Am. J. Hum. Genet. 46:994–999, 1990

Simulation of Mendelism Revisited: The Recessive Gene for Attending Medical School

Peter McGuffin and Philip Huckle

Department of Psychological Medicine, University of Wales College of Medicine, Cardiff

Summary

Much of the recent confusion concerning studies of complex phenotypes such as neuropsychiatric disorders may derive from the inappropriate assumption of simple Mendelian transmission. This has sometimes led to unrealistic expectations regarding the potential benefits of linkage studies. To investigate how Mendelism may be simulated, we collected data on a common familial behavioral trait, attendance at medical school, among the relatives of 249 preclinical medical students. The "risk" of first-degree relatives going to medical school was approximately 61 times that of the general population. Complex segregation analysis carried out under a unified model provided strong evidence of vertical transmission. The results were compatible with transmission of a major effect, and a recessive model provided as satisfactory a fit as a general single-locus model. Moreover, a commonly applied test, allowing the transmission probability parameter (τ_2) to deviate from its Mendelian value, did not give a significant improvement of fit. Only a more general model where all three transmission probabilities (τ_1, τ_2, and τ_3) were unrestricted resulted in a significantly better fit than did the recessive model.

Introduction

Familial aggregation is a common feature of many diseases and developmental abnormalities as well as of many normal human traits and characteristics. Such aggregation may be the product of shared genes, shared environment, or a mixture of the two. It is well established that, in certain inherited characteristics in animals measured on a continuous scale, the genetic contribution derives from the additive effects of multiple genes at different loci and that the same is also true of other characteristics which appear discontinuous or quasi-continuous (Gruneberg 1952). It is likely that similar mechanisms operate in some human diseases where liability to develop the disorder is not continuously distributed in the population and is due to the mainly additive combination of multiple genes and environmental effects. Only those individuals whose liability at some stage exceeds a certain threshold become affected (Falconer 1965). When familial correlations in liability are high and the disorder is common, the resultant pattern of inheritance may simulate that expected for single-factor inheritance (Edwards 1960). This may be particularly important where the phenotype is poorly defined, so that the inclusion of mild, uncertain, or borderline cases may lead to a classical segregation ratio which is absent when only definite cases are considered. Similarly, a deliberate selection only of "loaded" pedigrees can spuriously provide a "Mendelian appearance."

There is undoubtedly a genetic contribution to the familial aggregation of neuropsychiatric disorders such as schizophrenia (Gottesman et al. 1987), manic depressive illness (McGuffin 1988), and probably Alzheimer disease (Kay 1989). However, in none of these conditions is a simple Mendelian pattern of segregation usual. Therefore, with regard to the mode of transmission, researchers performing linkage studies are forced to make assumptions which may not be correct. The usual strategy is to concentrate on multiply affected families in which transmission is assumed to be via an incom-

Received September 27, 1989; final revision received January 18, 1990.

Address for correspondence and reprints: Peter McGuffin, M.B., Ph.D., Department of Psychological Medicine, University of Wales College of Medicine, Heath Park, Cardiff, South Wales CF4 4XN, United Kingdom.

0002-9297/90/4605-0020$02.00

pletely penetrant dominant gene. This has led to a perplexing pattern of results (Barnes 1989). There have been reports of an X-chromosome gene locus linked to a gene for major depressive disorders (Baron et al. 1987; Mendlewicz et al 1987), to a gene for familial Alzheimer disease on chromosome 21 (St. George-Hyslop et al. 1987), to a gene predisposing to manic depression on chromosome 11 (Egeland et al. 1987), and to a susceptibility locus for schizophrenia on chromosome 5 (Sherrington et al. 1988). Unfortunately, other researchers have either failed to replicate these findings or have even effectively excluded such linkages in their family material (Gershon et al. 1980; Detera-Wadleigh et al. 1987; Hodgkinson et al. 1987; Kennedy et al. 1988; Schellenberg et al. 1988). The one promising exception so far concerns early-onset Alzheimer-type dementia, for which a recent report provides independent replication of linkage with chromosome 21q markers and gives a plausible resolution of the apparent disagreement between earlier publications (Goate et al. 1989).

