

Testing for Hardy-Weinberg Equilibrium in Samples With Related Individuals

Catherine Bourgain,^{*,1} Mark Abney,[†] Daniel Schneider,[†] Carole Ober[†] and Mary Sara McPeck^{†,‡}

^{*}INSERM U535, 94817 Villejuif Cedex, France and [†]Department of Human Genetics and [‡]Department of Statistics, University of Chicago, Chicago, Illinois 60637

Manuscript received May 28, 2004

Accepted for publication September 8, 2004

ABSTRACT

When the classical χ^2 goodness-of-fit test for Hardy-Weinberg (HW) equilibrium is used on samples with related individuals, the type I error can be greatly inflated. In particular the test is inappropriate in population isolates where the individuals are related through multiple lines of descent. In this article, we propose a new test for HW (the QL-HW test) suitable for any sample with related individuals, including large inbred pedigrees, provided that their genealogy is known. Performed conditional on the pedigree structure, the QL-HW test detects departures from HW that are not due to the genealogy. Because the computation of the QL-HW test becomes intractable for very polymorphic loci in large inbred pedigrees, a simpler alternative, the GCC-HW test, is also proposed. The statistical properties of the QL-HW and GCC-HW tests are studied through simulations considering a sample of independent nuclear families, a sample of extended outbred genealogies, and samples from the Hutterite population, a North American highly inbred isolate. Finally, the method is used to test a set of 143 biallelic markers spanning 82 genes in this latter population.

WHEN there are no disturbing forces such as selection, mutation, or migration that would change allele frequencies over time and when there is random mating in a very large population, the two alleles of an individual are known not to be associated. As a consequence, the genotype frequencies are given by the product of the allele frequencies and the population is said to be in Hardy-Weinberg equilibrium (HWE). Although it is one of the most basic and oldest tests in genetics, HWE testing has lately regained serious attention from human geneticists. NIELSEN *et al.* (1999) proposed using it to fine map susceptibility loci. Indeed, as shown by these authors, when considering case-only samples and thus conditioning on phenotypes, departure from HW is expected at susceptibility loci acting in a nonmultiplicative manner. More recently, LEE (2003) proposed using the test as a genome-wide screening tool for association. As outlined by WEINBERG and MORRIS (2003), this latter strategy should be considered very carefully. It has important caveats that include little power to detect some plausible genetic models and nonnegligible probability of detecting departures from HW (HWD) due to other reasons, which include selection not driven by the phenotype under study, population stratification, and genotyping errors. In the present context of intensive use of SNP markers, for which the proportion of genotyping errors that are not detectable by routine Mendelian incompatibility checks is fairly significant, as

outlined by DOUGLAS *et al.* (2002), testing for HWE is an additional tool with which to identify genotyping errors. XU *et al.* (2002) have shown how systematic use of HW testing in case samples and in control samples may help to avoid false detection of association signals driven by genotyping errors in only one of the two samples (either cases or controls).

Classical approaches to testing for HWE include Pearson's χ^2 goodness-of-fit test (Gof-HW) and the corresponding likelihood-ratio test for unrelated individuals. However, these tests are not suitable for use in samples containing related individuals as they assume independence among the observations. Neglecting these correlations may lead to spurious conclusions (as we demonstrate in RESULTS). Further, permutation-based tests where alleles are randomly reassigned to genotypes to get null distributions are not suitable. Indeed, for such tests to be valid, the alleles must be exchangeable under the null hypothesis, which is not generally the case when considering related individuals. In this article, we propose a method for solving the problem of testing for HWE in a sample with any kind of related individuals, provided that their relationship is known. The method is designed to be suitable for large inbred pedigrees in which the relatedness among individuals may be considerable with many or all individuals related through multiple lines of descent. In addition to requiring correction for nonindependence among the individuals, testing for HWE in large inbred pedigrees also requires precise specification of the hypothesis to be tested. This is because, by definition, in an inbred individual i , a correlation exists between i 's two alleles, which is measured by

¹Corresponding author: INSERM U535, Hopital Paul Brousse, Batiment Leriche, BP 100, 94817 Villejuif Cedex, France.
E-mail: bourgain@vjf.inserm.fr

TABLE 1

Notations

N	No. of individuals in the sample
a	No. of distinct alleles at the locus
M	No. of possible distinct genotypes
h_i	Inbreeding coefficient of individual i
Δ_k^{ij}	k th identity coefficient between individuals i and j
p_k	Frequency of allele k
r	Fixation index
$Y = (Y_1, \dots, Y_N)$	$N(M-1)$ vector of genotype indicators
$\mu = (\mu_1, \dots, \mu_N)$	$N(M-1)$ vector of genotype indicator expectations
D_p	$[N(M-1) \times (a-1)]$ matrix with k th column equal to $\partial\mu/\partial p_k$
D_r	$N(M-1)$ vector equal to $\partial\mu/\partial r$
$p = (p_1, \dots, p_{(a-1)})$	Vector of the $(a-1)$ -independent allele frequencies
Σ	$[N(M-1) \times N(M-1)]$ covariance matrix of Y when $r = 0$
\mathbf{K}	Σ in the absence of inter-individual correlations
\hat{p}_0	Estimate of p when $r = 0$
$\hat{D}_p, \hat{D}_r, \hat{\mu}_0, \hat{\Sigma}, \hat{\mathbf{K}}$	$D_p, D_r, \mu, \Sigma, \mathbf{K}$ computed at $r = 0$ and $p = \hat{p}_0$

the inbreeding coefficient h_i . When the genealogy of the population is available, the inbreeding coefficient can be computed for any individual so that the correlation between the two alleles of any individual due to the pedigree structure can be measured. We propose a test performed conditional on the relationship among the genotyped individuals specified by the pedigree. In this case, the null hypothesis of the test is that, conditional on the pedigree information, the random variable representing assignment of genotypes to genotyped individuals follows the distribution that would be obtained if the founders' alleles were i.i.d. draws from a distribution (p_1, \dots, p_a) , where p_i is the frequency of allele i , $p_i > 0$, and $\sum p_i = 1$ and if, at each meiosis in the pedigree, one of the two parental alleles was passed on with chance $1/2$ each (notation defined throughout the article is summarized in Table 1). Deviation from this hypothesis could arise if, within the population, conditional on relationship, mate choice further depended on the genotypes at a given locus. It could also arise due to genotyping errors, selection, mutation, or, at a susceptibility locus, conditioning on phenotype information.

We have recently proposed (BOURGAİN *et al.* 2003) a valid test for case-control association studies in samples with related individuals that is suitable for large inbred pedigrees too complex for exact likelihood computations. Inference is based on the quasi-likelihood score function proposed by MCPEEK *et al.* (2004) and a quasi-likelihood score (QLS) statistic for case-control association is constructed. Under certain model assumptions and regularity conditions, this test is asymptotically the locally most powerful test of a general class of linear tests, which includes the χ^2 test for case-control association corrected for the correlation within and between the individuals. Here, we use a similar idea to construct

a quasi-likelihood score test for HWE (QL-HW test). Because it is based on genotypes rather than on alleles, the QL-HW statistic is more complex than the QLS. Allele frequencies do not factor out of the mean and derivative vectors and out of the covariance matrix, as they do in the QLS. Further, whereas the QLS requires only the inbreeding and kinship coefficients to summarize the correlation within an individual and between any pair of individuals, the QL-HW requires the values of the nine condensed identity coefficients for any pair of individuals. We use the efficient algorithm of ABNEY *et al.* (2000) to compute these coefficients in complex inbred pedigrees. Because the computation of the QL-HW test may become intractable for very polymorphic loci in large inbred pedigrees, we propose a simpler alternative, the generalized corrected χ^2 test for Hardy-Weinberg (GCC-HW test). In the case of an outbred sample with a biallelic locus, the GCC-HW test is equivalent to the Gof-HW test with a variance correction that appropriately accounts for inter-individual correlations. In the case of an inbred sample or when considering loci with more than two alleles, the GCC-HW test is a generalization of the corrected χ^2 test. GCC-HW is also performed conditional on the pedigree structure.

In what follows we first describe the QL-HW and GCC-HW statistics. Type I error of the two tests is assessed in samples of outbred nuclear or extended families and in samples of Hutterites from South Dakota, a North American religious and highly inbred isolate with known genealogy. The inflated type I error introduced in the classical Gof-HW test by the presence of related individuals, using either the χ^2 distribution or a permutation procedure to get significance, is illustrated in the outbred samples. Power of the QL-HW and GCC-HW tests to detect deviations from HWE in samples of inbred

TABLE 2

Definition of the model chosen for $E(Y) = \mu$ in the case of a locus with a different alleles

Genotype	$E(Y_{ig}) = \mu_{ig}$
k/k	$(1 - h_i - r)p_k^2 + (h_i + r)p_k$
a/a	$(1 - h_i - r)(1 - \sum_{j=1}^{a-1} p_j)^2 + (h_i + r)(1 - \sum_{j=1}^{a-1} p_j)$
k/l	$2(1 - h_i - r)p_k p_l$
a/l	$2(1 - h_i - r)p_l(1 - \sum_{j=1}^{a-1} p_j)$

h_i is the inbreeding coefficient of individual i , p_k is the frequency of allele k , and r is the fixation index.

and outbred individuals is evaluated. Finally, the QL-HW test is applied to a set of 143 intragenic biallelic markers typed in samples of Hutterites.

