Psychiatric Genetics (1993), **3**, 63 - 71

Segregation analysis of panic disorder

V.J. Vieland, S.E. Hodge, J.D. Lish, P. Adams and M.M. Weissman

Columbia University and the New York State Psychiatric Institute, Division of Clinical & Genetic Epidemiology, Unit 14, 722 W 168th Street, New York, NY 10032, USA

Correspondence to: V.J. Vieland at above address

We performed a simple segregation analysis of panic disorder, using 30 two- and three-generation pedigrees. Pedigrees were singly ascertained, either through the Epidemiologic Catchment Area study (seven probands), or as a consecutive series from an anxiety disorders clinic (23 probands). All probands were required to meet DSM-III panic disorder criteria, without comorbid major depression. Relatives (n = 189) were required to meet DSM-III criteria for panic disorder, with or without comorbid major depression. We fitted a single major dominant and a single major recessive model to the data, allowing for an age-of-onset distribution. Under the dominant model, we obtained the following parameter estimates: gene frequency = 0.01; (lifetime) susceptibility for gene carriers = 0.5; susceptibility for non-gene carriers = 0.01. Under the recessive model, we obtained the following parameter estimates: gene frequency = 0.2; susceptibility for gene carriers = 0.7; susceptibility for non-gene carriers = 0.01. The best-fitting dominant and best-fitting recessive models had equally high likelihoods. Discrepancies between our results and earlier reports are discussed, as are implications of these results for linkage analyses of panic disorder.

Keywords: Genetics – Panic disorder – Segregation analysis

INTRODUCTION

Panic disorder (PD), as defined by the DSM-III, appears to be a familial disorder, and has been a target of previous genetic investigation (see, for example, Crowe *et al.*, 1988). However, as is the case with other psychiatric disorders suspected of genetic etiology, no particular mechanism of inheritance has been established.

Evidence in favor of a genetic basis for PD comes from several sources. Family studies have consistently found evidence of familial aggregation of the disorder and/or transmission from parents to offspring (Cloninger *et al.*, 1981; Crowe *et al.*, 1983; Hopper *et al.*, 1990). Estimates of rates of PD in first-degree relatives of panic probands in these studies have ranged from 7% to 25%, while rates in the general population are estimated to be in the 1-2% range (Robins and Regier, 1991). The two twin studies of PD in the literature have suggested genetic contributions to the disorder as well, although the magnitude of these contributions may be small (Torgersen, 1983; Kendler *et al.*, 1993).

Two published genetic analyses have supported involvement of a major gene. Pauls *et al.* (1980) analyzed a dataset consisting of 19 probands with DSM-III PD and multiple family members per proband. They were unable to reject a dominant Mendelian model, and their best-fitting model yielded the following parameter estimates: frequency of the disease allele = 0.014; penetrance for gene carriers = 0.75; and a phenocopy rate ("penetrance" for non-gene carriers) of essentially 0.

In a second analysis, Crowe *et al.* (1983) examined 41 probands with DSM-III PD and 278 first-degree relatives. This dataset included the original 19 probands from Pauls *et al.* (1980). Crowe *et al.* obtained a best-fitting single major locus model which was again roughly dominant (they estimated separate penetrances for each of the three possible genotypes, and obtained higher estimates of penetrance for the homozygous (disease) genotype than the heterozygous). Their model yielded the following parameter estimates: frequency of the disease allele = 0.05; average penetrance for gene carriers = 0.36 (we have taken the average penetrance over their separate estimates for males, females, and the two susceptible genotypes); and a phenocopy rate of 0.

In both of these previous segregation analyses, probands have been required to meet DSM-III criteria for PD, but may also have met criteria for major depressive disorder (MDD) (Crowe *et al.* specified that probands who were comorbid for PD and MDD had to have primary PD, i.e. onset of PD prior to onset of MDD). While studies have consistently shown high rates of comorbidity between PD and MDD (see for example Weissman, 1989), there is also evidence that probands with PD plus MDD may represent a heterogeneous group, not entirely similar to PD

probands without MDD. For example, Weissman *et al.* (1992) have shown that rates of familial aggregation of PD among PD probands without MDD are higher than rates of familial aggregation among PD + MDD probands.

In the current study, we have analyzed a dataset consisting of 30 DSM-III PD probands, none of whom have comorbid MDD, and 189 first-degree relatives. By accepting as PD probands only those who do not meet criteria for comorbid MDD, we hope to focus on a more (genetically) homogeneous group. We have estimated parameters for simple genetic models, and attempted to replicate (or reject) the findings of the earlier segregation analyses in a new dataset. Our results have implications for understanding the genetics of PD, and some important implications for the conduct of linkage studies of the disorder.

