MDPI

Article

XCMAX4: A Robust X Chromosomal Genetic Association Test Accounting for Covariates

Youpeng Su †, Jing Hu †, Ping Yin, Hongwei Jiang, Siyi Chen, Mengyi Dai, Ziwei Chen and Peng Wang *

Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China; yp_s@hust.edu.cn (Y.S.); jing_h2019@163.com (J.H.); pingyin2000@126.com (P.Y.); jhwccc0@sina.com (H.J.); chen_siyi@hust.edu.cn (S.C.); daimengyi0505@163.com (M.D.); happyczw1998@163.com (Z.C.)
* Correspondence: pengwang_stat@hust.edu.cn
† These authors contributed equally to this work.

Abstract: Although the X chromosome accounts for about 5% of the human genes, it is routinely excluded from genome-wide association studies probably due to its unique structure and complex biological patterns. While some statistical methods have been proposed for testing the association between X chromosomal markers and diseases, very a few of them can adjust for covariates. Unfortunately, those methods that can incorporate covariates either need to specify an X chromosome inactivation model or require the permutation procedure to compute the *p* value. In this article, we proposed a novel analytic approach based on logistic regression that allows for covariates and does not need to specify the underlying X chromosome inactivation pattern. Simulation studies showed that our proposed method controls the size well and has robust performance in power across various practical scenarios. We applied the proposed method to analyze Graves' disease data to show its usefulness in practice.

Keywords: X chromosome; logistic regression; covariates; robust; Graves' disease

1. Introduction

Many diseases exhibit a gender preference, such as autoimmune diseases, cardiovascular diseases, psychiatric diseases, and cancer, implying that genetic variants on the X chromosome play an important role in sex differences [1–5]. However, most genome-wide association studies (GWAS) routinely exclude the analysis of X-chromosomal variants probably because the X chromosome has a unique structure and complex biological patterns [6–8]. Females have one more X chromosome than males, and to balance gene expression on the X chromosome with that of males, one of the female X chromosomes is inactivated in the early embryo [9]. Usually, the process of X chromosome inactivation (XCI) is considered random (XCI-R) [10], i.e., for an X-linked gene, nearly 50% of the cells have the paternal allele active while the rest cells have the maternal allele active. However, studies have shown that skewed XCI (XCI-S) is more biologically plausible [11]. XCI-S is a non-random process, which has been defined as a significant deviation from XCI-R, for instance, the inactivation of one of the alleles in more than 75% of cells [12]. In addition, up to 25% of X-linked genes can escape from XCI (XCI-E) [9]. Both alleles in the genes under XCI-E will be active, which are similar to autosomal genes.

To account for the unique characteristics of the X chromosome, several statistical methods have been developed for testing the association between X chromosomal markers and diseases [13–18]. However, very a few of them can adjust for covariates. In large-scale GWAS, spurious associations may occur due to the influence of additional covariates, such as sex, age, and population structure [19,20]. Particularly on the X chromosome, if the sex ratios differ between cases and controls, then sex will be a confounder when the allele frequency of females is unequal to that of males. In practice, a natural way to adjust for

covariates is to build a regression model, and logistic regression is generally adopted for binary traits. Based on the logistic regression framework, Gao et al. [15] integrated four tests (FM_{01}, FM_{02}, FM_F, and FM_S) in the software toolset XWAS. In FM_{01} and FM_{02}, three genotypes of females are both coded by 0, 1, and 2, while two genotypes of males are coded by 0 and 1 for FM_{01} and by 0 and 2 for FM_{02}. In the latter, males are treated as homozygous females to reflect the dosage compensation relationship between the two sexes. Hence, FM_{01} and FM_{02} assume that the underlying XCI patterns are XCI-E and XCI-R, respectively. On the other hand, FM_F and FM_S build logistic regressions for females and males separately and then combine the two *p* values using Fisher's and Stouffer's methods, respectively. However, these two methods do not take any XCI patterns into consideration and thus may suffer from substantial power loss if the test marker is undergoing XCI. Wang et al. [14] proposed another approach (denoted by maxLR) that can consider four special XCI patterns simultaneously: XCI-R, XCI-E, XCI-S fully toward the normal allele (XCI-SN), and XCI-S fully toward the risk allele (XCI-SR). In their method, three genotypes of females are coded as 0, γ, and 2 under XCI, where $\gamma \in (0, 2)$ measures the degree of skewness of XCI. For instance, $\gamma = 0$ (2) represents that all the risk (normal) alleles are inactivated in heterozygous females, which corresponds to the XCI-SN (XCI-SR) pattern. While maxLR has robust performance in power, its *p* value is evaluated based on the permutation procedure, which is very computationally intensive, especially in GWAS. Hence, it is still desirable to develop a robust method that can both adjust for covariates and analytically calculate the *p* value.

To fill this gap, this article proposed a novel statistical method to test the association between X chromosomal markers and a specific disease. Our method, which is also based on logistic regression, is robust because it does not require assigning a specific XCI pattern. Further, our method can compute the *p* value without the resample procedure by directly using the rhombus formula. We implemented an extensive simulation study to compare the performance of our approach with the existing ones. Simulation results showed that our method controls the size well and can maintain relatively high power across a variety of scenarios. Finally, we applied our proposed approach to the Graves's disease data to demonstrate its practical use.

2. Method

Consider an X-linked SNP with deleterious allele *A* and normal allele *a*. Then, there are three possible genotypes for females: *aa*, *Aa*, and *AA*, and two for males: *a* and *A*. We assume a binary variable *D* for the disease of interest with $D = 1$ (0) representing individuals with (without) the disease. $\boldsymbol{X} = \left[x_1, \cdots, x_p \right]'$ denotes the *p* covariates that need to be adjusted in the model, where $x_1 \equiv 1$ is the model intercept and x_2 represents the binary variable with 1 being female and 0 being male. We further assume that the relationship between the phenotype and genotype for individual *i* can be constructed by the following logistic regression model:

$$\log\left(\frac{\Pr(D_i = 1 | G_i, \boldsymbol{X_i})}{\Pr(D_i = 0 | G_i, \boldsymbol{X_i})} \right) = \boldsymbol{X_i}' \boldsymbol{\alpha} + \beta G_i \qquad (1)$$

where the subscript *i* denotes the *i*th individual, *G* is the genotypic score, $\boldsymbol{\alpha} = \left(\alpha_1, \alpha_2, \cdots, \alpha_p \right)'$, and β represents the regression coefficients for the covariates and the genotypic score. Note that the genotypic score depends on the underlying XCI pattern. According to the coding strategy by Wang et al. [14], G_i can be written in the following uniform form

$$G_i(Z_1, Z_2) = 2I_i(AA) + Z_1 I_i(Aa) + Z_2 I_i(A),$$

where $I(.)$ is the indicator function, and Z_1 and Z_2 are unknown parameters depending on the underlying XCI pattern. For instance, when the SNP is undergoing XCI, Z_1 and Z_2 can be assigned by γ and 2, respectively. In this coding strategy, γ is a measure of the skewness of XCI, and males are treated as homozygous females to reflect the dosage compensation.