The most popular explanation of disparate or contradictory findings is that most common major mental disorders are genetically heterogeneous. However, the fundamental methodological issue—i.e., that likelihood methods of linkage analysis (Morton 1955; Ott 1985) require that the mode of transmission of both the main trait and the marker trait be known—cannot be overlooked. If incorrect assumptions are made about the main trait, then incorrect results may be produced. Recent studies show that misspecification of the mode of transmission (or random misclassification of affecteds) is not likely to lead to spurious detection of linkage (Clerget-Darpoux et al. 1986; Greenberg and Hodge 1989). Nevertheless, the search for major genes for mental disorders is premised not on any strong evidence that such genes exist but, rather, on the absence of any compelling evidence against their existence. A salutary example of how a rather different behavioral trait may stimulate Mendelism was provided by Lilienfield (1959), who showed, using a simple binomial test, that the familial distribution of attending medical school was consistent with autosomal recessive inheritance. We set out to determine whether we could replicate this finding by using modern, more powerful methods of complex segregation analysis.

Methods

First- and second-year medical students at the University of Wales College of Medicine were asked to complete a detailed questionnaire concerning the higher educational attainments of their first- and second-degree relatives. Two hundred forty-nine students (85% of the sample) provided complete data. Table 1 shows the percentage of relatives over the age of 18 who had attended medical school. The overall percentage of first-degree relatives attending medical school was 13.4%, compared with approximately 0.22% of the general population (Medical Directory 1988), giving a relative risk of 61.

Table 1

Frequency of Attending Medical School in Adult Relatives of Medical Students

Category of Adult Relatives (*N*)	Attending Medical School (%)
Fathers (249)	16.1
Mothers (249)	6.0
Siblings (137)	21.9
Grandparents (598)	2.8
Uncles/aunts (1,313)	2.1

Segregation analysis was performed on the nuclear families of our students by using the computer program POINTER and by applying a unified model approach (Lalouel et al. 1984), so called because it unifies the Morton and MacLean (1974) mixed model with the Elston and Stewart (1971) concept of transmission probabilities. Under the mixed model (fig. 1) it is assumed that a trait x results from the additive contributions of a major transmissible effect g, a multifactorial transmissible effect c, and a random nontransmitted environ-

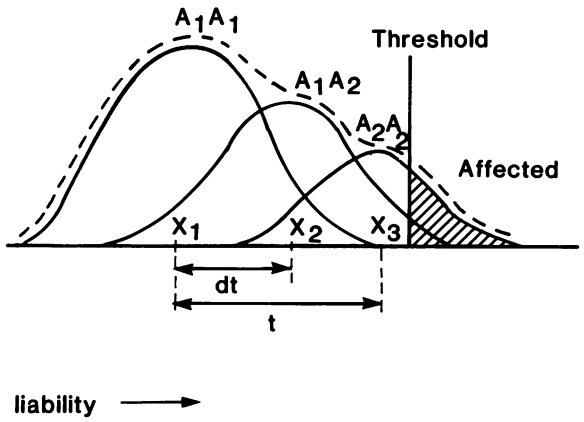

Figure 1 Mixed model. The three genotypes A_1A_1, A_1A_2, and A_2A_2 differ in their mean liabilities as shown. The multifactorial "background" is represented by the broken line. Only individuals whose liability exceeds the threshold (shaded area) exhibit the trait.

