Theoretical and Applied Genetics 43, 35–38 (1973)

On the Optimum Structure of Progeny Tests for Recessive Alleles

J. F. KIDWELL and G. W. HAGY

Division of Biological and Medical Sciences, Brown University, Providence, Rhode Island (U.S.A.)

Summary. A method and an alternative for determining the optimum structure of a progeny test program designed to reduce the frequency of recessive alleles are described. Some of the limitations and possible applications are discussed.

It is well known that when all homozygous recessive individuals are eliminated from a large random mating population, the equilibrium frequency of recessive alleles maintained by a balance between mutation and selection is approximately the square root of the mutation rate (Falconer, 1960). Consequently, the frequency of most recessive deleterious, undesirable or lethal alleles is expetced to be very low in most populations of domestic animals. This is most often true, but sometimes, for reasons that are not always obvious, their frequency becomes high enough to require a planned counter selection program. These programs are often based on progeny testing prospective sires. The breeder must determine when to progeny test for recessive alleles and the structure of the test. Schaible (1968) considered some aspects of these problems; but his formulation is not valid, and consequently his recommendations cannot be followed (Kidwell, 1970). In this paper we describe an optimum test and an alternative, both based on a suggestion first made by Kempthorne (1957).

The Number of Males Tested

Let n_i be the number of test progeny, all wild type, produced by the i^{th} tested male, and α_i be the probability that the i^{th} tested male will be accepted when he is, in fact, heterozygous; then α_i is the probability of a type I error, and

$$\alpha_i = (1 - p/2)^{n_i},$$

where p is the probability that a gamete produced by a tester female is heterozygous (Kidwell, 1951, 1970). Let s be the number of new sires required per unit time, and let $\hat{s}$ be the number of males that must be tested to provide at least s acceptable sires (males whose test mates produce no recessive progeny) with a probability of at least β. The probability that the i^{th} tested male is accepted, a_i, is equal to the *a priori* probability that he is heterozygous, b_i, multiplied by the conditional probability that he produces no recessive test progeny, given that he is heterozygous, α_i, plus the probability that he is homozygous.

$$a_i = \alpha_i\, b_i + 1 - b_i\,.$$

When the a_i are equal for all males, the number that must be tested is approximately the value of $\hat{s}$ that satisfies the expression

$$\sum_{x=s}^{\hat{s}} C_x^{\hat{s}}\, a^x\, (1 - a)^{\hat{s}-x} \geq \beta\,.$$

If, for example, 5 new sires are required, the breeder desires the probability of locating at least 5 acceptable sires to be at least .95, and all $a_i = .9$, seven males must be tested. The probability that at least 5 are acceptable is .9743.

It will more often happen that there are one or a few small sized groups that are otherwise preferred, but the males within each group have equal and low values of a_i, and a single larger group of otherwise less desirable males with a higher value of a_i. The method of calculating $\hat{s}$ is straightforward but tedious and is perhaps best illustrated with a simple example.

Suppose there are three preferred males, each with $a = .5$, a much larger group with $a = .9$, and the breeder requires the probability of locating at least 5 acceptable males to be at least .95. He will test all three with $a = .5$ and as many as are needed from the larger group. The probability distribution for the group of 3 is

Number found acceptable:	0	1	2	3
Probability:	.125	.375	.375	.125

We "guess" that five will be required from the larger group. The probability distribution is

Number found acceptable:	0	1	2	3	4	5
Probability:	.000	.0004	.0081	.0729	.3280	.5905

The number of ways of obtaining a total of 5, 6, 7, or 8 acceptable sires and the associated probabilities are as follows.

Number of acceptable males												
5			6			7			8			
Number accepted		Prob.	Number accepted		Prob.	Number accepted		Prob.	Number accepted		Prob.	
a_i=.5	a_i=.9		a_i=.5	a_i=.9		a_i=.5	a_i=.9		a_i=.5	a_i=.9		
3	2	.001	3	3	.009	3	4	.041	3	5	.074	
2	3	.027	2	4	.123	2	5	.221				
1	4	.123	1	5	.221							
0	5	.074										

The probability of at least 5 acceptable sires is the sum of the probabilities or .915. (The calculation was made using 4 decimal places.) Since this is too small, the calculation is made taking 6 from the larger group, and the probability of obtaining at least five acceptable sires is .979. Hence, $\hat{s} = 3 + 6 = 9$.

Evaluation of $\check{s}$ requires values of α_i, which depend on the n_i to be determined. For practical purposes, however, an arbitrary value (e.g., $.01 \leq \alpha_i \leq .05$) can be used.

