Genet Sel Evol (1996) 28, 345-358

Original article

Power and parameter estimation of complex segregation analysis under a finite locus model

P Uimari, BW Kennedy, JCM Dekkers

Department of Animal and Poultry Science, Centre for Genetic Improvement of Livestock, University of Guelph, Guelph, ON N1G 2W8, Canada

(Received 20 October 1995; accepted 7 May 1996)

Summary – Power and parameter estimation of segregation analysis was investigated for independent nucleus family data on a quantitative trait generated under a finite locus model and under a mixed model. For the finite locus model, gene effects at ten loci were generated from a geometric series. Additionally, linkage between a major locus and other loci was considered. Two different methods of segregation analysis were compared: a mixed model and a finite polygenic mixed model. Both statistical methods gave similar power to detect a major gene and estimates of parameters. An exception was a situation where two major loci had an equal effect on phenotype: the mixed model had a higher power than the finite polygenic mixed model, but estimates of the parameters from the mixed model were more biased than estimates from the finite polygenic mixed model. Segregation analysis was more powerful in detecting a major gene when data were generated under the finite locus model than under the mixed model. When a major gene was linked to another gene, a major gene was more difficult to detect than without such linkage. Segregation of two major genes created biased estimates. Bias increased with linkage when parents were not a random sample from a population in linkage equilibrium.

parameter estimation / power / major gene / segregation analysis

Résumé – Puissance et estimation des paramètres dans l'analyse de ségrégation complexe avec un modèle à nombre fini de locus. *La puissance de l'analyse de ségrégation et l'estimation des paramètres ont été étudiées sur des familles nucléaires indépendantes pour un caractère quantitatif déterminé soit par un nombre fini de locus soit selon un modèle d'hérédité mixte, impliquant un gène majeur et un résidu polygénique infinitésimal. Dans le modèle à nombre fini de locus, le nombre de locus supposé était de dix et leurs effets suivaient une loi de distribution géométrique. En outre, la possibilité de liaison génétique entre un locus majeur et d'autres locus était envisagée. Deux méthodes d'analyse de ségrégation ont été comparées, utilisant soit un modèle d'hérédité mixte, soit un modèle d'hérédité avec un nombre fini de locus. Les deux méthodes statistiques présentaient des puissances similaires pour détecter un gène majeur et estimer les paramètres correspondants. À l'exception toutefois d'une situation avec deux locus majeurs ayant le même effet sur le phénotype. Le modèle à hérédité mixte avait alors une puissance supérieure à celle du modèle à nom-*

bre fini de locus, mais les estimées des paramètres à partir du modèle mixte étaient plus biaisées que celles du modèle à nombre fini de locus. L'analyse de ségrégation était plus puissante pour détecter un gène majeur dans le cas d'un caractère déterminé par un nombre fini de locus que dans une situation d'hérédité mixte. Un gène majeur lié à un autre gène était plus difficile à détecter qu'en l'absence de liaison génétique. La ségrégation de deux gènes majeurs créait des biais d'estimation. Les biais étaient encore accrus en cas de liaison génétique quand les parents n'étaient pas tirés d'une population en équilibre gamétique pour les deux locus majeurs.

estimation de paramètre / puissance / gène majeur / analyse de ségrégation

INTRODUCTION

Statistical methods used to determine the mode of inheritance of a quantitative trait in detection of major genes rely on phenotypic information. In addition, methods can utilize information on genetic markers, which are now numerous. In both cases, the most common statistical methods to detect a major gene are based on maximum likelihood theory. Maximum-likelihood-based complex segregation analysis was introduced by Elston and Stewart (1971) and Morton and MacLean (1974). Complex segregation analysis combines three factors into a mixed model for analysis of phenotypes for a quantitative trait: a gene which explains a detectable part of genetic variance (major gene); residual polygenic variance, for which individual gene effects are not of direct interest or detectable; and environment. Recently a finite polygenic mixed model, which explains the polygenic part of inheritance by a finite number of loci, was proposed by Fernando et al (1994) as an alternative formulation for the mixed model. To make the finite polygenic mixed model computationally feasible it is assumed that loci which explain the polygenic part of inheritance are unlinked, biallelic, codominant, and have equal gene effects and equal frequencies of favourable alleles (0.5) across loci (Fernando et al, 1994).

Power of segregation analysis of independent nucleus family data (full-sib families) with the mixed model was investigated by MacLean et al (1975) and Borecki et al (1994) and for half-sib data by Le Roy et al (1989) and Knott et al (1991). In all cases, data were simulated according to the mixed model of inheritance. The general conclusion from these studies was that the best chance to detect a major gene is if it is dominant with moderate to low frequency in the population. By increasing data size (number of families and size of the families), major genes with smaller effects can be detected.