METHODS

Sample

Details of the sampling procedure and diagnostic methods are given in Weissman *et al.* (1992). A brief description follows. The sample was drawn from two sources: 23 probands were ascertained as a consecutive series from an anxiety disorders clinic; seven probands were obtained from a population sample drawn through the Epidemiologic Catchment Area (ECA) study. In no case was there more than one proband per family, nor is there any reason to suspect non-independence of ascertainments across families. Thus, the data appear to conform to true single ascertainment (Morton, 1959).

Diagnoses were obtained on probands, their parents, their siblings, their spouses, and their offspring, provided the offspring were at least 6 years of age. Diagnostic information was obtained by direct interview using the SADS-LA (Mannuzza *et al.*, 1989), the K-SADS-E (Orvaschel *et al.*, 1982) for subjects between the ages of 6 and 17, or by family history using the FH-RDC (Andreasen *et al.*, 1977). When the FH-RDC was administered, if an informant indicated any anxiety or phobic avoidance symptoms, a specific symptoms checklist for PD was included. All probands received direct interviews, as did approximately 65% of the 189 relatives. The mean number of FH-RDCs collected per individual was 3.4.

Probands were required to meet full criteria for DSM-III PD based on direct interview data, and if the diagnosis was not confirmed by two independent, blind, best-estimators, probands were dropped from the sample. Probands included in the present study were required to show no evidence of MDD ever, either preceding, concomitant with, or following PD, using modified RDC criteria for MDD (Mazure and Gershon, 1979). See Weissman *et al.* (1992) for details on criteria and collection of age of onset information.

Diagnoses in relatives were also made by best-estimate, based on all available information on each relative, including direct interview if available, and family history from multiple informants. The best-estimators were blind to proband diagnosis and status of subject (proband or relative), and were not involved in the data collection. Relatives were considered affected if they either (1) met full criteria for PD, or (2) met most criteria for PD, were missing some relevant information, but could be assigned a diagnosis of PD with no significant doubt in the judgement of the best-estimators. In either case, relatives were considered affected regardless of MDD status. In comparison with a normal control group, there was no increase in the rate of MDD without PD among relatives of the PD probands. The rate of PD without MDD in the relatives of the PD probands was 7.8%; the rate of PD with MDD was 6.4%. The overall odds ratio for PD with or without MDD, compared to the normal control group and adjusted for age and sex, was 18.1 (Weissman *et al.*, 1992).

Twenty of the probands were female, and 10 were male. The probands were ethnically homogeneous, and all were from the same geographical area. Their ages ranged from 22 to 84 (mean 43.4, S.D. 12.5).

In preparation for genetic modeling, the data were checked to make sure that they met certain assumptions of the model. The distribution of age of onset, which was available for all affected individuals, was examined for differences by sex and found to be virtually identical in the two groups (for females, the mean was 25.3 with S.D. 8.6; for males, the mean was 25.8 with S.D. 9.8; the age distribution of affected individuals did not differ by sex).

The age-of-onset distribution was also examined for equivalence of the mean, median, and mode, and for skewness and kurtosis. It appeared to be normally distributed.

We also wanted to be sure that the clinic sample and the ECA sample did not differ crucially. The two samples were compared and were found to contain similar proportions of multiplex families (43% ECA sample; 39% clinic sample). Age of onset was also examined for differences by source of ascertainment (ECA vs clinic), and the distributions were found to be very similar (ECA, $n = 12$: mean 28.3, S.D. 11.0; clinic, $n = 38$: mean 25.2, S.D. 8.8). Therefore we found no reason to treat the two samples separately. However, in order to be safe, we repeated all analyses omitting the ECA pedigrees. The results were essentially unchanged (results available upon request).

Figure 1 shows the 30 pedigree structures, the location of the proband within each pedigree, and diagnostic status.

Genetic modeling

All calculations were carried out using the REGTN program from the SAGE computer package (SAGE, 1992). REGTN performs segregation analysis of a trait with variable age of onset, where age of onset is assumed to be