The Test

Designate the cost of using the i^{th} sire when he is heterozygous as C_{1i} and the net cost of each test progeny as C_2. A total expected cost, K_t, per unit time can be defined.

$$K_t = \sum_{i=1}^{\hat{s}} \left\{ \frac{C_{1i}\,\alpha_i\,b_i}{\alpha_i\,b_i + 1 - b_i} + C_2\,n_i \right\} + \sum_{i=s+1}^{\hat{s}} C_2\,n_i\,.$$

The first sum on the right side refers to tested sires that are actually used. It is the sum of the *a posteriori* probabilities of heterozygosity multiplied by the cost of using a heterozygous sire plus the cost of testing the sire. The second term is the cost of testing males that are discarded. The optimum test minimizes K_t with respect to $\hat{s}$ and n_i for a given population structure. Since K_t always decreases as $\hat{s}$ decreases, and $\hat{s}$ decreases with decreasing b, the optimum strategy is to test those potential sires with the lowest *a priori* probability of heterozygosity. In many situations an individual with one or both parents known to be heterozygous will have a substantially higher *a priori* probability of heterozygosity than one chosen at random. However, it often happens that some of these males are considered most valuable, provided they are not heterozygous, on the basis of other criteria. In this circumstance the breeder can compare the increased expected cost due to the higher *a priori* probability with the value of the expected difference in improvement of the other traits. It can happen that optimum K_t is not minimum K_t in such cases.

For the sires that are used, the values of n_i that minimize the cost for the i^{th} sire can be calculated by setting $\frac{dK_i}{dn_i} = 0$ and solving for n_i. It can be shown that, setting

$$h_i = \frac{\alpha_i\,b_i}{(\alpha_i\,b_i + 1 - b_i)}$$

for the probability that the i^{th} male will be accepted and heterozygous, we have

$$\frac{dK_i}{dn_i} = \frac{-\,p\,C_{1i}}{2}\,h_i\,(1 - h_i) + C_{2i}\,.$$

This will equal zero when

$$h_i\,(1 - h_i) = \frac{2\,C_{2i}}{p\,C_{1i}}$$

from which the appropriate value of n_i may be determined.

If testing is sequential, the greatest reduction in K_t will be achieved by first assigning tester females to those males to be tested with the highest *a priori* probability of heterozygosity. The rationale is presented in the section on problems of application.

When to Test

If the males are not tested, the *a posteriori* probability of heterozygosity cannot be evaluated and all n_i are zero. The expected cost can then be defined as

$$K_0 = \sum_{i=1}^{s} b_i\,C_{1i}\,.$$

Testing is indicated when K_t is less than K_0. Suppose, as a simple example, that all $C_{1i} = \$20{,}000$, all $C_2 = \$100$, all $b_i = .5$, all $p = 1.0$, $\beta = .95$ and that 5 sires are required. Seven males must be tested, each with eight females. The total expected cost is \$5,989.11. If five males are used without testing, the total expected cost is \$50,000.00. Testing is clearly indicated.

Problems of Application

Although the foregoing theory and method seem appropriate for the circumstances considered, several problems likely to be met in practice can be anticipated. The method requires exact knowledge of the *a priori* probability of heterozygosity of the males. Only in a few special circumstances (e.g., the parents' genotypes are known) will exact knowledge be available. Most empirical estimates will have large sampling errors. the adverse effects of using an erroneous value can be large and important. There is no obvious way to avoid this difficulty.

In many instances the problem of providing tester females will be formidable. It may happen that the total number of test progeny that can be produced per unit time is less than required for minimum cost. A general procedure for directly obtaining the lowest cost possible within this restriction is not obvious. However, the breeder can use a "trial and error" procedure by calculating K_t for the various alternatives and using the one with the lowest expected total cost.

In more extreme instances there may be too few tester females to conduct a comprehensive progeny test program. The optimum strategy is to select, insofar as possible, those sires with the lowest *a priori* probability of heterozygosity and then allocate the available tester females to those with the *highest* values of b. The rationale may be deduced as follows.

It follows from the above formula for $\frac{dK_i}{dn_i}$ that tester females should be initially allocated to males with the highest value of h_i (which equals b_i at the start of testing) if all h values are below 0.5 and C_1 and C_2 are the same for all males. The optimum final allocation of tester females would be achieved when $\frac{dK_i}{dn_i}$ is equal for all males, when h_i, the conditioned probability of heterozygosity given acceptance would also be equal. Thus tester females should be allocated to males in order to equalize as far as possible this conditional probability.

Application of the method requires knowledge of the cost of using each male if he is heterozygous and of the cost of producing each test offspring. These will usually involve complex and interrelated biological and economic factors that may be exceedingly difficult to evaluate. Two obvious ones are: 1) the cost of introducing an allele previously absent from the population, i.e., the cost of changing the breeder's status from "clean" to "carrier"; and 2) the cost of maintaining the tester females, either by diverting part of the existing resources or establishing a new facility.

