Am. J. Hum. Genet. 56:319–326, 1995

Inferring a Major Gene for Quantitative Traits by Using Segregation Analysis with Tests on Transmission Probabilities: How Often Do We Miss?

I. B. Borecki,[1] M. A. Province,[1,2] and D. C. Rao[1,2]

[1]Division of Biostatistics and [2]Departments of Psychiatry and Genetics, Washington University School of Medicine, St. Louis

Summary

In an effort to safeguard against false inference of a major gene in segregation analysis, it has become common practice to require nonrejection of the Mendelian-transmission hypothesis (Mendelian τ's) and rejection of the no-transmission hypothesis (equal τ's). However, it is not known how often one would actually infer a major gene, when one exists, by using these criteria. A simulation study was undertaken to investigate this issue. Segregation of a Mendelian gene under a variety of models was simulated in families with both parents and three children. The data were analyzed by using POINTER; the assumptions under the generating and analysis models were identical. By design, the power to reject the no-major-effect hypothesis ($q = 0$) was >60% for all models considered; tests on the transmission probabilities were carried out only when $q = 0$ was rejected, using $\alpha = 0.05$ for all tests. The rates of Mendelian inference were mostly in the range of 22%–50% under recessive inheritance, versus 60%–99% under dominant inheritance. Notably, it was not possible to resolve the transmission (from among Mendelian τ's, equal τ's, and general unconstrained τ's) in ~20%–70% of the cases under recessive models, versus 3%–15% under dominant models. Therefore, while tests on transmission probabilities can serve to reduce rates of false inference of a major gene, it is also possible to fail to infer a major gene when one indeed exists, especially under recessive inheritance.

Introduction

Segregation analysis of human traits is a means by which genes influencing quantitative variation can be detected and, subsequently, identified and mapped. As certain traits are associated with specific diseases or increased morbidity, genetic determinants of normal variation, as well as of extreme variation, are relevant in building an understanding of the complex processes of human metabolism and physiology. While it is important to utilize safeguards against spurious evidence of monogenic inheritance, it is equally important to understand how well the analysis method performs in providing evidence of single-gene determinants when they exist.

Detection of major-gene effects in contemporary applications consists of a two-step process: (1) rejection of the null hypothesis of "no-major-effect," while minimally allowing for an alternative source of familial resemblance and, potentially, incorporating a variety of genetic or environmental effects as extensions of the basic model, and (2) performing tests on the transmission characteristics of the putative gene. The former relies primarily on phenotypic distributional information and thus is sensitive to departures from normality, while the latter focuses on transmission frequencies conditional on the presence of distributional commingling. Specifically, compatibility of the data with Mendelian expectations and rejection of the hypothesis that the major effect is not transmitted in families are usually required (Lalouel et al. 1983; Demenais et al. 1993). Studies carried out in recent years have demonstrated the utility of tests on the transmission probabilities in reducing the rate of false inference of a major gene when the data are simply skewed, for example (Go et al. 1978; Demenais et al. 1986). However, while the power to detect a major-gene effect on quantitative traits in randomly ascertained families has been investigated to some extent (MacLean et al. 1975; Borecki et al. 1994), the rate at which the major gene is corroborated by tests on the transmission probabilities has not been addressed.

Recently, we have documented estimates of the power to reject the "no-major-gene" null hypothesis, by using segregation analysis of randomly ascertained nuclear families with both parents and three children over a variety of samples sizes and model variations (Borecki et al. 1994). In the present investigation, we extend these studies to consider the proportion of cases in which the major gene is corroborated by tests on the transmission probabilities, by using simulated data. We restrict our attention to randomly ascertained families with phenotypes simulated on the basis of a simple genetic model.

Received March 16, 1994; accepted for publication September 29, 1994.

Address for correspondence and reprints: Dr. Ingrid B. Borecki, Division of Biostatistics, Washington University School of Medicine, Box 8067, 660 South Euclid Avenue, St. Louis, MO 63110.