mental component e, such that $x = g + c + e$. The major effect results from segregation at a single locus of two alleles A_1, A_2, leading to three genotypes with mean liabilities of exhibiting the trait of x_1, x_2, and x_3 as shown in the figure. If it is assumed that liability has both an overall mean of zero and a unit variance, the major gene effect is parameterized in terms of the following: q, the gene frequency of A_2; t, the displacement between x_1 and x_3; and d, a dominance parameter. The transmissible multifactorial component has a single parameter H, the proportion of total variance in liability contributed by multifactorial transmission. Under the unified model, random mating is assumed, and the parameters τ_1, τ_2, and τ_3 denote the probabilities of transmitting allele A_1 for genotypes A_1A_1, A_1A_2, and A_2A_2, respectively. The Mendelian expectations are $\tau_1 = 1$, $\tau_2 = \frac{1}{2}$, $\tau_3 = 0$. Pairwise comparison of models can be carried out using the likelihood ratio test where minus twice the difference in log likelihoods is asymptotically a χ^2 with df equal to the difference in the number of parameters (Cavalli-Sforza and Bodmer 1971).

Attending medical school is comparatively rare in the population as a whole, and there were no instances of multiple ascertainment of the same family. Therefore, we assumed that our procedure for obtaining the sample was close to single ascertainment through children and chose a low value for the probability of ascertainment, which was arbitrarily set at .001.

As mentioned above, we applied segregation analysis to nuclear-family data only. Because none of our probands had children, the sample of relatives encompassed three categories: (1) parents, (2) siblings within the usual "age at risk" for attending medical school, which in Britain is age 18 years or older, and (3) siblings younger than age 18 years. For simplicity, we reduced these to two liability classes. Parents and older siblings, i.e., groups (1) and (2), were considered to be in the first liability class, where the population risk of attending medical school is that for British adults, about 0.22%. Younger siblings, i.e., group (3), were considered to be in a second liability class. Here the population risk is much smaller and difficult to estimate accurately, but certainly it is not zero. From our own experiences, fewer than one in 40 students gain entry to medical school before the usual age, and therefore the risk for the second liability class was set at 0.005%. A further simplifying assumption was that sex differences could be ignored. Again, this might present a potential problem, since inspection of table 1 shows a higher proportion of "affected" fathers than "affected" mothers. However, the sex difference disappears in siblings (not shown in table 1), reflecting the fact that there are now equal numbers of men and women attending British medical schools.

Results

From inspection of the findings summarized in table 1, it is clear that we are dealing with a trait which is strongly familial (if we accept the population risk figures of 0.22%). But there is not, on the face of it, a simple pattern of inheritance. It could be argued that the 22% frequency of the trait in siblings over age 18 years (excluding probands) is not significantly different from the expectation under an autosomal recessive hypothesis (Pearson $\chi^2 = 0.33$, NS). However, a segregation ratio of 3:1 would of course only be expected in heterozygous × heterozygous matings in which both parents appear unaffected, whereas in fact the observed frequency of the trait in parents, particularly in fathers, is high. We therefore proceeded to complex segregation analysis using the program POINTER.

The results are summarized in table 2, where likelihood ratios are also listed for comparison of several models. A "mixed" model of a major gene plus multifactorial component has significantly more support than does multifactorial transmission alone ($\chi^2 = 14.4$, df = 3, $P < .005$). There is, however, no significant difference either between the mixed model and the major-locus model ($\chi^2 = 0.01$, df = 1, NS) or between the general single-locus model and a recessive model where "d" is fixed at 0 ($\chi^2 = 0$).

A "null" model of no transmission was much less satisfactory than even the multifactorial model ($\chi^2 = 149.17$, df = 1, $P = .000$). Similarly, a hypothesis of no major effect ($q = 0$) could be rejected, as could a model in which all transmission probabilities were set to be equal. Comparing the recessive model with a more general vertical transmission model in which τ_2 was allowed to deviate from its Mendelian value resulted in no significant change of the likelihood ($\chi^2 = 2.37$, df = 1, NS). However, when all transmission probabilities τ_1, τ_2 and τ_3 were unrestricted and a search was performed on these together with q and t, a likelihood ratio test comparison with the recessive model was significant at the 0.05% level ($\chi^2 = 8.92$, 3 df). We can therefore conclude the following:

a) We have very strong evidence of familial transmission since the null model is much less satisfactory than *any* of the transmission models. Similarly, we