METHODS

Model: Consider a sample of N individuals and a locus with a different alleles. Define $M = a(a + 1)/2$, the number of possible distinct genotypes. Let $Y = (Y_1, \dots, Y_N)^T$ be a vector of length $N(M - 1)$ (a $[N(M - 1)]$ vector) with $Y_i = (Y_{i1}, \dots, Y_{i(M-1)})^T$ a $(M - 1)$ vector of indicators, where Y_{ig} equals 1 if individual i has genotype g and 0 otherwise. Define $E(Y) = \mu = (\mu_1, \dots, \mu_N)^T$ with $\mu_i = (\mu_{i1}, \dots, \mu_{ig}, \dots, \mu_{i(M-1)})^T$. As recognized by WEIR (1996), departures from HWE can be characterized in several ways, including use of fixation indices (correlations between any pair of alleles within individuals) and disequilibrium coefficients (difference between a frequency and its value expected when there is no association between alleles). We propose considering a model based on a fixation index, the model of heterozygote excess and deficiency (the HED model) studied by ROUSSET and RAYMOND (1995), and we extend it to explicitly include N different inbreeding coefficients (h_i) for the N individuals. Table 2 gives the corresponding expression of μ_i , the i th component of μ . This model has a distinct parameters: the $(a - 1)$ independent allele frequencies $p = (p_1, \dots, p_k, \dots, p_{a-1})$ and the fixation index r with the following constraints on these parameters,

$$0 < p_k, \quad \sum_{k=1}^{a-1} p_k < 1 \quad (1)$$

$$1 - \min_i h_i - \frac{1}{1 - \min_k p_k} \leq r \leq 1 - \max_i h_i, \quad (2)$$

where the constraints on r are derived from the constraint that each genotype frequency lies between 0 and 1. Under the null hypothesis, $r = 0$ while $r \neq 0$ under the alternative hypothesis. p is a nuisance parameter to be estimated.

In the simple case of outbred individuals and a biallelic marker, where there are three possible genotypes

whose frequencies must sum to 1, our model for μ is equivalent to the fully parameterized model that has any two of the three genotype frequencies as free parameters. With a biallelic marker and inbred individuals, there is not a unique two-parameter model for μ that takes into account inbreeding. Either the HED model or the model based on a disequilibrium coefficient could be a natural choice. However, it turns out that in the biallelic case with inbreeding, these two models result in strictly equivalent score tests, so there is no need to choose between them. Thus, in the context of SNP studies where most deviations from HWE are likely to result from genotyping errors, our model is very general in allowing for deviation from HWE.

For multiallelic markers, our choice of model for μ narrows the alternative hypothesis to excess of all kinds of homozygous genotypes and lack of all kinds of heterozygous genotypes when $r > 0$ or the contrary when $r < 0$. A primary reason for choosing such a model is that it is relatively parsimonious in that it has only a free parameters. Furthermore, it is motivated by previous findings using microsatellites in which excess of all kinds of heterozygotes has been observed, in particular in the HLA region (OBER *et al.* 1997; ROBERTSON *et al.* 1999). Testing for an excess or a deficit of homozygosity is also of interest for detection of genotyping errors (SOBEL *et al.* 2002; see DISCUSSION). We believe that the HED model, in which the departure from HWE for homozygous genotypes depends on the frequency of the corresponding allele, is better suited to detection of genotyping errors than is the model based on a disequilibrium coefficient, in which the departure from HWE is the same for all homozygous genotypes. More complex models involving more than a parameters may be chosen when evidence for such alternative models is available. Otherwise, the increase in degrees of freedom of the test will reduce its power when the true alternative is indeed in the direction specified in our model.

Covariance matrix: For a pair of individuals in an inbred pedigree there are 15 possible ways in which the four alleles of the pair at a particular locus can be identical by descent (IBD; see Figure 1 and also LANGE 1997, Chap. 5). When the paternal and the maternal allele of an individual do not need to be distinguished, these 15 configurations reduce to 9, referred to as identity states. Following JACQUARD (1974), for a pair of individuals (i, j) , we let Δ_k^{ij} , $k = 1, \dots, 9$, denote the conditional probability of identity state k given the relationship between the individuals i and j . These Δ_k 's are called condensed identity coefficients. In outbred individuals, Δ_7^{ij} , Δ_8^{ij} , and Δ_9^{ij} are, respectively, the probability of sharing two, one, or zero alleles IBD given the pedigree, and all other $\Delta_k^{ij} = 0$. ABNEY *et al.* (2000) have implemented an efficient algorithm to compute the condensed identity coefficients in large complex pedigrees. Note that the inbreeding coefficients of individu-

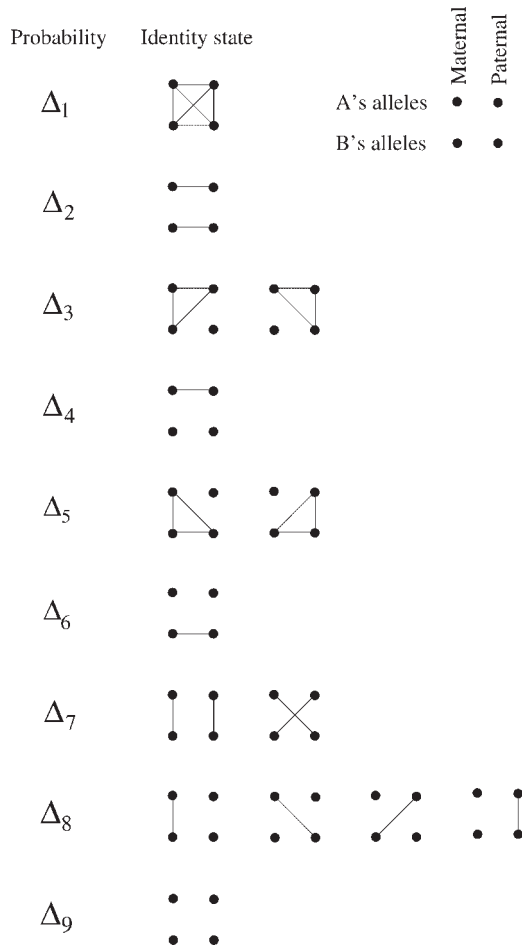


FIGURE 1.—Fifteen possible identity states for individuals A and B, grouped according to their nine condensed states. Lines indicate alleles that are IBD (from ABNEY *et al.* 2000).

als i and j can be obtained from the identity coefficients: $h_i = \Delta_1^{ij} + \Delta_2^{ij} + \Delta_3^{ij} + \Delta_4^{ij}$ and $h_j = \Delta_1^{ij} + \Delta_2^{ij} + \Delta_3^{ij} + \Delta_6^{ij}$. Let $\Sigma = \text{var}_0(Y)$ be the $[N(M-1) \times N(M-1)]$ covariance matrix of Y under the null hypothesis when $r = 0$. Table 3 gives the entries of Σ as a function of p and the 9 condensed identity coefficients. Define Ω to be the covariance matrix of Y under the alternative hypothesis, where Ω may depend on both p and r . To derive a test statistic, we require only that Ω be differentiable and invertible and that, under the null hypothesis, $\Omega = \Sigma$.

Quasi-score test for Hardy-Weinberg equilibrium (QL-HW test): We propose basing inference on the quasi-score function U (see, *e.g.*, McCULLAGH and NELDER 1989, Chap. 9):

$$U(r, p) = \begin{pmatrix} U_r(r, p) \\ U_p(r, p) \end{pmatrix} = \begin{pmatrix} D_r^T \Omega^{-1} (Y - \mu) \\ D_p^T \Omega^{-1} (Y - \mu) \end{pmatrix},$$

$$\text{where } D_r = \frac{\partial \mu}{\partial r} \text{ and } D_p = \frac{\partial \mu}{\partial p}.$$

Here D_r is a $[N(M-1)]$ vector and D_p is a $[(N(M-1)) \times (a-1)]$ matrix with the k th column equal to

$(\partial \mu / \partial p_k)$. Table 4 gives the expressions for $(\partial \mu_{ig} / \partial r)$ and $(\partial \mu_{ig} / \partial p_k)$. To construct our Hardy-Weinberg test statistic, we use a similar idea to that used in BOURGAIN *et al.* (2003) for the problem of case-control association testing. In analogy with the classical likelihood score statistic when there is a nuisance parameter (see, *e.g.*, COX and HINKLEY 1974, Chap. 9), we use U to build a quasi-score statistic $W_{\text{QL-HW}}$:

$$W_{\text{QL-HW}} = U_r^T(r_0, \hat{p}_0) I^r(r_0, \hat{p}_0) U_r(r_0, \hat{p}_0).$$