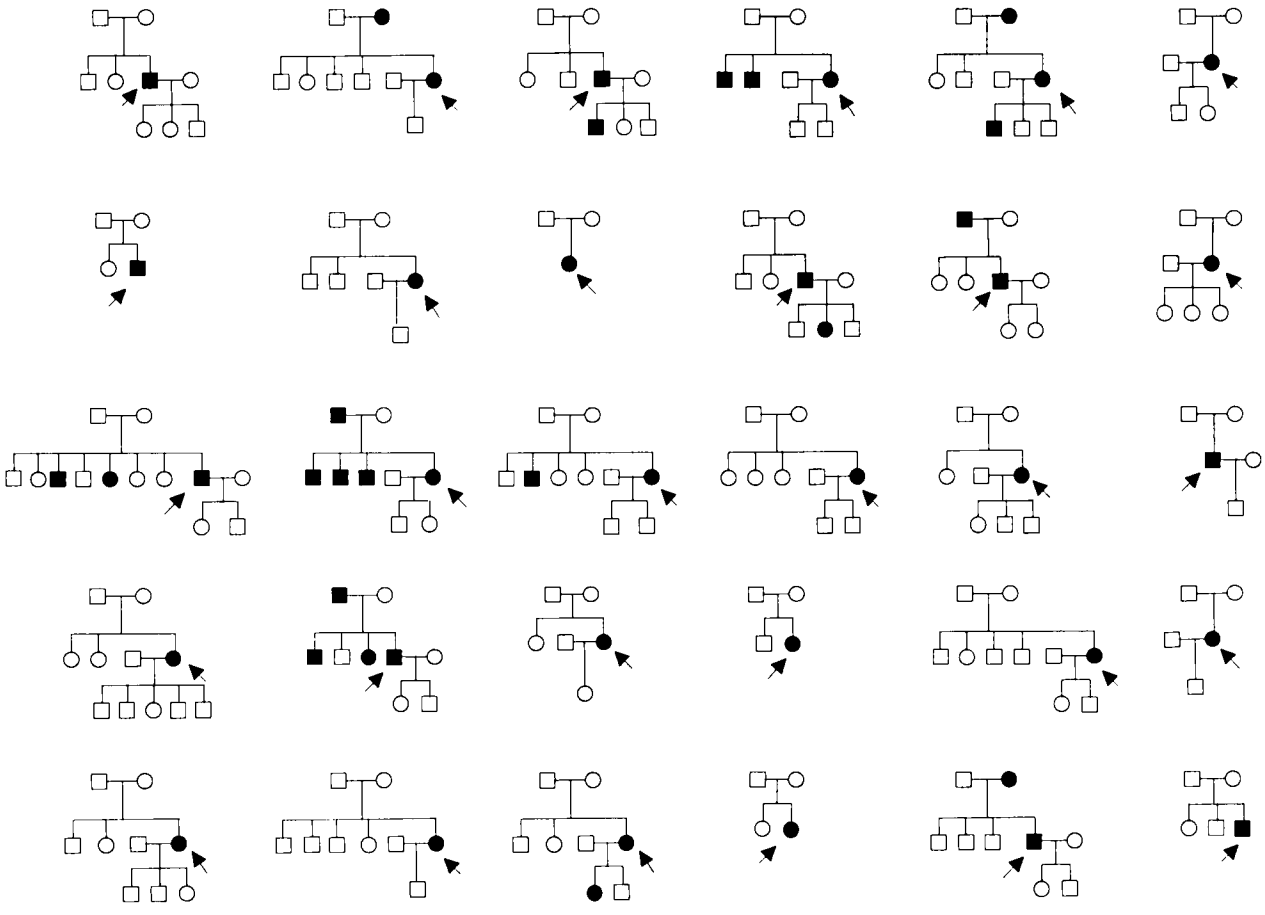

FIG. 1. All pedigrees used in the analysis. Shaded individuals are affected. Arrows refer to probands.

normally distributed. Ascertainment is corrected for by conditioning the likelihood of each family on the phenotype of the proband in that family (Cannings and Thompson, 1977). We analyzed the data under REGTN's Model 2, parameterized to represent a single major locus model, in which genotype is presumed to influence susceptibility to the disorder, but not to influence the parameters of the age-of-onset distribution. The full model contains 12 parameters: q = frequency of the disease allele a; Ψ_{aa}, Ψ_{ab}, Ψ_{bb} = population frequencies of the three genotypes respectively (we assumed Hardy-Weinberg equilibrium, hence these three parameters are determined by q); τ_{aa}, τ_{ab}, τ_{bb} = transmission probabilities (probability of transmitting the a allele) for the three (parental) genotypes (fixed at 1.0, 0.5, and 0.0 respectively, to correspond to Mendelian inheritance); the mean and standard deviation of the age of onset distribution (fixed at their sample values of $\bar{x} = 25.5$ and $s = 9.1$); and γ_{aa}, γ_{ab}, γ_{bb} = the susceptibilities (lifetime probability of becoming affected) for the three genotypes (under a simple dominant model γ_{aa} is constrained to be $= \gamma_{ab}$, with $\gamma_{aa} = \gamma_{ab} >> \gamma_{bb}$; while under a recessive model, γ_{ab} is constrained to be $= \gamma_{bb}$, with $\gamma_{aa} >> \gamma_{ab} = \gamma_{bb}$).