Table 2

Application of a Unified Model

Model	Iterated Parameters	Maximum-Likelihood Parameter Estimates							− 2 ln Likelihood + Constant
		d	t	q	H	τ_1	τ_2	τ_3	
Mixed (M)	$d\ t\ q\ H$	.087	4.044	.089	.008	[1]	[.5]	[0]	−2879.97
General single locus (GSL)	$d\ t\ q$	.651	3.899	.088	[0]	[1]	[.5]	[0]	−2879.86
Recessive (R)	$t\ q$	[0]	7.619	.088	[0]	[1]	[.5]	[0]	−2879.86
Vertical transmission 1 (V1)	$t\ q\ \tau_2$	[0]	6.363	.129	[0]	[1]	.143	[0]	−2882.23
Vertical transmission 2 (V2)	$t\ q\ \tau_1\ \tau_2\ \tau_3$	[0]	7.254	.038	[0]	1[a]	.668	.637	−2888.78
Multifactorial (MF)	H	[0]	[0]	[0]	.845	[1]	[.5]	[0]	−2865.57
Null (N)		[0]	[0]	[0]	[0]	[1]	[.5]	[0]	−2716.40
No major affect (NM)	t	[0]	.7619	[0]	[0]	[1]	[.5]	[0]	−2716.39
No transmission (NT)	$t\ q$	[0]	3.201	.113	[0]	.887	.887	.887	−2716.40

Note.—Likelihood ratios: M-MF = 14.40, df = 3, $P < .005$; M-GSL = 0.11, df = 1, NS; GSL-R = 0.0, df = 1, NS; V1-R = 2.37, df = 1, NS; V2-R = 8.92, df = 3, $P < .05$; MF-N = 149.17, df = 1, $P = .000$; R-NM = 163.47, df = 1, $P = .000$; GSL-NT = 163.46, df = 1, $P = .000$. Fixed parameter values are shown in brackets, e.g., [.5].

[a] Fixed at a bound during iteration.

can reject $q = 0$ and a "no-transmission" model where all three τ parameters are constrained to be equal.

b) We have strongly suggestive evidence of a major effect, in that a mixed model has more support than a multifactorial model but does not have an advantage over a single-locus alone.

c) On grounds of parsimony a recessive-gene hypothesis offers a more satisfactory explanation of the transmission of the trait than does a general single-locus model, since constraining the dominance parameter to the recessive-model value (i.e., $d = 0$) did not produce a reduction of the likelihood.

d) Allowing the transmission probability τ_2 to deviate from its Mendelian value of .5 did not result in a significantly improved fit. This is sometimes considered the most relevant test of Mendelian transmission (Lalouel et al. 1984). However, when we went on to apply an even more general model, allowing all transmission probabilities to be unrestricted, a significantly better fit was obtained, allowing rejection of the recessive model.

Discussion

Both intelligence and personality factors influence the choice of medicine as a career and the ability to gain entry to medical school. Both intelligence, as measured by IQ tests, and personality, as assessed by various scales, are moderately heritable (Henderson 1982). Therefore, it is probable that genetic factors do, in an indirect way, contribute to the familiality of attending medical school; but the major-gene hypothesis is, on commonsense grounds, highly implausible. It is far more likely that the major source of family resemblance for this trait derives from family culture and shared environment than from shared genes.

