

# An Unconditional Exact Test for the Hardy-Weinberg Equilibrium Law: Sample-Space Ordering Using the Bayes Factor

Luis E. Montoya-Delgado,\* Telba Z. Irony,<sup>†</sup> Carlos A. de B. Pereira<sup>‡</sup> and Martin R. Whittle<sup>§</sup>

\*Universidad del Cauca, Popayan, Cauca, Colombia, <sup>†</sup>Division of Biostatistics, Center for Devices and Radiological Health, Food and Drug Administration, Rockville, Maryland 20850, <sup>‡</sup>Instituto de Matematica e Estatística, Universidade de São Paulo, 05008-090 São Paulo, SP, Brazil and <sup>§</sup>Genomic Engenharia Molecular Ltda, 01332-903 São Paulo, SP, Brazil

Manuscript received September 17, 1999

Accepted for publication March 1, 2001

## ABSTRACT

Much forensic inference based upon DNA evidence is made assuming that the Hardy-Weinberg equilibrium (HWE) is valid for the genetic loci being used. Several statistical tests to detect and measure deviation from HWE have been devised, each having advantages and limitations. The limitations become more obvious when testing for deviation within multiallelic DNA loci is attempted. Here we present an exact test for HWE in the biallelic case, based on the ratio of weighted likelihoods under the null and alternative hypotheses, the Bayes factor. This test does not depend on asymptotic results and minimizes a linear combination of type I and type II errors. By ordering the sample space using the Bayes factor, we also define a *significance* (evidence) index, *P* value, using the weighted likelihood under the null hypothesis. We compare it to the conditional exact test for the case of sample size  $n = 10$ . Using the idea under the method of  $\chi^2$  partition, the test is used sequentially to test equilibrium in the multiple allele case and then applied to two short tandem repeat loci, using a real Caucasian data bank, showing its usefulness.

ONE of the major uses of data from human multiallelic DNA loci is forensic inference. Because of the increasing use of variable number of tandem repeats (VNTR) and short tandem repeats (STR) loci, the importance of the Hardy-Weinberg equilibrium (HWE) has been reinforced (DEVLIN *et al.* 1991; GEISSER and JOHNSON 1992; DEKA *et al.* 1995; AYRES and BALDING 1998; SHOEMAKER *et al.* 1998) by being a useful assumption in the analysis of DNA evidence, as used in human identification and paternity studies. The conclusions reached by analyzing such evidence depend on the probabilistic evaluation of them and this evaluation is simplified if HWE is shown to be valid.

To test the HWE, usually the  $\chi^2$  test, the conditional exact test, and the likelihood-ratio test are used. For a complete discussion on these procedures, including some comparisons, see, for instance, HERNANDEZ and WEIR (1989), GUO and THOMPSON (1992), MAISTE (1993), WEIR (1996), and LAZZERONI and LANGE (1997). The conditional exact test is analogous to Fisher's exact test for contingency tables. There is a sufficient statistic, under the null hypothesis, that is considered to be known. Hence the *p* value is based on the conditional probabilities of the sample points given the value of the statistic.

MAISTE (1993) and MAISTE and WEIR (1995) contrasted these tests and claimed to show that the exact

conditional test has a better performance. A problem with this test is to define an order in the sample space to calculate the *p* value. In fact, this is the main difficulty when dealing with high-dimension sample spaces (KEMPTHORNE and FOLKS 1971). In addition, EMIGH (1980) makes a useful comparison of the various equilibrium tests.

In cases where the number of alleles per locus or the sample size is large, a technique to generate a Markov chain is used. The objective is to estimate the *p* values for the exact conditional test. For more details see GUO and THOMPSON (1992) and LAZZERONI and LANGE (1997). Today it is possible to perform the exact conditional test (or at least an approximation of it) under any sample size.

The  $\chi^2$  test is highly dependent on asymptotic results. In addition to being inefficient when the sample is insufficiently large, it may fail whenever there are categories (genotypes) of low expected frequencies. In the problem we consider, these genotypes are present in large numbers because of the inherent genetic structure and so it is not unusual to find an allele that appears only once or twice in a database, even when there are a considerable total number ( $\approx 5000$ ) of points.

Being a nonasymptotic test, the conditional exact test is, on the other hand, useful in cases where small samples are dealt with. It depends on a multinomial distribution and requires the ranking of all the possible samples with the same frequencies of alleles and same sample size. However, this ranking of the possible samples is problematic as we enlarge the size of the sample. Another disadvantage of this test is that it imposes unrealis-

Corresponding author: Martin R. Whittle, Genomic Engenharia Molecular Ltda, Rua Itapeva 500, cj 5AB, 01332-903 São Paulo, SP, Brazil. E-mail: mwhittle@genomic.com.br

tic probabilities on the data points whose probabilities had already been determined from conditioning. Recall that, by conditioning, the sample space is severely reduced and the probabilities may be drastically increased.

AYRES and BALDING (1998) propose a computational methodology to estimate the inbreeding coefficient, which allows one to measure the deviation from HWE under a model of inbreeding. SHOEMAKER *et al.* (1998) describe a Bayesian methodology to study the independence between pairs of alleles in a given locus; in this case, they consider the inbreeding coefficient and the disequilibrium coefficient (HERNANDEZ and WEIR 1989).

The problem of estimating the allelic frequencies, under the Bayesian perspective, was considered by GUNEL and WEARDEN (1995). CHOW and FONG (1992) studied this same problem but as a particular case of the simultaneous estimation of the related proportions.