Here I^r is the (r, r) component of $[E(UU^T)]^{-1}$, where $E(UU^T)$ is analogous to the information matrix. In our case, the value of r under the null hypothesis is $r_0 = 0$. \hat{p}_0 is the quasi-likelihood estimate of the nuisance parameter p when $r = r_0$. In other words, $p = \hat{p}_0$ is the solution of

$$U_p(r_0, p) = [D_p^T \Sigma^{-1} (Y - \mu)]_{r=r_0} = 0 \quad (3)$$

subject to constraints (1), assuming that there is exactly one such solution. Because we cannot derive a closed-form solution for Equation 3, we obtain \hat{p}_0 by use of the Newton-Raphson method with Fisher scoring as presented by McCULLAGH and NELDER (1989) for the quasi-likelihood framework. \hat{p}_0 is obtained by iterating

$$\hat{p}_0^{i+1} = \hat{p}_0^i + [(\hat{D}_p^T \hat{\Sigma}^{-1} \hat{D}_p)^{-1} \hat{D}_p^T \hat{\Sigma}^{-1} (Y - \hat{\mu}_0)]_{p=\hat{p}_0^i} \quad (4)$$

until convergence occurs. Here, \hat{p}_0^i is the approximation to \hat{p}_0 at step i , $\hat{\Sigma}$, \hat{D}_p , and $\hat{\mu}_0$ are computed at $r = 0$ and $p = \hat{p}_0^i$. To reduce the number of iterations before convergence, we start the iteration procedure with \hat{p}_0^1 equal to the best linear unbiased estimator (BLUE) of p proposed by McPECK *et al.* (2004). This estimator is based on allele information and does not make use of the additional information of genotypes, *i.e.*, which alleles are paired within the individuals. Finally,

$$\begin{aligned} W_{\text{QL-HW}} &= (Y - \hat{\mu}_0)^T \hat{\Sigma}^{-1} \hat{D}_r \\ &\times [\hat{D}_r^T \hat{\Sigma}^{-1} \hat{D}_r - \hat{D}_r^T \hat{\Sigma}^{-1} \hat{D}_p (\hat{D}_p^T \hat{\Sigma}^{-1} \hat{D}_p)^{-1} \hat{D}_p^T \hat{\Sigma}^{-1} \hat{D}_r]^{-1} \\ &\times \hat{D}_r^T \hat{\Sigma}^{-1} (Y - \hat{\mu}_0), \end{aligned} \quad (5)$$

where $\hat{\Sigma}$, \hat{D}_p , \hat{D}_r , and $\hat{\mu}_0$ are evaluated at $r = 0$ and $p = \hat{p}_0$. As demonstrated in HEYDE (1997), under the null hypothesis, the quasi-score statistic should follow a χ^2 distribution with 1 d.f. provided that $\text{Var}_0(U)^{-1/2} U \sim \mathcal{MVN}(0, I)$ under the null hypothesis. We tested the accuracy of the χ^2 approximation to the null distribution of $W_{\text{QL-HW}}$ (see *Null distribution and power study of the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ statistics*).

To efficiently perform the calculations required in Equation 4, at each step, we take the Cholesky decomposition of $\hat{\Sigma}$, *i.e.*, we find an upper triangular matrix \mathbf{B} such that $\mathbf{B}^T \mathbf{B} = \hat{\Sigma}$, using an algorithm that simultaneously computes $b_0 = \mathbf{B}^{-T} \hat{D}_p$ and $b_1 = \mathbf{B}^{-T} (Y - \hat{\mu}_0)$ at little extra cost (GRAYBILL 1976). Then Equation 4 becomes $\hat{p}_0^{i+1} = \hat{p}_0^i + (b_0^T b_0)^{-1} b_0^T b_1$, where $b_0^T b_0$ is an $[(a-1) \times (a-1)]$ matrix, also inverted using its Cholesky decom-

TABLE 3
Expressions for the covariances among the genotype indicators Y_{ig}

Genotypes		$\text{cov}(Y_{ig1}, Y_{ig2})$
Covariances between indicators of genotypes g_1 and g_2 in individual i		
$g_1, g_2 = g_1$		$\mu_{i,g_1}^0(1 - \mu_{i,g_1}^0)^a$
$g_1, g_2 \neq g_1$		$-\mu_{i,g_1}^0 \mu_{i,g_2}^0$
g_1	g_2	$\text{cov}(Y_{ig1}, Y_{ig2})$
Covariances between indicators of genotype g_1 in individual i and genotype g_2 in individual j		
k/k	k/k	$p_k \Delta_1^{ij} + p_k^2(\Delta_2^{ij} + \Delta_3^{ij} + \Delta_5^{ij} + \Delta_7^{ij}) + p_k^3(\Delta_4^{ij} + \Delta_6^{ij} + \Delta_8^{ij}) + p_k^4 \Delta_9^{ij} - \mu_{i,k/h}^0 \mu_{j,k/h}^0$
k/k	l/l	$p_k p_l \Delta_2^{ij} + p_k p_l^2 \Delta_4^{ij} + p_k^2 p_l \Delta_6^{ij} + p_k^2 p_l^2 \Delta_8^{ij} - \mu_{i,k/h}^0 \mu_{j,l/l}^0$
k/k	k/l	$\Delta_3^{ij} p_k p_l + 2\Delta_4^{ij} p_k^2 p_l + \Delta_8^{ij} p_k^2 p_l + 2\Delta_9^{ij} p_k^3 p_l - \mu_{i,k/h}^0 \mu_{j,k/l}^0$
k/k	l/m	$2\Delta_4^{ij} p_k p_l p_m + 2\Delta_9^{ij} p_k^2 p_l p_m - \mu_{i,k/h}^0 \mu_{j,l/m}^0$
k/l	l/l	$\Delta_5^{ij} p_k p_l + 2\Delta_6^{ij} p_k^2 p_l + \Delta_8^{ij} p_k p_l^2 + 2\Delta_9^{ij} p_k^3 p_l - \mu_{i,k/l}^0 \mu_{j,l/l}^0$
k/m	l/l	$2\Delta_6^{ij} p_k p_m p_l + 2\Delta_9^{ij} p_k p_m p_l^2 - \mu_{i,k/m}^0 \mu_{j,l/l}^0$
k/m	k/m	$2\Delta_7^{ij} p_k p_m + \Delta_8^{ij}(p_k p_m + p_k p_m^2) + 4\Delta_9^{ij} p_k^2 p_m^2 - \mu_{i,k/m}^0 \mu_{j,k/m}^0$
k/m	k/l	$\Delta_8^{ij} p_k p_m p_l + 4\Delta_9^{ij} p_k^2 p_m p_l - \mu_{i,k/m}^0 \mu_{j,k/l}^0$
k/m	b/l	$4\Delta_9^{ij} p_k p_m p_b p_l - \mu_{i,k/m}^0 \mu_{j,b/l}^0$

^a The expressions for $\mu_{i,g}^0$ expectation of Y_{ig} under the null, are given in Table 2 setting $r = 0$.

position. $W_{\text{QL-HW}}$ is computed with a similar technique in Equation 5.

We conclude this section by describing the form of the QL-HW test statistic in the case of N individuals belonging to F independent families. For each f , $1 \leq f \leq F$, let N_f be the number of sampled individuals in family f , and let Y_f , μ_f , D_{ff} , D_{ff} , and $\hat{\Sigma}_f$ be the vectors or matrices previously described but restricted to the N_f members of family f . Allele frequencies are then computed using Equation 4, where $\hat{D}_f^T \hat{\Sigma}_f^{-1} \hat{D}_f = \sum_{j=1}^F \hat{D}_{ff}^T \hat{\Sigma}_f^{-1} \hat{D}_{ff}$ and $\hat{D}_f^T \hat{\Sigma}_f^{-1} (Y - \mu_0) = \sum_{j=1}^F \hat{D}_{ff}^T \hat{\Sigma}_f^{-1} (Y_f - \mu_{0f})$. Similarly $W_{\text{QL-HW}}$ is computed using Equation 5, where all terms of the form $X^T \hat{\Sigma}^{-1} B$ (where X and B are either vectors or matrices) in the formula are computed as $X^T \hat{\Sigma}^{-1} B = \sum_{j=1}^F X_f^T \hat{\Sigma}_f^{-1} B_j$.

GCC-HW test: As the number of alleles per locus and the number of individuals in a family increase, the calculations of the Cholesky decompositions required in (4) and (5) become much more computationally intensive. In practice, with this method, we were not able to analyze Hutterite samples of 400 individuals for loci with more than five alleles. To overcome this limitation, we developed an alternative test, the GCC-HW test, which is computationally simpler than the QL-HW test and therefore suitable for an even wider set of situations (larger samples, larger number of alleles per locus). For outbred samples in the biallelic case, the GCC-HW test is Pearson's Gof-HW corrected using the variance that appropriately accounts for inter-individual correlations. In inbred samples or when considering multiallelic loci, the GCC-HW test is a generalization of this approach. We start by presenting this alternative test in the case of a biallelic locus in an outbred sample with related individuals and present its general expression in a second step. We then clarify how it differs from the QL-HW test.

Special case—corrected χ^2 test for HW for biallelic loci in outbred samples with related individuals: Let us first consider a biallelic locus genotyped in an outbred sample with independent individuals. In this case, the QL-HW test is the same as the Gof-HW test, which can be written as $\text{Gof-HW} = (1/N)[(N\hat{p}_n^2 - N_{11})/\hat{p}_n(1 - \hat{p}_n)]^2$, where N_{ij} is the observed count for genotype ij and \hat{p}_n is the allele frequency estimate of allele 1 obtained by naive counting [$\hat{p}_n = (2N_{11} + N_{12})/2N$].