Table 1 lists the genotypic scores for all five genotypes and the corresponding values of Z_1 and Z_2 under the four special XCI patterns.

Table 1. The genotypic scores for five genotypes and their corresponding values of Z_1 and Z_2 under the four special XCI patterns.

XCI Pattern	*aa*	*Aa*	*AA*	*a*	*A*	Z_1	Z_2
XCI-SN	0	0	2	0	2	0	2
XCI-R	0	1	2	0	2	1	2
XCI-SR	0	2	2	0	2	2	2
XCI-E	0	1	2	0	1	1	1

We chose the score statistic to test the null hypothesis: $\beta = 0$ because the association tests for all the SNPs share the same null model. For a total sample size of n, the score function can be derived as

$$U(Z_1, Z_2) = \sum_{i=1}^{n} \{ G_i(Z_1, Z_2)[D_i - \Pr(D_i = 1|\boldsymbol{X_i})] \},$$

where $\Pr(D_i = 1|\boldsymbol{X_i}) = \frac{exp^{\boldsymbol{X_i}'\hat{\alpha}}}{1+exp^{\boldsymbol{X_i}'\hat{\alpha}}}$ is the disease probability estimated for individual i without considering the genotype (details of the derivation are given in Appendix A). The information matrix of (1) can be written as follows:

$$I(Z_1, Z_2) = \begin{pmatrix} I_\beta(Z_1, Z_2) & I_{\beta\boldsymbol{\alpha}}(Z_1, Z_2) \\ I_{\beta\boldsymbol{\alpha}}(Z_1, Z_2)' & I_{\boldsymbol{\alpha}} \end{pmatrix},$$

where

$$I_\beta(Z_1, Z_2) = \sum_{i=1}^{n} G_i(Z_1, Z_2)^2 [1 - \Pr(D_i = 1|\boldsymbol{X_i})]\Pr(D_i = 1|\boldsymbol{X_i}),$$

$$I_{\beta\boldsymbol{\alpha}}(Z_1, Z_2) = \left(\sum_{i=1}^{n} X_{i1} G_i(Z_1, Z_2)[1 - \Pr(D_i = 1|x_i)]\Pr(D_i = 1|x_i), \cdots, \sum_{i=1}^{n} X_{ip} G_i(Z_1, Z_2)[1 - \Pr(D_i = 1|x_i)]\Pr(D_i = 1|x_i) \right),$$

and

$$I_{\boldsymbol{\alpha}} = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^2 [1 - \Pr(D_i = 1|\boldsymbol{X_i})]\Pr(D_i = 1|\boldsymbol{X_i}) & \cdots & \sum_{i=1}^{n} X_{i1} X_{ip} [1 - \Pr(D_i = 1|\boldsymbol{X_i})]\Pr(D_i = 1|\boldsymbol{X_i}) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} X_{ip} X_{i1} [1 - \Pr(D_i = 1|\boldsymbol{X_i})]\Pr(D_i = 1|\boldsymbol{X_i}) & \cdots & \sum_{i=1}^{n} X_{ip}^2 [1 - \Pr(D_i = 1|\boldsymbol{X_i})]\Pr(D_i = 1|\boldsymbol{X_i}) \end{bmatrix}.$$

Under the null hypothesis: $\beta = 0$, we have

$$W_S = U(Z_1, Z_2) V(Z_1, Z_2)^{-1} U(Z_1, Z_2) \sim \chi_1^2,$$

where $V(Z_1, Z_2) = I_\beta(Z_1, Z_2) - I_{\beta\boldsymbol{\alpha}}(Z_1, Z_2) I_{\boldsymbol{\alpha}}^{-1} I_{\beta\boldsymbol{\alpha}}(Z_1, Z_2)'$ is estimated as the variance of $U(Z_1, Z_2)$. Therefore, the statistic

$$S(Z_1, Z_2) = \frac{U(Z_1, Z_2)}{\sqrt{V(Z_1, Z_2)}}$$

asymptotically follows a standard normal distribution under the null hypothesis.

Note that the calculation of the test statistic relies on the underlying XCI pattern. Unfortunately, this is generally unknown for a specific SNP. We thereby proposed a robust

test referred to as XCMAX4 to account for the four special XCI models. The XCMAX4 statistic is defined as follows:

$$XCMAX4 = \max(|S(0,\,2)|, |S(1,\,2)|, |S(2,\,2)|, |S(1,\,1)|)$$

Due to the correlation between the four score tests, XCMAX4 does not follow any classical distributions. We assume that $S(0,\,2)$, $S(1,\,2)$, $S(2,\,2)$, and $S(1,\,1)$ jointly asymptotically follow a multivariate normal distribution $N(\mathbf{0},\,\mathbf{\Sigma})$, where $\mathbf{0}$ is a four-dimensional vector with all elements being 0, and $\mathbf{\Sigma}$ is the correlation matrix with

$$\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho_{(0,2),(1,2)} & \rho_{(0,2),(2,\,2)} & \rho_{(0,2),(1,1)} \\ \rho_{(1,2),(0,2)} & 1 & \rho_{(1,2),(2,\,2)} & \rho_{(1,2),(1,1)} \\ \rho_{(2,\,2),(0,2)} & \rho_{(2,\,2),(1,2)} & 1 & \rho_{(2,\,2),(1,1)} \\ \rho_{(1,1),(0,2)} & \rho_{(1,1),(1,2)} & \rho_{(1,1),(2,\,2)} & 1 \end{bmatrix}.$$

In the above correlation matrix, $\rho_{(z_{11},z_{21}),(z_{12},z_{22})}$ is the correlation coefficient between $S(Z_{11}, Z_{21})$ and $S(Z_{12}, Z_{22})$. Given $\mathbf{\Sigma}$, we can analytically derive the *p* value of XCMAX4. Particularly, let $f(\mathbf{y}, \mathbf{0},\, \mathbf{\Sigma})$ be the density function of the multivariate normal distribution $N(\mathbf{0},\,\mathbf{\Sigma})$; then, for a given $z > 0$, the *p* value of XCMAX4 is calculated by

$$P_r(XCMAX4 > z) = 1 - \int \int \int \int_{-z}^{z} f(\mathbf{y}, \mathbf{0}, \mathbf{\Sigma}) d\mathbf{y}.$$