The aim of this study is to develop an exact test on the basis of the comparison between weighted likelihoods (DICKEY and LIENTZ 1970) under the null and alternative hypotheses. The ratio of these two functions is the Bayes factor (BF). A distribution of the BF under the null hypothesis defines a natural order in the sample space. Therefore the test is exact and unconditional and does not depend on asymptotic results. In addition, the test is desirable, in terms of decision theory, in that it minimizes a linear combination of type I and type II errors. The approach suggested by BARNARD (1945) may be used to construct a non-Bayesian test that is unconditional and exact. In fact it considers suprema in the place of weighted averages.

The weight used to calculate the weighted likelihoods under each of the hypotheses is based on the *a priori preferences*, which are derived from the choice of an *a priori* distribution over the parametric space. The dimension of the subspace defined by the hypothesis of HWE is smaller than that of the original parametric space. Therefore, to calculate the weighted likelihood under the null hypothesis, we use line integrals. In the same manner, type I and type II errors are weighted average errors (IRONY and PEREIRA 1995). In other words, suppose  $\alpha_1$  and  $\beta_1$  are the errors associated with the Neyman-Pearson test of simple hypotheses  $H_0, \theta = \theta_0$  vs.  $H_1, \theta = \theta_1$ , where  $\theta_0$  is a parametric element of the equilibrium curve and  $\theta_1$  is outside the curve. Consider now the set of all such pairs. The weighted errors for the unconditional exact test are the weighted averages of these Neyman-Pearson errors.

Usually, a test of hypothesis consists of comparing the supremum of the likelihood in the subset of the parameter space corresponding to the null hypothesis with the supremum in the whole parameter space. The test presented in this article consists of comparing the averages of the likelihoods of these sets. That is, instead of comparing suprema the test compares averages. The most important property of the test is that it minimizes linear combinations of the two kinds of errors. For in-

stance, if  $\alpha$  and  $\beta$  are errors of the first and second types, the test presented here minimizes  $\alpha + \beta$  (DEGROOT 1989). Thus for a fixed  $\alpha$ , it maximizes  $(1 - \beta) - \alpha$  and consequently also the power of the test.

We define the significance level of the test by ordering the sample space using the Bayes factor. The BF, following GOOD (1983), is the ratio of the weighted likelihood under the null hypothesis and the weighted likelihood under the alternative hypothesis. To compare two sample points,  $s$  and  $t$ , we calculate the BF of these points. The order of the sample points follows the order of the BFs. After ordering the sample space, consider  $s$  as the sample observation. We define the  $P$  value as the sum of the weighted likelihood under the null hypothesis over the set of points smaller or equal to  $s$ . The idea of this Bayesian significance level is not new: it was suggested by KEMPTHORNE and FOLKS (1971). Note that the significance level as such uses the whole sample space in its calculus and then may not follow the likelihood principle (ROYALL 1997). Hence our  $P$  value may not be considered a full Bayesian procedure. A full Bayesian significance test for equilibrium and for contingency tables can be found in PEREIRA and STERN (1999).

The significance level, the  $P$  value, takes into consideration the alternative hypothesis in its calculus, which controls the type II error. Recall that the Bayes factor is the ratio between the probabilities under the two hypotheses. Usually this is not considered for the standard  $P$  value, which can cause problems such as rejecting the null hypothesis even when, under the alternative hypothesis, the observed sample has a lower probability (PEREIRA and WECHSLER 1993).

A program to compute our  $P$  value is available for the MatLab environment at the website <http://www.ime.usp.br/~cpereira/signifpr.html>.

A more complex test environment where two loci are considered simultaneously is presented by DEVLIN *et al.* (1996). A future useful project will be to construct a Bayesian alternative test for this situation and extend it for more than two loci.

## METHODS

To exemplify the use of the test (see *Examples*), data from two STR loci were analyzed: D17S250 (WEBER *et al.* 1990) and MYC (POLYMERPOULOS *et al.* 1992). Genomic DNA was obtained from unrelated, predominantly Caucasian individuals undergoing paternity testing nationwide by Genomic Engenharia Molecular Ltda. Alleles were amplified by PCR in the presence of [ $\alpha$ - $^{32}$ P]dCTP, separated on DNA sequencing gels by electrophoresis, and visualized by autoradiography. Allele sizing was done by running adjacent M13 sequence ladders.

**Definition and presentation of the exact unconditional test**

Consider a single autosomal biallelic locus, comprising alleles  $a_1$  and  $a_2$ , which does not undergo mutation. Let

$$(a_1, a_1), (a_1, a_2), (a_2, a_2)$$

be the genotypes at this locus, and we denote  $p_1, p_2,$  and  $p_3$  as the respective proportions of these genotypic classes in the population ( $p_1 + p_2 + p_3 = 1$ ). Let us suppose that the system is codominant; that is, distinct genotypic classes define distinct phenotypic classes. In this way, in a sample of size  $n$ , the frequencies of members in each class  $n_1, n_2,$  and  $n_3$ , satisfying the condition  $n_1 + n_2 + n_3 = n$ , can be observed.

In a panmictic population obeying Mendelian rules, equilibrium is attained in one reproductive generation and this assures the existence of a real number  $p \in (0, 1)$ , such that the genotypic proportions satisfy the relations

$$p_1 = p^2, \quad p_2 = 2p(1 - p), \quad p_3 = (1 - p)^2. \quad (1)$$

Hence, to decide as to the existence or not of equilibrium, it is necessary to test the null hypothesis  $H_0, \omega = (p_1, p_2, p_3) \in \Omega_0$ , where

$$\Omega_0 = \{(x_1, x_2, x_3) | 0 \leq x_1 \leq 1; x_2 = 2\sqrt{x_1}(1 - \sqrt{x_1}); x_3 = (1 - \sqrt{x_1})^2\},$$

against the alternative hypothesis  $H_1, \omega \in \Omega_1 = \Omega - \Omega_0$ , where

$$\Omega = \{(x_1, x_2, x_3) | x_i \geq 0, i = 1, 2, 3; \sum_{j=1}^3 x_j = 1\}.$$

Consequently, the statistical problem of interest is the construction of a procedure to test the following two alternative hypotheses,

$$H_0: \omega \in \Omega_0$$

and

$$H_1: \omega \in \Omega_1 = \Omega - \Omega_0,$$

where  $\Omega, \Omega_0,$  and  $\Omega_1$  are defined as above (see Figure 1).

