Methods of segregation analysis for animal breeding data: parameter estimates

S. A. KNOTT, C. S. HALEY* & R. THOMPSON*

Institute of Cell, Animal and Population Biology, University of Edinburgh, Zoology Building, West Mains Road, Edinburgh EH9 3JT and AFRC Institute of Animal Physiology and Genetics Research, Edinburgh Research Station, Roslin, Midlothian, Scotland EH25 9PS, UK*

Three approximations to the combined model likelihood (including variation due to both polygenes and a major gene) used in segregation analysis were compared in their ability to estimate the parameters describing the model. Paternal half-sib data simulated under four different combined models were used. In general, when a combined model provides the best explanation of the data the parameter estimates from all three approximations investigated are, on average, in good agreement with the values simulated. The use of Hermite integration to replace the integration in the combined model likelihood provided the parameter estimates closest to those simulated. The other two approximations, based on extensions of linear mixed-model theory, gave estimates further from the expected values, however, these latter two methods can estimate transmitting abilities for the sires directly from the analyses. The three approximations are similar in their ability to genotype sires at the major locus. On average, with the models studied here, about 65 per cent of sires can be successfully genotyped.

Keywords: animal breeding, major genes, maximum likelihood, segregation analysis.

Introduction

Segregation analysis provides a tool for the detection of a segregating major gene (Morton & MacLean, 1974) by the comparison of the likelihood of the data under the combined model (containing genetic variation from both polygenes and a major gene) with the likelihood under the polygenic model (containing only polygenic genetic variation). For a sire model the exact combined model likelihood contains an integration over the transmitting ability for each sire which is computationally intensive to calculate even with only a few offspring per sire. Three approximations to the combined model likelihood have been described by Knott *et al.* (1992). The first (Herm) approximates this integration over transmitting ability with a summation using the Hermite polynomial to provide abscissae and weights. The second and third are extensions of mixed-model methods (in the statistical sense, including fixed and random effects), which are already familiar to animal breeders. One replaces the integration with a single estimate of the mode of each sire's transmitting ability distribution (ME1) and the other estimates three modes, one for each possible major genotype of the

sire (ME3). These approximations have been compared in their ability to detect a segregating major gene in balanced half-sib data (Knott *et al.*, 1992). The use of a maximum likelihood (ML) technique means that estimates of the parameters involved in the model can be obtained. In this paper the ability of these approximations to estimate the effect and frequency of a major gene segregating in a polygenic background, using phenotypic records from paternal half-sibs, will be investigated.

In order to utilize the knowledge that an allele with large effect on the trait of interest is segregating in the population, identification of the genotype of each individual at that locus is required. Mating schemes could then be designed to optimize improvement. The ability of the methods to identify correctly the major genotype of individuals and to predict the polygenic component will also be considered.

Analyses

A simulation study was carried out to investigate the ability of the three approximations to the combined model likelihood to estimate the parameters involved in the model. The simulated data have been described

Correspondence: S. A. Knott.

Table 1 Parameter values used to simulate the combined (major gene and polygenic) models

Model	$p(A)$	μ_{AA}	μ_{Aa}	μ_{aa}	σ_u^2	σ_w^2	σ_{mg}^2	h_{poly}^2	h_{mg}^2
Additive	0.2	0.5	20	10	0	5	95	50	0.2	0.33
Additive	0.4	0.5	20	10	0	10	90	50	0.4	0.33
Dominant	0.5	20	20	0	5	95	75	0.2	0.43
Rare	0.2	20	10	0	5	95	32	0.2	0.24

$p(A)$ = the frequency of the high-scoring allele in the parent population.

μ_d = the effect of major genotype d, relative to the effect of the low-scoring homozygote (μ_{aa}).

σ_u^2 = the additive polygenic sire variance component.

σ_w^2 = the residual variance.