0002-9297/95/5601-0041$02.00

Material and Methods

Major-Gene Model

A major-gene model, as parameterized in the unified mixed model (Lalouel et al. 1983) and implemented in the computer program POINTER (Lalouel and Morton 1981; Morton et al. 1983), was used to generate and analyze simulated data sets. The mixed model assumes that a phenotype is composed of the independent and additive contributions from a major effect, a multifactorial background, and a normally distributed residual. The major effect is assumed to result from the segregation at a single locus having two alleles (i.e., *A* and *a*). The total variance in the phenotype (V) can be decomposed as $V = G + C + E$, where G is the variance attributable to the major gene, C is the multifactorial variance, and E is the residual. The multifactorial effect (C) is modeled after polygenic inheritance, which generates residual correlations among family members; for a given level of heritability ($H = C/V$), the expected value of the parent-offspring and sibling correlations are the same and equal to $H/2$. There are six parameters in the model: the overall variance (V); the overall mean (u); the frequency of the allele "*a*" at the major locus leading to higher trait values (q); the displacement between the two homozygous means (t); the relative position of the heterozygote, or dominance (d); and the multifactorial heritability (H). For the present, data are generated assuming the simple cases of complete recessivity ($d = 0$, where the upper distribution includes *aa* individuals) or includes dominance ($d = 1$, where the upper distribution includes *Aa* and *aa* individuals). The three remaining relevant parameters describe the probability of transmission of the *A* allele from individuals carrying the *AA*, *Aa*, and *aa* genotypes: τ_1, τ_2, and τ_3 are 1, 1/2, and 0, respectively, under Mendelian transmission.

Simulation Procedure

Data were simulated by using Monte Carlo methods for a varying number of nuclear families composed of two parents and three offspring each. The distribution of genotypes in the parental generation was simulated under Hardy-Weinberg equilibrium, where the frequency of the "high" trait distribution is q^2 under recessive inheritance and $q^2 + 2q(1 - q)$ under dominant inheritance. Segregation of alleles into the offspring followed Mendelian transmission probabilities. Phenotypes were assigned on the basis of generating models, including genotype-specific means and residuals with a covariance structure within families specified by the heritability and assuming no spouse correlation, such that the distribution of the quantitative phenotype had an overall mean and variance of 0 and 1, respectively. Three levels of trait prevalence were considered—5%, 10%, and 20%—leading to different gene frequencies, depending on whether recessive or dominant inheritance was assumed. The displacement between the two phenotypic distributions is in units of SD relative to the total phenotypic variance, and, for each level of trait frequency, various displacements were considered, increasing in units of 0.25. For each combination of parameters for the major gene, two levels of background polygenic variation were simulated: none and moderate ($H = 30\%$). A minimum of 50 families were entertained, up to a maximum of 300. Choice of the specific experimental points was guided by our previous study (Borecki et al. 1994), wherein the power to detect a major gene solely on the basis of rejection of the $q = 0$ null hypothesis was evaluated. For each set of generating parameter values for the major gene, the sample size (in terms of number of families) yielding as close as possible to 90% power to reject $q = 0$ was chosen. In addition, the next smaller sample size also was evaluated, usually representing a decrement of 50 families.

Analysis of a simulated data set was carried out by using the computer program POINTER. Parameters are estimated by the method of maximum likelihood, calculating the joint probability of observing the offspring phenotypes and parental phenotypes (i.e., joint likelihood). Four separate hypotheses were fitted for each simulated data set: (1) the Mendelian mixed model estimating V, u, d, t, q, and H (with τ_1, τ_2, and τ_3 fixed at their Mendelian values); (2) the null hypothesis of no major effect, estimating V, u, and H (with $d = t = q = 0$); (3) the general transmission model, estimating V, u, d, t, q, and H and τ_1, τ_2, and τ_3; and (4) the no-transmission model, estimating V, u, d, t, q, H, and one common τ (i.e., $\tau_1 = \tau_2 = \tau_3$). Each hypothesis was tested, using the likelihood-ratio criterion, which is minus twice the difference between the log-likelihood obtained under a more general model and that under a reduced model. The likelihood ratio is asymptotically distributed as a χ^2, with df equal to the number of independent parameter constraints. First, the hypothesis of no major effect was tested (model 2 vs. model 1), conservatively assuming 3 df. Only in those samples in which this null hypothesis is rejected were additional tests on the transmission probabilities carried out, namely, compatibility with Mendelian transmission (model 1 vs. model 3; 3 df), and tenability of the no-transmission hypothesis (model 4 vs. model 3; 2 df). A Mendelian inference was made if the Mendelian hypothesis was *not* rejected and the no-transmission hypothesis *was* rejected; all inferences were based on a 5% level of significance. For each experimental condition, the proportion of replications with evidence for a major gene in which the Mendelian inference is made is called the *rate* of Mendelian inference.