Assuming that the sample elements are obtained independently, by using a Bernoulli multivariate process, prefixing the sample size  $n$ , and representing the data by  $d = (n_1, n_2, n_3)$ , we have that the likelihood function is given by

$$v[\omega|d] = n! \prod_{i=1}^3 \frac{p_i^{n_i}}{n_i!}, \quad (2)$$

where  $\omega \in \Omega$ .

Let us represent the researcher prior preferences by a Dirichlet density function (see WILKS 1968) with parameter vector  $(\alpha_1, \alpha_2, \alpha_3), \alpha_i > 0$ . That is, if  $\Gamma[\cdot]$  represents the Gamma function,

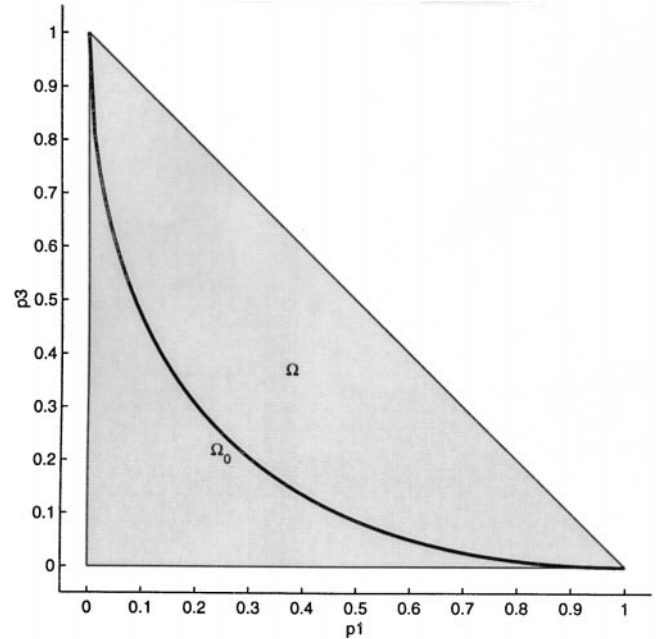


FIGURE 1.—The HWE curve. The complete parametric space is shown by the shaded area and the Hardy-Weinberg equilibrium is represented by the curve.

$$g[\omega] = \Gamma \left[ \sum_{i=1}^3 \alpha_i \right] \prod_{i=1}^3 \frac{p_i^{\alpha_i - 1}}{\Gamma[\alpha_i]} \quad (3)$$

is the function that defines the prior preferences of the possible parameter points  $\omega = (p_1, p_2, p_3)$ . From the Bayesian perspective, to choose this density is to choose a conjugate prior since the posterior density will also be a Dirichlet density with parameters  $(A_1, A_2, A_3)$ , where  $A_i = \alpha_i + n_i, i = 1, 2,$  or  $3$ .

Considering now (3) as the weighing system, the weighted likelihood average over  $\Omega$  is given by

$$f_1[d] = \frac{\int_{\Omega} v[\omega|d] g[\omega] d\omega}{\int_{\Omega} g[\omega] d\omega} = n! \frac{\Gamma[\sum_{i=1}^3 \alpha_i]}{\Gamma[\sum_{i=1}^3 A_i]} \prod_{i=1}^3 \frac{\Gamma[A_i]}{\Gamma[\alpha_i]}. \quad (4)$$

Also note that  $\Omega_0$  is a line inside the simplex  $\Omega$  and hence the weighted likelihood average over  $\Omega_0$  is the ratio of two line integrals as follows:

$$f_0[d] = \frac{\int_{\Omega_0} v[\omega|d] g[\omega] d\omega}{\int_{\Omega_0} g[\omega] d\omega} \quad (5)$$

$$= \frac{n! 2^{n_2}}{\prod_{i=1}^3 n_i!} \frac{\int_0^1 (\sqrt{1 - 3x(1 - x)}) x^{2A_1 + A_2 - 3} (1 - x)^{2A_3 + A_2 - 3} dx}{\int_0^1 (\sqrt{1 - 3x(1 - x)}) x^{2\alpha_1 + \alpha_2 - 3} (1 - x)^{2\alpha_3 + \alpha_2 - 3} dx}. \quad (6)$$

Let us suppose that the *a priori* density over  $\Omega$  is uniform. Thus, making  $\alpha_i = 1$  for  $i = 1, 2, 3$  in Equations 4 and 6, this means that the exact values of the weighted likelihoods over  $\Omega_1$  and  $\Omega_0$  are given, respectively, by

$$f_i[d] = \frac{2}{(n + 1)(n + 2)} \tag{7}$$

and

$$f_0[d] = \frac{n!2^{n_2}}{\prod_{i=1}^3 n_i!} \frac{\int_0^1 \sqrt{1 - 3x(1 - x)} x^{2n_1+n_2} (1 - x)^{2n_3+n_2} dx}{\int_0^1 \sqrt{1 - 3x(1 - x)} dx}.$$

Consequently if we assume that *a priori* the two hypotheses have equal probabilities, 0.5, we obtain

$$BF[d] = 2^{n_2-1} \frac{(n + 2)!}{\prod_{i=1}^3 n_i!} \frac{\int_0^1 \sqrt{1 - 3x(1 - x)} x^{2n_1+n_2} (1 - x)^{2n_3+n_2} dx}{\int_0^1 \sqrt{1 - 3x(1 - x)} dx}. \tag{8}$$

A test for the hypothesis of HWE consists of comparing BF [d] with unity. In this case we have a test that minimizes the sum of the average of the two types of errors.