Next, we need to accurately estimate the correlation matrix $\mathbf{\Sigma}$. To this end, we first build a new model that contains two parameters representing genetic effects as follows:

$$\log\left(\frac{\Pr(D_i = 1 | G_i, \mathbf{X}_i)}{\Pr(D_i = 0 | G_i, \mathbf{X}_i)}\right) = \mathbf{X}_i'\boldsymbol{\alpha} + \beta_1 G_i(Z_{11}, Z_{21}) + \beta_2 G_i(Z_{12}, Z_{22}). \tag{2}$$

The information matrix of (2) can be expressed as follows:

$$I(Z_{11}, Z_{21}, Z_{12}, Z_{22}) = \begin{pmatrix} I_{\beta_1\beta_2}(Z_{11}, Z_{21}, Z_{12}, Z_{22}) & I_{\beta_1\beta_2\boldsymbol{\alpha}}(Z_{11}, Z_{21}, Z_{12}, Z_{22}) \\ I_{\beta_1\beta_2\boldsymbol{\alpha}}(Z_{11}, Z_{21}, Z_{12}, Z_{22})' & I_{\boldsymbol{\alpha}} \end{pmatrix},$$

where

$$\begin{aligned} &I_{\beta_1\beta_2}(Z_{11}, Z_{21}, Z_{12}, Z_{22}) \\ &= \begin{pmatrix} \sum\limits_{i=1}^{n} G_i(Z_{11},\, Z_{21})^2 [1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) & \sum\limits_{i=1}^{n} G_i(Z_{11},\, Z_{21})G_i(Z_{12},\, Z_{22})[1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) \\ \sum\limits_{i=1}^{n} G_i(Z_{11},\, Z_{21})G_i(Z_{12},\, Z_{22})[1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) & \sum\limits_{i=1}^{n} G_i(Z_{12},\, Z_{22})^2 [1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) \end{pmatrix}, \end{aligned}$$

and

$$I_{\beta\boldsymbol{\alpha}}(Z_{11}, Z_{21}, Z_{12}, Z_{22}) = \begin{pmatrix} \sum\limits_{i=1}^{n} X_{i1} G_i(Z_{11}, Z_{21})[1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) & \cdots & \sum\limits_{i=1}^{n} X_{ip} G_i(Z_{11}, Z_{21})[1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) \\ \sum\limits_{i=1}^{n} X_{i1} G_i(Z_{12}, Z_{22})[1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) & \cdots & \sum\limits_{i=1}^{n} X_{ip} G_i(Z_{12}, Z_{22})[1 - \Pr(D_i = 1|\mathbf{X}_i)]\Pr(D_i = 1|\mathbf{X}_i) \end{pmatrix}$$

Under the null hypothesis $\beta_1 = \beta_2 = 0$, the statistic

$$W_S = [U(Z_{11}, Z_{21}), U(Z_{12}, Z_{22})]C(Z_{11}, Z_{21}, Z_{12}, Z_{22})^{-1}[U(Z_{11}, Z_{21}), U(Z_{12}, Z_{22})]'$$

asymptotically follows a chi-square distribution with two degrees of freedom, where

$$C(Z_{11}, Z_{12}, Z_{21}, Z_{22}) = I_{\beta_1\beta_2}(Z_{11}, Z_{21}, Z_{12}, Z_{22}) - I_{\beta_1\beta_2\boldsymbol{\alpha}}(Z_{11}, Z_{21}, Z_{12}, Z_{22})I_{\boldsymbol{\alpha}}^{-1}I_{\beta_1\beta_2\boldsymbol{\alpha}}(Z_{11}, Z_{21}, Z_{12}, Z_{22})'$$

is the covariance matrix of $U(Z_{11}, Z_{12})$ and $U(Z_{21}, Z_{22})$. Therefore, the correlation coefficient between $S(Z_{11}, Z_{12})$ and $S(Z_{21}, Z_{22})$ can be estimated as

$$\frac{(1,0)C(Z_{11}, Z_{21}, Z_{12}, Z_{22})(0,1)'}{\sqrt{(1,0)C(Z_{11}, Z_{21}, Z_{12}, Z_{22})(1,0)' \times (0,1)C(Z_{11}, Z_{21}, Z_{12}, Z_{22})(0,1)'}}$$

Once Σ is estimated, we can calculate the p value of XCMAX4. Although the four-dimensional integral can be calculated in the commonly used software (e.g., the mvtnorm package in R, https://cran.r-project.org/web/packages/mvtnorm/index.html, (accessed on 10 April 2022)), the algorithm based on the Quasi-Monte-Carlo procedure needs a lot of computing resources to achieve relatively high accuracy. Hence, it would be still desirable to obtain its analytic form if possible. Fortunately, we can use the rhombus formula [13,21] to obtain the upper bound of the p-value of XCMAX4 as follows:

$$P(\text{XCMAX4} > z) \leq 2[\Phi(z) - \Phi(-z) - 1] + \frac{4\phi(z)}{z} \sum_{i=1}^{3} \left[\Phi\left(\frac{L_{i(i+1)}z}{2}\right) + \Phi\left(\frac{(\pi - L_{i(i+1)})z}{2}\right) - 1 \right],$$

where $\Phi(x)$ and $\phi(x)$ denote the cumulative distribution function and probability density function of the standard normal distribution, respectively, and $L_{i(i+1)} = \arccos\left(\rho_{i(i+1)}\right)$, where $\rho_{i(i+1)}$ is the correlation efficient between ith and $(i+1)$th score statistics. Note the order of four test statistics $S(0,2)$, $S(1,2)$, $S(2,2)$, and $S(1,1)$ is not specified in the above formula, so 12 kinds of upper bounds can be obtained. Therefore, only the smallest bound among them is adopted as an approximation of the p value. As shown in Wang et al. [13], such approximation is very accurate for small p values, which would be quite useful in GWAS because the significance level is generally very stringent (e.g., 5×10^{-8}) in such studies.