The reported rates of Mendelian inference are not the power estimates, per se, since the Mendelian inference involves two components, each of which is subject to a different type of error. If we define α as the type I error rate associated with the test of Mendelian transmission and define β as the type II error rate associated with the test of no

Table 1

Power to Reject $q = 0$ under Recessive Inheritance

No. of Families	Trait Frequency = 5% (q = .224)								Trait Frequency = 10% (q = .316)							
	t = 1.75 (%V_{mg} = 14.6)		t = 2.0 (%V_{mg} = 19.1)		t = 2.25 (%V_{mg} = 24.1)		t = 2.5 (%V_{mg} = 29.8)		t = 1.5 (%V_{mg} = 20.2)		t = 1.75 (%V_{mg} = 27.5)		t = 2.0 (%V_{mg} = 36.0)		Trait Frequency = 20% (q = .447); t = 1.5 (%V_{mg} = 36.0)	
	$H = 0$	$H = .3$	$H = 0$	$H = .3$	$H = 0$	$H = .3$	$H = 0$	$H = .3$	$H = 0$	$H = .3$	$H = 0$	$H = .3$	$H = 0$	$H = .3$	$H = 0$	$H = .3$
50					.594 (202)	.667 (204)	.847 (203)	.883 (205)				.444 (205)	.855 (200)	.856 (215)		
100			.624 (202)	.660 (203)	.955 (200)	.921 (202)					.822 (202)	.778 (203)			.861 (201)	.776 (214)
150			.905 (201)	.835 (206)							.920 (200)	.935 (200)				.923 (209)
200	.629 (202)	.642 (201)							.749 (203)	.659 (205)						
300	.860 (200)	.841 (207)							.910 (201)	.812 (202)						

NOTE.—Shown in parentheses after power estimates is number of valid replications.

transmission, the combined rate of Mendelian inference is $(1 - \alpha) \times (1 - \beta)$. Another complication is that we have restricted tests on the transmission probabilities to cases where the primary null hypothesis of no major effect ($q = 0$) is rejected, as is usually done in practice. Thus, the results presented here should be taken as a conditional rate of Mendelian inference reflecting a mixture of type I and type II errors. However, this conditional rate can be multiplied by the power to reject $q = 0$ for any experimental condition to obtain an estimate of the unconditional rate.

Occasional numerical difficulties with model fitting were encountered, especially with respect to the general transmission model. A strategy was developed to maximize the number of converged solutions (according to usual numerical criteria), by utilizing restarts and perturbations of the starting values. Whenever a converged solution was not achieved, the iterative process was restarted, up to three times, using the most recent estimates from the previous stalled run as initial parameter values. Boundary solutions were accepted as they occurred, since our previous experience suggested that, indeed, the maximum likelihood did occur at the boundary in the expected cases (e.g., when data generated under $H = 0$ were analyzed, the heritability estimate often moved toward the zero boundary). Difficulties in obtaining interior solutions with the parameters describing the major gene effect were not encountered, with the exception of the dominance parameter, since we chose experimental conditions and sample sizes where there was good power to detect the resulting commingling. In order to count any particular replication as valid, we required convergence for each of the four hypotheses; additional replications were simulated and analyzed until a minimum of 200 replications, all with converged solutions, were obtained. In general, a greater number of replications are used to assess the $q = 0$ hypothesis than the transmission hypotheses, since, for the former, valid solutions are required only for the respective null hypothesis (model 2, described above) and for the Mendelian mixed model (model 1). The number of replications used to assess the rate of Mendelian inference is restricted, first, by the presence of invalid solutions for any of the four hypotheses or numerical anomalies in obtaining the likelihood ratios and, second, by the condition that $q = 0$ is rejected.

Results

The proportion of replications in which the null hypothesis $q = 0$ was rejected for each experiment under recessive models is shown in table 1. As was the case with our previous study (Borecki et al. 1994), there were minimal problems fitting the relevant models for this test, with usually none or <1% of the replications being invalid. Since an independent series of seeds was utilized for this investigation, there was some variation in the estimate of the power, as compared with our previous study, although the rates are quite comparable.

In the cases where the no-major-effect hypothesis ($q = 0$) was rejected, we determined whether a Mendelian gene would be supported by tests on the transmission probabilities involving nonrejection of Mendelian segregation and rejection of no transmission of the major effect. The conditional rates of Mendelian inference, conditional on evidence of a major effect, are shown in table 2. In general, the proportion of cases in which a Mendelian gene is inferred increases with sample size and effect size of the major gene. Overall, the rates are rather low. For example, in those cases where the frequency of the recessive homozygote is low (5%), the Mendelian inference is achieved in only 25%–35% of the samples drawn under conditions characterized by ~84%–95% power to detect the major effect. The rates are generally lower (mostly, 22%–26%) under conditions characterized by lower power to detect the major-gene effect (59%–66%), albeit the tests on transmission probabilities are conditional on evidence of the major effect. As the homozygous recessive genotype frequency doubles to 10%, the rates of Mendelian inference improve to 33%–60% for those conditions characterized by >80% power to detect the major effect. The highest