Sometimes the exact calculation of Equation 8 is not feasible and so it is useful to show approximations to its determination. An approximation to Equation 8 (using Taylor’s expansion) is given by

$$BF[d] = \frac{(n + 2)! (2n_1 + n_2)! (2n_3 + n_2)! 2^{n_2-1}}{n_1! n_2! n_3! (2n + 1)! 3} \times \left[ 5 - 12 \frac{(2n_1 + n_2 + 1)(2n_3 + n_2 + 1)}{(2n + 2)(2n + 3)} \right].$$

Note that although we have no closed form for general Dirichlet priors, using Equations 5 and 6, and numerical integration, we can easily compute the Bayes factor for any choice of the prior parameters. In computing the *P* value using the program mentioned above for general Dirichlet priors, one needs only to adjust the data input. Instead of inputting the vector  $(x_1, x_2, x_3)$ , one must use  $(A_1 - 1, A_2 - 1, A_3 - 1)$ .

In this discussion we emphasize the use of uniform priors only for the purpose of a fair comparison with the alternative classical methods. Recall that with the uniform prior, the posterior is the normalized likelihood function. In the next section we provide a comparison of this proposed test for HWE to the conditional exact test.

**Comparison between the unconditional and the conditional exact tests**

In this section we compare the exact unconditional test proposed in the previous section to the traditional conditional exact test. Considering all possible samples of size  $n = 10$ , we calculate the *P* value for each of these samples. To pinpoint the two different ways of computing the probabilities, we refer to *p* value in the conditional test and to *P* value in the unconditional one (see PEREIRA and WECHSLER 1993). Table 1 lists the BF and the *P* values. Tables 2 and 3 show, respectively, the *P* values and *p* values, multiplied by 100, for the unconditional and for the conditional exact tests.

From Tables 2 and 3 it can be seen that the conditional exact test is more conservative than the unconditional exact test. The unconditional test is observed to minimize the sum of the average of the two types of errors. For a sample where  $(n_1, n_2, n_3) = (1, 8, 1)$  the unconditional exact test rejects the hypothesis of equilibrium ( $P = 0.036744 \approx 0.04$ ). Meanwhile the conditional exact test will not reject this hypothesis ( $p = 0.20$ ). The *p* value of the conditional exact test can be obtained from the weighted likelihood ratios. For example, let us fix the sample size at  $n = 10$  and suppose that the total number of observed elements that show the allele  $a_1$  is 9. That is,  $T = 2n_1 + n_2 = 9$ . To compute the *p* value of the conditional exact test, it is enough to consider all possible sample points for which  $T = 9$ . To determine the conditional probability of a sample point  $(n_1, n_2, n_3)$  given that  $2n_1 + n_2 = 9$  we have only to divide the BF obtained in this point by the sum of the BF of all points having  $T = 9$ . With these probabilities calculated we compute the *p* values in the usual manner, adding to the probability of each point all the smaller ones. Table 4 illustrates this calculation. Note that column 4, which is equal to column 3, is the second column divided by its sum.

In the general case, we maintain the sample size at  $n$  and the total number of observed elements that have the allele  $a_1$  at  $T = t$ . Considering the statistic  $T$  defined by  $T = 2n_1 + n_2$ , where  $n_1$  and  $n_2$  are random variables that denote the number of individuals observed in the sample who have the genotype  $(a_1, a_1)$  and  $(a_1, a_2)$ , respectively, this means that for each  $d = (n_1, n_2, n_3)$  with  $2n_1 + n_2 = t$  and  $\sum_{i=1}^3 n_i = n$ , whatever the value of *p* at  $(0, 1)$ , the conditional probability  $\Pr[d|T = t]$  is given by the equation

$$\Pr[d|T = t] = \frac{BF[d]}{\sum_{u \in \mathcal{D}} BF[u]},$$

where

$$\mathcal{D} = \left\{ u = (n_1, n_2, n_3) \mid \sum_{i=1}^3 n_i = n \text{ and } 2n_1 + n_2 = t \right\}.$$

Note that the sample space for  $n = 10$  has a total of 66 sample points, in contrast to the conditional test, where  $T = 10$ , which considers only a sample set of 6 points:  $(0, 10, 0)$ ,  $(1, 8, 1)$ ,  $(2, 6, 2)$ ,  $(3, 4, 3)$ ,  $(4, 2, 4)$ , and  $(5, 0, 5)$ .

**Hierarchical sequential testing for multiple alleles**

The ideal situation to build the significance test for the multiallelic case would be to consider an ordering in the whole space. However, the dimensionality of the parameter space is incredibly large. For instance, consider a locus with 20 alleles. For this example, the parameter space will have dimension 210. Hence the number of possible sample points increases drastically. Theoretically, substituting line integrals with surface integrals,



**TABLE 1**  
**Bayes factor and unconditional exact test *P* value**

BF	$n_1$	$n_3$	<i>P</i> value	BF	$n_1$	$n_3$	<i>P</i> value
0.003294	5	5	0.000050	0.668971	1	2	0.114822
0.004180	4	6	0.000177	0.681555	2	5	0.135475
0.008585	3	7	0.000437	0.878364	3	3	0.148784
0.013385	0	0	0.000640	0.944451	1	7	0.177404
0.029514	2	8	0.001534	0.961614	0	4	0.206543
0.036584	4	5	0.002642	1.003370	2	4	0.236949
0.056796	3	6	0.004364	1.054040	2	2	0.252919
0.074330	0	1	0.006616	1.070260	1	3	0.285351
0.137487	2	7	0.010782	1.170700	2	3	0.320827
0.164693	4	4	0.013278	1.283220	1	6	0.359712
0.193783	1	9	0.019149	1.363110	1	4	0.401019
0.200674	3	5	0.025231	1.442420	1	5	0.444728
0.229342	0	2	0.032181	1.539860	0	5	0.491391
0.301153	1	1	0.036744	2.203720	0	6	0.558170
0.360605	2	6	0.047671	2.879690	0	7	0.645433
0.487797	3	4	0.062453	3.488090	0	8	0.751133
0.519280	0	3	0.078188	3.960460	0	9	0.871147
0.539942	1	8	0.094550	4.252150	0	10	1.000000

we can proceed exactly as in the biallelic case but at an extremely high computational cost. Next we present a sequential procedure that loses in precision in benefit of cost.