σ_{mg}^2 = the variance contributed by the major gene.

$$h_{poly}^2 = \frac{4\sigma_u^2}{\sigma_u^2 + \sigma_w^2}.$$

$$h_{mg}^2 = \frac{\sigma_{mg}^2}{\sigma_u^2 + \sigma_w^2 + \sigma_{mg}^2}.$$

in Knott *et al.* (1992) and consist of phenotypes from 20 half-sib progeny from each of 50 sires. Four combined models were considered (see Table 1) with 100 replicates of each model.

Briefly, the mean effect of the low scoring homozygous genotype (μ) at the major locus and the deviation from this mean of the other two major genotype means (μ_{AA} and μ_{Aa}) were estimated in the analysis. The population was assumed to be in Hardy–Weinberg equilibrium and a single allele frequency $p(A)$ was estimated. Two analyses of each dataset were carried out, the first assuming that the polygenic heritability was known and estimating just the residual variance, and the second estimating the polygenic heritability as well as the residual variance.

The maximization procedure used for the three approximations is described in Knott *et al.* (1992). The test statistic is provided by twice the difference between the natural logarithms of MLs under the combined model and the polygenic model.

The genotypes of animals at the major locus cannot be determined with complete certainty but the most probable genotype can be obtained assuming that the mode of inheritance is known and that all parameters in the model are known. In essence, the probability of an individual being a particular genotype is the ratio of the likelihood of the pedigree assuming that the individual is that genotype to the total likelihood for the pedigree. For the sire model being considered the inclusion of a polygenic component and a major gene is straightforward (Elsen *et al.*, 1988). Sires are assumed

to be independent of each other and, hence, the conditional probability of each genotype for each sire is dependent on the phenotypes of his half-sib offspring only. Using the notation defined in Knott *et al.* (1992), and assuming that the trait is controlled by both a major gene and a polygenic component, the conditional probability for sire i having genotype c using the exact likelihood is:

$$q_i(c) = \frac{p(c) \int_{-\infty}^{+\infty} h(u_i) \prod_{j=1}^{n} \sum_{d=1}^{m} \text{trans}(d|c) k_d(y_{ij}|\mu,\mu_d,u_i,\sigma_w^2) \cdot du_i}{\sum_{c'=1}^{m} p(c') \int_{-\infty}^{+\infty} h(u_i) \prod_{j=1}^{n} \sum_{d=1}^{m} \text{trans}(d|c') k_d(y_{ij}|\mu,\mu_d,u_i,\sigma_w^2) \cdot du_i} \quad (1)$$

where $p(c)$ is the frequency of major genotype c in the population of sires, $\text{trans}(d|c)$ is the probability of an offspring having major genotype d given that the sire has genotype c, $h(u_i)$ is the likelihood that sire i has polygenic transmitting ability u_i, $k_d(y_{ij}|\mu,\mu_d,u_i,\sigma_w^2)$ is the conditional likelihood that offspring j of sire i has phenotype y_{ij} when offspring j has major genotype d and the polygenic component contributed from sire i is u_i.

In practice the parameters required to describe the effect and frequency of the alleles at the major locus and the polygenic and residual variances will be replaced by their ML estimates. Expressions for the offspring can be obtained in a similar way. For the three approximations the integration is replaced in the same way as described above. For the ME3 and ME1 methods the use of an EM (Expectation Maximization) algorithm means that these probabilities are calculated each iteration and the values from the final iteration can be used.

Using the ME1 approximation, an estimate of each sire's transmitting ability is obtained immediately from the analysis. It is a weighted estimate over all possible major genotype combinations for the sire and half-sib family. The ME3 approximation provides a transmitting ability estimate for each major genotype of each sire. To obtain a single transmitting ability for each sire the three estimates can be weighted according to the conditional frequency of the sire being each of the genotypes. The use of the Herm likelihood does not immediately give transmitting ability estimates. An estimate of the transmitting ability can be obtained by weighting the estimate used in each point of the summation by the probability of that estimate. This

could be calculated, either for each major genotype of the sire, or as given in the equation below, for all three genotypes considered together.