Many aspects that might affect robustness of segregation analysis with the mixed model have been studied also (MacLean et al, 1975; Go et al 1978; Demenais et al, 1986). The main concern has been false detection of a major gene with skewed data. To overcome this problem, power transformation of the data was proposed (MacLean et al, 1976). The optimal solution for skewed data is to make the transformation simultaneously with estimation of other parameters (MacLean et al, 1984). Removing skewness may, however, lead to reduced power to detect a major gene (Demenais et al, 1986).

Other common assumptions in segregation analysis include homogeneous variance within major genotypes, independence between the major gene and polygenic effects, no genotype by environmental correlation, and no correlation between environment of parent and offspring (MacLean et al, 1975).

One basic assumption of segregation analysis, which has received less attention, is normality of the residual distribution (polygenic + environmental) within a major genotype. This assumption is met if the polygenic part is controlled by infinite number of genes that each have only a small effect on phenotype, ie, the infinitesimal model (Bulmer, 1980), and if the environmental factor is normally distributed. However, the infinitesimal model might not be the best model for the distribution of gene effects. A model where few genes with a large effect and several genes with small effects control a quantitative trait may be closer to the real nature of the distribution of gene effects. Evidence from *Drosophila melanogaster* supports this hypothesis (Shrimpton and Robertson, 1988; Mackay et al, 1992). Such a distribution of gene effects can be approximated by a geometric series (Lande and Thompson, 1990).

If gene effects follow a geometric series, the distribution within major genotype may not be normal, as with the infinitesimal model. This violates the assumption of a normally distributed polygenic part of the mixed model commonly used in segregation analysis. Two or more loci with large effects can also lie in a cluster on a chromosome, which would link the major gene to other genes and thus violate the assumption of independent segregation of a major gene and polygenes.

The objective of this paper was to study the effect of violation of the two assumptions of the underlying model in segregation analysis, namely a skewed polygenic distribution and linkage between a major gene and polygenes, on the power of detecting a major gene and on parameter estimation. Behavior of the mixed model of segregation analysis (Morton and MacLean, 1974) was compared to the finite polygenic mixed model (Fernando et al, 1994). The methods were compared under an independent nucleus family data structure.

MATERIALS AND METHODS

Balanced data on a quantitative trait were simulated for 25 independent full-sib families, with a sire, dam, and ten offspring. All parents were assumed to be unrelated and were generated from a population under Hardy–Weinberg and linkage equilibria. Genotypes of parents were generated under a ten-locus model (finite locus model) or under a mixed model (from now on this will be called the mixed generating model, whenever necessary, to distinguish between models used for generating and for analyzing the data).

Under the finite locus model, the gene with largest effect had a substitution effect of 1.0 (the difference between two homozygotes is twice the substitution effect) and the gene with the second largest effect had a substitution effect of 0.25, 0.5 or 1.0. Gene effects of the eight other loci followed the geometric series 0.25, 0.125, 0.0625, where one locus had an effect of 0.25, three loci an effect of 0.125 and four loci an effect of 0.0625. Gene frequencies were 0.5 for all loci except for the major locus, for which frequency of the dominant allele was either 0.1, 0.5, or 0.9. Two alleles per locus were simulated. The three loci with largest effect were completely dominant

and other loci were additive. Genotypes of progeny were generated using either independent segregation of loci or the two loci with the largest effect were linked with a recombination rate of 0.1. In the case of linkage, linkage phase of the parents was either random or all parents were double heterozygotes for the two linked loci (favourable alleles on same chromosome).

For every finite locus scenario, corresponding genotypes were also generated with a mixed model. Under the mixed-generating model, a major gene with a substitution effect of 1.0 was simulated, along with a polygenic part, which was simulated from a normal distribution with 0 mean and genetic variance equal to the total genetic variance (additive + dominance) of the other nine loci in the corresponding finite locus model. The polygenic effect of progeny was generated from a normal distribution with mean equal to the average of polygenic effects of the parents and variance equal to half of the polygenic variance.

Phenotypes were generated for both the finite locus and the mixed-generating model by adding an environmental effect to the genotypic effects. Environmental effects were simulated from a normal distribution with mean 0 and variance corresponding to one minus the broad sense heritability (H^2, total genetic variance over phenotypic variance), which was equal to 0.4. A summary of the genetic scenarios that were simulated is given in table I.

Table I. Parameters used for simulating data.