Let us consider a single autosomal locus with multiple codominant alleles. Let  $(a_i, a_j)$  be the genotype referring to the alleles  $a_i$  and  $a_j$ ,  $i = 1, \dots, m$ ;  $j = i, \dots, m$ , and

$$p_{((i-1)(2m-i) + 2j)/2}, \quad i = 1, \dots, m; \quad j = i, \dots, m,$$

is the proportion of the genotypic class  $(a_i, a_j)$  in the population.

Because the system is codominant, in a sample of size  $n$ , the frequencies of the elements in each genotypic class  $n_1, n_2, \dots, n_{m(m+1)/2}$ , with  $\sum_{i=1}^{m(m+1)/2} n_i = n$  can be identified.

Assuming that the sample elements are obtained inde-

pendently, by using a Bernoulli (multivariate) process, we see that the likelihood function is proportional to

$$\prod_{i=1}^{m(m+1)/2} p_i^{n_i},$$

where

$$p_i \geq 0, \quad i = 1, \dots, \frac{m(m+1)}{2} \quad \text{and} \quad \sum_{i=1}^{m(m+1)/2} p_i = 1.$$

In a manner analogous to the case of the biallelic locus, the condition of equilibrium given in Equation 1 is characterized in the general case by Equation 9, because of the following statement: in a panmictic population that obeys Mendelian laws, equilibrium is attained in one reproductive generation and there are  $u_1, u_2, \dots$ ,

**TABLE 2**  
**Unconditional exact test (two-tailed): *P* value multiplied by 100**

	10	100										
	9	87	3									
	8	75	9	0								
	7	65	18	1	0							
	6	56	36	5	0	0						
$n_3$	5	49	44	14	3	0	0					
	4	21	40	24	6	1	0	0				
	3	8	29	32	15	6	3	0	0			
	2	3	11	25	32	24	14	5	1	0		
	1	1	4	11	29	40	44	36	18	9	2	
	0	0	1	3	8	21	49	56	65	75	87	100
		0	1	2	3	4	5	6	7	8	9	10
							$n_1$					

**TABLE 3**  
**Unconditional exact test (two-tailed):  $p$  value multiplied by 100**

	10	100										
	9	100	5									
	8	100	16	1								
	7	100	31	5	0							
	6	100	48	13	2	0						
$n_3$	5	100	100	48	8	2	0					
	4	47	100	57	25	8	2	0				
	3	22	100	100	56	25	8	2	0			
	2	17	52	100	100	57	48	13	5	1		
	1	5	20	52	100	100	100	48	31	16	5	
	0	0	5	17	22	47	100	100	100	100	100	100
		0	1	2	3	4	5	6	7	8	9	10
							$n_1$					

$u_m \in (0, 1)$  with  $\sum_{j=1}^m u_j = 1$ , such that they obey the relation

$$p((i - 1)(2m - i) + 2j)/2 = \begin{cases} u_i^2 & \text{if } j = i, i = 1, \dots, m \\ 2u_i u_j & \text{if } i = 1, \dots, m - 1, \text{ and } j = i + 1, \dots, m. \end{cases} \tag{9}$$

Although we can define the exact unconditional test using the BF (as in the case with two alleles), great difficulties arise when calculating the surface integrals in this case. Upon examining the population HWE determined by Equation 9, we can prove that the following statement, a property of the multinomial distribution, is true. This comprises the basis of the procedure that we propose to test the hypothesis of equilibrium in a situation of multiple alleles.

**Statement:** Let  $a_1, a_2, \dots, a_m$  be the alleles under consideration. If for any  $i = 1, 2, \dots, m$ , the condition of HWE given in Equation 1, is satisfied by the biallelic system  $a_i, A_i$ , where

$$I_1 = \{1, 2, \dots, m\} - \{i\}, \quad A_{I_1} = \{a_1, a_2, \dots, a_m\} - \{a_i\},$$

and if for every  $j = 2, \dots, m - 1$ , the condition of HWE is satisfied by the biallelic system  $a_j, A_j$ , whatever  $i_j$  in  $I_{j-1}$ , where

$$I_j = I_{j-1} - \{i_j\}, \quad A_j = A_{I_{j-1}} - \{a_j\},$$

then the system with alleles  $a_1, a_2, \dots, a_m$  satisfies the condition of HWE given by Equation 9.

Thus if a system with  $n$  alleles,  $A_0 = \{a_1, a_2, \dots, a_m\}$ , obeys the law of HWE, this law is obeyed by “any system” obtained by partitioning  $A_0$  into at least two nonempty subsets and upon considering each element of this partition as “an allele.”

The idea under the HWE test for the multiallelic case is based on the chi-square partition in contingency tables (see EVERITT 1977, for instance). Consider a data bank, sample  $S$ , for a specific locus and consider also that there are  $m$  (a positive integer) different alleles,  $a_1, a_2, \dots, a_m$ . The order of testing, starting from the smallest allele frequency to the highest one, has the objective of working, in each step, with the biggest possible sample. The reason for this is to try to work, in all steps, with the smallest possible errors.