$$u_i = \frac{\sum\limits_{g=1}^{G} u_g \exp\left[-\dfrac{u_g^2}{2\sigma_u^2}\right] \sum\limits_{c=1}^{m} p(c) \prod\limits_{j=1}^{n} \sum\limits_{d=1}^{m} \text{trans}(d\,|\,c) \exp\left[-\dfrac{(y_{ij}-\mu-\mu_d-u_g)^2}{2\sigma_w^2}\right]}{\sum\limits_{g=1}^{G} \exp\left[-\dfrac{u_g^2}{2\sigma_u^2}\right] \sum\limits_{c=1}^{m} p(c) \prod\limits_{j=1}^{n} \sum\limits_{d=1}^{m} \text{trans}(d\,|\,c) \exp\left[-\dfrac{(y_{ij}-\mu-\mu_d-u_g)^2}{2\sigma_w^2}\right]}$$

where G is the number of points in the summation, $u_g = V_i x_g + l_i$, V_i is the scaling parameter (here used as σ_u), l_i is the location parameter for sire i (here used as 0) and x_g are the abscissae.

In the combined model analyses the transmitting ability is the predicted mean of the polygenic component in the half-sib offspring of the sire. The expected mean contribution of the major gene to his progeny can be calculated as:

$$\sum\limits_{c=1}^{m} q_i(c) \sum\limits_{d=1}^{m} \text{trans}(d\,|\,c)\mu_d.$$

The estimated transmitting abilities and the predicted mean major gene performance in the progeny were compared with the simulated values for 20 sets of data containing an additive major gene with equal allele frequencies and a polygenic heritability of 0.2 (additive

0.2 model) and 20 with a dominant major gene (dominant model).

Results and discussion

The results are based on the analysis of each dataset that produced the highest likelihood. The high scoring homozygote was defined as genotype AA.

If the estimates from the analyses are unbiased, the mean parameter estimates over the 100 simulations should give good estimates of the population parameters, i.e. those used to simulate the data. Table 1 gives the parameter values used to simulate the four combined models. For many of the analyses the ME1 and ME3 likelihoods give zero test statistics (see Table 2) with the major genotype means equal to zero and the allele frequency cannot be estimated. We are concerned here with the estimation of the major gene parameters, rather than the total genetic variation, hence these analyses have been omitted from the summaries of results. In some cases, the high number of analyses resulting in a polygenic model means that results are based on rather few observations.

Variances

The mean and standard deviation of the variance estimates for the analyses which resulted in a non-polygenic model are given in Table 3. The sire and residual variances are those estimated in the analyses. The major gene variance is calculated as:

$$p(\text{AA})\mu_{\text{AA}}^2 + p(\text{Aa})\mu_{\text{Aa}}^2 - [p(\text{AA})\mu_{\text{AA}} + p(\text{Aa})\mu_{\text{Aa}}]^2$$

using the estimates for the major genotype means and genotype frequencies obtained from the analyses,

Table 2 Number of analyses from the 100 replicates resulting in mixed (MM), major gene (MG) and polygenic (poly) models for the three approximations to the mixed model likelihood

Model		Herm			ME1			ME3		
		MM	MG	poly	MM	MG	poly	MM	MG	poly
Additive 0.2	(f)	100	—	0	20	—	80	95	—	5
	(e)	80	20	0	0	95	5	2	88	10
Additive 0.4	(f)	100	—	0	2	—	98	43	—	57
	(e)	93	7	0	0	37	63	1	35	64
Dominant	(f)	100	—	0	99	—	1	100	—	0
	(e)	92	8	0	9	81	0	4	96	0
Rare	(f)	100	—	0	56	—	44	97	—	0
	(e)	91	9	0	1	82	17	21	64	15

(f) with polygenic heritability fixed in the analyses.
(e) with polygenic heritability estimated in the analyses.