Locus	*Gene effects in all loci*								
	Set 1			*Set 2*			*Set 3*		
	AA	Aa	aa	AA	Aa	aa	AA	Aa	aa
1	1.0	1.0	−1.0	1.0	1.0	−1.0	1.0	1.0	−1.0
2	0.25	0.25	−0.25	0.5	0.5	−0.5	1.0	1.0	−1.0
3	0.25	0.25	−0.25	0.25	0.25	−0.25	0.25	0.25	−0.25
4,5,6	0.125	0.0	−0.125	0.125	0.0	−0.125	0.125	0.0	−0.125
7,8,9,10	0.0626	0.0	−0.0625	0.0625	0.0	−0.0625	0.0625	0.0	−0.0625

	Variances			
	$\sigma_g^{2\,a}$	$\sigma_U^{2\,b} - \sigma_e^{2\,c}$		
		Set 1	*Set 2*	*Set 3*
$p^d = 0.1$	0.6156	0.1250–1.1109	0.2656–1.3218	0.8281–2.1656
$p = 0.5$	0.7500	0.1250–1.3125	0.2656–1.5234	0.8281–2.3672
$p = 0.9$	0.0396	0.1250–0.2469	0.2656–0.4578	0.8281–1.3016

[a] Variance of the major gene (locus 1). [b] Total genetic variance of the polygenic part of inheritance including additive and dominance variances (loci 2 to 10). Loci 2 and 3 where completely dominant, the rest were additive. [c] Environmental variance. [d] The frequency of the dominant allele at the major locus (locus 1). Frequency of alleles at all other loci equal to 0.5.

Simulated data sets were analyzed by two computer packages. The Pedigree Analysis Package (PAP Rev 4.02, Hasstedt, 1982, 1994) was used to compute the likelihood of the mixed model and SALP (segregation and linkage analysis for pedigrees, Stricker et al, 1994) to compute the likelihood of the finite polygenic mixed model. Only one major locus was fitted in SALP. Mendelian transmission probabilities, equal variances within genotypes and no power transformation were used in PAP. Downhill simplex method is used for maximization in SALP and Gemini (Lalouel, 1979) in PAP. Because Gemini does not allow maximization at boundaries of the parameter space (gene frequency and heritability have boundaries at 0 and 1) the program occasionally stopped. In those cases, the parameter that reached the boundary was fixed close to the boundary (0.0001 or 0.9999 for gene frequency and 0.0001 for heritability) and other parameters were maximized conditional on that. Because the major gene was simulated with complete dominance, μ_{AA} was fixed to be equal to μ_{Aa} in all maximum likelihood analyses. Input values for simulation were used as starting values for the maximization process. Likelihood ratio test statistic was calculated by comparing a general model to a model with equal means ($\mu_{AA} = \mu_{Aa} = \mu_{aa}$).

Because SALP and PAP use different parameterization of effects, parameters were converted to two genotypic means (μ_{AA} and μ_{aa}), gene frequency of the dominant allele (p), and polygenic (σ^2_u) and environmental (σ^2_e) variances. Instead of polygenic and environmental variances, PAP estimates heritability (h^2) and the phenotypic standard deviation conditional on major genotype; for the finite polygenic mixed model SALP estimates a scaling factor ($= \sqrt{[\sigma^2_u/(q(1-q)k)]}$, where q is the allele frequency at polygenic loci, which was fixed at 0.5, and k is twice the number of polygenic loci, which was fixed at ten), and phenotypic variance.

Each simulated major gene scenario (table I) was replicated 50 times. Empirical power of the mixed model of analysis was measured as the proportion of cases in which the likelihood ratio test statistic exceeded the X^2 distribution with 2 *df* at 5% significance level.

Because the likelihood test statistic is only asymptotically distributed according to the X^2 distribution (Wilks, 1938), 200 replicates of six data sets without a major gene were generated based on the infinitesimal model and the proportion of test statistics which supported the major gene hypothesis was calculated for both the mixed model and the finite polygenic mixed model. Polygenic and environmental variances of the examples corresponded to sets 2 and 3 (table I) without a major gene. The proportion of false detection is expected to be 5% when a 5% type I error level is used.

Empirical power of the mixed model was measured as the proportion of cases in which the major gene hypothesis was accepted. Under the mixed-generating model, the power corresponds to the probability of detecting the simulated major gene. This is not the case when data are simulated under the finite locus model; instead of detecting the first locus as a major gene, the power indicates the probability of detecting any of the simulated loci as a major gene.

RESULTS

Power of the likelihood ratio test

The proportions of false detection of major gene when no major gene effect was generated, but the likelihood ratio between the mixed model and the polygenic model was compared to the X^2 table value with two degrees of freedom at 5% significance level, were 4, 3 and 6% for set 2 distribution of gene effects (table I) and 4, 3 and 5% for set 3 distribution of gene effects with gene frequencies of 0.1, 0.5, and 0.9, respectively. Using the finite polygenic mixed model and its sub-model the corresponding values were 4, 3, 4 and 4, 4, 3%, for set 2 and set 3, respectively. Thus the true power of detecting a major gene for the data structure used here can be somewhat higher for both methods than reported in table II.