The sequential procedure, to test the hypothesis of HWE, is as follows.

**Procedure:**

*Step 1:* Without loss of generality, call  $a_1$  the least frequent allele in sample  $S$ .

*Step 2:* Divide the sample  $S$  into three mutually exclusive sets:

- $S_{11}$ , all individuals with genotype  $(a_1, a_1)$ ;
- $S_{i.}$ , all individuals with genotype  $(a_i, a_1)$ ,  $i \neq 1$ ; and
- $S_{..}$ , all individuals not having the allele  $a_1$ .

**TABLE 4**  
**Conditional  $p$  value based on BF for the case where  $n = 10$  and  $T = 9$**

$d = (n_1, n_2, n_3)$	BF[ $d$ ]	Pr[ $d T = 9$ ]	BF[ $d$ ]/ $\Sigma$	$p$ value
(0, 9, 1)	0.074330	0.030483	0.030483	0.045487
(1, 7, 2)	0.668971	0.274351	0.274351	0.519886
(2, 5, 3)	1.170700	0.480114	0.480114	1.000000
(3, 3, 4)	0.487797	0.200048	0.200048	0.245535
(4, 1, 5)	0.036584	0.015004	0.015004	0.015004
$\Sigma$	2.438382			

**TABLE 5**  
**Testing results for Example 1 ( $p$  value for  $\chi^2$  test)**

$i$	Allele (bp)	$n_{i1}$	$n_{i2}$	$n_{i3}$	$BF_i$	$P$ value	$p$ value
1	91	0	2	5712	2587	99.93	99.99
2	85	0	3	5709	2585	99.91	99.98
3	93	0	3	5706	2583	99.91	99.98
4	95	0	5	5701	2579	99.85	99.95
5	123	0	5	5696	2577	99.85	99.95
6	97	0	6	5690	2573	99.82	99.92
7	99	0	7	5683	2569	99.80	99.89
8	89	0	12	5671	2553	99.65	99.68
9	87	0	22	5649	2505	99.36	98.93
10	101	2	21	5626	0.74	0.390	0.000
11	121	7	35	5584	$9.6 \times 10^{-9}$	0.000	0.000

*Step 3:* Apply the unconditional exact test for the biallelic case in the partition  $(S_{i1}, S_{i2}, \dots)$ . If HWE is rejected stop and declare the population to be in disequilibrium. If HWE is accepted go to step 4.

*Step 4:* If  $S_{i2}$  is composed of elements with only one allele involved, stop and declare the population to be in equilibrium. If more than one allele is involved in the elements of  $S_{i2}$ , rename  $S_{i2}$  as  $S$  and go to step 1.

**Examples**

In the examples we illustrate the sequence in which the tests for equilibrium were performed and present the values of the Bayes factors with respective  $p$  values and the values of the  $p$  values for the corresponding  $\chi^2$  tests.

Let  $m_a$  be the number of alleles present in the locus being studied. For each  $i = 1, \dots, m_a - 1$ , we define

- $a_i$ , the allele chosen to carry out the  $i$ th test;
- $(a_i)^c$ , the set made up of the remaining alleles to carry out the  $i$ th test;
- $n_{i1}$ , the number of genotypes  $(a_i, a_i)$ ;
- $n_{i2}$ , the number of genotypes  $(a_i, (a_i)^c)$ ;
- $n_{i3}$ , the number of genotypes  $((a_i)^c, (a_i)^c)$ ; and
- $BF_i$ , the Bayes factor corresponding to the  $i$ th test.

**EXAMPLE 1:** The following data were obtained from the STR locus MYC in which 19 alleles are observed and  $n_a = 5714$  (Table 5).

In this case, therefore, the hypothesis of HWE should be rejected.

**EXAMPLE 2:** Here the data were obtained from the STR locus D17S250 in which 21 alleles are seen and  $n_a = 5592$  (Table 6).

The hypothesis is not rejected in this example.

**DISCUSSION**

A Bayesian test of hypothesis was presented in this article. However, its evaluation and comparison with the

alternative conditional test are done under a classical perspective. To start this discussion we recall that the generalized Neyman-Pearson (GNP) test is an optimal test under both perspectives, classical and Bayesian (DEGROOT 1989, Section 8.2). In the GNP situation, one compares two probability (density) functions,  $f_0$ , the probability function under the null hypothesis, against  $f_1$ , the one under the alternative hypothesis. Having chosen a constant  $k > 0$ , if  $f_0 > (\geq) k f_1$ , then the null hypothesis is accepted. On the other hand, if  $f_0 \leq (<) k f_1$ , then the null hypothesis is rejected. Note that, in fact, we compare the values of the two likelihoods at the sample point that effectively occurred, to make the decision. If  $\alpha$  and  $\beta$  are, respectively, the probabilities of the two kinds of errors, the GNP is the test that minimizes the linear combination  $\alpha - k\beta$ . Considering adequately the choice of  $k$  as a function of losses and prior probabilities, this linear combination is the minimum expected loss, which makes the test optimal under the Bayesian perspective. PRESS (1989) presents a complete description of the Bayesian method.

The HWE case is different in that both hypotheses are composite. That is, each hypothesis is represented as a set of probability functions. One of the difficulties is that the two sets have different dimensions. The alternative hypothesis is bidimensional although the null hypothesis is unidimensional. The idea of the test, under the classical point of view, is to define as the two probability functions,  $f_0$  and  $f_1$ , the averages of the likelihood over the parametric sets defined by the null and the alternative hypotheses, respectively. Having now two probability functions, we apply the GNP procedure to define the critical region. To compute the average  $p$  value we have to order the sample space. We say that a sample point  $x_i$  is higher than  $x_j$ , denoted  $x_j \leq x_i$ , if  $f_0(x_i)/f_1(x_i) > f_0(x_j)/f_1(x_j)$ . To define the  $P$  value (not  $p$  value) we consider the sum of  $f_0(x_j)$  over all sample points  $x_j \leq x_0$ , where  $x_0$  is the sample point effectively observed. Since we cover the whole sample space, we have called the test an *unconditional exact test*.