Despite this, our analysis not only supported vertical transmission, which could be compatible with either cultural or genetic inheritance, but also suggested transmission of a major "recessive-like" effect. Thus under mixed-model analysis, a major effect, which did not differ significantly from a recessive pattern of inheritance, receives greater support than does the more intuitively plausible hypothesis of multifactorial transmission. Under a unified model we compared the hypothesis of $\tau_2 = \frac{1}{2}$ with the hypothesis of an unrestricted τ_2. This has been suggested by the original authors of the unified model as the most relevant test of Mendelian segregation (Lalouel et al. 1984). Again the recessive-gene hypothesis withstood the test, and thus far our analysis, using modern, more sophisticated methods, replicated findings published more than 30 years ago (Lilienfield 1959). Searching on all three transmission probabilities of Elston and Stewart (1971) provides a yet more stringent test of Mendelian transmission. This is not always routinely performed within the context of a unified model analysis (and we suspect from our own experience that this is at least in part because of practical problems of achieving convergence). However, when we compared the completely unrestricted τ model with the recessive model, we were finally able to reject the latter, albeit at just under the 5% level of significance (likelihood ratio $\chi^2 = 8.91$, 3 df).

Our study is not without its methodological imperfections. Reliance on a general population risk figure derived from the British *Medical Directory* is not ideal, and a control population taking into account socioeconomic status and educational level would be desirable. A sex effect is also apparent in our data in that medical students' fathers are more likely to have attended medical school than are their mothers. This, of course, is not a biological phenomenon but, rather, is a reflection of the previous selection policies (now abandoned) in British medical schools of deliberately choosing more men than women. We chose to ignore parental sex effect in our analyses, since no sex effect was present in siblings. This is certainly an oversimplification, but it is not one likely to have biased our results in the direction of detecting a recessive gene. Therefore, despite these caveats, the findings on all but one test are compatible with a recessive-gene hypothesis and partially replicate results obtained 3 decades earlier (Lilienfield 1959).

The recessive-gene hypothesis did fail the final and most stringent test of comparison, i.e., with a completely unrestricted τ model. However, even here the significance level was not high, and it could be argued that, because of the multiple significance tests carried out, we have not provided convincing refutation of the recessive-gene hypothesis. It could therefore be further argued that we have more consistent, and somewhat more persuasive, evidence of a major gene for attending medical school than for any of the neuropsychiatric disorders recently investigated in linkage studies.

We do not suggest that such linkage strategies should be abandoned; indeed, we consider that genetic marker studies of neuropsychiatric disorders have a definite place in disorders where there is strong prior evidence of an important genetic component (McGuffin 1988). Nor do we suggest that genetic approaches are out of place in studies of complex common disease. Indeed, we have elsewhere advocated genetic strategies as a means of refining phenotypes (Farmer et al. 1987), and we strongly favor the view that genetic studies provide the most consistent clues to the etiology of common neuropsychiatric disorders (e.g., see McGuffin et al. 1987). However, our findings can perhaps serve as a reminder of the dangers of an overeager acceptance of simple explanations of the transmission of complex phenotypes and provide a caution against the wholesale application of genetic linkage investigations into *any* behavioral trait which shows familial aggregation. As we have discussed elsewhere (McGuffin and Sturt 1986), it would be mistaken to regard either segregation or linkage analysis as a panacea for the problems of complex phenotypes or to disregard the lessons to be learned from the classical methods of clinical family, twin, and adoption studies. Instead, an effort continues to be necessary to improve diagnostic criteria and phenotype measurement and definition. Such an effort in combination with the careful and circumspect interpretation of the results of segregation and linkage analysis ultimately holds real promise.

References

Barnes DM (1989). Troubles encountered in gene linkage land. Science 243:313–314

Baron M, Risch N, Hamburger R, Mandel H, Kushner S, Newman M, Drumer M, et al (1987) Genetic linkage between X chromosome markers and manic depression. Nature 326:806–808

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations. WH Freeman, San Francisco

Clerget-Darpoux F, Bonaiti-Pellie C, Hochez J (1986) Effects of misspecifying genetic parameters in lod score analysis. Biometrics 42:393–399