3. Simulation Study

3.1. Simulation Settings

We conducted comprehensive simulation studies to compare the performance of XCMAX4 with FM_{01}, FM_{02}, FM_F, and FM_S, all of which can adjust covariates. Note that we did not include the maxLR in our simulations because this method is a permutation-based approach, which would be too time-consuming for GWAS. The data are simulated from the following model:

$$\log\left(\frac{\Pr(D_i = 1 | G_i, x_{i2}, x_{i3})}{\Pr(D_i = 0 | G_i, x_{i2}, x_{i3})}\right) = \alpha_1 + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \beta G_i, \tag{3}$$

where x_2 is the binary covariate sex, x_3 is a continuous covariate, which is sampled from the uniform distribution $U(0,1)$, and G is the genotype score. The ratio of males to females is assumed to be 1 : 1 in the general population, so x_2 follows a binomial distribution $B(0.5)$. Further, we assume that the genotype of females (aa, Aa, AA) follows a trinomial distribution with probabilities $\left(q_{f0}, q_{f1}, q_{f2}\right)$, while the genotype of males (a, A) follows a binomial distribution $(1 - q_m, q_m)$. Let q_f and F be the respective risk allele frequency and the inbreeding coefficient for females. Then, we have $q_{f0} = \left(1 - q_f\right)^2 + Fq_f\left(1 - q_f\right)$, $q_{f1} = 2(1 - F)q_f\left(1 - q_f\right)$, and $q_{f2} = q_f^2 + Fq_f\left(1 - q_f\right)$. The values of q_f and q_m are both set to be 0.1, 0.2, and 0.3, so there are nine combinations in total. F is assigned to be 0 and 0.05, where the former implies Hardy–Weinberg equilibrium (HWE) and the latter represents a scenario of Hardy–Weinberg disequilibrium (HWD). The intercept α_1 is fixed at -5. For the coefficients x_2 and x_3, we consider two cases for each of them: $\alpha_2 = (0.4005, -0.4005)$ and $\alpha_3 = (0.5, 1.5)$. The genetic effect β is set to be 0, 0.1116, 0.15, and 0.1858, where $\beta = 0$ means no association between the SNP and the disease status, and

the other three values of β indicate that the odds ratios of females with genotype AA are about 1.25, 1.35, and 1.45. Obviously, the case of $\beta = 0$ is used to study the size, while the empirical power is investigated in the non-zero β cases.

Note that, when studying the power, we only choose three combinations of q_f and q_m: (0.3, 0.3), (0.3, 0.2), and (0.2, 0.3) for convenience. The scenarios that the SNP undergoes XCI or escapes from XCI are both considered. For the former, we let γ range from 0 to 2 in increments of 0.5. As such, we have considered various XCI patterns, including XCI-SN, XCI-R, and XCI-SR. Once the XCI pattern is assumed, we can assign the corresponding value for the genotypic score G.

Given the covariates, the genotypic score, and the regression coefficients, we can generate the disease status from the binomial distribution for a large population. Then, we randomly sample 2500 cases and 2500 controls from this population. We find that when $\alpha_2 = \pm 0.4005$, the proportions of females in cases varied from 40% to 60% in the simulated data. The size is estimated at three nominal levels: $\alpha = 1 \times 10^{-3}$, 1×10^{-4}, and 1×10^{-5} based on 1,000,000 replicates, while the power is only estimated at the nominal level $\alpha = 1 \times 10^{-4}$ based on 10,000 replicates. The p value of XCMAX4 is evaluated by using the rhombus formula.

3.2. Results

3.2.1. Size

Table 2 shows the estimated type I error rate at the nominal significance level $\alpha = 1 \times 10^{-4}$ when HWE holds in the female population. As expected, all the methods controlled the size well in all the scenarios. Although XCMAX4 appears slightly conservative in some scenarios, its p values are similar to the nominal level. We also simulated the scenarios of HWD ($F = 0.05$). However, we observed that the performances of all the tests were similar to those of Table 2, and HWD in females had little impact on the size. Therefore, the simulation results with non-zero F are presented in the Supplementary Material (Table S1). The results of type I error rates estimated at the nominal level $\alpha = 1 \times 10^{-3}$ and $\alpha = 1 \times 10^{-5}$ are also given in Supplementary Material (Tables S2–S5). As can be seen, XCMAX4 still had the correct size in general, except being slightly conservative at $\alpha = 1 \times 10^{-3}$.

Table 2. Estimated type I error rate $(\times 10^{-4})$ at the nominal significance level 1×10^{-4} for XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S against q_f, q_m, α_2, and α_3 based on 1,000,000 replicates under HWE.

| q_f	q_m	α_3	$\alpha_2 = 0.4005$					$\alpha_2 = -0.4005$				
			XCMAX4	FM_{01}	FM_{02}	FM_F	FM_S	XCMAX4	FM_{01}	FM_{02}	FM_F	FM_S
	0.1		1.03	0.74	0.86	0.95	0.82	0.87	0.93	0.93	0.66	0.98
0.1	0.2		0.88	0.96	0.84	0.86	0.96	1.07	1.02	1.10	0.99	1.02
	0.3		0.91	0.87	0.84	1.10	0.84	0.86	1.02	0.88	0.95	1.00
	0.1		0.94	0.96	0.98	0.93	0.95	0.86	0.85	0.74	0.78	0.79
0.2	0.2	0.5	1.02	1.02	0.93	1.12	1.00	1.12	1.43	1.24	1.22	1.39
	0.3		0.90	1.09	0.99	0.76	1.02	1.10	0.99	1.03	0.98	1.01
	0.1		0.87	0.88	0.87	0.88	0.87	0.79	1.01	0.93	0.96	0.91
0.3	0.2		0.93	1.01	1.04	0.77	0.99	0.83	0.94	0.89	0.91	0.93
	0.3		0.83	1.13	0.92	0.95	0.91	0.96	1.15	1.05	1.14	1.06
	0.1		0.88	1.06	0.93	0.79	1.01	0.88	1.00	0.83	0.77	0.92
0.1	0.2		0.84	0.77	0.82	0.79	0.75	0.92	0.87	0.96	0.98	0.86
	0.3		0.88	1.22	1.08	0.96	1.16	0.99	1.17	1.14	1.07	1.09
	0.1		0.92	0.93	1.06	0.92	1.03	0.81	1.03	0.91	0.96	0.91
0.2	0.2	1.5	0.84	1.01	0.92	0.91	0.98	0.86	0.80	0.88	0.85	0.78
	0.3		0.94	1.08	1.14	1.08	1.08	0.85	0.94	0.98	0.98	0.97
	0.1		1.00	1.08	1.03	0.93	1.00	0.99	0.85	1.01	0.96	0.94
0.3	0.2		0.88	1.08	0.91	0.93	0.92	0.93	0.76	0.90	0.83	0.94
	0.3		0.82	0.95	0.86	1.01	0.88	0.98	1.06	0.97	1.10	1.09