The unconditional test is opposite to the one that

TABLE 6  
Testing results for *Example 2* ( $p$  value for  $\chi^2$  test)

$i$	Allele (bp)	$n_{i1}$	$n_{i2}$	$n_{i3}$	$BF_i$	$p$ value	$p$ value
1	131	0	4	5588	2529.7	99.88	99.96
2	139	0	5	5583	2526.4	99.85	99.94
3	137	0	6	5577	2522.5	99.82	99.92
4	171	0	8	5569	2515.7	99.76	99.86
5	135	0	10	5559	2507.2	99.70	99.78
6	133	0	11	5548	2499.9	99.67	99.73
7	145	0	21	5527	2454.9	99.37	99.01
8	169	0	31	5496	2384.2	99.08	97.84
9	143	1	107	5388	773.2	95.76	81.93
10	141	1	111	5276	788.5	95.85	86.66
11	147	1	134	5141	857	96.43	99.11
12	167	2	183	4956	599.8	94.91	97.30
13	165	9	402	4545	278.7	90.05	99.94
14	163	18	506	4021	158.7	84.27	88.81
15	161	21	511	3489	118	82.10	88.72
16	159	19	502	2968	107	81.67	90.40
17	149	31	498	2439	53.5	68.27	61.56
18	157	51	549	1839	20.59	46.42	41.64
19	155	105	709	1025	10.19	21.51	47.46
20	151	236	497	292	11.12	27.94	68.17

considers as the likelihood under the null hypothesis the conditional probability function given the observed value of  $T = 2n_1 + n_2$ , which is a sufficient statistic under the null hypothesis (HALDANE 1954; CHAPCO 1976; ELSTON and FORTHOFFER 1977). To compute the  $p$  value in this case one must look only for the sample points with the same value of  $2n_1 + n_2$  obtained by the observed one. This is the reason to call this test a conditional test. Tentatively, GEISSER and JOHNSON (1992) presented an unconditional test that was based on quantiles. However, it seems to be not appropriate as discussed by DEVLIN *et al.* (1993a,b). CANNINGS and EDWARDS (1969) without conditioning presented a way to estimate a deviation from the HWE. However, they did not discuss hypothesis testing.

Turning here to our procedure, from the Bayesian point of view, where a posterior density is defined over the parametric space, one could say that the test is fully conditional. The reason is obvious because we compute a conditional density for the parameter given the observed sample point. Considering the uniform prior in the parametric space and 0.5 as the prior probability of the null hypothesis to be true, the ratio of the average likelihoods, as presented, is the posterior odds, which compared with a chosen  $k$  would define the testing procedure presented. This is a test that minimizes  $\hat{\alpha} + k\hat{\beta}$ , where  $\hat{\alpha}$  and  $\hat{\beta}$  are the average of the probabilities of errors of types I and II, respectively (PEREIRA 1985).

As far as we know, PEREIRA and ROGATKO (1984) presented the first Bayesian article for testing the HWE. However, ALTHAM (1971) presented a Bayesian estimation of a parameter that can be used to evaluate the

HWE. It does not mention the HWE because it is described in a different context. PEREIRA and ROGATKO (1984) defined an *ad hoc* way to define the likelihoods, which could not be properly supported. They also presented credible sets that could be used to test HWE. The value of the credibility was used to define the size of the first kind of error.

LINDLEY (1988) considers a Bayesian estimation for equilibrium parameters in the case of two alleles. The parameters studied are obtained from an alternative parameterization that on the one hand allows the use of Gaussian priors, but on the other complicates the interpretation at the moment of assessing the *a priori* distribution.

The hierarchical sequential procedure described in this article is based on intuition. Recall that multinomial likelihoods can be factorized using partitions on the set of categories. Also, the HWE is a special kind of association among alleles at a specific locus. Whenever we conclude that a specific allele is in HWE association with all the others, we believe we do not have to use it again when testing the remaining alleles. It could be argued that, by using this procedure, the probability of rejecting HWE may increase as alleles are being eliminated. However, since the sample size is decreasing, the power of the tests will decrease. Hence, it is reasonable to believe that there is a compensation and that the procedure will do the job fairly. Note also that the sequence order depends on the rarity of the alleles in such a way that the sample size reduction occurs as slowly as possible.

Today, the use of Bayesian ideas in genetics is a reality.



More recently, an interesting article by VIELAND (1998) suggested that the future of genetic data analysis is strongly related to the Bayesian paradigm.

The views expressed are those of the authors and not necessarily those of the FDA.