Thus we were left with three parameters to be estimated: frequency of the disease allele (q); susceptibility for (disease) gene carriers ($\gamma_{aa} = \gamma_{ab} \equiv \gamma_{a-}$ for the dominant model, γ_{aa} for the recessive model); and susceptibility for non-gene carriers (γ_{bb} for the dominant model, $\gamma_{ab} = \gamma_{bb} \equiv \gamma_{-b}$ for the recessive model). However, with three parameters in the model we had convergence problems using Newton-Raphson iteration, and this remained the case even when we fixed q and simply tried to estimate the other two parameters.

For this reason, we did not rely on the program's maximization routine, but rather, used REGTN to calculate likelihoods across a grid of (fixed) parameter values, which could then be directly inspected for global maxima, and used for comparing models. We felt that a reasonable restriction on any plausible genetic model would be that it should predict a realistic prevalence of PD. Since the lifetime prevalence of panic disorder is estimated to be between 1% and 2% (Robins and Regier, 1991), the parameter values for which likelihoods were computed

TABLE I. Grid values used in computing likelihoods for the dominant, recessive model

Dominant grid values: likelihoods were calculated at all those grid points shown here that also met the criterion $0.01 \leq K \leq 0.04$, for K = population prevalence

q = 0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1
γ_{a-} = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0
γ_{bb} = 0.0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1

Recessive grid values: likelihoods were calculated at all those grid points shown here that also met the criterion $0.01 \leq K \leq 0.04$, for K = population prevalence.

q = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4
γ_{aa} = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1.0
γ_{-b} = 0.0, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1, 0.15

were restricted to those predicting a population prevalence between 1% and 4%. Table I shows the grids of parameter values used for the dominant and the recessive models.

Throughout the Results section, we will use relative support (differences in ln likelihoods) to compare models (Edwards, 1972; Conneally *et al.*, 1985). Since specific distributional assumptions about the parameters (e.g. that they conform to multivariate normality or that their estimates conform to asymptotic expectations) may be unwarranted, classical hypothesis testing may be inappropriate for the comparisons we wish to make. Relative support affords a direct and distribution-free measure for evaluating alternative models.

RESULTS

Maximum likelihood parameter estimation

We obtained maximum likelihood parameter estimates under both the dominant and recessive models. These sets of estimates will be referred to as the "best-fit dominant" and the "best-fit recessive" models respectively.

(1) Estimation of parameters for the dominant model (best-fit dominant). Table IIa-c shows the results for the dominant model. Under the dominant model, the maximum likelihood within the grid occurred at the point $q = 0.01$, $\gamma_{a-} = 0.5$, $\gamma_{bb} = 0.01$ (ln likelihood = −233.09). These values correspond to a population prevalence of about 2%. The bold figures in the table show the approximate 2-unit support region around the (overall) maximum. (The interval is approximate because only a crude grid of values, and not a continuum of values, has been examined.) As can be seen, a 2-unit support region spans a wide range of parameter values.

(2) Estimation of parameters for the recessive model (best-fit recessive model). Table IIIa-c shows the results for the recessive model, with the approximate 2-unit support region around the (overall) maximum shown by the bold figures. Under the recessive model, the maximum likelihood along the grid occurred at the point $q = 0.20$, $\gamma_{aa} = 0.7$, $\gamma_{-b} = 0.01$ (ln likelihood = −233.10). These values correspond to a population prevalence of about 4%.

TABLE II. Results for the dominant model (results are shown for the vicinity of the overall maximum only) (ln likelihoods)

$\gamma_{aa} = \gamma_{ab}$	γ_{bb}			
	0.00	0.01	0.02	0.03
	(a) Frequency of the disease allele = 0.005			
0.2	—	−246.93	−247.62	−246.36
0.3	—	−240.06	−241.98	−241.95
0.4	—	−236.19	−238.26	−238.73
0.5	—	**−234.30**	−236.00	−236.56
0.6	—	**−233.86**	**−234.88**	−235.27
0.7	—	**−234.64**	**−234.72**	**−234.74**
0.8	—	−236.64	−235.46	**−234.92**
1.0	—	−252.42	−243.06	−239.12
	(b) Frequency of the disease allele = 0.01			
0.2	—	−242.75	−244.30	−243.95
0.3	—	−236.64	−238.64	−239.20
0.4	—	**−233.83**	−235.41	−236.16
0.5	—	**−233.09***	−233.89	−234.46
0.6	−243.47	**−233.86**	**−233.67**	—
0.7	−251.04	−235.92	**−234.54**	—
0.8	−262.02	−239.30	−236.48	—
1.0	−482.10	−261.04	−248.47	—
	(c) Frequency of disease allele = 0.02			
0.2	—	−239.32	−240.97	−241.24
0.3	**−233.75**	**−234.54**	−235.92	—
0.4	−235.11	**−233.11**	**−233.71**	—
0.5	−238.58	**−233.74**	**−233.37**	—
0.6	−244.00	−235.90	—	—
0.7	−251.70	−239.40	—	—
0.8	−262.81	—	—	—
1.0	−486.23	—	—	—