Table 2

Rate of Mendelian Inference under Recessive Models for the Set of Replications in Which $q = 0$ Was Rejected

No. of Families	Trait Frequency = 5% (q = .224)								Trait Frequency = 10% (q = .316)						Trait Frequency = 20% (q = .447); t = 1.5 ($\%V_{mg}$ = 36.0)	
	t = 1.75 ($\%V_{mg}$ = 14.6)		t = 2.0 ($\%V_{mg}$ = 19.1)		t = 2.25 ($\%V_{mg}$ = 24.1)		t = 2.5 ($\%V_{mg}$ = 29.8)		t = 1.5 ($\%V_{mg}$ = 20.2)		t = 1.75 ($\%V_{mg}$ = 27.5)		t = 2.0 ($\%V_{mg}$ = 36.0)			
	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3
50					.250 (120)	.236 (127)	.267 (165)	.251 (167)				.299 (87)	.447 (170)	.335 (158)		
100			.220 (123)	.258 (124)	.364 (187)	.282 (174)					.444 (160)	.358 (151)			.725 (171)	.613 (150)
150			.320 (172)	.301 (163)							.511 (184)	.464 (181)				.593 (182)
200	.387 (119)	.228 (123)							.460 (150)	.389 (131)						
300	.339 (168)	.331 (166)							.592 (179)	.394 (160)						

NOTE.—Shown in parentheses after power estimates is number of valid replications.

rates are associated with the highest gene frequency investigated in this study.

Since in an appreciable number of cases the tests on transmission probabilities did not support the Mendelian inference under recessive inheritance, the question remains, what conclusion would be drawn? As shown in table 3, it was not possible to resolve the transmission for the bulk of the remaining replications, that is, neither the Mendelian nor the no-transmission hypotheses could be rejected. While these rates of nonresolution are comparatively modest for models assuming a 20% trait frequency, there was an apparent lack of statistical information to distinguish among the transmission hypotheses in 27%–69% of the cases under various other models with lower trait frequencies. The inference that there was no transmission of the major effect (i.e., rejection of Mendelian transmission and nonrejection of no transmission) was supported, on average, ~8% of the time (range 1%–13%), and both transmission hypotheses were rejected in the remaining small number of instances.

The power to detect a major effect under dominant models is reported in table 4. As noted elsewhere (Borecki et al. 1994), the power to detect major gene effects is greater for dominant models, as compared with recessive models, and the presence of a polygenic background is associated with a decrease in power for all dominant models, with not much discernible effect in recessive models. In our previous study, we suggested that the greater power to detect dominant genes must be attributable to actual segregation information, since the comparisons can be matched precisely on the characteristics of the commingling information. This supposition is strongly supported by the results shown in table 5, where the rate of Mendelian inference is >80% in all but one instance and is generally >90% for those combinations of parameter values and sample sizes yielding power to detect a major effect in the 85%–95% range. Inference of no transmission of major effect was rare and, for many conditions, was not the conclusion in any replicate. No conclusion could be drawn regarding the transmission patterns in ~3%–15% of the replications (i.e., nonrejection of both transmission hypotheses), and both transmission hypotheses were rejected in the remaining few cases. Thus, in general, it would appear that there is a much better chance to detect and infer a dominant Mendelian gene influencing a quantitative trait than a recessive one.

Table 3

Proportion of Replications under Recessive Inheritance in Which Neither Mendelian Nor No-Transmission-of-Major-Effect Hypotheses Could Be Rejected

No. of Families	Trait Frequency = 5% (q = .224)								Trait Frequency = 10% (q = .316)						Trait Frequency = 20% (q = .447); t = 1.5 ($\%V_{mg}$ = 36.0)	
	t = 1.75 ($\%V_{mg}$ = 14.6)		t = 2.0 ($\%V_{mg}$ = 19.1)		t = 2.25 ($\%V_{mg}$ = 24.1)		t = 2.5 ($\%V_{mg}$ = 29.8)		t = 1.5 ($\%V_{mg}$ = 20.2)		t = 1.75 ($\%V_{mg}$ = 27.5)		t = 2.0 ($\%V_{mg}$ = 36.0)			
	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3
50					.667 (120)	.685 (127)	.576 (165)	.563 (166)				.598 (87)	.465 (170)	.532 (158)		
100			.667 (123)	.677 (124)	.471 (187)	.557 (174)					.406 (160)	.490 (151)			.170 (171)	.213 (150)
150			.506 (172)	.528 (163)							.272 (184)	.348 (181)				.148 (182)
200	.504 (119)	.642 (123)							.460 (150)	.473 (131)						
300	.482 (168)	.512 (166)							.274 (179)	.350 (160)						

NOTE.—Shown in parentheses after power estimates is number of valid replications.