We next give an equivalent form for Gof-HW in terms of a vector-valued function

$$S(r, p) = \begin{pmatrix} S_r(r, p) \\ S_p(r, p) \end{pmatrix},$$

which can be left unspecified except that it has the property that

$$S(r_0, p) = \begin{pmatrix} D_r^T \mathbf{K}^{-1} (Y - \mu_0) \\ D_p^T \mathbf{K}^{-1} (Y - \mu_0) \end{pmatrix},$$

where $r_0 = 0$. Here D_r , D_p , Y , and μ_0 are the vectors and matrix described in *Quasi-score test for Hardy-Weinberg equilibrium (QL-HW test)*. \mathbf{K} is the $[(M-1)N \times (M-1)N]$ covariance matrix of Y . Because all the individuals are outbred and independent, \mathbf{K} is a block-diagonal matrix with $\mathbf{K} = \mathbf{I}_N \otimes \mathbf{K}_{\text{ind}}$, where \mathbf{K}_{ind} is a 2×2 matrix whose entries depend only on p and are given in the first part of Table 3. It turns out that the Gof-HW test is equal to W_{GFW}^2 , where

$$W_{\text{GFW}}^2 = (\hat{\Sigma}_r^T [\widehat{\text{var}}_0(S)^{-1}]_r \hat{\Sigma}_r). \quad (6)$$

Here, $\hat{\Sigma}_r$ is computed at $(r, p) = (r_0, \hat{p}_0)$. $p = \hat{p}_0$ is the solution of $S_p(r_0, p) = 0$ and is actually equal to \hat{p}_n . $[\widehat{\text{var}}_0(S)^{-1}]_r$ denotes the (r, r) entry of $\text{var}(S)^{-1}$ computed at $(r, p) = (r_0, \hat{p}_0)$. It is actually equal to $1/N$.

TABLE 4
Expressions for $(\partial\mu/\partial p_k)$ and $(\partial\mu/\partial r)$ in the case of a locus with a different alleles

Genotype	$\partial\mu_{ig}/\partial p_k$	$\partial\mu_{ig}/\partial r$
k/k	$h_i + r + 2(1 - h_i - r)p_k$	$p_k(1 - p_k)$
l/l	0	$p_l(1 - p_l)$
a/a	$-2(1 - h_i - r)(1 - \sum_{j=1}^{a-1} p_j) - h_i - r$	$(1 - \sum_{j=1}^{a-1} p_j)\sum_{j=1}^{a-1} p_j$
k/l	$2(1 - h_i - r)p_l$	$-2p_k p_l$
m/l	0	$-2p_m p_l$
a/l	$-2(1 - h_i - r)p_l$	$-2p_l(1 - \sum_{j=1}^{a-1} p_j)$
a/k	$2(1 - h_i - r)(1 - \sum_{j=1}^{a-1} p_j - p_k)$	$-2p_k(1 - \sum_{j=1}^{a-1} p_j)$

One way to extend (6) so that it is valid when the individuals are related is to consider the same $S(r, p)$ function (with $\mathbf{K} = \mathbf{I}_N \otimes \mathbf{K}_{\text{ind}}$, thus neglecting the inter-individual correlations) and to recompute $[\widehat{\text{var}}_0(S)^{-1}]_r$ to take the inter-individual correlations into account. We have $\text{var}_0(Y) = \Sigma$, with elements given in Table 3, where all $h_i = 0$ in this case. Using $\text{var}_0(S) = \hat{E}_0(SS^T)$ and noting that $D_r^T \mathbf{K}^{-1} D_p = 0$ for outbred individuals, we get $[\widehat{\text{var}}_0(S)^{-1}]_r = [\hat{D}_r^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_r]^{-1}$, where $\hat{\Sigma}$, $\hat{\mathbf{K}}$, and \hat{D}_r are computed at $(r, p) = (r_0, \hat{p}_0)$ (here again, $\hat{p}_0 = \hat{p}_n$). We assess the accuracy of the χ^2 approximation to the null distribution of this corrected χ^2 ($W_{\text{CC-HW}}$) statistic in *Null distribution and power study of the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ statistics*.

General form of GCC-HW test: There does not appear to be a Pearson χ^2 test corresponding to the HED model for multiallelic loci. Furthermore, in the case of inbreeding, while one can extend the Gof-HW test to allow for a mean level of inbreeding, it does not take into account the different inbreeding coefficients for different individuals. To generalize the $W_{\text{CC-HW}}$ statistic, we consider the following framework. Let

$$C(r, p) = \begin{pmatrix} C_r \\ C_p \end{pmatrix}$$

and define (\hat{r}_p) to be the solution of $C(r, p) = 0$ subject to constraints (1) and (2), assuming that there is exactly one such solution. Although the framework is more general, we consider here C to be the form

$$C_i = D_i^T \mathbf{L}^{-1} (Y - \mu) \quad \text{with } i = r, p, \mu = E(Y), \text{ and } D_i = \partial\mu/\partial i, \quad (7)$$

for some invertible matrix \mathbf{L} that may depend on r and p , with D_p , \mathbf{L} , and μ not depending on Y . Assume that (\hat{r}_p) is consistent for the parameters (r_p) of the HED model and is asymptotically normal. Following the arguments used by COX and HINKLEY (1974, Chap. 9) to derive the score test, under certain regularity conditions, we consider the expansion of $C(r, p)$ in a Taylor series about (\hat{r}_p)

$$C(r, p) \approx C(\hat{r}, \hat{p}) + C'(\hat{r}, \hat{p}) \begin{pmatrix} r - \hat{r} \\ p - \hat{p} \end{pmatrix} \approx -E(C'(r, p)) \begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix},$$

where

$$C'(r, p) = \begin{pmatrix} \frac{\partial C_r}{\partial r} & \frac{\partial C_r}{\partial p} \\ \frac{\partial C_p}{\partial r} & \frac{\partial C_p}{\partial p} \end{pmatrix}.$$

Then,

$$\begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \approx -E(C'(r, p))^{-1} C(r, p) \quad (8)$$

and

$$\text{var} \begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \approx E(C'(r, p))^{-1} \text{var}[C(r, p)] E(C'(r, p))^{-1}. \quad (9)$$

Under the assumptions made on (\hat{r}_p) ,

$$W = (\hat{r} - r)^T \left[\text{var} \begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \right]^{-1} (\hat{r} - r) \quad (10)$$

should have an asymptotic χ^2 distribution with $\dim(r)$ d.f. Here

$$\text{var} \begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \Big|_r$$

is the (r, r) entry of $\text{var}(\frac{\hat{r}-r}{\hat{p}-p})$. To test the null hypothesis $r = r_0$, we plug in $r = r_0$ in (10). p is a nuisance parameter and we use its restricted estimator \hat{p}_0 , where $p = \hat{p}_0$ is the solution of $C_p(r_0, p) = 0$, subject to constraints (1), assuming that there is exactly one such solution. Under certain regularity conditions, this is asymptotically equivalent to the case when the true p is used.

The $W_{\text{GCC-HW}}$ statistic has the general form defined in (10), where $C(r, p)$ is given by (7) with \mathbf{L} set to \mathbf{K} , the matrix defined in the subsection *Special case—corrected χ^2 test for HW for biallelic loci in outbred samples with related individuals*.

$$(\hat{r} - r) \quad \text{and} \quad \left[\text{var} \begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \right]_r^{-1}$$

are approximated with Equations 8 and 9, respectively. Doing so, we can consider the same HED model as the one used in the QL-HW test, which allows a distinct h_i for each individual, by simply using for μ the model described in Table 2. To test for HW, we use $r_0 = 0$. The \mathbf{K} matrix

used in $C(r, p)$ is the covariance matrix of Y where the intra-individual correlations are taken into account but where the inter-individuals correlations are neglected. Thus, \mathbf{K} is still a block diagonal matrix but with distinct matrix elements \mathbf{K}_i of size $(M - 1) \times (M - 1)$, whose entries depend on p and h_b , and \mathbf{K}^{-1} is a block-diagonal matrix with matrix element \mathbf{K}_i^{-1} . We show in the APPENDIX that $W_{\text{GCC-HW}}$ reduces to $W_{\text{GCC-HW}} = \hat{C}_r^2 / \text{den}_{\text{GCC}}$, where $\hat{C}_r = C_r(\hat{r}_0, \hat{p}_0)$ and

$$\begin{aligned} \text{den}_{\text{GCC}} = & [\hat{D}_i^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_i - 2 \hat{D}_i^T \hat{\mathbf{K}}^{-1} \hat{D}_p (\hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{D}_p)^{-1} \hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_i \\ & + \hat{D}_i^T \hat{\mathbf{K}}^{-1} \hat{D}_p (\hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{D}_p)^{-1} \hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_p (\hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{D}_p)^{-1} \hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{D}_i] \end{aligned} \quad (11)$$

computed at $(r, p) = (\hat{r}_0, \hat{p}_0)$. In outbred samples, \hat{p}_0 corresponds to the naive counting estimator in which allele frequencies are estimated by their observed proportions. In inbred samples, $C_p(\hat{r}_0, p) = 0$ may not have a closed-form solution, in which case we use the Newton-Raphson method, as we did for the QL-HW test: \hat{p}_0 is obtained by iteration using Equation 4, where Σ is replaced by \mathbf{K} . We start the iteration procedure with \hat{p}_0^1 equal to the BLUE (McPECK *et al.* 2004) computed under the pretense that all the individuals have their correct inbreeding coefficients but are independent. Note that, because $D_i^T \mathbf{K}^{-1} D_p = 0$ when considering a biallelic locus in an outbred sample, $W_{\text{CC-HW}}$ is a special case of $W_{\text{GCC-HW}}$. Whereas the computation of $W_{\text{QL-HW}}$ requires the Cholesky decomposition of Σ , a covariance matrix of size $[(M - 1)N \times (M - 1)N]$, the computation of $W_{\text{GCC-HW}}$ requires only the Cholesky decompositions of the $N \mathbf{K}_i$ matrices of size $[(M - 1) \times (M - 1)]$. It is thus much less computationally intensive. Similarly to the $W_{\text{QL-HW}}$ quasi-score statistic, we assess the accuracy of the χ^2 approximation to the null distribution of $W_{\text{GCC-HW}}$ in RESULTS.

