of the eight test statistics on the genotype spectrum when there is a large proportion of male cases and the control cohort is not matched by sex (see the leftmost panel of Fig. 1).

**RESULTS**

We have selected 567 of the 4,050 simulation designs to illustrate the main results of our study, all of which use a disease prevalence of $K = 1/1,000.$ Since $K$ simply acts as a scalar multiplier in determining the penetrance parameters, the effect in our simulations of increasing $K$ is to increase the power of all tests by a similar level and so does not alter the conclusions of our study (see Supplementary Material, Sections 6.1 and 6.2 for the $K = 1/10$ and $K = 1/100$ simulation results). The effect of matching cases and controls by sex is similarly minor and so we focus on those simulations using a control cohort that is 50% female (see Supplementary Material 6.1, 6.2 and 6.3 for the simulations using cases and controls matched by sex). These 567 simulation designs are summarised in Table I and have been chosen to highlight the effect of:

- How the frequency of the risk allele in the population affects the size and power of each test.
- How the sex ratios of the case and control cohorts affect the size and power of each test.
- How the underlying disease model affects the power of each test.

**SIZE OF EACH TEST**

Figure 1 reports the empirical size of each test, which we calculate as the proportion of times the test statistic exceeds the relevant $\gamma^{2}$ quantile corresponding to the nominal type 1 error of 0.0001. Using the Wilson [1927] method, we determine that type 1 error rates falling within the 99% confidence interval $[0.00005, 0.00022]$ are not significantly different from the nominal size of 0.0001.

Most tests are of the correct size, most of the time; the only systematic pattern we observe is that when tests are of the incorrect size they are conservative, not anticonservative, and that this generally occurs near the boundaries of the allele frequency spectrum. The female-only $\chi^{2}$ test is particularly prone to this due to the smaller sample size used by this test. The $S_{2}$ test records slightly conservative results right across the allele frequency spectrum when there is a large proportion of male cases and the control cohort is not matched by sex (see the leftmost panel of Fig. 1).

**POWER OF EACH TEST**

The power of each test is the proportion of times the test reports a significant result out of the 100,000 replicates, where the nominal significance level is $\alpha = 0.0001.$ Figures 2 and 3 show the power curves for each of the eight-test statistics at low-risk loci ($t = 1.25$) and mid-risk loci ($t = 1.5$), respectively.

The most obvious result is that the female-only tests, $\chi_{f}^{2}$ and $\text{CA}_{f},$ are vastly underpowered when compared to any of Clayton’s or Zheng et al.’s tests. Only when an allele confers a high level of risk, and the study includes a large proportion of females, do the female-only tests have comparable power to tests 1–6 (see, for example, the rightmost panel of Supplementary Fig. 22). There is little power under all disease models to detect associations at SNPs with allele frequency $b$ near the boundaries of the $[0,1]$ interval, as is to be expected.

For tests 1–6, which make use of both male and female data, there is generally greater power to detect X chromosome associations when the case cohort is sampled with a male-bias, regardless of the underlying disease model. For instance, in Figure 2 the power of the $S_{1}$ statistic to detect an association at a SNP following the $0.50^{\text{–}}\gamma_{0}^{\text{add}}$ model can vary from 11% to 32% depending on whether the case cohort is 25% or 75% male. This same pattern holds true regardless of whether the case and control cohorts are matched by sex (the range is 14% to 27% when cases and controls are matched by sex for the $0.50^{\text{–}}\gamma_{0}^{\text{rec}}$ model, see Supplementary Fig. 19).

There appears to be little power advantage in using a control cohort that is matched by sex to the case cohort as opposed to using a common control cohort comprised of an equal number of males and females. Indeed, in some instances the common control cohort approach is more statistically powerful than the sex-matched approach (e.g. compare models $\text{CA}_{\text{rec}} = \gamma_{1.5}^{\text{add}}$ to models $\gamma_{1.5}^{\text{add}}$ and $\gamma_{1.5}^{\text{add}}$ in Figure 2 and Supplementary Fig. 19, respectively). This result is reassuring to those wishing to use a genome bank control

---

**TABLE 1. Simulation parameters for the 567 simulation designs to be discussed in Results section**

| Parameter | Levels |
|-----------|--------|
| $b$       | 0.01, 0.02, ..., 0.05, 0.1, ..., 0.95, 0.96, ..., 0.99 |
| $\gamma$  | $\gamma_{0}^{\text{null}}, \gamma_{0}^{\text{add}}, \gamma_{0}^{\text{dom}}, \gamma_{0}^{\text{rec}}$ |
| $\%_{\text{case}}$ | $\%_{\text{null}}, \%_{\text{add}}, \%_{\text{dom}}, \%_{\text{rec}}$ |
| $\%_{\text{control}}$ | $\%_{\text{null}}, \%_{\text{add}}, \%_{\text{dom}}, \%_{\text{rec}}$ |

Genet. Epidemiol.
cohort that cannot be matched by sex to the case cohort, though of course there may well be other advantages in matching cases and controls by sex not explicitly related to X chromosome association testing.