Table 3 Mean and standard deviation of the variance estimates for the analyses with fixed and estimated polygenic heritability that resulted in a non-polygenic model. Expected means are given in Table 1. Also given are the regressions on (b) and the correlations (r) with the estimates obtained directly from the simulation for each dataset

Model	Statistic	Herm σ_w^2	σ_u^2	σ_{mg}^2	ME1 σ_w^2	σ_u^2	σ_{mg}^2	ME3 σ_w^2	σ_u^2	σ_{mg}^2
Fixed polygenic heritability										
Additive 0.2	Mean	94.36	4.97	50.80	99.78	5.25	46.43	108.99	5.74	33.30
	s.d.	11.95	—	14.27	15.99	—	19.94	11.88	—	13.81
	b	0.80	−0.08	1.63	1.37	0.01	4.79	0.91	−0.09	2.14
	r	0.29	−0.23	0.17	0.28	0.01	0.35	0.33	−0.28	0.23
Additive 0.4	Mean	88.70	9.86	51.36	77.85	8.65	71.37	111.03	12.34	22.90
	s.d.	13.82	—	17.94	1.68	—	6.73	11.99	—	16.14
	b	0.97	−0.11	2.35	—	—	—	0.65	−0.20	−0.13
	r	0.30	−0.20	0.23	—	—	—	0.22	−0.43	−0.04
Dominant	Mean	90.81	4.78	80.44	93.84	4.94	74.45	95.26	5.01	74.79
	s.d	8.06	—	11.17	9.36	—	12.78	8.10	—	11.06
	b	0.77	0.01	0.94	0.84	0.02	1.15	0.87	0.01	0.95
	r	0.46	0.07	0.44	0.44	0.08	0.47	0.53	0.03	0.45
Rare	Mean	90.52	4.76	34.55	101.87	5.36	18.25	100.52	5.29	22.44
	s.d.	10.56	—	12.62	11.68	—	13.76	10.24	—	10.53
	b	0.89	−0.04	1.39	1.13	−0.07	1.66	1.53	−0.04	0.94
	r	0.36	−0.15	0.31	0.40	−0.22	0.33	0.63	−0.17	0.25
Estimated polygenic heritability										
Additive 0.2	Mean	93.95	5.43	50.79	88.90	0.00	61.32	90.30	0.25	59.84
	s.d.	13.82	5.21	17.13	9.04	0.00	10.97	11.18	1.69	14.11
	b	0.58	0.89	2.56	0.67	0.00	1.65	0.66	0.12	3.16
	r	0.18	0.32	0.22	0.32	0.00	0.22	0.26	0.13	0.33
Additive 0.4	Mean	87.96	10.18	52.30	76.15	0.00	76.64	82.22	0.76	66.11
	s.d.	13.47	6.56	18.06	7.44	0.00	12.06	9.21	4.58	14.93
	b	0.88	0.74	0.71	0.24	0.00	1.89	0.62	0.10	2.66
	r	0.28	0.30	0.07	0.13	0.00	0.27	0.28	0.05	0.36
Dominant	Mean	90.05	5.03	80.92	92.07	0.05	83.74	93.06	0.80	81.81
	s.d.	5.03	4.07	10.85	9.75	0.46	12.00	13.44	9.21	11.90
	b	0.88	1.02	0.88	1.05	0.03	0.82	0.92	0.52	0.88
	r	0.49	0.57	0.42	0.53	0.15	0.36	0.49	0.51	0.39
Rare	Mean	90.09	5.00	35.08	87.00	0.87	42.21	90.40	2.02	37.53
	s.d.	11.26	3.80	13.79	11.29	2.78	14.86	10.79	4.15	15.67
	b	0.89	1.03	1.40	1.06	0.14	1.75	0.82	0.58	1.61
	r	0.33	0.56	0.29	0.39	0.10	0.34	0.31	0.27	0.30

assuming Hardy–Weinberg equilibrium. Although the expected parameter estimates are the same for each of the 100 simulations of the same genetic model, the realized values will vary due to sampling. Hence, for each dataset the variance estimates from segregation analysis can be compared with the estimates obtained directly from the simulation. The simulated sire and residual variances have been calculated by analysis of variance using the simulated data with the effect of the major gene subtracted. The major gene variance has been calculated using the equation given above with the realized major genotype frequencies for the offspring obtained in the simulation and the major geno-

type means used to simulate the data. The regression of the estimates from segregation analysis on, and the correlation with, the simulated values for the same set of data are also given in Table 3.