When the N individuals belong to F independent families, $W_{\text{GCC-HW}}$ and \hat{p}_0 are computed using the same formulas as in the one-family case presented above but with all the terms of the form $X^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} B$ computed as $X^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} B = \sum_{j=1}^F X_j^T \hat{\mathbf{K}}_j^{-1} \hat{\Sigma}_j \hat{\mathbf{K}}_j^{-1} B_j$.

Explanation of the difference between QL-HW and GCC-HW: Because both have an appropriate variance correction, both QL-HW and GCC-HW are valid tests that are approximately χ^2 distributed in sufficiently large samples under regularity conditions. The difference is that, in the definition of GCC-HW, the matrix \mathbf{K} is plugged in for \mathbf{L} in Equation 7, while in the definition of QL-HW, the matrix Σ is plugged in for \mathbf{L} . This difference potentially affects the power of the statistics. For instance, by an argument similar to that given in BOURGAIN *et al.* (2003), the QL-HW should be asymptotically more powerful than GCC-HW for local alternatives (*i.e.*, alternatives close to the null hypothesis) under certain regularity conditions, assuming that the HED model holds, and, in fact, should be asymptotically locally most powerful of all statistics of the type defined by Equation

7. We compare actual power of GCC-HW and QL-HW under simulated scenarios in RESULTS.

Null distribution and power study of the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ statistics: To test the accuracy of the χ^2 approximation to the null distributions of the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ statistics, we assessed their distributions under the null hypothesis using (i) a sample of 30 outbred independent nuclear families with two parents and two sibs typed (sample nuc), (ii) a sample of 456 individuals from 10 extended outbred genealogies (sample gaw), and (iii) two samples of 611 (sample H1) and 306 (sample H2) Hutterites. The 10 extended genealogies were simulated for the Genetic Analysis Workshop 12 (ALMASY *et al.* 2001). The different genealogies are independent but the mean within-genealogy kinship coefficient is 0.0415 ([0.030–0.054]). The Hutterites are a North American religious isolate originating from Tyrol whose entire population can be traced back to 90 ancestors in the 1700s/1800s. The complete genealogy of the sample could be constructed from a $\geq 12,000$ -person Hutterite pedigree. This yielded a 1623-person pedigree that included all known ancestors of the individuals in samples H1 and H2. The nine identity coefficients for all pairs of individuals, from which the inbreeding coefficients can also be obtained, were computed using the efficient algorithm implemented by ABNEY *et al.* (2000). Genotype information for a marker was simulated by randomly drawing alleles independently from the same allele frequency distribution for (i) the parents of the nuclear families or (ii) the founders of the 10 genealogies or (iii) the founders of the 1623-person pedigree. Mendelian transmission of these alleles was then simulated (i) in the nuclear families, (ii) in the 10 genealogies, or (iii) throughout the 1623-person pedigree. The validity of the χ^2 null distribution for $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ was assessed by comparing the proportion of simulations showing a statistic whose value is greater than the χ^2 threshold for a nominal type I error and the value of this nominal type I error. For the outbred nuc and gaw samples, the empirical type I error of the classical Gof-HW test was also assessed, using either the χ^2 distribution or a permutation procedure (where alleles are randomly reassigned to genotypes 5000 times) to obtain significance.

For the Hutterite samples, in addition to the χ^2 approximation, we also had to develop a parametric bootstrap procedure to assess significance, as follows: for a given data set, we estimated the allele frequencies (\hat{p}_{data}) using Equation 4 (with $\hat{\Sigma}$ for $W_{\text{QL-HW}}$ and $\hat{\mathbf{K}}$ for $W_{\text{GCC-HW}}$) and computed the value of the statistic W_{data} (either $W_{\text{QL-HW}}$ or $W_{\text{GCC-HW}}$). \hat{p}_{data} was then used to simulate 5000 replicates of the original data set under the null hypothesis (following the procedure described in the previous paragraph). The statistic was computed in each replicate. The empirical significance was assessed as the proportion of simulated statistic values greater than W_{data} . In principle, one could double-check the type I error

TABLE 5

Parent/offspring transmission model used to mimic selection in the power computations

Parental genotypes (G_p)	Offspring genotype (G_o)	$\Pr(G_o G_p)^*$
$ii \times ii$	ii	1
$ii \times jj$	ij	1
$ii \times ij$	ii	$1 - s$
	ij	s
$ij \times ij$	ii	$(1 - s)/2$
	jj	$(1 - s)/2$
	ij	s

* Probability of offspring genotype conditional on parental genotypes.

of this parametric bootstrap procedure by performing a nested set of simulations, but we did not attempt this for computational reasons.

Simulations were also used to assess power. Even though HW testing is used mostly to identify potential genotyping errors in SNP studies, generating a genotyping error model would also have created Mendelian incompatibilities. As Mendelian incompatibilities are routinely checked before any other test, considering such a genotyping error model would have implied evaluation of a combined Mendelian checking/HW testing approach, whereas we wanted to focus solely on HW testing. Thus, we chose to consider a model without Mendelian incompatibilities. Further, we required this model to be suitable for samples with fixed genealogy. We simulated a selection pattern where the allele transmission from parent to offspring is not equally likely for the two parental alleles (we retain independence of children's genotypes conditional on parents' genotypes). For a biallelic marker, a single parameter $0 < s < 1$ is required to describe our model. Table 5 gives the corresponding transmission probabilities from parents to offspring as a function of s . Here, we considered the situation where, for parental couples who are compatible with both homozygous and heterozygous children, the probability that any given child is heterozygous is greater than would be expected under Mendelian inheritance: $s > 1/2$. Power was assessed in the samples nuc, H1 and H2 using 5000 replicates. For the Hutterite samples, founders' alleles were simulated assuming HWE in the founding population, with founding allele frequency taken to be 0.5, 0.7, or 0.85. In the case of nuc, we assume that the families are sampled from a population that is in a steady state for our selection model. We give in Table 6 the population genotype frequencies corresponding to this steady state in an infinite-size population under random mating. When $s > 1/2$, the genotype frequency distribution depends only on s and not on p (whatever the initial allele frequency, $p = 0.5$ in the steady state). The parental genotypes of the nuc families were sampled independently from the steady-state genotype distribution for various

choices of $s > 0.5$. In such a nuclear family sampling scheme, the classical Gof-HW test would consider only the 60 independent parents. To measure the gain in power obtained while using QL-HW on the global sample instead of the Gof-HW test on the parent sample, we derived analytically the power of the Gof-HW test using a noncentral χ^2 distribution.

Analysis of biallelic markers: We used the W_{QL-HW} statistic to test for HW in a set of 143 biallelic markers in 82 genes that were genotyped in samples of Hutterites. A variety of genotyping methods were used: multiplex PCR and immobilized probe linear array system (LAS) (MIREL *et al.* 2002) for 48 markers (34%), single base extension with fluorescent polarization (SBE-FP; CHEN *et al.* 1999) for 44 markers (31%), dot blot with allele-specific oligos (OBER *et al.* 2000) for 26 markers (18%), restriction fragment length polymorphism analysis (RFLP) for 13 markers (9%), size separation on acrylamide gel for 9 insertion/deletion (in/del) polymorphisms (6%), allele-specific (AS) PCR for 1 marker (0.7%), SSCP (GONEN *et al.* 1999) for 1 marker (0.7%), and denaturing high-performance liquid chromatography (DHPLC; OEFNER and UNDERHILL 1995) for 1 marker (0.7%). All markers were checked for Mendelian incompatibilities in the general pedigree using PEDCHECK (O'CONNELL and WEEKS 1998).