When data were generated under the mixed model, the highest power was achieved when frequency of the dominant allele was low and the lowest power with a rare recessive allele (table II). This pattern was consistent across different proportions of genetic variance explained by polygenes (sets 1, 2 and 3). Under the finite locus model, the pattern changed when two major loci had an equal effect on the trait (table II, set 3); the highest power for the mixed model was achieved when one of the genes was almost fixed in the population, however, the difference between cases of gene frequency of 0.5 and 0.9 for the finite polygenic mixed model was small (without linkage).

The effect of the proportion of total genetic variance that a major gene explained on the power was very clear under the mixed-generating model; the power was higher if the major gene explained a large proportion of total genetic variance, when compared within the same gene frequency (table II, sets 1, 2 and 3). The same pattern was true when data were generated under the finite locus model:

Table II. Power of the mixed model and the finite polygenic mixed model to detect a major gene based on 50 replicates.

p^b	*Data generating model*	*Set 1*[a]	*Set 2*			*Set 3*		
		NL (%)	*NL (%)*	*L (%)*	*LH (%)*	*NL (%)*	*L (%)*	*LH (%)*
0.1	Finite	90 / 88[c]	62 / 62	62 / 62		14 / 8	10 / 6	
	Mixed	82 / 82	68 / 66			18 / 18		
0.5	Finite	64 / 64	58 / 56	40 / 38	24 / 20	44 / 42	36 / 32	20 / 14
	Mixed	58 / 58	46 / 44			16 / 14		
0.9	Finite	50 / 52	36 / 34	34 / 34		66 / 40	66 / 40	
	Mixed	54 / 52	22 / 22			4 / 4		

Under the finite locus model data were generated without linkage (NL), with linkage (L), and with linkage when all parents were double heterozygotes for two loci with largest effects (LH). [a] Distribution of gene effects for finite locus model, see table I. [b] Frequency of the dominant allele at the major locus. [c] The first number indicates power of the mixed model and the second number indicates power of the finite polygenic mixed model.

power reduced when the effect of the second largest locus increased (table II, sets 1, 2 and 3). An exception was, again, a case when two major loci had an equal effect on the trait and frequencies of favourable alleles at the major loci were 0.5 and 0.9 (table II, set 3, $p = 0.9$). In most cases, the higher power of detecting a major gene was achieved when data were generated under the finite locus model than under the mixed model.

Violation of the assumption of independent segregation of the major gene and other genes had a negative effect on the power of the mixed model as well as on the power of the finite polygenic mixed model (table II). Even larger reductions in the power were observed when all parents were double heterozygotes for the two linked loci with largest effects (table II). In this case, not only the assumption of independent segregation of a major gene and polygenes was violated but also the assumption of Hardy–Weinberg equilibrium in the parental population; true probabilities for parents to be homozygotes were zero, not p^2 and $(1 - p)^2$, as was assumed in the analysis. The reduction in the power due to violation of Hardy–Weinberg equilibrium was confirmed by a simulation where all parents were heterozygous for the major locus (a finite locus model similar to set 2 with $p = 0.5$, no linkage). In this case, the power of the mixed model was 28% compared to 58% when the parent population was in Hardy–Weinberg equilibrium (table II, set 2, $p = 0.5$).

Parameter estimation

Mean estimates of parameters, with their empirical standard deviations based on 50 replicates, and true values are given in tables III and IV. The expected variance components for polygenes given in table III (results for the finite locus model) do not include dominance variance of the second and the third largest loci (smaller loci were additive), because the statistical methods studied here did not take polygenic dominance variance into account. As a result, dominance variance may be partly confounded with estimates of additive genetic variance and partly with estimates of residual variance.

For the first distribution of gene effects (set 1) and the finite locus model, both methods gave similar estimates (table III). In most cases, estimates agreed well with true values, although some discrepancies were found for variance components. The standard deviation of the estimate of the genotypic mean depended on the estimated gene frequency and was larger for low frequencies.

Going from the set 1 distribution of gene effects to set 2, with a larger second locus effect, variation of estimates increased (table III). More bias was also observed. For example, when gene frequency was 0.9, the difference between genotypes was underestimated (by about 0.25) by both methods and gene frequency was underestimated at 0.8.