## LITERATURE CITED

- ALTHAM, P. M. E., 1971 Exact Bayesian analysis of an intraclass  $2 \times 2$  table. *Biometrika* **58**: 679–680.
- AYRES, K. L., and D. J. BALDING, 1998 Measuring departures from Hardy-Weinberg: a Markov chain Monte Carlo method for estimating the inbreeding coefficient. *Heredity* **80**: 769–777.
- BARNARD, G. A., 1945 A new test for  $2 \times 2$  tables. *Nature* **156**: 177.
- CANNINGS, C., and A. M. F. EDWARDS, 1969 Expected genotypic frequencies in a small sample: deviation from the Hardy-Weinberg equilibrium. *Am. J. Hum. Genet.* **21**: 245–247.
- CHAPCO, W., 1976 An exact test of the Hardy-Weinberg law. *Biometrics* **32**: 183–189.
- CHOW, M., and D. K. H. FONG, 1992 Simultaneous estimation of the Hardy-Weinberg proportions. *Can. J. Stat.* **20**: 291–296.
- DEGROOT, M. H., 1989 *Probability and Statistics*, Ed. 2. Addison Wesley, Reading, MA.
- DEKA, R., L. JIN, M. SHRIVER and L. YU, 1995 Population genetics of dinucleotide ( $dC - dA$ ); ( $dG - dT$ )<sub>n</sub> polymorphisms in world populations. *Am. J. Hum. Genet.* **56**: 461–474.
- DEVLIN, B., N. RISCH and S. ROEDER, 1991 Forensic DNA tests and Hardy-Weinberg equilibrium. *Science* **253**: 1039–1041.
- DEVLIN, B., N. RISCH and S. ROEDER, 1993a Statistical evaluation of DNA fingerprinting: a critique of the NRC report. *Science* **259**: 748–749.
- DEVLIN, B., N. RISCH and S. ROEDER, 1993b NRC report on DNA typing. *Science* **260**: 1057–1058.
- DEVLIN, B., N. RISCH and S. ROEDER, 1996 Disequilibrium mapping: composite likelihood for pairwise disequilibrium. *Genomics* **36**: 1–16.
- DICKEY, J. M., and B. P. LIENTZ, 1970 The weighted likelihood ratio, sharp hypotheses about chances, the order of a Markov chain. *Ann. Math. Stat.* **41**: 214–266.
- ELSTON, R. C., and R. FORTHOFFER, 1977 Testing the Hardy-Weinberg equilibrium in small samples. *Biometrics* **33**: 536–542.
- EMIGH, J. M., 1980 A comparison of tests for Hardy-Weinberg equilibrium. *Biometrics* **36**: 627–642.
- EVERITT, B. S., 1977 *The Analysis of Contingency Tables*. Chapman & Hall, London.
- GEISSER, S., and W. JOHNSON, 1992 Testing Hardy-Weinberg equilibrium on allelic data from VNTR loci. *Am. J. Hum. Genet.* **51**: 1084–1088.
- GOOD, I. J., 1983 *Good Thinking: The Foundations of Probability and Its Applications*. University of Minnesota Press, Minneapolis.
- GUNEL, E., and S. WEARDEN, 1995 Bayesian estimation and testing of gene frequencies. *Theor. Appl. Genet.* **91**: 534–543.
- GUO, S., and E. THOMPSON, 1992 Performing the exact test for Hardy-Weinberg proportion for multiple alleles. *Biometrics* **48**: 361–372.
- HALDANE, J. B. S., 1954 An exact test for randomness of mating. *J. Genet.* **52**: 631–635.
- HERNÁNDEZ, J. L., and B. S. WEIR, 1989 A disequilibrium coefficient approach to Hardy-Weinberg testing. *Biometrics* **45**: 53–70.
- IRONY, T. Z., and C. A. PEREIRA, 1995 Bayesian hypothesis test: using surface integrals to distribute prior information among hypotheses. *Resenhas* **2**: 27–46.
- KEMPTHORNE, O., and L. FOLKS, 1971 *Probability, Statistics, and Data Analysis*. The Iowa State University Press, Ames, IA.
- LAZZERONI, L. C., and K. LANGE, 1997 Markov chains for Monte Carlo tests of genetic equilibrium in multidimensional contingency tables. *Ann. Stat.* **27**: 138–168.
- LINDLEY, D. V., 1988 Statistical inference concerning Hardy-Weinberg equilibrium. *Bayesian Stat.* **3**: 307–326.
- MAISTE, P. J., 1993 Comparison of statistical tests for independence at genetic loci with many alleles. Ph.D Thesis, North Carolina State University, Raleigh, NC.
- MAISTE, P. J., and B. S. WEIR, 1995 A comparison of tests for independence in the FBI RFLP data bases. *Genetica* **96**: 125–138.
- PEREIRA, C. A. de B., 1985 Testing hypotheses defined in spaces of different dimensions: Bayesian vision and classical interpretation. Associate Professor Thesis, Instituto de Matemática e Estatística, University of São Paulo, Brazil (in Portuguese).
- PEREIRA, C. A. de B., and A. ROGATKO, 1984 The Hardy-Weinberg equilibrium under a Bayesian perspective. *Braz. J. Genet.* **4**: 689–707.
- PEREIRA, C. A., and J. M. STERN, 1999 Evidence and credibility: full Bayesian significance test for precise hypotheses. *Entropy* **1**: 69–80.
- PEREIRA, C. A. de B., and S. WECHSLER, 1993 On the concept of P-value. *Braz. J. Prob. Stat.* **7**: 159–177.
- POLYMEROPOULOS, M. H., H. XIAO and C. R. MERRIL, 1992 Dinucleotide repeat polymorphism at the human c-myc oncogene locus (MYC). *Hum. Mol. Genet.* **1**: 65.
- PRESS, S. J., 1989 *Bayesian Statistics: Principles, Models, and Applications*. John Wiley, New York.
- ROYALL, R., 1997 *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall, London.
- SHOEMAKER, J., I. PAINTER and B. S. WEIR, 1998 A Bayesian characterization of Hardy-Weinberg disequilibrium. *Genetics* **149**: 2079–2088.
- VIELAND, V. J., 1998 Bayesian linkage analysis, or: how I learned to stop worrying and love the posterior probability of linkage. *Am. J. Hum. Genet.* **63**: 947–954.
- WEBER, J. L., A. E. KWITEC, P. E. MAY, F. S. COLLINS and D. H. LEDBETTER, 1990 Dinucleotide repeat polymorphism at the D17S250 and D17S261 loci. *Nucleic Acids Res.* **18**: 4640.
- WEIR, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.
- WILKS, S. S., 1968 *Mathematical Statistics*. Wiley, New York.

Communicating editor: G. A. CHURCHILL