RESULTS

Null Distribution of the W_{QL-HW} and W_{GCC-HW} statistics:

Table 7 presents the empirical type I error of the classical Gof-HW, W_{QL-HW} , and W_{GCC-HW} statistics in the outbred samples of nuclear families and in the extended outbred genealogies. The classical Gof-HW test, whatever the significance procedure used, either the χ^2 distribution or the permutation test, is anti-conservative. While the inflation in the type I error is moderate in the nuclear families (type I error of 7% instead of 5%), it becomes substantial in the extended GAW12 genealogies (type I error of 10%). In contrast, when the χ^2 distribution is used to assess significance of the W_{QL-HW} or W_{GCC-HW} statistics, the empirical type I error lies within the 95% confidence interval of the nominal type I error (5%) for all three allelic distributions considered in the founders. In the two Hutterite samples (results shown in Table 8), the empirical type I error for both W_{QL-HW} and W_{GCC-HW} is smaller than the nominal one. Use of the χ^2 approximations to the null distributions of the statistics would thus lead to conservative tests in the Hutterite sample. We note, however, that when simulating a sample consisting of 10 independent copies of the 1623-person Hutterite pedigree and running the two statistics on a sample of 6110 Hutterites (10 times sample H1), the empirical type I error for a biallelic locus with frequency 0.5 in the founders is not statistically different from the nominal one when assuming a χ^2 distribution. Similarly, when considering only 2 outbred genealogies out of the 10 from the gaw sample (corresponding to

TABLE 6

Steady-state genotype frequencies for the selection model presented in Table 5 under random mating, for a biallelic locus, as a function of s and of the initial allele frequency p

Genotypes	11	12	22
$s = \frac{1}{2}$	p^2	$2p(1 - p)$	$(1 - p)^2$
$s > \frac{1}{2}$	$(-\frac{1}{2} + \frac{1}{2}a)/(1 - 2s)$	$(2(1 - s) - a)/(1 - 2s)$	$(-\frac{1}{2} + \frac{1}{2}a)/(1 - 2s)$
$s < \frac{1}{2}$			
$p = \frac{1}{2}$	$(-\frac{1}{2} + \frac{1}{2}a)/(1 - 2s)$	$(2(1 - s) - a)/(1 - 2s)$	$(-\frac{1}{2} + \frac{1}{2}a)/(1 - 2s)$
$p < \frac{1}{2}$	0	0	1
$p > \frac{1}{2}$	1	0	0

With $a = \sqrt{4s^2 - 6s + 3}$.

106 individuals deriving from 41 founders), the χ^2 distribution provides valid type I errors for both the $W_{\text{QL-HW}}$ and the $W_{\text{GCC-HW}}$ tests. The asymptotic conditions required for the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ to have a χ^2 distribution under the null hypothesis should thus be met in most samples, except in the extreme case where all the individuals belong to a single inbred pedigree. In this case, the central limit theorem either may not apply or may require a larger sample size to be a good approximation, and the test may be conservative. We give in Table 8 the thresholds corresponding to a 5% nominal type I error, for three different allelic distributions, in the two Hutterite samples. These thresholds vary with sample sizes and allele frequencies. We would thus recommend use of empirical P -values obtained by the parametric bootstrap (as described in METHODS), for samples in which all the individuals come from a single inbred pedigree.

Power of the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ statistics: Figure 2 displays the power of $W_{\text{QL-HW}}$ in the nuc sample (30 nuclear families) for values of s varying between 0.5 and

0.75, where power was computed using the χ^2 approximation. Figure 2 also displays the power of the Gof-HW test in the sample of 60 independent parents that can be extracted from the nuclear families. The gain in power obtained while using $W_{\text{QL-HW}}$ instead of Gof-HW can be substantial: when $s = 0.65$, the power of $W_{\text{QL-HW}}$ is 79% instead of 55% for the Gof-HW test. There is still a notable increase in power at more realistic values of s close to 0.5. The power of $W_{\text{GCC-HW}}$ was also computed for the nuc sample, but it is not shown in the plot as the results were almost indistinguishable from those for $W_{\text{QL-HW}}$. Thus, the simpler $W_{\text{GCC-HW}}$ appears to be approximately as powerful as the $W_{\text{QL-HW}}$ in this sampling scheme. Because they allow one to use all the information available in the families and thus to increase the sample size, the QL-HW and GCC-HW tests may significantly enhance the power to detect HWD when familial data are available.

In the Hutterite sample, founder genotypes are not available, so the Gof-HW test in founders cannot be performed. Thus, we compare power only for the $W_{\text{QL-HW}}$

TABLE 7

Empirical type I error of the classical χ^2 goodness-of-fit test for HWE with significance assessed using either the χ^2 distribution or a permutation procedure and of the $W_{\text{QL-HW}}$ and $W_{\text{GCC-HW}}$ statistics with significance assessed using the χ^2 distribution, for a 5% nominal type I error

Allele frequency	Statistic:		$W_{\text{QL-HW}}$:		$W_{\text{GCC-HW}}$:
	Significance with:	χ^2 distribution	Permutation	χ^2 distribution	χ^2 distribution
Nuclear families					
0.50		0.073	0.075	0.051	0.052
0.70		0.072	0.074	0.053	0.053
0.85		0.070	0.073	0.046	0.044
GAW12 genealogies					
0.50		0.103	0.102	0.050	0.049
0.70		0.099	0.102	0.052	0.049
0.85		0.086	0.094	0.047	0.042

Estimates are based on 5000 simulations in a sample of 30 outbred nuclear families (nuc) and in a sample of 456 individuals from 10 extended genealogies of the GAW12.

TABLE 8

Empirical type I error of the W_{QL-HW} and W_{GCC-HW} statistics for a 5% nominal type I error and 5% corrected threshold, in two different Hutterite samples: H1 (611 individuals) and H2 (306 individuals)

Allele frequency ^a	Statistic: Sample:	Empirical type I error				5% corrected threshold ^b			
		W_{QL-HW}		W_{GCC-HW}		W_{QL-HW}		W_{GCC-HW}	
		H1	H2	H1	H2	H1	H2	H1	H2
0.50		0.017	0.029	0.020	0.026	2.61	3.00	2.83	3.00
0.70		0.017	0.015	0.014	0.024	2.55	2.71	2.58	2.85
0.85		0.008	0.018	0.003	0.009	2.21	2.59	1.88	2.18

Estimates are based on 5000 simulations.

^a Allele frequency distribution from which the alleles of the pedigree founders are drawn.

^b Uncorrected threshold is 3.84.

and W_{GCC-HW} tests. To see a power difference between the two tests, we set $s = 0.6$, even though this value is probably unrealistically high for selection in a human population. Table 9 displays the results, where significance is assessed using a correct 5% nominal threshold obtained by simulation under the true model (the thresholds are shown in Table 8). For the purpose of simulation studies, this method of assessing significance was used because it is much faster than the parametric bootstrap that we use in the real data examples. Because the same method is used with both statistics, we expect this to have a negligible impact on the power comparison. The W_{QL-HW} performs slightly better than the W_{GCC-HW} in the Hutterite samples; however, the power difference between the two tests is quite small. Thus, use of W_{GCC-HW} instead of W_{QL-HW} in situations in which W_{QL-HW} cannot be computed (*e.g.*, a highly polymorphic

marker in a pedigree with many individuals typed) should not have too strong an impact on power. Conversely, W_{QL-HW} should be preferred when computable. Note that the power of both methods in the Hutterite samples is expected to increase when the allele frequency becomes closer to 0.5. Indeed, as can be seen from Table 5, our model is such that selection acts only when there is at least one heterozygous parent. When the allele frequency becomes closer to 0.5, the proportion of heterozygous parents increases and the impact of selection increases as a result.

Marker testing for HWE in the Hutterites: The results for the 14 markers, of the 143 biallelic markers tested, that showed an associated P -value $< 5\%$ by the W_{QL-HW} are shown in Table 10. Four of these markers (HLA-G_689, LTA_26, OR12D2_2, OR2H3_587) are located within the HLA region, which has been associated with homozy-

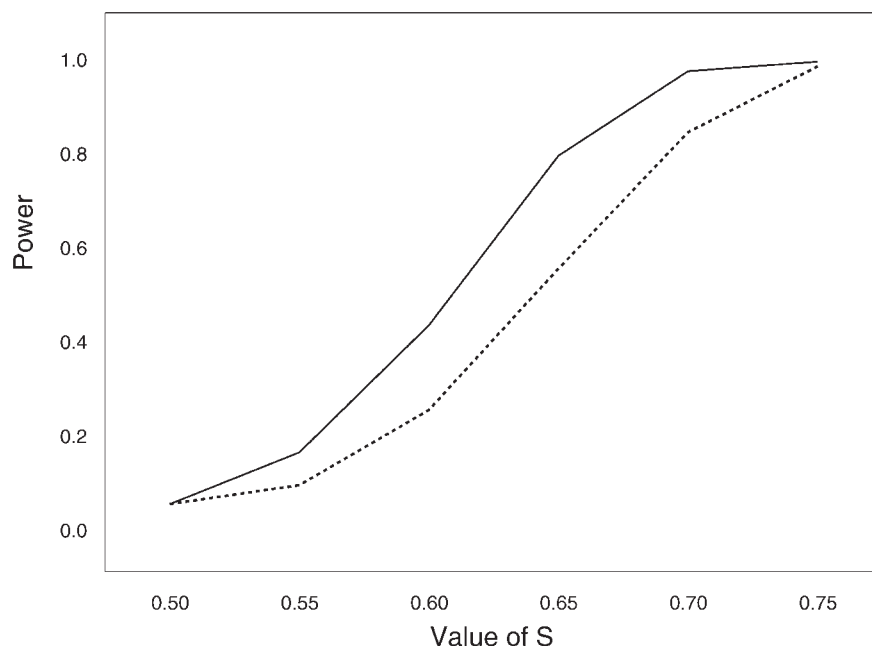


FIGURE 2.—Power of the QL-HW test in 30 nuclear families (solid line) and power of the Gof-HW test in the corresponding parent sample (dashed line), as a function of s , for a 5% type I error.