When two major genes with equal effect were simulated, parameter estimates were biased (table III, set 3). The difference between homozygotes was inflated by as much as 25% in the case of equal gene frequencies (0.5). Gene frequency estimates were also biased; with a simulated gene frequency of 0.1, the average estimate was around 0.15. Estimates were even more biased when the first major gene had a frequency 0.9. In that case, the mixed model gave estimates closer to 0.5 than

Table III. Finite locus data generating model.

Distribution[a]	True values	NL		L		LH	
		MM	FPMM	MM	FPMM	MM	FPMM
Set 1							
μ_{AA}	1.25	1.27 (0.28)	1.24 (0.25)				
μ_{aa}	−0.75	−0.77 (0.11)	−0.77 (0.11)				
p	0.10	0.11 (0.04)	0.11 (0.04)				
σ^2_u	0.09	0.08 (0.09)	0.09 (0.07)				
σ^2_e	1.11	1.12 (0.15)	1.24 (0.14)				
μ_{AA}	1.25	1.26 (0.15)	1.25 (0.15)				
μ_{aa}	−0.75	−0.74 (0.30)	−0.75 (0.31)				
p	0.50	0.49 (0.11)	0.49 (0.10)				
σ^2_u	0.09	0.05 (0.11)	0.10 (0.09)				
σ^2_e	1.31	1.34 (0.21)	1.31 (0.19)				
μ_{AA}	1.25	1.26 (0.06)	1.25 (0.06)				
μ_{aa}	−0.75	−0.62 (0.47)	−0.67 (0.48)				
p	0.90	0.88 (0.10)	0.88 (0.07)				
σ^2_u	0.09	0.09 (0.04)	0.09 (0.04)				
σ^2_e	0.25	0.26 (0.05)	0.26 (0.04)				
Set 2							
μ_{AA}	1.38	1.37 (0.38)	1.35 (0.32)	1.38 (0.52)	1.38 (0.42)		
μ_{aa}	−0.63	−0.67 (0.16)	−0.66 (0.16)	−0.63 (0.21)	−0.60 (0.15)		
p	0.10	0.11 (0.06)	0.11 (0.06)	0.10 (0.09)	0.09 (0.05)		
σ^2_u	0.19	0.18 (0.18)	0.20 (0.15)	0.22 (0.20)	0.23 (0.17)		
σ^2_e	1.32	1.36 (0.21)	1.35 (0.20)	1.37 (0.24)	1.37 (0.24)		
μ_{AA}	1.38	1.44 (0.20)	1.39 (0.22)	1.41 (0.21)	1.35 (0.22)	1.21 (0.23)	1.21 (0.22)
μ_{aa}	−0.63	−0.64 (0.39)	−0.68 (0.36)	−0.59 (0.44)	−0.65 (0.35)	−1.09 (0.71)	−1.14 (0.42)
p	0.50	0.48 (0.12)	0.50 (0.15)	0.48 (0.11)	0.47 (0.11)	0.78 (0.19)	0.76 (0.15)
σ^2_u	0.19	0.14 (0.19)	0.17 (0.17)	0.17 (0.21)	0.17 (0.21)	0.02 (0.05)	0.03 (0.07)
σ^2_e	1.52	1.53 (0.27)	1.54 (0.25)	1.57 (0.31)	1.57 (0.31)	2.43 (0.52)	2.42 (0.50)

Average estimates of parameters for the mixed model (MM) and for the finite polygenic mixed model (FPMM) and their standard