Table 4

Power to Reject $q = 0$, under Dominant Inheritance

No. of Families	Trait Frequency = 5% (q = .025)								Trait Frequency = 10% (q = .051)				Trait Frequency = 20% (q = .106); t = 1.5 (%V_{mg} = 36.1)	
	t = 1.75 (%V_{mg} = 14.4)		t = 2.0 (%V_{mg} = 18.8)		t = 2.25 (%V_{mg} = 23.8)		t = 2.5 (%V_{mg} = 29.3)		t = 1.5 (%V_{mg} = 20.1)		t = 1.75 (%V_{mg} = 27.4)			
	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3
50			.641 (234)	.467 (212)	.814 (236)	.735 (211)	.938 (209)	.877 (204)			.764 (212)	.608 (212)	.743 (202)	.617 (222)
100			.892 (223)	.808 (219)	.983 (230)	.952 (208)					.995 (208)	.848 (204)	.990 (204)	.916 (215)
150	.858 (211)	.635 (219)							.883 (206)	.707 (208)				
200	.947 (209)	.763 (224)							.960 (202)	.852 (210)				
300		.908 (218)												

NOTE.—Shown in parentheses after power estimates is number of valid replications.

As might be expected, some numerical difficulties were encountered in fitting the relevant transmission hypotheses and in carrying out the respective likelihood-ratio tests. A particular replication was unacceptable—and therefore the generated sample was rejected—if either of the following situations pertained: (*a*) any one of the four hypotheses did not converge owing to numerical difficulties, or (*b*) construction of the likelihood-ratio-test statistic resulted in a negative χ^2, after allowance was made for some tolerance for numerical variations. Of course, in practice, one would attempt to resolve both of these situations by altering the parameters of the numerical algorithm or by choosing alternative initial parameter estimates on the basis of interactions with the data or grid searches. However, in the context of a simulation study, it is extremely difficult to develop an algorithmic approach that recognizes all of the problems that could arise and that responds in the correct manner. Therefore, we took best advantage of what information was available to us to maximize the return of admissible solutions at the global maximum—namely, that the exact model and generating parameters are known. Despite our use of excellent initial values, some proportion of samples were rejected.

The mean sample-rejection rate for tests on the transmission probabilities over all the recessive models we considered was 4.3%, with a range of 0%–10.7% and a distribution skewed to the right. Under otherwise identical conditions, sample-rejection rates were higher for generating models having a polygenic background than for those with no residual familial resemblance. Considering the complexity of the hypotheses being fit, we consider this to be a tolerable rate.

In contrast, sample-rejection rates were much higher for dominant models—an average of 21.5% of the replications were excluded for possible reasons cited above, with a range of 3.1%–40.7%. In order to investigate this situation further, we chose to examine an experiment in detail—one with good power to detect the major-gene effect (.908) and with a large number of families and, thus, good segregation information. We refer to the experiment for a 5% trait, t = 1.75 and H = 0.3; the following results are fairly representative of those for other experiments with substantial sample-rejection rates. Of 209 replications in which it was possible to reject the q = 0 hypothesis, 69 (33%) of those samples were rejected, and the problem was with the general transmission model. There were numerical difficulties in the iterative process in approximately one-quarter of these rejected replications, resulting in non-

Table 5

Rate of Mendelian Inference under Dominant Models for Set of Replications in Which $q = 0$ was Rejected

No. of Families	Trait Frequency = 5% (q = .025)								Trait Frequency = 10% (q = .051)				Trait Frequency = 20% (q = .106); t = 1.5 (%V_{mg} = 36.1)	
	t = 1.75 (%V_{mg} = 14.4)		t = 2.0 (%V_{mg} = 18.8)		t = 2.25 (%V_{mg} = 23.8)		t = 2.5 (%V_{mg} = 29.3)		t = 1.5 (%V_{mg} = 20.1)		t = 1.75 (%V_{mg} = 27.4)			
	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3	H = 0	H = .3
50			.830 (112)	.780 (82)	.843 (140)	.893 (122)	.936 (172)	.943 (158)			.850 (147)	.838 (105)	.869 (145)	.883 (111)
100			.891 (147)	.940 (133)	.988 (166)	.974 (151)	.977 (171)				.943 (193)	.934 (152)	.938 (195)	.902 (174)
150	.864 (147)	.840 (100)							.919 (173)	.933 (135)				
200	.926 (162)	.936 (125)							.936 (188)	.956 (160)				
300		.943 (140)												