TABLE 9

Power of the W_{QL-HW} and W_{GCC-HW} statistics for different allele frequency distributions in two Hutterite samples (H1 and H2) with $s = 0.60$

Sample	H1			H2		
Allele frequency ^a	0.5	0.7	0.85	0.5	0.7	0.85
W_{QL-HW}	0.92	0.88	0.76	0.78	0.76	0.63
W_{GCC-HW}	0.90	0.88	0.76	0.78	0.73	0.62

Nominal type I error is 5%; estimates are based on 5000 simulations.

^a Allele frequency distribution from which the alleles of the pedigree founders are drawn.

gous deficiencies in this population (ROBERTSON *et al.* 1999). The genotyping methods used to type the 14 markers in Table 10 were not very different from those used to type the 143 markers except that the only marker typed with allele-specific PCR was among the 14 markers with a P -value ≤ 0.05 , markers typed by LAS were underrepresented (34% of all markers but only 14% of the markers with a P -value ≤ 0.05), and markers typed by dot blot were overrepresented (18% of all markers but 28% of the markers with a P -value ≤ 0.05) (see Figure 3). Thus, it is unlikely that any particular genotyping method was more error prone or contributing

disproportionately. We further investigated the genotyping history of these 14 markers. Two markers, HLA-G_689 and OR12D2_2, produced an excess of Mendelian errors and had already been either discarded (for OR12D2_2) or retyped using another method (for HLA-G_689). Two additional markers, PSG4_1 and PLAUR_210, had a high rate of ambiguous genotypes. Overall, it is likely that the deviations from HWE for these 4 markers were due to genotyping errors, although only 2 were previously identified as problematic by Mendelian error checking.

Finally, the remaining 10 markers with a P -value ≤ 0.05 of 143 (6.9%) represent slightly more than what is expected just by chance (but the observed proportion is not significantly different from 5% at the 5% level). Given that the P -values associated with each of these 10 markers are not dramatically small and that we had at least 1 other marker typed in 9 of these genes that was not significant, we believe that the small deviations from expectation in these 9 genes might be due to chance. We cannot rule out, however, that these departures reflect the real action of evolutionary forces.

DISCUSSION

We demonstrated in this article that the classical Gof-HW test has inflated type I error in samples with related

TABLE 10

Results of the W_{QL-HW} test in the Hutterites for the 14 biallelic markers with an associated empirical P -value ≤ 0.05

Marker name	Genotyping method	rs no.	Sample size	Allele frequency	W_{QL-HW} P -value
PLAUR_210 ^a	Dot blot ^c	RS2302524	604	0.85	0.0002
PON1_55	LAS ^d	RS85460	646	0.74	0.003
PSG4_1 ^a	Dot blot ^c	RS4028448	306	0.52	0.015
C5_802	SBE-FP ^e	RS17611	394	0.51	0.015
ZFR_1	SBE-FP ^e	RS714932	758	0.67	0.023
IL4R_3223	AS PCR ^f	RS2057768	413	0.68	0.027
LTA_26 ^b	LAS ^d	RS1041981	671	0.83	0.028
NIDDM1_43	Dot blot ^c	RS3792267	675	0.68	0.034
ESE-3_50834	in/del ^g	RS4987414	504	0.51	0.036
IL4R_1597	Dot blot ^c	RS15987825	410	0.56	0.036
OR12D2_2 ^{a,b}	SBE-FP ^e	—	632	0.52	0.039
HLA-G_689 ^{a,b}	SBE-FP ^e	RS2735022	679	0.56	0.041
ZFR_3	RFLP ^h	RS889318	681	0.66	0.042
OR2H3_587 ^b	SBE-FP ^e	RS3129034	672	0.71	0.043

Genotyping methods and dbSNP rs numbers are shown. Exact P -values are computed with 5000 simulations.

^a Markers for which genotyping errors are suspected.

^b Markers in the HLA region.

^c Dot blot with allele-specific oligos.

^d Multiplex PCR and immobilized probe linear array system.

^e Single-base extension with fluorescent polarization.

^f Allele-specific PCR.

^g Size separation on acrylamide gel.

^h Restriction fragment length polymorphism analysis.

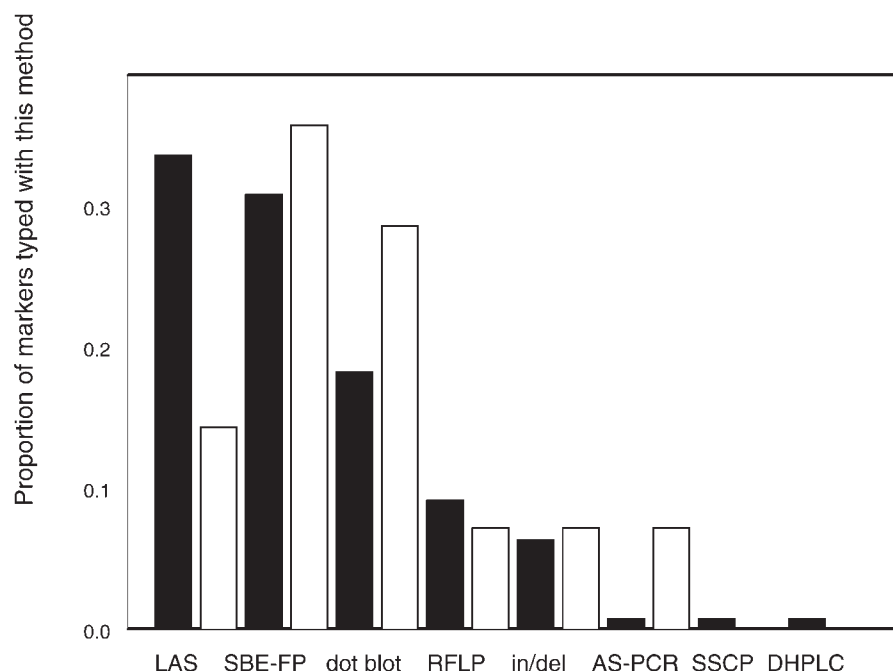


FIGURE 3.—Proportion of markers typed with the eight different genotyping methods for the set of 143 SNPs tested (solid bars) and for the 14 SNPs with an associated P -value ≤ 0.05 (open bars).

individuals when significance is assessed either with a χ^2 approximation or with a permutation test. SING and ROTHMAN (1975) previously discussed the problem of testing for HWE in small human populations with non-negligible correlations among the individuals. Whereas they corrected the expectation of the χ^2 in a very particular case of correlation among the observations, we propose in this article a general approach suitable for any sample with related individuals provided that their genealogy is known. Our QL-HW test models the correlation between the two alleles of each individual due to the pedigree structure and, by making use of the quasi-likelihood framework, takes into account the correlations among the individuals in the construction of the test.

We chose to base our test on the model of heterozygote excess or deficiency studied by ROUSSET and RAYMOND (1995), and we extended it to allow a different inbreeding coefficient for each individual in the sample. In the context of SNP studies, where deviations from HWE are likely to result from genotyping errors, our model is very general in allowing for deviations from HWE. For multiallelic markers, our model specifically tests for an excess or a deficiency of homozygosity. A primary reason for choosing such a model is parsimony. Furthermore, it is motivated by previous findings using microsatellites in which excess of all kinds of heterozygotes has been observed, in particular in the HLA region (OBER *et al.* 1997; ROBERTSON *et al.* 1999). This model is also relevant for detection of genotyping errors. Indeed, as noted by SOBEL *et al.* (2002), when automated software or trained technicians genotype by scoring bands on a gel, the most common error is false homozygosity. An error occurs in this case when an allele ampli-

fies insufficiently or a band falls outside a prescribed range. In addition, these authors also discuss the possibility of false heterozygosity when a stutter band is misinterpreted as a second allele. Thus, we expect our model to be useful for detecting genotype errors in the cases of both SNPs and multiallelic markers. Alternatively, a model involving more parameters could be incorporated into our quasi-likelihood score test framework, if the available evidence favored such a model.

We also propose a simpler alternative test, the GCC-HW test, to be used when the QL-HW test might not be computationally feasible, for instance, in the case of highly polymorphic loci in large inbred pedigrees. We found that for a variety of study designs, ranging from many nuclear families to two or more extended genealogies, the χ^2 approximation to the null distributions of both the QL-HW and the GCC-HW was very accurate. It was only in the extreme case of a single large inbred pedigree that the tests were rather conservative. In that case, we recommend using the parametric bootstrap to assess significance. Further, we showed that the use of the GCC-HW test in place of the QL-HW test has only a very minor effect on the power to detect deviation from HWE. Finally, we illustrated on nuclear family data how the QL-HW test can improve the power to detect departure from HWE over the classical Gof-HW test.