Note: bold values fall within the overall 2-unit support interval for the dominant model; the overall maximum is shown in Table IIb.
* Overall maximum for the dominant model.

Tests of models

We tested a series of alternative models against our best-fit models, by examining differences in support (ln L differences). Specifically, we tested the following models against our best-fit dominant: (1) a fully penetrant dominant with no sporadics; (2) a dominant with reduced penetrance but no sporadics; (3) the set of parameters estimated by Pauls *et al.* (1980) (the "Pauls *et al.* model"); (4) the set of parameters estimated by Crowe *et al.* (1983) (the "Crowe *et al.* model"). Additionally, we tested the following models against our best-fit recessive: (5) a fully penetrant recessive with no sporadics; (6) a recessive with reduced penetrance but no sporadics. Finally, (7) we compared our best-fit dominant with our best-fit recessive. Table IV summarizes the results of the seven tests described below.

(1) Dominant with full penetrance, no sporadics. Looking at the pedigrees, it is clear that a dominant model with full penetrance and no sporadic cases is incompatible with the data, i.e. that the true likelihood, when γ_{a-} is fixed at 1.0 and γ_{bb} is fixed at 0.0, is $-\infty$.

(2) Dominant with reduced penetrance, no sporadics. The maximum ln L for the dominant model when γ_{bb} is fixed at 0.0 is -233.75 (at $q=0.02$, $\gamma_{a-}=0.3$). Comparing this to the best-fit dominant yields a relative support of 0.66, which is well within a 2-unit support interval. Hence the data are not incompatible with the hypothesis that the phenocopy rate is in fact 0 under a dominant model, provided that the penetrance for gene carriers is low.

(3) Pauls et al. model. Fixing the parameter values at the estimates obtained by Pauls *et al.* ($q=0.014$, $\gamma_{a-}=0.75$, $\gamma_{bb}=0.0$), yields a ln $L=-256.26$, and a relative support of 23.2. Thus our data clearly reject the Pauls *et al.* estimates.

(4) Crowe et al. model. Fixing the parameter values at the estimates obtained by Crowe *et al.* ($q=0.05$, $\gamma_{a-}=0.36$, $\gamma_{bb}=0.0$), yields a ln $L=-234.86$, and a relative support of 1.77. Thus, the parameter estimates obtained by Crowe *et al.* do fall within a 2-unit support interval of our maximum likelihood for the best-fit dominant model.

TABLE III. Results for the recessive model (results are shown for the vicinity of the overall maximum only) (ln likelihoods)

γ_{aa}	$\gamma_{ab}=\gamma_{bb}$			
	0.00	0.01	0.02	0.03
	(a) Frequency of the disease allele = 0.1			
0.2	—	−254.68	−252.00	−249.04
0.3	—	−248.45	−247.39	−245.71
0.4	—	−244.12	−243.86	−242.85
0.5	—	−241.03	−241.15	−240.54
0.6	—	−238.80	−239.07	−238.68
0.7	—	−237.22	−237.49	−237.22
0.8	—	−236.12	−236.30	−236.07
1.0	−235.79	**−235.03**	**−234.82**	**−234.52**
	(b) Frequency of the disease allele = 0.2			
0.2	—	−244.44	−244.49	−243.66
0.3	−236.44	−238.76	−239.36	—
0.4	**−234.25**	−235.64	−236.25	—
0.5	**−233.33**	**−233.96**	**−234.42**	—
0.6	**−233.20**	**−233.22**	—	—
0.7	**−233.62**	**−233.10***	—	—
0.8	**−234.42**	—	—	—
1.0	—	—	—	—
	(c) Frequency of disease allele = 0.3			
0.2	−237.79	−239.30	−239.78	—
0.3	**−234.36**	−235.13	—	—
0.4	**−233.26**	—	—	—
0.5	—	—	—	—
0.6	—	—	—	—
0.7	—	—	—	—
0.8	—	—	—	—
1.0	—	—	—	—

* Overall maximum for the recessive model.