3.2.2. Power

Figures 1–3 plot the powers of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S under various XCI patterns when $\beta = 0.15$, $F = 0$ and $\left(q_f, q_m\right) = (0.3, 0.3)$, $(0.3, 0.2)$, and $(0.2, 0.3)$, respectively. These figures show that all four subfigures exhibited a similar pattern in power, indicating that the covariates had a very limited impact on the performance of all methods.
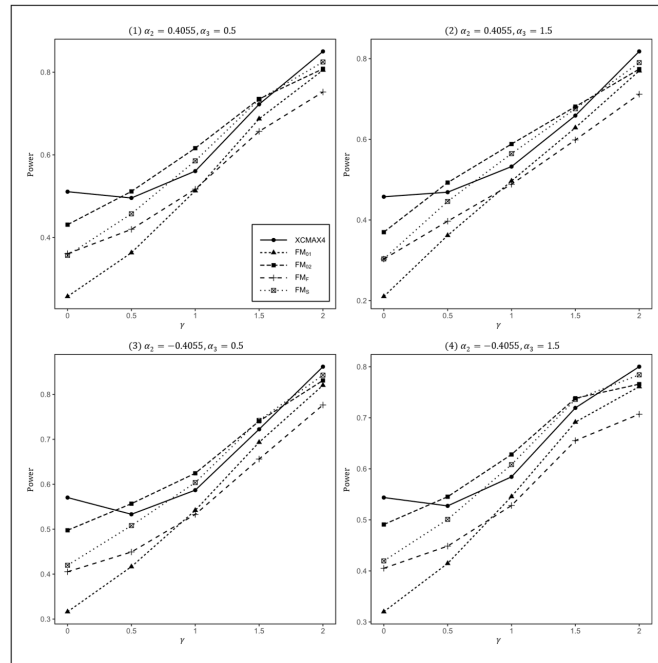
Figure 1. Estimated powers of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S under various XCI models. The simulation was based on 10,000 replicates with $\beta = 0.15$, $\alpha_1 = -5$, $F = 0$, and $q_f = q_m = 0.3$.
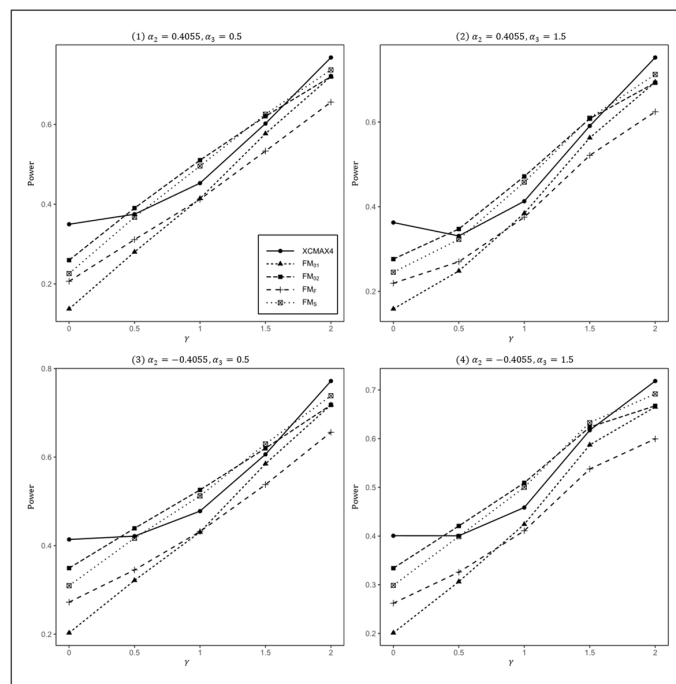
Figure 2. Estimated powers of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S under various XCI models. The simulation was based on 10,000 replicates with $\beta = 0.15$, $\alpha_1 = -5$, $F = 0$, $q_f = 0.3$, and $q_m = 0.2$.
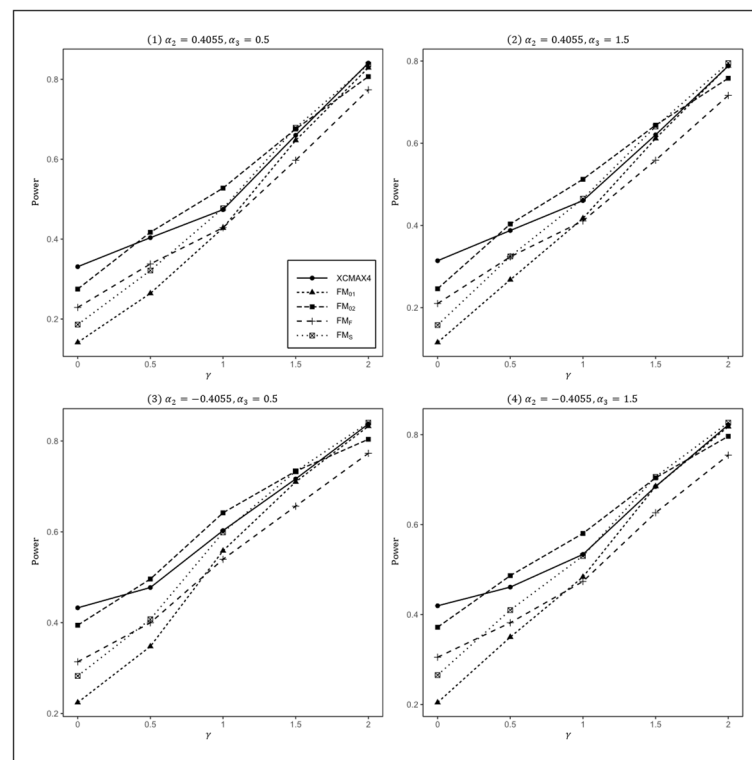
Figure 3. Estimated powers of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S under various XCI models. The simulation was based on 10,000 replicates with $\beta = 0.15$, $\alpha_1 = -5$, $F = 0$, $q_f = 0.2$, and $q_m = 0.3$.

In Figure 1, we can see that FM_{01} and FM_F were generally less powerful than other methods in all situations. XCMAX4 performed best when $\gamma = 0$ (XCI-SN) and 2 (XCI-SR). However, when $\gamma = 1$ (XCI-R), FM_{02} was the most powerful, followed by FM_S and XCMAX4. This was expected because FM_{02} is proposed exactly under XCI-R. We also observed that XCMAX4 had a better power than FM_S when $\gamma = 0.5$, while FM_S performed slightly better than XCMAX4 when $\gamma = 1.5$. In both scenarios, FM_{02} was still the most powerful method, but the differences in power between these three methods were generally very small. Notice that the results in Figures 2 and 3 are analogous to those in Figure 1, and thereby the allele frequencies of females and males did not apparently change the power profiles of all of the methods.