Table III. *Continued.*

Distribution[a]	True values	NL		L		LH	
		MM	FPMM	MM	FPMM	MM	FPMM
μ_{AA}	1.38	1.46 (0.15)	1.43 (0.14)	1.44 (0.12)	1.43 (0.12)		
μ_{aa}	−0.63	−0.27 (0.56)	−0.43 (0.57)	−0.51 (0.68)	−0.54 (0.65)		
p	0.90	0.78 (0.19)	0.81 (0.15)	0.82 (0.18)	0.83 (0.16)		
σ_u^2	0.19	0.15 (0.11)	0.17 (0.11)	0.15 (0.10)	0.15 (0.10)		
σ_e^2	0.46	0.48 (0.09)	0.48 (0.09)	0.49 (0.08)	0.49 (0.08)		
Set 3							
μ_{AA}	1.63	1.32 (0.90)	1.27 (0.56)	1.57 (1.09)	1.54 (0.81)		
μ_{aa}	0.38	0.56 (0.61)	0.46 (0.35)	0.38 (0.55)	0.36 (0.32)		
p	0.10	0.15 (0.17)	0.15 (0.10)	0.14 (0.23)	0.10 (0.06)		
σ_u^2	0.56	0.69 (0.41)	0.71 (0.37)	0.59 (0.42)	0.64 (0.45)		
σ_e^2	2.17	2.27 (0.45)	2.31 (0.44)	2.29 (0.55)	2.35 (0.47)		
μ_{AA}	1.63	1.83 (0.29)	1.77 (0.31)	1.77 (0.33)	1.72 (0.29)	1.59 (0.34)	1.52 (0.27)
μ_{aa}	−0.38	−0.75 (0.69)	−0.66 (0.53)	−0.69 (0.82)	−0.54 (0.29)	−1.20 (0.91)	−0.80 (0.40)
p	0.50	0.48 (0.17)	0.49 (0.13)	0.49 (0.16)	0.51 (0.14)	0.77 (0.21)	0.75 (0.14)
σ_u^2	0.56	0.28 (0.41)	0.32 (0.30)	0.33 (0.36)	0.38 (0.34)	0.01 (0.00)	0.02 (0.04)
σ_e^2	2.37	2.38 (0.45)	2.39 (0.39)	2.49 (0.45)	2.50 (0.59)	3.92 (0.93)	4.22 (0.74)
μ_{AA}	1.63	2.02 (0.24)	1.79 (0.28)	1.99 (0.24)	1.81 (0.25)		
μ_{aa}	−0.38	−0.16 (0.48)	−0.33 (0.26)	−0.19 (0.50)	−0.38 (0.24)		
p	0.90	0.59 (0.17)	0.75 (0.17)	0.46 (0.24)	0.75 (0.17)		
σ_u^2	0.56	0.16 (0.23)	0.42 (0.36)	0.20 (0.25)	0.39 (0.32)		
σ_e^2	1.30	1.29 (0.20)	1.35 (0.24)	1.33 (0.25)	1.36 (0.25)		

Data were generated without linkage (NL), with linkage (L), and with linkage when all parents were double heterozygotes for two loci with largest effects (LH). [a] Distribution of gene effects, see table I. See text for notations.

Table IV. Mixed data generating model.

	Set 1[a]			*Set 2*			*Set 3*		
	True values[b]	*MM*	*FPMM*	*True values*	*MM*	*FPMM*	*True values*	*MM*	*FPMM*
μ_{AA}	1.00	1.08 (0.41)	1.06 (0.34)	1.00	1.09 (0.51)	1.11 (0.46)	1.00	1.09 (0.84)	1.13 (0.66)
μ_{aa}	−1.00	−0.97 (0.12)	−0.97 (0.12)	−1.00	−1.00 (0.20)	-0.96 (0.19)	−1.00	−1.04 (0.36)	−0.99 (0.36)
p	0.10	0.10 (0.05)	0.10 (0.04)	0.10	0.11 (0.07)	0.10 (0.05)	0.10	0.11 (0.10)	0.10 (0.08)
σ^2_u	0.13	0.10 (0.09)	0.12 (0.09)	0.26	0.20 (0.17)	0.23 (0.16)	0.82	0.64 (0.46)	0.68 (0.38)
σ^2_e	1.11	1.10 (0.14)	1.13 (0.14)	1.32	1.37 (0.20)	1.37 (0.20)	2.16	2.24 (0.36)	2.24 (0.34)
μ_{AA}	1.00	0.99 (0.18)	0.98 (0.17)	1.00	1.08 (0.51)	0.98 (0.24)	1.00	1.17 (0.72)	1.12 (0.37)
μ_{aa}	−1.00	−1.00 (0.31)	−0.96 (0.34)	−1.00	−1.00 (0.20)	−0.97 (0.42)	−1.00	−1.03 (0.66)	−1.00 (0.50)
p	0.50	0.52 (0.14)	0.51 (0.12)	0.50	0.52 (0.19)	0.52 (0.14)	0.50	0.51 (0.21)	0.48 (0.16)
σ^2_u	0.13	0.10 (0.11)	0.12 (0.10)	0.26	0.21 (0.21)	0.25 (0.20)	0.82	0.62 (0.59)	0.60 (0.46)
σ^2_e	1.31	1.34 (0.20)	1.34 (0.21)	1.52	1.57 (0.23)	1.56 (0.25)	2.37	2.42 (0.36)	2.42 (0.38)
μ_{AA}	1.00	1.01 (0.07)	1.01 (0.07)	1.00	1.01 (0.11)	1.01 (0.10)	1.00	1.05 (0.23)	1.01 (0.19)
μ_{aa}	−1.00	−0.99 (0.28)	−1.16 (0.41)	−1.00	−1.00 (0.41)	−1.13 (0.60)	−1.00	−1.10 (0.62)	−1.18 (0.43)
p	0.90	0.91 (0.08)	0.91 (0.08)	0.90	0.92 (0.10)	0.91 (0.08)	0.90	0.90 (0.15)	0.89 (0.06)
σ^2_u	0.13	0.12 (0.05)	0.12 (0.05)	0.26	0.26 (0.11)	0.26 (0.11)	0.82	0.76 (0.33)	0.79 (0.32)
σ^2_e	0.25	0.25 (0.03)	0.25 (0.03)	0.46	0.46 (0.07)	0.45 (0.07)	1.30	1.28 (0.19)	1.30 (0.20)

Average estimates of parameters for the mixed model (MM) and for the finite polygenic mixed model (FPMM) and their standard deviations (in parentheses), based on 50 replicates. [a] Distribution of gene effects, see table I. [b] See text for notations.