For all models, both with the polygenic heritability estimated and fixed in the analyses, the mean major gene variance estimated using the Herm method was closest to the expected value. There is a positive linear relationship between these estimates and the value obtained directly from the simulation and the slope of the regression was around one. With the polygenic heritability fixed in the analyses, the ME1 and ME3 approximations always underestimate the variance

contribution from the major gene in comparison with Herm and, as a consequence, there is a consistent overestimation of the residual variance. With the polygenic heritability estimated, a large number of the ME1 and ME3 analyses resulted in a major gene model with the sire variance equal to zero and the residual variance was generally underestimated.

There is a very low or negative association between the sire variance obtained from segregation analysis and the value estimated from the simulation. This is because this variance is not being estimated directly but as a fixed proportion of the residual variance, as there is more information on the residual variance this

is estimated fairly well at the expense of the sire variance. The sire variance, with the heritability estimated, is now estimated directly and, generally for Herm, the correlation with, and the regression on, the values estimated from the simulation are close to one.

Major gene parameters

Table 4 gives the mean and standard deviation of the parameter estimates $[\mu_{AA}, \mu_{Aa}, p(A)]$ from those analyses that gave a non-polygenic maximum. These can be compared with the values used to simulate the data. The mean estimate of the variance contributed by

Table 4 Mean and standard deviation of the major gene parameter estimates for the analyses with fixed and estimated polygenic heritability that resulted in a non-polygenic model. Expected means are given in Table 1. Also given are the regression (b) of the ME1 and ME3 estimates on and the correlations (r) with the estimates obtained using Herm for each dataset

Model	Statistic	Herm			ME1			ME3		
		$p(A)$	μ_{AA}	μ_{Aa}	$p(A)$	μ_{AA}	μ_{Aa}	$p(A)$	μ_{AA}	μ_{Aa}
Fixed polygenic heritability										
Additive 0.2	Mean	0.49	18.75	9.11	0.47	19.20	9.33	0.49	16.22	8.03
	s.d.	0.13	4.65	5.16	0.22	4.12	5.44	0.18	3.68	3.66
	b				1.65	1.29	1.67	1.14	0.65	0.60
	r				0.88	0.87	0.90	0.82	0.70	0.78
Additive 0.4	Mean	0.49	18.85	9.52	0.54	22.98	9.13	0.48	16.02	7.08
	s.d.	0.16	5.51	5.15	0.10	0.31	2.82	0.26	5.84	5.00
	b				—	—	—	1.62	0.49	0.71
	r				—	—	—	0.81	0.28	0.50
Dominant	Mean	0.50	20.69	20.15	0.52	18.75	20.74	0.51	20.24	19.80
	s.d.	0.05	3.79	1.75	0.05	2.46	1.53	0.04	2.50	1.63
	b				0.58	0.57	0.83	0.66	0.61	0.90
	r				0.58	0.88	0.94	0.82	0.93	0.97
Rare	Mean	0.26	17.91	9.87	0.13	21.04	7.87	0.21	18.38	8.48
	s.d.	0.17	7.68	4.57	0.12	7.67	6.89	0.15	6.90	4.05
	b				0.39	0.68	0.96	0.70	0.80	0.59
	r				0.48	0.61	0.57	0.81	0.86	0.64
Estimated polygenic heritability										
Additive 0.2	Mean	0.49	19.24	9.13	0.50	22.10	10.96	0.51	21.86	10.82
	s.d.	0.16	6.13	6.62	0.09	2.13	1.71	0.10	2.46	2.18
	b				0.40	0.12	0.11	0.50	0.27	0.27
	r				0.71	0.33	0.40	0.70	0.55	0.67
Additive 0.4	Mean	0.48	19.31	9.75	0.49	24.59	12.04	0.50	23.16	11.79
	s.d.	0.16	5.44	5.33	0.08	2.51	2.41	0.12	2.39	3.63
	b				0.72	0.34	0.68	0.78	0.46	0.58
	r				0.82	0.49	0.86	0.86	0.66	0.78
Dominant	Mean	0.50	20.49	20.29	0.52	23.44	19.05	0.51	22.72	19.26
	s.d.	0.05	3.88	1.92	0.06	4.04	1.78	0.05	3.75	1.65
	b				0.77	0.66	0.63	0.93	0.70	0.67
	r				0.60	0.63	0.68	0.86	0.73	0.79
Rare	Mean	0.26	18.72	10.83	0.30	21.36	9.15	0.28	20.93	9.24
	s.d.	0.17	8.14	8.03	0.14	3.69	3.80	0.15	4.98	3.65
	b				0.57	0.38	0.35	0.49	0.44	0.39
	r				0.62	0.73	0.41	0.47	0.67	0.48