Figure 4 plots the powers of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S under the XCI-E pattern with $\beta = 0.15$. Based on this figure, FM_{01} was uniformly the most powerful in all scenarios as expected, followed by FM_S and XCMAX4. FM_{02} was generally less powerful than FM_{01}, FM_S, and XCMAX4, but still performed better than FM_F. The power results with $\beta = 0.15$, and $F = 0.05$ are provided in Supplementary Material (Figures S1–S4), which are similar to those in Figures 1–4, indicating that HWD in females had little effect on the power results. The power results with $\beta = 0.1116$ and 0.1858 are generally consistent with those in Figures 1–4, implying that the properties of XCMAX4 did not vary with the magnitude of the genetic effect (see Figures S5–S20 in Supplementary Material). As expected, when the value of β increased, the powers of all methods uniformly increased.

In conclusion, FM_{01} and FM_{02} can have high power if the underlying XCI pattern is modelled correctly but may be less powerful in other scenarios. In contrast, XCMAX4 retained a relatively good power across a variety of scenarios. Compared to XCMAX4, FM_S may suffer from power loss if the SNP is undergoing XCI but will be more powerful under XCI-E. FM_F had the overall worst performance and thus is not recommend. It should be noted that, FM_{01}, FM_{02}, FM_F, and FM_S adopted logistic regression, which is slightly more computationally intensive than XCMAX4 in GWAS because the implementation of the logistic regression requires additional iterations. Compared to the other four methods,

testing 2000 SNPs, XCMAX4 saved half the time. The details of time comparisons are given in Supplemental Material (Table S6).
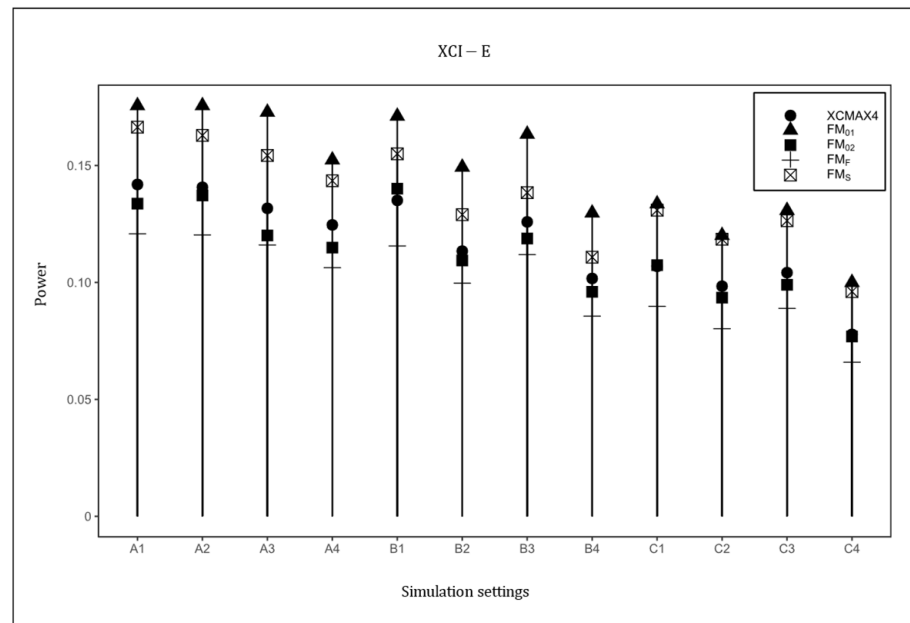
Figure 4. Powers of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S under XCI-E. The simulation was based on 10,000 replicates with $\beta = 0.15$, $\alpha_1 = -5$, and $F = 0$. In the horizontal coordinates, "A", "B", and "C" represent three combinations of (q_f, q_m): $(0.3, 0.3)$, $(0.3, 0.2)$, and $(0.2, 0.3)$, respectively, and the numbers 1–4 represent four combinations of (α_2, α_3): $(0.4055, 0.5)$, $(0.4055, 1.5)$, $(-0.4055, 0.5)$, and $(-0.4055, 1.5)$, respectively.

4. Application to Graves' Disease Data

Graves' disease (GD) is an autoimmune disease of hyperthyroidism that is four times more common in women than in men [22,23]. Substantial studies have shown that the genetic background explains about four-fifths of the susceptibility to GD.

Considering the distinct gender bias, it is highly reasonable to speculate that the genes on the X chromosome play an important role in the development of GD. Recently, two independent studies found that rs3827440, a non-synonymous SNP of the GRP174 gene on the X chromosome, was associated with GD. A two-stage GWAS, focused on the Han population in China, first reported this finding, which was further validated in two Caucasian cohorts. There are two alleles at rs3827400, with T being the risk allele and C being the normal one. Table 3 displays the four datasets about rs3827400 mentioned in these two studies. We applied XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S to each dataset; the results are shown in Table 4. Note that sex was included as a covariate when calculating the p values of XCMAX4, FM_{01}, and FM_{02}.

Table 3. Data of rs3827400 related to Graves' disease in two independent studies.

Dataset	Race	Female Case			Male Case		Female Control			Male Control	
		CC	TC	TT	C	T	CC	TC	TT	C	T
Chu et al. (stage I)	Han	163	508	444	109	232	219	541	367	172	186
Chu et al. (stage II)	Han	471	1606	1298	284	606	584	1344	957	396	526
Szymanski et al. (Warsaw)	Caucasian	146	205	85	53	51	188	229	81	146	104
Szymanski et al. (Gliwice)	Caucasian	58	78	30	20	11	71	73	27	20	10

Table 4. p values of the XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S tests from four datasets.

Dataset	XCMAX4	FM_{01}	FM_{02}	FM_F	FM_S
Chu et al. (stage I) $\left(\times 10^{-8}\right)$	1.573	9.513	0.507	1.832	1.731
Chu et al. (stage II) $\left(\times 10^{-15}\right)$	0.847	7.764	0.561	4.108	1.144
Szymanski et al. (Warsaw) $\left(\times 10^{-1}\right)$	1.083	0.491	0.395	1.038	0.410
Szymanski et al. (Gliwice) $\left(\times 10^{-1}\right)$	5.967	2.515	2.800	5.500	2.628

This table indicates that none of these methods uniformly performed the best across all four datasets. For the two datasets from the Chinese population, all methods consistently showed that rs3827400 was associated with GD at the 1×10^{-4} significance level. Among these tests, XCMAX4 consistently had the second smallest p values. However, the p values of all the methods from both Caucasian datasets suggested no such an association at the same significance level probably because of their relatively small sample size. We also observed that XCMAX4 appeared slightly conservative in these scenarios, but this was not surprising because the rhombus formula is less accurate when the p value is greater than 0.01.

Because both the Han population and the Caucasian population contained two datasets, we also tested such association at the population level by treating the data source as an additional covariate. The corresponding results are given in Table 5, which are similar to those in Table 4.

Table 5. p values of XCMAX4, FM_{01}, FM_{02}, FM_F, and FM_S tests from Han and Caucasian populations.