0.9 and the finite polygenic mixed model between 0.5 and 0.9. Overestimation of differences between genotypes led to underestimation of polygenic variance, because a larger proportion of total genetic variance was attributed to variance between genotypes.

With linkage between the two loci with largest effect, a significant inflation was observed in all estimates when the linked genes were of equal size (table III, set 3). When all base population parents were double heterozygotes for the two linked loci of large effect, parameter estimates were highly biased (table III). Estimates of the difference between the two genotypes was 0.8 units higher than the true difference between the genotypes in one locus when the two loci with the largest effect on phenotype had equal effects. Also in this case, gene frequency was higher than the expected 0.5 and the estimate of additive genetic variance was almost zero. Bias in estimates of the parameters was larger for the mixed model than for the finite polygenic mixed model.

More consistent estimates over the different genetic scenarios were achieved when data were generated under the mixed model than under the finite locus model (table IV). No important differences were found between the mixed model and the finite polygenic mixed model. The variance of estimates of all parameters increased when the proportion of genetic variance explained by the major gene decreased (going from set 1 to set 3), but average values of estimates were still close to expected values.

DISCUSSION AND CONCLUSIONS

The purpose of this paper was to study the sensitivity of complex segregation analysis to violation of some of the assumptions of the underlying model, in particular a normal distribution of polygenic effects and no linkage between a major gene and polygenes. Similarity in the power of both methods of segregation analysis (the mixed model and the finite polygenic mixed model) was observed, except when data were generated based on the finite locus model with two major genes. Similar results for both methods can be expected because the computer package (SALP), which maximized the finite polygenic mixed model used equal allele frequencies (0.5) and additive gene action for all genes except the major gene, which created an approximate normal genetic distribution within major genotypes. The finite polygenic mixed model with one major locus is a closer approximation of a mixed model (Fernando et al, 1994) than an oligogenic model, which explains inheritance by a few independent loci and estimates the effect of the each locus separately (Elston and Stewart, 1971). Performance of the oligogenic model or a finite polygenic mixed model with several major loci was not studied, but might have been better than the methods studied here when data are generated from a finite number of loci.

Type I error rate was checked only for the mixed generation model and was around (or below) the expected 5%. The true type I error rate under the finite locus model is unknown. Thus, the power given in table II under the finite locus model is the probability of rejecting a pure polygenic model when the likelihood ratio test statistic is compared to the X^2 table value with two degrees of freedom.

The nature of polygenic variance (ie, the finite locus model versus the mixed-generating model) had a significant impact on power of major gene detection. In the mixed model, the polygenic component inherited by progeny has an expected value equal to the average of the polygenic values of the parents (or midparent breeding value), which is not valid if any of the genes contributing to the polygenic component are dominant. The discrepancy of progeny from the expected midparent polygenic value increases with an increase in the relative magnitude of dominant loci over all polygenic loci. In addition, with dominance, the genetic variance of offspring conditional on parental polygenotype is not equal to half of the additive genetic variance but also contains dominance variance, which is relatively large compared with additive variance when a large recessive gene with low frequency segregates in the population. These discrepancies from assumptions of the mixed model should have a negative impact on its power in cases where data were simulated under a finite locus model compared with a mixed generating model. However, no negative effect on the power was observed. Instead, in most cases the power was higher under the finite locus model than under the mixed-generating model (table II). In the case of two loci with major effect (table II, set 3) and to a lesser extent with sets 1 and 2, the methods had a chance to detect either of the major genes, which may explain the higher power under the finite locus model. In contrast, when the same situation was generated using the mixed model, a major gene explained only a small proportion of the total genetic variance, the detection of the major gene was difficult. Which of the genes was detected as a major gene under the finite locus model was not investigated, but based on intermediate estimates for gene frequency, it seems that in some families the gene from the first locus was detected as a major gene, and in other families the gene from the second locus (or other loci) was detected.