the major gene was closest to the expected value when using the Herm approximation (see Table 3) and the mean estimates for the effect of the major genotypes and the frequency of the alleles are in good agreement with the expected values. Hence, the parameter estimates obtained using the Herm method have been used as a baseline against which to compare the major gene parameters from the three approximations for each dataset. Table 4 gives the slope of the regression of the ME1 and ME3 major gene parameter estimates on, and the correlation with, the Herm estimates for the same set of data.

With the polygenic heritability fixed in the analyses, those simulations which resulted in a non-zero test statistic for the ME1 method tended to be those that produced the largest major gene estimates in the Herm method and, hence, the mean ME1 major genotype mean estimates tend to be larger than the mean Herm estimates for all analyses. For individual datasets the ME1 method produces smaller estimates than the Herm method.

The mean effect of the major genotypes was overestimated for the analyses using the ME1 and ME3 likelihoods with the polygenic heritability estimated. This is because of the large number of analyses resulting in a major gene model with a sire variance estimate of zero. Hence, all the genetic variance is explained by the major gene effect.

In general, both when analysing with fixed polygenic heritability and when estimating the heritability, the regression and correlation coefficients obtained comparing the ME3 estimates with those obtained using the Herm method are closer to one than the coefficients obtained comparing the ME1 to the Herm estimates. With the heritability estimated the difference

between these two approximations is less than with the heritability fixed because the resulting major gene models are identical for the two methods.

Genotyping sires at the major locus

The conditional probability of each sire being each major genotype (from equation 1) was calculated using the maximum likelihood estimates. Two criteria for genotyping have been considered, one allocating sires to the major genotype with the highest probability and the other allocating sires to a genotype only if the conditional probability was greater than 0.75. For each analysis the percentage of sires correctly genotyped was calculated and the results are given in Table 5 as the mean percentage correctly genotyped over the 100 analyses of each model, excluding those that resulted in a polygenic model.

Using 0.75 as the minimum value of the conditional probability required to genotype a sire, a proportion of sires (between 60 and 25 per cent using the Herm method) could not be genotyped. However, of those genotyped the proportion incorrect was lower than when all sires were genotyped based on the major genotype with the highest conditional probability. More sires were correctly genotyped when the simulated major gene was rare or had a dominant effect, especially using the high probability as a criterion. In addition, using 0.75 as the conditional probability required for genotyping, homozygotes were more easily genotyped than heterozygotes (results not shown).

The results from the ME1, and to a lesser extent the ME3 method, will be biased because of the omission of the analyses that resulted in a polygenic model. These analyses, in general, correpsond to models with small

Table 5 Percentage of sires correctly genotyped (and the percentage incorrect in brackets) for two cases: (i) where the conditional probability has to be greater than 0.75 for a sire to be genotyped and (ii) where all sires are genotyped based on their highest conditional probability. Only analyses resulting in non-polygenic models have been included