NOTE.—Shown in parentheses after power estimates is number of valid replications.

converged solutions. However, in three-quarters of these cases, all solutions were apparently converged, but negative χ^2s were encountered. Specifically, the Mendelian model had a better likelihood than did the general transmission model in which all three transmission probabilities were estimated. These particular data sets were examined by hand, and, in all these cases, another solution could be found for the general transmission model with a better likelihood, resulting in a Mendelian inference. This experience suggests that the presence of multiple maxima may be a relatively common phenomenon under the general transmission model.

It was difficult to assess from these data whether the estimated rates of Mendelian inference were biased in some way on account of exclusion of the faulty replications, since the observed rate of Mendelian inference was high at 94%, and, indeed, most of the faults would resolve to that conclusion. Of greatest concern was the effect of multiple maxima, since this problem accounted for the majority of rejected replications under dominant models. Thus, the question becomes, What would be the distribution of outcomes if the rate of sample rejection could be reduced? Further, multiple maxima also could potentially bias our results under recessive models, even if there are no outward signs of problems in convergence or negative χ^2s. Recall that the results under recessive models were characterized by moderately low rates of Mendelian inference and by moderately high rates of nonresolution of transmission hypotheses—if there is another solution with a better likelihood for the general transmission model in some proportion of cases, then one might expect a decrease in the number of replications falling into the latter category and an increase in any of the three remaining possible inference categories, depending on the magnitude of the improvement.

In an attempt to assess the potential direction and magnitude of bias, if any, we reexamined four experiments under each of dominant and recessive inheritance, using an alternative strategy to evaluating transmission hypotheses. We preferentially included experiments that had high sample-rejection rates under dominant inheritance, as follows: (I) 5% trait, $t = 1.75$, $H = .3$, $N = 200$ families; (II) 5% trait, $t = 2.25$, $H = 0$, $N = 50$ families; (III) 5% trait, $t = 2.25$, $H = .3$, $N = 50$ families; and (IV) 10% trait, $t = 1.75$, $H = .3$, $N = 50$ families.

The corresponding models under recessive inheritance were evaluated also. The idea of the alternative approach was that reducing the dimensionality of the general transmission model would improve the chances of obtaining converged solutions at the global maximum. Thus, we restricted our attention to the specific Mendelian hypothesis of $\tau_2 = 1/2$, as originally suggested by Lalouel et al. (1983), since most information on transmission comes from segregation from heterozygotes anyway. Now the test of the Mendelian hypothesis is a 1 df test where τ_2 alone is estimated in the general transmission model, with τ_1 and τ_3 fixed at 1 and 0, respectively. The approach to testing the no-transmission model also had to be modified, since the equal-τ model is no longer nested. The comparison was carried out using the Akaike (1974) information criterion; since the two models have the same number of estimated parameters, the likelihoods were compared directly: if the modified general transmission model was ≥ 0.5 points better (on the scale of twice the log likelihood), then the no-transmission hypothesis was rejected.

The results are shown in table 6. For reference, the power to reject the null hypothesis of $q = 0$ is given for each experiment. First, it is apparent that, in all but one case, the number of valid replications ("valid" being defined as acceptably converged solutions with positive χ^2s for the likelihood-ratio tests), of the total in which $q = 0$ was rejected, increases with the alternative approach of restricting our attention to τ_2 alone, especially under dominant models. Consistent with the predictions outlined above, the rate of Mendelian inference under recessive models increased from 23%–36% to 46%–54%, and corresponding drops in the rate of ambiguous situations where neither transmission hypothesis could be rejected and modest increases in the rates of rejection of both transmission hypotheses were found. The results under dominant inheritance appear to be more robust: slight decreases in the rates of Mendelian inference are accompanied by small increases in the rates of rejecting both transmission hypotheses.