For the $\gamma_{\text{add}}$ disease models, statistical power is greatest around the centre of the $b \in [0,1]$ interval; however, it is not quite symmetrical around $b = 0.5$ with maximal power occurring nearer to $b = 0.4$. The $S_1$ statistic is typically the most powerful for these $\gamma_{\text{add}}$ models, though the $Z_{\text{mfG}}$ test is the better choice when the case cohort is strongly female-biased.

The power curves of the test statistics under the $\gamma_{\text{dom}}$ and $\gamma_{\text{rec}}$ models are more-or-less reflected around $b = 0.5$. The intuitive explanation for this is that the “dominant constraint” with a common allele ($b = 0.5$) is the same as the “recessive constraint” with a rare allele ($b \ll 0.5$)—there is little power to detect associations at SNPs under the “recessive constraint” when the risk allele is rare in the case cohort.

Fig. 1. Along the x-axis is the frequency of the $B$ allele in males and females at a population level. The dashed red line indicates the nominal size of 0.0001 and the dashed black lines indicate the lower and upper bounds of the corresponding Wilson 99% confidence interval [0.00005, 0.00022].

Fig. 2. Along the x-axis is the frequency of the $B$ allele in males and females at a population level. The top rows are the results for the additive constraint, the middle rows are for the dominant constraint and the bottom rows for the recessive constraint. The proportion of cases that are female increases from the leftmost column to the rightmost column (25%, 50%, 75%).
population \((b \ll 0.5)\); conversely, there is little power to detect associations at SNPs under the “dominant constraint” when the risk allele is so common in the population that most people have it regardless of phenotype \((b \gg 0.5)\).

Similar to the \(\gamma^{add}\) simulations, the \(S_1\) statistic is again among the most powerful for the \(\gamma^{dom}\) and \(\gamma^{rec}\) simulations with the \(Z_{mfG}\) statistic a better choice when there is a large proportion of female cases. This suggests the \(Z_{mfG}\) statistic gives greater weight to association signals detected in females than in males and vice-versa for the \(S_1\) statistic.

The \(S_2\) statistic is on 2 df and is not particularly powerful under the \(\gamma^{add}\) simulations. However, the additional df provides flexibility that results in the \(S_2\) statistic being of comparable, or greater, power than both \(S_1\) and \(Z_{mfG}\) under the \(\gamma^{dom}\) and \(\gamma^{rec}\) simulations. The other 2 df test, \(Z_{C\alpha}\), does not see quite the same performance increase, though it is the most powerful test for high risk, dominant loci (\(\gamma^{dom}\) simulations) where \(b\) is large (see Supplementary Figs. 10, 17 and 22).

As \(t\) increases, Clayton’s and Zheng et al.’s tests have near identical power at the \(\alpha = 0.0001\) significance levels across most values of \(b\) while the female-only tests, \(\chi_f^2\) and \(CA_f\), still lag behind.

The \(S_1\) statistic is among the most powerful statistics for all our simulations. When the case cohort is sampled with a female bias the \(Z_{mfG}\) statistic has a slight advantage and when the disease model does not satisfy the “additive constraint” the additional df can make the \(S_2\) statistic a more powerful test.

**DISCUSSION**

The main difficulty in comparing genetic association test statistics via a simulation study is simulating data that is reflective of the true underlying biology. This is particularly true for the X chromosome as there have been very few X chromosome genotype-phenotype associations published in the literature and their biological mechanisms are not well understood. Ideally we would compare these eight-test statistics on real GWAS data, containing a known X chromosome association; however, a search of recent GWAS literature and databases, such as the Human Genome Variation database (http://www.hgvsbaseg2p.org), failed to identify any suitable, freely available data set.

We believe the three classes of disease model that we have considered (\(\gamma^{add}, \gamma^{dom}, \gamma^{rec}\)) capture a large proportion of realistic scenarios, though we are not aware of any biological study to explicitly support these. Zheng et al. [2007] use a different methodology to simulate X chromosome genotypes by equating \(p_0^x = p_1 + p_1^2/2\) and similarly for \(q_0^x, q_1^x\) and \(q_1^x\). An implication of this simulation model is that a male hemizygous for the risk allele is at greater risk than a female that is homozygous for the risk allele, a scenario that does not seem to us to be intuitively obvious. Zheng et al.’s model is also formed without reference to any biological study to support it.

Clayton’s two score test statistics for X chromosome association testing were published subsequently to Zheng et al. [2007]. We have investigated Clayton’s score test statistics and shown these to be generally more powerful than the methods considered in Zheng et al. [2007].

There are several other important differences between our simulation study and that published in Zheng et al. [2007], namely:

- We have considered a much more realistic GWAS sample size than Zheng et al. [2007] which considered a GWAS with only 200 cases and 200 controls.
- We extensively investigated the effect of the sex ratios within each cohort on the statistical power to identify X
chromosome associations, whereas Zheng et al. [2007] only considered a GWAS with an equal number of males and females.