Linkage between a major gene and polygenes reduced power but did not have a large impact on parameter estimates if the linked genes were not of equal size and if the parents were a random sample from a population in linkage equilibrium. Furthermore, based on one simulation example, violation of the assumption of Hardy–Weinberg equilibrium in the parental generation reduced power substantially. Therefore, it is recommended to test a model that assumes Hardy–Weinberg equilibrium against a model with free genotypic frequencies for the parental generation.

The results given here are restricted to data from independent nucleus families. Based on results by Fernando et al (1994), the finite polygenic mixed model is a closer approximation of the mixed model under an example data set with three generations than PAP if data are generated with a mixed model. How these methods perform under the finite locus model when information from more than two generations are available or when nucleus families are not independent was not studied. Thus, the natural area for future studies is the performance of methods under multigenerational data when data are generated under the finite locus model.

In conclusion, both segregation analysis methods studied here gave similar power to detect a major gene and estimates of parameters under different genetic scenarios. The only distinguishable difference between methods was under the finite locus model when two major genes had equal effect on a trait. In that case, the mixed model (or PAP, when used as a mixed model) was more powerful than the finite

polygenic mixed model (or SALP) in rejecting the polygenic model, but the finite polygenic mixed model gave estimates with less bias than the mixed model. The finite locus model did not have a negative effect on the power compared with the mixed generating model. Instead, the power of the methods was often higher under the finite locus model than when data were generated under the mixed model. Segregation of two major genes in a population caused biased estimates. Linkage had a negative effect on the power, but parameter estimates remained unbiased if the parents were a random sample from a large population in linkage equilibrium and if the major gene had a substantially larger effect on the trait than the other genes.

ACKNOWLEDGMENTS

This research was funded by the Natural Sciences and Engineering Research Council of Canada and the Academy of Finland which are greatly acknowledged. We thank two anonymous reviewers for the comments on the paper.

REFERENCES

Borecki IB, Province MA, Rao DC (1994) Power of segregation analysis for detection of major gene effects on quantitative traits. *Genet Epidemiol* 11, 409-418

Bulmer MG (1980) *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford, UK

Demenais F, Lathrop M, Lalouel JM (1986) Robustness and power of the unified model in the analysis of quantitative measurements. *Am J Hum Genet* 38, 228-234

Elston RC, Stewart J (1971) A general model for the genetic analysis of pedigree data. *Hum Hered* 21, 523-542

Fernando RL, Stricker C, Elston RC (1994) The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theor Appl Genet* 88, 573-580

Go RCP, Elston RC, Kaplan EB (1978) Efficiency and robustness of pedigree segregation analysis. *Am J Hum Genet* 30, 28-37

Hasstedt SJ (1982) A mixed model likelihood approximation for large pedigrees. *Comput Biomed Res* 15,295-307

Hasstedt SJ (1994) PAP: Pedigree Analysis Package, Rev 4.02, Department of Human Genetics, University of Utah, Salt Lake City, UT, USA

Knott SA, Haley CS, Thompson R (1991) Methods of segregation analysis for animal breeding data: a comparison of power. *Heredity* 66, 299-311

Lalouel JM (1979) GEMINI: A computer program for optimization of general nonlinear function. Technical Report no 14, Salt Lake City, Department of Medical Biophysics and Computing, University of Utah, UT, USA

Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124, 743-756

LeRoy P, Elsen JM, Knott S (1989) Comparison of four statistical methods for detection of a major gene in a progeny test design. *Genet Sel Evol* 21, 341-357

Mackay TFC, Lyman RF, Jackson MS (1992) Effects of *P*-element insertions on quantitative traits in *Drosophila melanogaster*. *Genetics* 130, 315-332

MacLean CJ, Morton NE, Lew R (1975) Analysis of family resemblance. IV. Operational characteristics of segregation analysis. *Am J Hum Genet* 27, 365-384

MacLean CJ, Morton NE, Elston RC, Yee S (1976) Skewness in commingled distribution. *Biometrics* 32, 695-699

MacLean CJ, Morton NE, Yee S (1984) Combined analysis of genetic segregation and linkage under oligogenic model. *Comput Biomed Res* 17, 471-480

Morton NE, MacLean CJ (1974) Analysis of family resemblance. III. Complex segregation of quantitative traits. *Am J Hum Genet* 26, 489-503

Shrimpton AE, Robertson A (1988) The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II. Distribution of third chromosome bristle effects within chromosome sections. *Genetics* 118, 445-459

Stricker C, Fernando RL, Elston RC (1994) SALP: Segregation and Linkage Analysis for Pedigrees, Release 1.0, Computer Program Package. Swiss Federal Institute of Technology ETH, Institute of Animal Sciences, Zurich, Switzerland

Wilks SS (1938) The large sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat* 9, 60-62