Population	XCMAX4	FM_{01}	FM_{02}	FM_F	FM_S
Han $\left(\times 10^{-22}\right)$	1.275	55.347	0.285	1.792	1.444
Caucasian $\left(\times 10^{-2}\right)$	6.932	2.571	2.553	5.795	1.993

5. Discussion

This paper proposed a novel robust method, XCMAX4, to test the association between the marker on the X chromosome and a specific disease for case-control design. Our method is an extension of the CMAX3 [24] test on the X chromosome, which can both incorporate the information of XCI and allow for covariates. Unlike the maxLR proposed by Wang et al., XCMAX4 is construted by using the score test, which is more efficient in GWAS because we only need to fit the null model once. Moreover, the maxLR requires permutation to calculate the p value, which makes it unappealing in GWAS. In contrast, the p value of XCMAX4 can be computed analytically by using the rhombus formula. On the other hand, although FM_{01}, FM_{02}, FM_F, and FM_S can also adjust for covariates, they do not take various XCI models into consideration and thus may suffer from substantial power loss in some scenarios. However, XCMAX4 can retain a relatively high power by accounting for four special XCI patterns simultaneously. Simulation results showed that XCMAX4 controlled the size well and had robust performance in power. Therefore, we recommend using XCMAX4 for its effectiveness, robustness, and generality. Finally, to help implement XCMAX4 in practice, we provide an R function XCMAX4, which is available at https://github.com/YoupengSU/XCMAX4.git (accessed on 12 April 2022).

and J.H.; visualization S.C., Z.C., and M.D.; supervision, P.W.; funding acquisition, P.Y. and H.J. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The real data used in this study are available from two published papers at https://dx.doi.org/10.1136%2Fjmedgenet-2013-101595, and https://doi.org/10.1111/tan.12259 (assessed on 7 April 2022).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Detailed Derivation of the Score Statistic

The log-likelihood function of Model (1) can be written as

$$l(\beta, \boldsymbol{\alpha}') = \log\left(\prod_{i=1}^{n} D_i{}^{\mu_i}(1-D_i)^{1-\mu_i}\right) = \sum_{i=1}^{n}\left\{D_i \log \mu_i + (1-D_i)\log(1-\mu_i)\right\}$$

$$= \sum_{i=1}^{n}\left\{D_i \log \frac{\mu_i}{1-\mu_i} + \log(1-\mu_i)\right\}$$

$$= \sum_{i=1}^{n}\left\{D_i(\boldsymbol{X_i}'\boldsymbol{\alpha} + \beta G_i) - \log\left(1 + e^{\boldsymbol{X_i}'\boldsymbol{\alpha} + \beta G_i}\right)\right\},$$

where $\mu_i = \frac{exp^{\boldsymbol{X_i}'\boldsymbol{\alpha}+\beta G_i}}{1+exp^{\boldsymbol{X_i}'\boldsymbol{\alpha}+\beta G_i}}$ representing the probability of having disease for ith individual.

Assume that $\hat{\boldsymbol{\theta}}_0 = (0, \hat{\boldsymbol{\alpha}}')'$ is the restricted maximum likelihood estimate of $\boldsymbol{\theta} = (\beta, \boldsymbol{\alpha}')'$ under the condition $\beta = 0$, then the score function and Fisher's information matrix can be given as

$$U(\hat{\boldsymbol{\theta}}_0) = \left(\frac{\partial l(\beta, \boldsymbol{\alpha}')}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}\right) = \left(\frac{\partial l(\boldsymbol{\alpha},\beta)}{\partial \beta}, \frac{\partial l(\boldsymbol{\alpha},\beta)'}{\partial \boldsymbol{\alpha}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}\right)' = \left(\frac{\partial l(\boldsymbol{\alpha},\beta)}{\partial \beta}, \mathbf{0}'\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}\right)',$$

$$= \left(\sum_{i=1}^{n}\left\{G_i\left[D_i - \frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}\right]\right\}, \mathbf{0}'\right)' = (U(Z_1, Z_2), \mathbf{0}')',$$

and

$$I(\hat{\boldsymbol{\theta}}) = -E\left(\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_0}\right) = \begin{pmatrix} \sum\limits_{i=1}^{n}\left(\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}\right)G_iG_i & \cdots & \sum\limits_{i=1}^{n}\left(\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}\right)X_{ip}G_i \\ \vdots & \ddots & \vdots \\ \sum\limits_{i=1}^{n}\left(\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}\right)G_iX_{ip} & \cdots & \sum\limits_{i=1}^{n}\left(\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}\right)X_{ip}X_{ip} \end{pmatrix}_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix},$$

where

$$I_{11} = \sum_{i=1}^{n}\left(\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}\right)G_iG_i,$$

$$I_{12} = \left(\sum_{i=1}^{n} X_{i1}G_i\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}, \cdots, \sum_{i=1}^{n} X_{ip}G_i\frac{e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}}{\left(1+e^{\boldsymbol{X_i}'\hat{\boldsymbol{\alpha}}}\right)^2}\right),$$

$$I_{21} = I_{12}',$$

and

$$
I_{22} = \begin{pmatrix} \sum_{i=1}^{n} \frac{e^{X_i'\hat{\alpha}}}{\left(1+e^{X_i'\hat{\alpha}}\right)^2} X_{i1} X_{i1} & \cdots & \sum_{i=1}^{n} \frac{e^{X_i'\hat{\alpha}}}{\left(1+e^{X_i'\hat{\alpha}}\right)^2} X_{i1} X_{ip} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} \frac{e^{X_i'\hat{\alpha}}}{\left(1+e^{X_i'\hat{\alpha}}\right)^2} X_{ip} X_{i1} & \cdots & \sum_{i=1}^{n} \frac{e^{X_i'\hat{\alpha}}}{\left(1+e^{X_i'\hat{\alpha}}\right)^2} X_{ip} X_{ip} \end{pmatrix}.
$$

By Cox et. al. [25], we can obtain the score test statistic as

$$
W_S = U(\hat{\boldsymbol{\theta}}_0)' I(\hat{\boldsymbol{\theta}})^{-1} U(\hat{\boldsymbol{\theta}}_0) = U(\hat{\boldsymbol{\theta}}_0)' I(\hat{\boldsymbol{\theta}})^{-1} U(\hat{\boldsymbol{\theta}}_0) = U(\hat{\boldsymbol{\theta}}_0)' \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}^{-1} U(\hat{\boldsymbol{\theta}}_0)
$$

$$
= U(\hat{\boldsymbol{\theta}}_0)' \begin{pmatrix} \left(I_{11} - I_{12} I_{22}^{-1} I_{21}\right)^{-1} & -\left(I_{11} - I_{12} I_{22}^{-1} I_{21}\right)^{-1} I_{12} I_{22}^{-1} \\ -I_{22}^{-1} I_{21} \left(I_{11} - I_{12} I_{22}^{-1} I_{21}\right)^{-1} & I_{22}^{-1} + I_{22}^{-1} I_{21} \left(I_{11} - I_{12} I_{22}^{-1} I_{21}\right)^{-1} I_{12} I_{22}^{-1} \end{pmatrix} U(\hat{\boldsymbol{\theta}}_0)
$$

$$
= U(Z_1, Z_2)' \left(I_{11} - I_{12} I_{22}^{-1} I_{21}\right)^{-1} U(Z_1, Z_2),
$$

which asymptotically follows a chi-square distribution with degrees of freedom being 1. In Model (2), β becomes a two-dimensional vector, and the proofs are similar, so the details are omitted.