TABLE IV. Summary of tests of models

	Dominant model tests			
	q	γ_{a-}	γ_{bb}	Relative support[a]
Best-fit dominant	0.01	0.5	0.01	—
Reduced penetrance, no sporadics	0.02	0.3	[0.0][b]	0.7
Crowe *et al.* model	[0.05]	[0.36]	[0.0]	1.8
Pauls *et al.* model	[0.014]	[0.75]	[0.0]	23.2
Full penetrance, no sporadics	—	[1.0]	[0.0]	$-\infty$
	Recessive model tests			
	q	γ_{aa}	γ_{-b}	Relative support[c]
Best-fit recessive	0.2	0.7	0.01	—
Reduced penetrance, no sporadics	0.2	0.6	[0.0]	0.1
Full penetrance, no sporadics	0.1	[1.0]	[0.0]	2.7

[a] Numbers in this column represent relative support (differences in ln likelihoods) in favor of the best-fit dominant model.
[b] Numbers in square brackets represent values fixed when computing likelihoods.
[c] Numbers in this column represent relative support in favor of the best-fit recessive model.

(5) Recessive with full penetrance, no sporadics. The maximum ln L for the recessive model when γ_{aa} is fixed at 1.0 and γ_{-b} is fixed at 0.0 is -235.79. Comparing this to the best-fit recessive model yields a relative support of 2.69, corresponding to a likelihood ratio of approximately 15:1. Thus, a fully penetrant recessive model without sporadic cases falls somewhat outside the 2-unit support region. Under the restrictions set on prevalence, only one gene frequency allows this possibility at all ($q=0.1$).

(6) Recessive with reduced penetrance, no sporadics. The maximum ln L for the recessive model when γ_{-b} is fixed at 0.0 is -233.20 (at $q=0.2$, $\gamma_{aa}=0.6$), and the relative support is 0.10. Hence the data are compatible with the hypothesis that the phenocopy rate is in fact 0 under a recessive model.

(7) Best-fit dominant vs best-fit recessive. Although we cannot test the best-fit dominant and best-fit recessive models against one another directly, we can note that each is equally likely (-233.09 cf. -233.10) relative to the most general model (whose likelihood we do not know).

Hence, whatever the true point of maximum likelihood may be (allowing all relevant parameters in the general model to be estimated simultaneously), if we could test the fit of our models against the full model, and if we were to reject the recessive model, we would also reject the dominant model. (Also vice-versa, if we did not reject the best-fit dominant model, we would also fail to reject the best-fit recessive model.)

Also, note that our best-fit recessive model is strongly preferred over the Pauls *et al.* (dominant) model, and also somewhat preferred over the Crowe *et al.* (dominant) model.

DISCUSSION

There is evidence in the literature for a genetic mechanism for at least some cases of PD. While this study cannot test whether genes play a role in the development of PD, we can estimate parameters under the assumption of a major dominant or recessive gene. The results of doing so have implications both for an understanding of the possible etiology of PD itself, and also for the conduct of linkage studies of panic disorder.

Summary of findings

Assuming a single dominant or recessive gene underlying at least some cases of PD, we find that the two best-fitting models cannot be distinguished from one another in terms of likelihood. One is a dominant model, with $q=0.01$, $\gamma_{a-}=0.5$, and $\gamma_{bb}=0.01$. The other is a recessive model, with $q=0.2$, $\gamma_{aa}=0.7$, and $\gamma_{-b}=0.01$. In both cases, a 2-unit support interval spans a range of values for all three parameters.

Our findings do not appear to be compatible with a simple Mendelian mechanism (fully penetrant dominant or recessive, with no sporadic cases), nor with the specific parameter estimates of Pauls *et al.* But we cannot reject the estimates obtained by Crowe *et al.* While our 2-unit support intervals do include parameter sets for which the phenocopy rate is 0, our maximum likelihood estimate of a phenocopy rate of 0.01 under either (dominant or recessive) model represents a substantive departure from the earlier findings (see below for additional discussion of this point).

Clinical considerations

It is possible that discrepancies between our findings and those of the earlier studies are related to our exclusion of probands comorbid for PD and MDD. As discussed in the Introduction, inclusion of comorbid probands in a segregation analysis may introduce some degree of heterogeneity into the data that would not otherwise be present. We have deliberately used only "pure" PD probands, and thereby eliminated one potential source of noise from the data.