Discussion

Historically, the use of transmission frequencies has been part of segregation analysis from its inception, originally, in the form of segregation ratios for fully penetrant dichotomous traits. Extensions to accommodate quantitative traits also modeled the segregation of alleles under Mendelian assumptions, and it was Elston and Stewart (1971) who first suggested that tests on the transmission probabilities (τ's) should be required in addition to evidence of admixture before concluding the existence of a major gene. Go et al. (1978) investigated the properties of segregation analysis under the generalized single-locus model, specifically addressing issues of robustness of the model against false inference of a major gene in the presence of polygenic inheritance with distributional skewness and kurtosis, and environmentally based sibling correlation. Go et al. concluded that inclusion of the transmission probabilities in the model allows one to distinguish between the Mendelian hypothesis and environmentally caused skewness. Evidence of an admixture of distributions was used to distinguish between polygenic and single-gene inheritance; however, while it was expected that the transmission probabilities would be consistent with the Mendelian hypothesis when the true underlying mech-

Table 6

Comparison of Results, Using Alternative Strategies for Testing Transmission Models

Experiment	Power ($q = 0$)	Rate of Mendelian Transmission		Rate of General Transmission		Rate of Ambiguous Transmission	
		τ_1, τ_2, τ_3 estimated	τ_2 estimated	τ_1, τ_2, τ_3 estimated	τ_2 estimated	τ_1, τ_2, τ_3 estimated	τ_2 estimated
Recessive:							
I	.642	.228 (123/129)	.461 (128/129)	.033	.078	.642	.414
II	.955	.364 (187/190)	.487 (189/190)	.070	.079	.471	.407
III	.921	.282 (174/186)	.462 (184/186)	.029	.082	.557	.386
IV	.444	.299 (87/91)	.544 (90/91)	.026	.111	.598	.311
Dominant:							
I	.763	.936 (125/171)	.914 (162/171)	.008	.086	.056	0
II	.983	.988 (166/226)	.870 (208/226)	0	.130	.012	0
III	.952	.974 (151/198)	.907 (193/198)	.007	.093	.020	0
IV	.608	.838 (105/129)	.728 (81/129)	.019	.259	.143	.012

NOTE.—Data are the proportion of replications in which either of Mendelian, general, or ambiguous transmission was inferred, followed in parentheses by (no. of valid replications)/(total no. of replications in which $q = 0$ was rejected).

anism was a major gene, the issue of the "power" of this test was addressed only cursorily: it was suggested that examination of the point estimates of the transmission probabilities would be useful in indicating that the likelihood was flat in those dimensions if the estimates were inexplicably different from the Mendelian values with little loss of fit. Caution in concluding a major gene effect in this case was advised, because the focus was very much on minimizing the type I error.

In the meanwhile, segregation analysis under the so-called mixed model, developed by Morton and MacLean (1974), incorporated a multifactorial background and sibling correlation, in addition to a single locus. While specifically allowing for these effects improved the robustness of the model in some circumstances, accumulating experience suggested that other unaccounted determinants or sources of variation could cause distributional non-normality or false evidence of a major gene. Thus, transmission probabilities were introduced by Lalouel et al. (1983) into the "unified model," to aid in distinguishing true Mendelian effects. Simulation studies carried out by Demenais et al. (1986) again demonstrated dramatically the utility of tests on the transmission probabilities, in preventing false inference of a major gene, for data simulated under a polygenic model with skewness: the frequency of false inference of a major gene was reduced from 80%–100% of the simulations to 10%–40%. Moreover, these authors showed that the conservative approach of applying a power transformation to ameliorate skewness was associated with a decrease of 55% in the power to detect a major gene. On balance, these observations led to the important recommendation that, to prevent false inference, it may be preferable to rely on tests on the transmission probabilities, including both nonrejection of the Mendelian hypothesis and rejection of the hypothesis of no transmission of major effect. Still, the emphasis was on controlling type I error.

Many of the quantitative phenotypes of interest to investigators in the present are likely to be etiologically complex, and their analysis poses new challenges. Application of traditional segregation analysis still represents a reasonable first approach to such traits; however, it may be the case that the simple mixed model is not sufficient to explain the observed familial resemblance. In at least two cases, modification of the basic segregation model to include genotype-specific age and gender effects resulted in resolution of Mendelian-segregation patterns for both a putative gene for systolic blood pressure (Pérusse et al. 1991) and another for the body-mass index (Borecki et al. 1993), suggesting that transmission probabilities are sensitive to violations of model assumptions. Further, it has been demonstrated that even one paternity exclusion can be responsible for strong rejection of Mendelian transmission probabilities for a patently Mendelian trait (Bonaïti-Pellié et al. 1992). Thus, the question arises: how often do we fail to detect a true major gene because the tests on the transmission patterns do not satisfy the stated requirements?