- Our simulations have considered the full spectrum of allele frequencies, from 0.01 to 0.99, while Zheng et al. [2007] dealt only with allele frequencies of 0.1 and 0.3.
- Our simulations also considered disease prevalences of \( K = 1/100 \) and \( K = 1/1000 \) in addition to the value of \( K = 1/10 \) considered by Zheng et al. [2007].

Extensions to our simulation study include allowing the allele frequency to differ between males and females (i.e. departures from HWE) and allowing different trait prevalences in males and females. If allele frequencies differ between the sexes then it is generally necessary to stratify the analysis by sex, otherwise sex will be confounded with the association. In the situation considered in our simulation study, where the sex ratio varies between the case and control cohorts and allele frequencies do not differ between males and females, sex is not truly a confounder [Clayton, 2009]. In this situation, stratification (e.g. by using the logistic sex options in PLINK) may lead to a considerable loss in power, whereas Clayton’s tests do not require stratification if there is no difference in allele frequencies between the sexes. Clayton [2009] provides a comprehensive discussion of when sex may be a confounding variable and when it is necessary to stratify the analysis by sex.

We believe a caveat is warranted to the claim in Clayton [2008] that \( S_1 \) and \( S_2 \) do not assume HWE. It has previously been shown that HWE on the X chromosome requires that (i) allele frequencies in males and females are equal in one generation and (ii) HWE holds in females in that generation in the traditional sense [Zheng et al., 2007]. While neither \( S_1 \) or \( S_2 \) assume (ii), it is clear that both tests require (i) and so violation of equal allele frequencies between the sexes would be cause for concern. The effect of both (i) and (ii) on the test statistics \( Z_A^2, Z_C^2, Z_{mg}^2, \) and \( Z_{mgA}^2 \) has been previously studied [Zheng et al., 2007]; however, the effect of these assumptions on \( S_1 \) and \( S_2 \) is yet to be explored.

It is important to bear in mind that while many diseases do display a sex-bias, this may be due to reasons other than genetic variation on the sex chromosomes; it could be, for example, due to sex differences in the endocrine and immune systems [Ober et al., 2008]. However, we believe an improved analysis of X chromosome data, such as by using Clayton’s association testing methods, will be beneficial to many GWAS studies not just studies of those diseases that display a sex-bias.

We have compared the performance of eight-test statistics to assess the size and power of these tests to detect associations between X chromosome genotypes and phenotype across a wide range of realistic disease models, allele frequencies and study designs. We have shown that several existing methods that make use of both male and female samples are uniformly more powerful that methods that only use female data. In particular our results suggest Clayton’s \( S_1 \) statistic as having strong all round performance across a wide range of disease models, sex ratios and allele frequencies.

**ACKNOWLEDGMENTS**

We thank the reviewers of both this and an earlier version of the manuscript for many helpful comments and suggestions. We also thank Richard Huggins for helpful discussions. Melanie Bahlo is supported by a Career Development Award from The National Health and Medical Research Council of Australia (NHMRC) and the NHMRC Medical Genomics Bioinformatics and Proteomics Program.

**REFERENCES**

Amos-Landgraf JM, Cottile A, Plenge RM, Friez M, Schwartz CE, Longshore J, Willard HF. 2006. X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. Am J Hum Genet 79:493–499.

Armitage P. 1955. Tests for linear trends in proportions and frequencies. Biometrics 11:375–386.

Chow JC, Yen Z, Ziesche SM, Brown CJ. 2005. Silencing of the mammalian X chromosome. Annu Rev Genomics Hum Genet 6:69–92.

Clayton D. 2008. Testing for association on the X chromosome. Biostatistics 9:593–600.

Clayton DG. 2009. Sex chromosomes and genetic association studies. Genome Med 1:110.

Clayton D, Leung H. 2010. \( \text{suppMatrix}: \) The \text{supp.matrix} and \text{X.snp.matrix} classes. R package version 1.12.0 http://www-gene.cimr.cam.ac.uk/clayton/software/

Cochran W. 1954. Some methods for strengthening the common \( \chi^2 \) tests. Biometrics 10:417–451.

Lettre G, Lange C, Hirschhorn JN. 2007. Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol 31:358–62.

Ober C, Loisel DA, Gilad Y. 2008. Sex-specific genetic architecture of human disease. Nat Rev Genet 9:911–922.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker, PIW, Daly MJ, Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, ISBN 3-900051-07-0.http://www.R-project.org

Slager SL, Schaid DJ. 2001. Case-control studies of genetic markers: power and sample size approximations for Armitage’s test for trend. Hum Hered 52:149–153.

Wilson EB. 1927. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22:209–212.

Zheng G, Joo J, Zhang C, Geller NL. 2007. Testing association for markers on the X chromosome. Genet Epidemiol 31:834–843.