A further possible source of discrepancy between our sample and previously reported samples concerns the source of ascertainment. We have not included probands from clinics other than an anxiety disorders clinic. A recent study (P. Wickramaratne, personal communication) suggests that there may be substantial differences between panic probands ascertained through anxiety disorders clinics, and panic probands ascertained through depression clinics. The study found that the rate of panic in relatives of panic probands comorbid for MDD, and ascertained through an anxiety disorders clinic, was three times the rate in relatives of PD/MDD comorbid probands ascertained through a depression clinic (the difference was not however statistically significant). This suggests that panic probands ascertained through depression clinics may be a more heterogeneous group than those ascertained through anxiety disorders clinics. By restricting our sample to non-comorbid probands *and* probands ascertained through an anxiety disorders clinic (or the ECA), we may have further restricted heterogeneity in our dataset, in comparison to datasets used in the earlier reports.

Two major implications of our results for linkage studies

Allowing for phenocopies in linkage analyses of PD. It may not be a good idea to depend only on a linkage analysis model that does not allow for phenocopies. Our estimate of the phenocopy rate (0.01 for both the dominant and recessive models) represents a substantial proportion of cases. Let x = sporadic rate, defined as the proportion of all cases that are phenocopies (i.e. the probability that an affected individual does not carry the disease genotype). For our best-fit dominant model, we obtain $x=0.49$; for the best-fit recessive, we get $x=0.26$. That is, if the dominant model is correct, 49% of all cases represent a non-genetic form of the disorder. If the recessive model is correct, this proportion is 26%. This constitutes a large amount of etiologic heterogeneity among clinically undifferentiated cases of PD.

Of course in a linkage study, where families are usually selected for aggregation of the disorder, the proportion of phenocopies among cases may be substantially lower. Nevertheless, failing to allow for the possibility that some cases may be phenocopies can easily lead to erroneous attributions of recombination. Additionally, since a substantial phenocopy rate will reduce the power of a linkage study, our results suggest that the power to detect linkage for PD may be lower for currently available datasets than had previously been supposed (Vieland *et al.*, 1993a). To date, most published linkage studies of PD have been conducted without allowance for phenocopies (Crowe *et al.*, 1987, 1990; Mutchler *et al.*, 1990).

Analyzing PD linkage data under both dominant and recessive models. The second, and critical, implication

for linkage studies is that both dominant and recessive mechanisms must be considered. There is now ample evidence in the literature that 2-point linkage analysis is robust to misspecification of parameters, provided that the mode of inheritance is correctly specified (Clerget-Darpoux *et al.*, 1986; Greenberg and Hodge, 1989; Vieland *et al.*, 1992). In particular analyzing extended pedigree data selected for familial aggregation under the wrong mode of inheritance can result in dramatic diminution of lod scores (Vieland *et al.*, 1993b). Failure to analyze PD linkage data under a recessive model appears to be unwarranted by family data at this point, and could prove disastrous for power to detect linkage.

Alternative genetic models

While we have looked only at single major locus models (dominant, recessive), other possibilities exist for genetic mechanisms. Crowe *et al.*, for example, considered a multifactorial threshold model and also obtained an adequate fit to the data.

Another possibility worth considering is that of a two-locus epistatic model. For these models, the phenotype is determined by the genotype at each of two loci, where the mode of inheritance at each is simple Mendelian. For instance, a recessive-dominant epistatic model specifies that in order to be affected, an individual must inherit both the recessive genotype at one locus (say, aa, where "a" is the disease allele at the first locus) and the dominant genotype at the other (BB or Bb, where "B" is the disease allele at the second locus). A simple procedure exists for checking the compatibility of population prevalence and segregation ratio data with two-locus epistatic transmission graphically (Greenberg, 1981). Using the Greenberg graphs to check the fit of a two-locus epistatic model, we find that a recessive-recessive model and a recessive-dominant model can both be fitted to our data, while a dominant-dominant model cannot.

The possibility of two-locus epistasis is not in and of itself an impediment to linkage studies of PD. If PD is in fact governed by two epistatic loci under a recessive-recessive model, single-locus linkage analysis will provide an excellent approximation to a correct two-locus linkage analysis, provided that the data are analyzed under a single-locus recessive model (Vieland *et al.*, 1992, 1993b). The same rule applies for detection of linkage to the recessive locus in a recessive-dominant model. However, while there is no specific evidence that PD is in fact a two-locus disorder, we once again have a strong argument in favor of analyzing linkage data under both dominant and recessive models. As long as the true mode of inheritance is unknown, there is nothing to lose by considering both models, and potentially a great deal to be lost in not considering both models.

Limitations of this study

Finally, the results of this study should be interpreted in light of certain limitations. These in turn suggest some fruitful areas of research for future genetic modeling.