Model		Herm > 0.75	Herm Highest	ME1 > 0.75	ME1 Highest	ME3 > 0.75	ME3 Highest
Additive 0.2	(f)	31.0 (11.1)	63.5 (36.5)	46.1 (25.1)	59.3 (40.7)	22.4 (11.3)	59.5 (40.5)
	(e)	35.0 (16.6)	59.6 (40.4)	50.8 (18.3)	67.6 (32.4)	50.2 (18.2)	67.3 (32.7)
Additive 0.4	(f)	26.8 (12.8)	57.3 (42.7)	58.0 (32.0)	63.0 (37.0)	18.5 (20.8)	48.7 (51.4)
	(e)	29.1 (16.5)	56.8 (43.2)	54.5 (24.8)	65.2 (34.8)	49.1 (27.4)	61.4 (38.6)
Dominant	(f)	48.5 (8.8)	74.4 (25.6)	52.5 (10.0)	74.7 (25.3)	47.4 (7.6)	75.4 (24.6)
	(e)	48.9 (9.9)	73.8 (26.2)	55.1 (15.9)	71.8 (28.2)	54.2 (13.9)	73.1 (26.9)
Rare	(f)	47.3 (11.4)	68.8 (31.2)	63.7 (20.3)	72.0 (28.0)	48.9 (10.8)	71.0 (29.0)
	(e)	49.8 (13.7)	69.5 (30.5)	54.0 (20.2)	66.4 (33.6)	53.3 (19.3)	66.8 (33.2)

(f) with the polygenic heritability fixed in the analyses.
(e) with the polygenic heritability estimated in the analyses.

estimated major gene effects with the Herm method and sires are more difficult to genotype. The conditional probabilities for the ME1 method are more extreme, being nearer to 1 or 0, than those calculated using the Herm method and this results in more sires being correctly genotyped at a probability of 0.75. With the heritability estimated in the analyses, both the ME1 and ME3 methods were improvements over the Herm method when using 0.75 as the conditional probability required for genotyping. When taking the genotype with the highest conditional probability, the three methods were similar. The results from the additive model with a polygenic heritability of 0.4 for ME1 are based on very few analyses, in which evidence for the major gene was large and the sires relatively easy to genotype.

Predicting progeny performance

The transmitting ability estimates and predicted progeny major gene effect, obtained using segregation analysis, were compared with the values simulated. The correlation of the estimate with the simulated value for the same dataset is given in Table 6 for the analyses using the Herm method. The results are based on analyses of 20 replicate datasets generated under each of the additive (additive 0.2) and dominant models with the polygenic heritability estimated. The ability to predict the total genetic effect in the offspring was considered. The data were also analysed assuming a polygenic model. The transmitting ability estimates

Table 6 Mean correlations (with standard deviations in brackets) of the estimated components of the progeny phenotypes with those simulated from analyses of 20 datasets simulated under additive 0.2 and 20 under dominant using Herm

Components	Mixed model	Polygenic model
Additive major gene		
$g, \hat{u}$	0.35 (0.21)	0.47 (0.10)
$mg, \hat{mg}$	0.69 (0.17)	
$g + mg, \hat{u} + \hat{mg}$	0.90 (0.02)	0.91 (0.02)
$\bar{y}, \hat{u} + \hat{mg}$	0.99 (0.01)	1.00 (0.00)
Dominant major gene		
$g, \hat{u}$	0.55 (0.17)	0.46 (0.11)
$mg, \hat{mg}$	0.84 (0.04)	
$g + mg, \hat{u} + \hat{mg}$	0.91 (0.02)	0.91 (0.02)
$\bar{y}, \hat{u} + \hat{mg}$	0.97 (0.02)	1.00 (0.00)

g = simulated progeny polygenic mean.
$\hat{u}$ = estimated sire transmitting ability.
mg = simulated mean effect of the major gene in the progeny.
$\hat{mg}$ = estimated mean effect of the major gene in the progeny.
$\bar{y}$ = simulated progeny phenotype mean.

obtained from these analyses have been compared with the mean of the simulated polygenic component for the progeny, the mean total genetic effect and the mean phenotype.