We believe that, by making the HWE test available for a larger variety of samples and by increasing sample sizes through the inclusion of related individuals that are typically discarded, both the QL-HW and GCC-HW tests are of crucial interest in several relevant situations for geneticists. In particular, as illustrated by the data analysis presented here, HWE testing is an additional tool for genotyping error detection. Indeed, of the four

markers with HWD and strong suspicion of typing errors after scrupulous checking, only two had been previously detected by regular Mendelian error checking. We thus strongly recommend that HW tests be routinely performed to detect genotyping errors, particularly while working with SNPs. Finally, used very cautiously, the QL-HW and GCC-HW tests could also be considered as possible tools to detect association in samples with related individuals. The program to run the QL-HW and GCC-HW tests is available on request from C.B. (dbSNP home page, <http://www.ncbi.nlm.nih.gov/SNP/index.html>).

The authors acknowledge Dina Newman for technical assistance and two anonymous reviewers for their thoughtful comments. The GAW12 data are available thanks to the Genetic Analysis Workshop grant GM31575. This study was supported by National Institutes of Health grants HL56399, HL66533, HG001645, and DK55889.

LITERATURE CITED

- ABNEY, M., M. S. McPECK and C. OBER, 2000 Estimation of variance components of quantitative traits in inbred populations. *Am. J. Hum. Genet.* **66**: 629–650.
- ALMASY, L., J. D. TERWILLIGER, D. NIELSEN, T. D. DYER, D. ZAYKIN *et al.*, 2001 GAW12: simulated genome scan, sequence, and family data for a common disease. *Genet. Epidemiol.* **21**: S332–S338.
- BOURGAIN, C., S. HOFFJAN, R. NICOLAE, D. NEWMAN, L. STEINER *et al.*, 2003 Novel case-control test in a founder population identifies P-selectin as an atopy susceptibility locus. *Am. J. Hum. Genet.* **73**: 612–626.
- CHEN, X., L. LEVINE and P. Y. KWOK, 1999 Fluorescence polarization in homogeneous nucleic acid analysis. *Genome Res.* **9**: 492–498.
- COX, D. R., and D. V. HINKLEY, 1974 *Theoretical Statistics*. Chapman & Hall, London.
- DOUGLAS, J. A., A. D. SKOL and M. BOEHNKE, 2002 Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am. J. Hum. Genet.* **70**: 487–495.
- GONEN, D., J. VEENSTRA-VANDERWEELE, Z. YANG, B. L. LEVENTHAL and E. H. J. COOK, 1999 High throughput fluorescent CE-SSCP SNP genotyping. *Mol. Psychiatry* **4**: 339–343.
- GRAYBILL, F. A., 1976 *Theory and Application of the Linear Model*. Duxbury Press, North Scituate, MA.
- HEYDE, C., 1997 *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer-Verlag, New York.
- JACQUARD, A., 1974 *The Genetic Structure of Populations*. Springer-Verlag, New York.
- LANGE, K., 1997 *Mathematical and Statistical Methods for Genetic Analysis*. Springer-Verlag, New York.
- LEE, W. C., 2003 Searching for disease-susceptibility loci by testing for Hardy-Weinberg disequilibrium in a gene bank of affected individuals. *Am. J. Epidemiol.* **158**: 379–400.
- MCCULLAGH, P., and J. A. NELDER, 1989 *Generalized Linear Models*, Ed. 2. Chapman & Hall, London.
- MCPECK, M. S., X. WU and C. OBER, 2004 Best linear unbiased allele-frequency estimation in complex pedigrees. *Biometrics* **60**: 359–367.
- MIREL, D. B., A. M. VALDES, L. C. LAZZERONI, R. L. REYNOLDS, H. A. ERLICH *et al.*, 2002 Association of IL4R haplotypes with type 1 diabetes. *Diabetes* **51**: 3336–3341.
- NIELSEN, D., M. G. EHM and B. WEIR, 1999 Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am. J. Hum. Genet.* **63**: 1531–1540.
- OBER, C., L. R. WEITKAMP, N. COX, H. DYTCH, D. KOSTYU *et al.*, 1997 HLA and mate choice in humans. *Am. J. Hum. Genet.* **61**: 497–504.
- OBER, C., S. A. LEAVITT, A. TSALENKO, T. D. HOWARD, D. M. HOKI *et al.*, 2000 Variation in the interleukin 4 receptor gene confers susceptibility to asthma and atopy in ethnically diverse populations. *Am. J. Hum. Genet.* **66**: 517–526.
- O'CONNELL, J. R., and D. E. WEEKS, 1998 PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am. J. Hum. Genet.* **63**: 259–266.
- OEFNER, P. J., and P. A. UNDERHILL, 1995 Comparative DNA sequencing by denaturing high-performance liquid chromatography (DHPLC). *Am. J. Hum. Genet.* **57**: A266.
- ROBERTSON, A., D. CHARLESWORTH and C. OBER, 1999 Effect of inbreeding avoidance on Hardy-Weinberg expectations: examples of neutral and selected loci. *Genet. Epidemiol.* **17**: 165–173.
- ROUSSET, F., and M. RAYMOND, 1995 Testing heterozygous excess and deficiency. *Genetics* **140**: 1413–1419.
- SING, C. F., and E. D. ROTHMAN, 1975 A consideration of the chi-square test of Hardy-Weinberg equilibrium in a non-multinomial situation. *Ann. Hum. Genet.* **39**: 141–145.
- SOBEL, E., J. C. PAPP and K. LANGE, 2002 Detection and integration of genotyping errors in statistical genetics. *Am. J. Hum. Genet.* **70**: 496–508.
- WEINBERG, C., and R. W. MORRIS, 2003 Invited commentary: testing for Hardy-Weinberg disequilibrium using a genome single-nucleotide polymorphism scan based on cases only. *Am. J. Epidemiol.* **158**: 401–403.
- WEIR, B. S., 1996 *Genetic Data Analysis 2: Methods for Discrete Population Genetic Data*, Ed. 2. Sinauer Associates, Sunderland, MA.
- XU, J., A. TURNER, J. LITTLE, E. R. BLEECKER and D. A. MEYERS, 2002 Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: Hint for genotyping error? *Hum. Genet.* **111**: 573–574.

Communicating editor: K. W. BROMAN

APPENDIX: EXPRESSION OF W_{GCC-HW}

To derive W_{GCC-HW} , we need to evaluate $(\hat{r} - r)$ and

$$\left[\text{var} \begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \right]^{-1}$$

at $(r, p) = (r_0, \hat{p}_0)$ using approximations (8) and (9), respectively, which require the computation of $\mathbf{A} \equiv \hat{E}[C'(r, p)]_{(r_0, \hat{p}_0)}^{-1}$ and $\mathbf{B} \equiv \widehat{\text{var}}[C(r, p)]_{(r_0, \hat{p}_0)}$. We use the notation

$$\mathbf{A} = - \begin{pmatrix} \alpha_{rr} & \alpha_{rp} \\ \alpha_{rp}^T & \alpha_{pp} \end{pmatrix}, \quad \mathbf{A}^{-1} = - \begin{pmatrix} \alpha^{rr} & \alpha^{rp} \\ \alpha^{rpT} & \alpha^{pp} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \beta_{rr} & \beta_{rp} \\ \beta_{rp}^T & \beta_{pp} \end{pmatrix}.$$

It follows from (7) with $\mathbf{L} = \mathbf{K}$ that $\alpha_{rr} = \hat{D}_r^T \hat{\mathbf{K}}^{-1} \hat{D}_r$, $\alpha_{rp} = \hat{D}_r^T \hat{\mathbf{K}}^{-1} \hat{D}_p$, $\alpha_{pp} = \hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{D}_p$, $\beta_{rr} = \hat{D}_r^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_r$, $\beta_{rp} = \hat{D}_r^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_p$, and $\beta_{pp} = \hat{D}_p^T \hat{\mathbf{K}}^{-1} \hat{\Sigma} \hat{\mathbf{K}}^{-1} \hat{D}_p$. Because $C_p(r_0, \hat{p}_0) = 0$ and letting $\hat{C}_r = C_r(r_0, \hat{p}_0)$, (8) implies that $(\hat{r} - r)|_{(r_0, \hat{p}_0)} \approx \alpha^{rr} \hat{C}_r$. Similarly, (9) implies

$$\widehat{\text{var}} \left[\begin{pmatrix} \hat{r} - r \\ \hat{p} - p \end{pmatrix} \right]_{(r_0, \hat{p}_0)} = \alpha^{rr} \beta_{rr} \alpha^{rr} + 2 \alpha^{rp} \beta_{rp}^T \alpha^{rr} + \alpha^{rp} \beta_{pp} \alpha^{rpT}.$$

Note that $\alpha^{rr} = \alpha_{rr}^{-1} [1 + \alpha_{rp} \mathbf{J} \alpha_{rp}^T \alpha_{rr}^{-1}]$ and $\alpha^{rp} = -\alpha_{rr}^{-1} \alpha_{rp} \mathbf{J}$, where $\mathbf{J} = [\alpha_{pp} - \alpha_{rp}^T \alpha_{rr}^{-1} \alpha_{rp}]^{-1}$. Plugging these expressions in (10) gives, after some algebra, the final expression of den_{GCC} given in (11).