An Alternative Test

Many other models and methods of optimizing a progeny test are undoubtedly possible. Kempthorne (1957), for example, considered the test of a single male whose test mates are all homozygous recessive ($p = 1$) and wrote, "The cost of concluding that an Aa sire is AA could be denoted by C_1, the cost of each offspring by C_2. The total risk with n offspring is therefore $\left(\frac{1}{2}\right)^n C_1 + n\,C_2$. The experimenter might well choose n to make this risk a minimum . . .". To conclude that a heterozygous male is homozygous is to commit a type I error. The cost of making a type I error is not necessarily the same as the cost of using a heterozygous sire, although they surely have many common elements and will be similar. Summing over all males tested, a total expected cost can be defined,

$$K_t' = \sum_{i=1}^{\hat{s}} (C_{1i}'\alpha_i + C_2\, n_i) .$$

The "primes" are used to distinguish this from the previous test. Using the same methods as before, the values of n_i' that minimize K_t' are

$$n_i' = \frac{ln(C_{1i}'/C_2) + ln\,\{ln\,[2/(2-p)]\}}{ln\,[2/(2-p)]}$$

This test has the advantage that the b_i are required only to determine the number of sires that must be tested, hence errors in estimates of b_i are less critical. It may be much more difficult to evaluate the cost of making a type I error than the cost of using a heterozygous male.

If, as seems likely, the C_{1i} and C_{1i}' are similar, the total costs will also be similar. If, for example, we set $C_{1i} = C_{1i}'$ and consider the simple example used to illustrate the comparison of the cost of testing with not testing, seven males must be tested, each with 8 females, as before, and the total cost is \$6,146.88 rather than \$5,989.11.

Hypothetical Applications

1. Many dairy cattle artificial insemination centers regularly progeny test young potential sires for milk and butterfat production by mating them to approximately 300 females (usually first calf heifers) with the expectation of obtaining first lactation records on 40 to 50 daughters. These bulls are then "laid of" until their daughters' records are complete. It would be possible to progeny test these bulls for all recessive alleles by mating each to a sample of his daughters. (Obviously a few heterozygous sires will be identified by the initial progeny and discarded without further test.) This would have no effect on the daughters' first lactation record unless the calf was homozygous for a recessive allele resulting in early abortion, or something similar. These females would not be available for production testing younger males, and this might contribute to the cost of the test but is not likely to pose a serious management or breeding problem. The inbreeding of the calves might also add to the cost, but not a large amount.

A realistic estimate of the b_i's, the cost of using a heterozygous bull, making a type I error or producing test progeny is not available, so the optimum number of progeny cannot be calculated for either test. It is of interest, however, to determine what value the ratio C_{1i}'/C_2 must have for optimum n_i' to be such that the probability of a type I error is some arbitrary value, for example .05 or .01. In this case $p = 1/4$, and 23 and 35 test progeny, respectively, are required. These would be the optimum values of n'_i if the ratio of the two costs were 372 and 1,846, respectively. Thus if the cost of a type I error were 372 times the net cost of producing each test calf, the optimum test structure would result in the production of 23 test calves, and the probability of deciding that a heterozygous bull is homozygous would be .05. Although no data are available to us, it seems highly likely that in many instances the ratio of the costs will be large enough to justify progeny testing for all recessives, as Wriedt (1930) suggested in a different context long ago. It would also provide an accurate estimate of the number of loci at which an average bull is heterozygous for a recessive allele.

2. During the 1950's the University of Nevada's Holstein-Freisian herd produced several potential herd sires. The recessive red allele was segregating in the herd, and each bull was progeny tested prior to being proven for production. The test was conducted by mating to Hereford cows in the commercial herd, hence $p = 1$. A reasonable estimate is \$20,000 for the C'_{1i} and \$10.00 for C_2. The optimum value of n_i is 10.44 or 11. The expected cost is \$119.76 for bulls that were used and \$110.00 for those discarded. The value of K_0 was \$10,000. Actually only 5 to 7 test calves were produced by each bull, hence the test was not optimum.

These hypothetical examples suggest that application of these methods may reveal a rather large number of circumstances in which progeny testing for recessive alleles in domestic animals is warranted. They certainly provide a logical basis for reaching a reasoned decision.

The authors are Professor and Associate Professor in the Division of Biological and Medical Sciences, Brown University, Providence, Rhode Island 02912. They are indebted to Professor Alan Robertson for numerous valuable comments and suggestions. This work was supported by U.S. Public Health Service Grant GM-00329-10.

Literature

1. Falconer, D. S.: Introduction to Quantitative Genetics, p. 37. Edinburgh: Oliver and Boyd Ltd. 1960. — 2. Kempthorne, O.: An Introduction to Genetic Statistics, pp. 142—148. New York: John Wiley and Sons 1957. — 3. Kidwell, J. F.: The number of progeny required to test a male for heterozygosity. J. Hered. **42**, 215—216 (1951). — 4. Kidwell, J. F.: The use of *a priori* probability in progeny tests for recessive alleles. J. Hered. **61**, 55—58 (1970). — 5. Schaible, R. H.: Progeny testing for recessive genes. Roles of parental types and gene frequencies. J. Hered. **59**, 120—125 (1968). — 6. Wriedt, S.: Heredity in Livestock, p. 137. London: Macmillan 1930.

Received December 20, 1971
Communicated by A. Robertson

J. F. Kidwell
G. W. Hagy
Division of Biological and Medical Sciences
Brown University
Providence, Rhode Island 02912 (USA)