The balanced structure of the data means that, when analysed assuming a polygenic model, the correlation between the progeny phenotype means and the estimated transmitting ability is one. The mean of the total simulated genetic component for the offspring is equally well predicted by the combined and the polygenic models with correlations of at least 0.9 on average. With the simulated dominant major gene the combined model analyses can distinguish the major gene and the polygenic components. Hence, analyses under the combined model provide, on average, better estimates of the simulated polygenic component than analyses under the polygenic model. Analysing the simulated data containing an additive major gene, the two genetic components cannot be so easily separated. In five of the 20 analyses under the combined model a major gene model resulted, which gave correlations of zero between the simulated and estimated polygenic components and caused the low mean correlation observed. When these analyses are omitted a mean correlation of 0.46 (with standard deviation of 0.21) is obtained which is similar to the value from analyses assuming a polygenic model. With the dominant major gene only one analysis resulted in a major gene model, omitting this gives a correlation between the simulated and estimated mean polygenic component in the offspring of 0.58 (with standard deviation of 0.11).

The two genetic components are not well separated using the ME1 and ME3 approximations and when the heritability is estimated most analyses result in a major gene model. The correlations obtained with these approximations were, on average, lower than those given in Table 6 from analyses with the Herm method.

Conclusions

In general, where a major gene effect was detected all methods could provide reasonable estimates of its effect and the allele frequency. When the polygenic heritability was estimated in the analysis, the actual proportion of total genetic to environmental variation in the data could be due to either major genetic or polygenic variation or to these two effects combined. There is difficulty, especially using the ME1 and ME3 approximations, in distinguishing the two sources of genetic variance and the analyses result in a model containing either major gene or polygenic variation. The parameter results obtained from these analyses that resulted in a major gene model, however, in general, gave a good indication as to the effect of the

gene, whether it was additive or dominant, and the frequencies of the alleles in the population.

In general with all the models simulated, there was a weak, positive correlation between the variance components estimated in the analyses and the estimates of the variances generated in the simulation. Where the phenotypic distributions were non-symmetric (i.e. for the dominant and rare models) the Herm method overestimated the contribution of the major gene to the phenotypic variance and underestimated the residual variance to a similar extent. When the heritability was fixed, the ME1 and ME3 analyses, on average, overestimated the residual variance and consequently underestimated the effect of the major gene compared with the results from the Herm analyses.

To optimize selection, the genotype of each individual, both at the major locus and for the polygenic component, is required. Information from the two components can then be combined as desired. The results presented here suggest that, if sires are assigned the genotype with the highest conditional probability, most sires (between 57 and 74 per cent for the Herm method) are correctly genotyped. However, if the criterion for assigning a major genotype to individuals is that the conditional probability of being that genotype is high (reducing the number of sires incorrectly genotyped), only a small proportion of sires can be successfully genotyped (between 26 and 50 per cent for the Herm method). As the number of offspring per sire increases the proportion of sires correctly genotyped at the high probability increases.

The mean major gene effect in the offspring was well estimated with the combined model. The total genetic performance was equally well estimated from the combined and the polygenic model. The models assumed that the polygenic component was additive and when the simulated major gene was dominant the two genetic components could be more easily distinguished. In this situation analyses assuming a combined model provided better estimates of the polygenic contribution to the mean progeny phenotype than assuming a polygenic model.

In conclusion, if evidence for a major gene is found in the data the estimates of its effect and frequency are in good agreement with the expected values. Using the ME1 and ME3 approximations the polygenic contribution to the variance is not well estimated as a variance component of zero often results when a major gene is found.

Acknowledgements

References

ELSEN, J. M., VU TIEN KHANG, J. AND LE ROY, P. 1988. A statistical model for genotype determination at a major locus in a progeny test design. *Genet., Select., Evol.,* **20**, 211–226.

KNOTT, S. A., HALEY, C. S. AND THOMPSON, R. 1992. Three approximations to the segregation analysis likelihood: a comparison of power. *Heredity,* **68**, 299–311.

MORTON, N. E. AND MACLEAN, C. J. 1974. Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am. J. Hum. Genet.,* **26**, 489–503.