Detera-Wadleigh SD, Berretini WH, Goldin LR, Boorman D, Anderson S, Gershon ES (1987) Close linkage of c-Harvey ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees. Nature 325:806–807

Edwards JH (1960) The simulation of Mendelism. Acta Genet 10:63–70

Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, et al (1987) Bipolar affective disorder linked to DNA markers on chromosome 11. Nature 325:783–787

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

Falconer DS (1965) The inheritance of liability to certain diseases, estimated from the incidence among relatives. Ann Hum Genet 29:51–76

Farmer AE, McGuffin P, Gottesman II (1987) Twin concordance for DSM-III schizophrenia: scrutinising the validity of the definition. Arch Gen Psychiatry 44:634–641

Gershon ES, Mendlewicz J, Gastpar M, Bech P, Goldin LR, Kielholz P, Rafaelson OJ, et al (1980) A collaborative study of genetic linkage of bipolar manic depressive illness and red/green color blindness. Acta Psychiatr Scand 61:319–338

Goate AM, Haynes AR, Owen MJ, Farrall M, James LA, Lai YC, Mullan MJ, et al (1989) Predisposing locus for Alzheimer's disease on chromosome 21. Lancet 1:352–355

Gottesman II, McGuffin P, Farmer AE (1987) Clinical genetics as clues to the 'real' genetics of schizophrenia. Schiz Bull 13:23–48

Greenberg DA, Hodge SE (1989) Linkage analysis under "random" and "genetic" reduced penetrance. Genet Epidemiol 6:259–264

Gruneberg H (1952) Genetical studies on the skeleton of the mouse. IV. Quasi-continuous variations. J Genet 51:95–101

Henderson ND (1982) Human behaviour genetics. Annu Rev Psychol 33:403–440
Hodgkinson S, Sherrington R, Gurling HMD, Marchbanks R, Reeders S, Mallet J, McInnis M, et al (1987) Molecular genetic evidence for heterogeneity in manic depression. Nature 325:805–806
Kay DWK (1989) Genetics, Alzheimer's disease and senile dementia. Br J Psychiatry 154:311–320
Kennedy JL, Giuffra LA, Moises HW, Cavalli-Sforzas LL, Pakstis AJ, Kidd JR, Castiglione CM, et al (1988) Evidence against linkage of schizophrenia to markers on chromosome 5 in a northern Swedish pedigree. Nature 336:167–170
Lalouel JM, Rao DL, Morton NE, Elston RL (1984) A unified model for complex segregation analysis. Am J Hum Genet 26:484–503
Lilienfield AM (1959) A methodological problem in testing a recessive genetic hypothesis in human disease. Am J Public Health 49:199–204
McGuffin P (1988) Major genes for major affective disorder. Br J Psychiatry 153:591–596
McGuffin P, Murray RM, Reveley AM (1987) Genetic influences on the functional psychoses. Br Med Bull 43:531–556
McGuffin P, Sturt E (1986) Genetic markers in schizophrenia. Hum Hered 36:65–88
Medical directory (1988) Churchill Livingstone, Edinburgh
Mendlewicz J, Sevy S, Brocas H, Simon P, Charon F, Legros S, Vassar G (1987) Polymorphic DNA marker and X chromosome and manic depression. Lancet 2:1230–1232
Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7:277–318
Morton NE, MacLean CJ (1974) Analysis of familial resemblance. III. Complex segregation analysis of quantitative traits. Am J Hum Genet 26:484–503
Ott J (1965) Analysis of human genetic linkage. Johns Hopkins University Press, Baltimore
St. George-Hyslop PH, Rudolph ET, Polinski RJ, Haines JL, Nee L, Watkins PC, Myers RH, et al (1987) The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science 235:885–890
Schellenberg GD, Bird TD, Wijsmen WM (1988) Absence of linkage of chromosome 21q21 markers to familial Alzheimer's disease. Science 241:1507–1510
Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dadleston K, Barraclough B, Wasmuth J, et al (1988) Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336:164–167