References

1. Voskuhl, R. Sex differences in autoimmune diseases. *Biol. Sex Differ.* **2011**, *2*, 1. [CrossRef]
2. Appelman, Y.; van Rijn, B.B.; Ten Haaf, M.E.; Boersma, E.; Peters, S.A. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis* **2015**, *241*, 211–218. [CrossRef]
3. Riecher-Rossler, A. Sex and gender differences in mental disorders. *Lancet Psychiatry* **2017**, *4*, 8–9. [CrossRef]
4. Dong, M.; Cioffi, G.; Wang, J.; Waite, K.A.; Ostrom, Q.T.; Kruchko, C.; Lathia, J.D.; Rubin, J.B.; Berens, M.E.; Connor, J.; et al. Sex Differences in Cancer Incidence and Survival: A Pan-Cancer Analysis. *Cancer Epidemiol. Biomark. Prev.* **2020**, *29*, 1389–1397. [CrossRef] [PubMed]
5. Erol, A.; Winham, S.J.; McElroy, S.L.; Frye, M.A.; Prieto, M.L.; Cuellar-Barboza, A.B.; Fuentes, M.; Geske, J.; Mori, N.; Biernacka, J.M.; et al. Sex differences in the risk of rapid cycling and other indicators of adverse illness course in patients with bipolar I and II disorder. *Bipolar Disord.* **2015**, *17*, 670–676. [CrossRef] [PubMed]
6. Lu, Z.; Carter, A.C.; Chang, H.Y. Mechanistic insights in X-chromosome inactivation. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2017**, *372*, 356. [CrossRef] [PubMed]
7. Lyon, M.F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **1961**, *190*, 372–373. [CrossRef] [PubMed]
8. Carrel, L.; Willard, H.F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **2005**, *434*, 400–404. [CrossRef] [PubMed]
9. Fang, H.; Disteche, C.M.; Berletch, J.B. X Inactivation and Escape: Epigenetic and Structural Features. *Front. Cell Dev. Biol.* **2019**, *7*, 219. [CrossRef]
10. Disteche, C.M. Dosage compensation of the sex chromosomes and autosomes. *Semin. Cell Dev. Biol.* **2016**, *56*, 9–18. [CrossRef]
11. Cantone, I.; Fisher, A.G. Human X chromosome inactivation and reactivation: Implications for cell reprogramming and disease. *Philos. Trans. R Soc. Lond. B Biol. Sci.* **2017**, *372*, 358. [CrossRef] [PubMed]
12. Wang, P.; Zhang, Y.; Wang, B.Q.; Li, J.L.; Wang, Y.X.; Pan, D.; Wu, X.B.; Fung, W.K.; Zhou, J.Y. A statistical measure for the skewness of X chromosome inactivation based on case-control design. *BMC Bioinform.* **2019**, *20*, 11. [CrossRef] [PubMed]
13. Wang, P.; Xu, S.Q.; Wang, B.Q.; Fung, W.K.; Zhou, J.Y. A robust and powerful test for case-control genetic association study on X chromosome. *Stat. Methods Med. Res.* **2019**, *28*, 3260–3272. [CrossRef] [PubMed]
14. Wang, J.; Yu, R.; Shete, S. X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation. *Genet. Epidemiol.* **2014**, *38*, 483–493. [CrossRef]
15. Gao, F.; Chang, D.; Biddanda, A.; Ma, L.; Guo, Y.; Zhou, Z.; Keinan, A. XWAS: A Software Toolset for Genetic Data Analysis and Association Studies of the X Chromosome. *J. Hered.* **2015**, *106*, 666–671. [CrossRef]
16. Clayton, D. Testing for association on the X chromosome. *Biostatistics* **2008**, *9*, 593–600. [CrossRef]
17. Zheng, G.; Joo, J.; Zhang, C.; Geller, N.L. Testing association for markers on the X chromosome. *Genet. Epidemiol.* **2007**, *31*, 834–843. [CrossRef]
18. Chen, Z.; Ng, H.K.; Li, J.; Liu, Q.; Huang, H. Detecting associated single-nucleotide polymorphisms on the X chromosome in case control genome-wide association studies. *Stat. Methods Med. Res.* **2017**, *26*, 567–582. [CrossRef]
19. Clayton, D.G. Sex chromosomes and genetic association studies. *Genome Med.* **2009**, *1*, 110. [CrossRef]
20. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [CrossRef]

21. Li, Q.Z.; Zheng, G.; Li, Z.H.; Yu, K. Efficient approximation of P-value of the maximum of correlated tests, with applications to genome-wide association studies. *Ann. Hum. Genet.* **2008**, *72*, 397–406. [CrossRef] [PubMed]

22. Chu, X.; Shen, M.; Xie, F.; Miao, X.J.; Shou, W.H.; Liu, L.; Yang, P.P.; Bai, Y.N.; Zhang, K.Y.; Yang, L.; et al. An X chromosome-wide association analysis identifies variants in GPR174 as a risk factor for Graves' disease. *J. Med. Genet.* **2013**, *50*, 479–485. [CrossRef] [PubMed]

23. Szymanski, K.; Miskiewicz, P.; Pirko, K.; Jurecka-Lubieniecka, B.; Kula, D.; Hasse-Lazar, K.; Krajewski, P.; Bednarczuk, T.; Ploski, R. rs3827440, a nonsynonymous single nucleotide polymorphism within GPR174 gene in X chromosome, is associated with Graves' disease in Polish Caucasian population. *Tissue Antigens* **2014**, *83*, 41–44. [CrossRef] [PubMed]

24. Chen, Z.; Zang, Y. CMAX3: A Robust Statistical Test for Genetic Association Accounting for Covariates. *Genes* **2021**, *12*, 1723. [CrossRef]

25. Cox, D.R.; Hinkley, D.V. *Theoretical Statistics*; CRC Press: Boca Raton, FL, USA, 1979.