Before considering other variations either of the model or of the study design, we endeavored to establish a baseline rate referring only to the effects of sampling variation; for this task, we chose to investigate randomly sampled nuclear families. Indeed, when traits are common (as considered here), it becomes unnecessary to enrich the sample by selecting through clusters of affected individuals, and, insofar as segregation analysis is concerned, an appropriate ascertainment correction is difficult to specify under many high-density sampling schemes. Greenberg (1992) also has

suggested that there may be substantive advantages to the study of nuclear families (as opposed to the relatively less common multiplex pedigrees), even for linkage studies, since nuclear families (i) are readily available, (ii) usually represent the common form of the disease or trait, and (iii) can provide information on heterogeneity by subdivision of the sample on the basis of clinical presentation or criteria. Therefore, although we considered randomly sampled nuclear families as a simple starting point, this type of sample also can have considerable appeal in certain circumstances.

The studies conducted here are not exhaustive, but they suggest that dominant major genes are detected quite successfully under a variety of models and that the conditional rate of Mendelian inference is usually >80% in the cases we examined. Those sets of conditions with greater power to reject the $q = 0$ null hypothesis also are characterized by higher rates of Mendelian inference. These general conclusions hold even after our approach to hypothesis testing was modified to examine τ_2 alone, although it is possible that the "true" rates of Mendelian inference may be slightly lower than those indicated in table 5. In contrast, the situation appears much worse for detecting recessive major loci. The rate of Mendelian inference is quite low, suggesting that this methodology may not be very efficient for detecting recessive major loci—a large part of the remaining replications were characterized by nonresolution of transmission patterns. Even after reduction of the number of solutions at local maxima, the rate of Mendelian inference increased from ~20%–30% to only 46%–54%. It should be noted that the "true" rates of Mendelian inference may be higher than those indicated in table 2 on account of undetected multiple maxima; however, the estimated rates in our modified approach are still sufficiently low to warrant concern. Slight differences were noted between models without and with a moderate polygenic background under recessive inheritance only: presence of a polygenic component was associated both with slightly lower rates of Mendelian inference and with slightly higher rates of ambiguous transmission (where neither transmission hypothesis could be rejected).

In summary, tests on the transmission probabilities in segregation analysis of quantitative traits remain useful in safeguarding against false inference of major genes in a variety of circumstances. However, the present study shows that, by adhering strictly to those tests, quite often one will fail to correctly infer a major gene and that the error is more frequent for recessive than for dominant traits, largely on account of a lack of segregation information. It should be remembered that these conclusions apply to samples of nuclear families as considered here, and it is not known whether the conclusions apply to extended pedigrees. In general, resolution of Mendelian segregation may be sought by increasing the sample size, extending the basic model to include interactions, utilizing a linked marker, or simultaneously analyzing another (other) quantitative trait(s) influenced by the same underlying gene. For the present, caution should be exercised in rejecting the Mendelian inference in cases indicating recessive inheritance.

Acknowledgments

The authors gratefully acknowledge the advice and assistance of Dr. Reimut Wette in the implementation of the simulation module. This work was partly supported by NIH grants GM-28719, 1 RO1 HL47317, and NOI-HC-25108 and by Medical Research Council of Canada grant PG-11811.

References

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automatic Control AC 19:716–723

Bonaïti-Pellié C, Poisson N, Bechtel Y, Bechtel P (1992) Sensitivity of transmission probabilities to paternity exclusion in segregation analysis. Genet Epidemiol 9:67–71

Borecki IB, Bonney GE, Rice T, Bouchard C, Rao DC (1993) Influence of genotype-dependent effects of covariates on the outcome of segregation analysis of the body mass index. Am J Hum Genet 53:676–687

Borecki IB, Province MA, Rao DC (1994) Power to detect major gene effects for quantitative traits by using segregation analysis. Genet Epidemiol 11:409–418

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

Demenais F, Martinez M, Andrieu N (1993) The transmission probability model is useful to prevent false inference. Am J Hum Genet 52:441–442

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

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

Greenberg D (1992) There is more than one way to collect data for linkage analysis. Arch Gen Psychiatry 49:745–750

Lalouel JM, Morton NE (1981) Complex segregation analysis with pointers. Hum Hered 31:312–321

Lalouel JM, Rao DC, Morton NE, Elston RC (1983) A unified model for complex segregation analysis. Am J Hum Genet 35: 816–826

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

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

Morton NE, Rao DC, Lalouel JM (1983) Methods in genetic epidemiology. Karger, Basel

Pérusse L, Moll PP, Sing CF (1991) Evidence that a single gene with gender- and age-dependent effects influences systolic blood pressure determination in a population-based sample. Am J Hum Genet 49:94–105