Sex differences in rates. Population data show substantially higher rates of PD in females than males (Crowe *et al.*, 1983; Robins *et al.*, 1984). The 2:1 ratio of female to male probands in our data exemplifies this discrepancy. One explanation sometimes put forward for such a difference in rates by sex is a polygenic threshold model. Such a model postulates an underlying continuous distribution of some factor which contributes to liability to the disorder. Individuals with a level of that factor beyond some threshold value succumb to the disorder. Sex differences in rates are accounted for in this model by postulating a higher threshold for one sex than for the other. So, for PD, the suggestion would be that males possess a higher threshold for developing PD than do females.

This model makes a specific prediction about rates in relatives, namely that relatives of male probands should show higher rates of the disorder than relatives of female probands (because the male probands must themselves have higher levels of the susceptibility factor, hence a higher genetic "loading", than the female probands).

In our data, rates in siblings of female probands are essentially the same as rates in siblings of male probands ($6/37 = 0.16$ siblings of female probands are affected; $4/23 = 0.17$ siblings of male probands are affected). Crowe *et al.* report a similar finding for their data. Thus, the data do not appear to support a polygenic threshold model.

Under the single major locus model, the explanation for such a sex difference in rates under the dominant model is either (a) γ_{a-} for females does not equal γ_{a-} for males; (b) γ_{bb} for females does not equal γ_{bb} for males; or (c) both of the above (similarly for the recessive model). Crowe *et al.* estimated that the female penetrance was 46% and the male penetrance 25% for heterozygotes (the great majority of affecteds) under the single major locus model. The size of our sample precluded separate estimation of these parameters for females and males.

Genetic influences on age of onset. We analyzed the data under the assumption that genotype influences susceptibility to PD, but not the parameters of the age of onset distribution. However, it is possible for genotype to influence both susceptibility and the age of onset distribution. This would be the case if, for instance, in addition to gene carriers having higher susceptibility than non-gene carriers, gene carriers also had a lower mean age of onset than non-gene carriers. This is a common pattern for genetic disorders, and can be fruitfully exploited in linkage studies (Mérette *et al.*, 1992).

Such a pattern does not appear to hold for PD. Epidemi-

ologic data fail to show bimodality in the age of onset of PD, and in at least one study, family data do not show a decrease in familial aggregation with increased age of onset in probands (P. Wickramaratne, personal communication).

In this dataset, a crude attempt may be made to assess the possibility of an influence of genotype on the age of onset distribution by comparing the mean age of onset in multiplex families (families with more than one affected) and in simplex families (families where the proband is the sole affected individual). This approach is crude in that, while it separates the data into families more likely to be transmitting the putative gene and families more likely to represent phenocopies, both subsets of the data may contain either type of case. We do not in fact find a difference by family type in the mean age of onset [multiplex families: mean = 25.0 (9.5); simplex families: mean = 27.6 (9.1)].

Restrictions on prevalence in the calculation of the likelihoods. We calculated likelihoods across a fixed grid of parameter values, and it is in principle possible that the true parameter values lie outside the range considered. However, the only substantive limitation we placed upon these parameter values was a restriction on the prevalence to the range 1-4%. Since there is widespread agreement among epidemiologic studies that rates of PD range from 1% to 2%, and since our restriction on prevalence leaves room for a large amount of undercounting, the requirement that any acceptable genetic model predict prevalences in this range seems quite reasonable.

Acknowledgements

This research was supported by NIMH grants MH-00884-K21, MH-48858, MH-28274, MH-43878, MH-46981, and NIH grant DK-31813. Results were obtained using the program package SAGE, which is supported by a US Public Health Service Resource grant (1 P41 RR03655) from the Division of Research Resources. We would like to thank David Greenberg for helpful discussion, Tovah Hollander for programming assistance, and Gary Heiman for assistance in preparation of the manuscript.

REFERENCES

Andreasen NC, Endicott J, Spitzer RL and Winokur G (1977) The family history method using diagnostic criteria: reliability and validity. *Archives of General Psychiatry*, **34**, 1229-1235.

Cannings C and Thompson EA (1977) Ascertainment in the sequential sampling of pedigrees. *Clinical Genetics*, **12**, 208-212.

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

Cloninger CR, Martin RL, Clayton P and Guze SB (1981) A blind follow-up and family study of anxiety neurosis: Preliminary analysis of the St. Louis 500. In: *Anxiety: New Research and Changing Concepts* (Eds DR Klein and GJ Rabkin), pp. 137-154. Raven Press, New York.

Conneally PM, Edwards JH, Kidd KK, Lalouel JM, Morton NE, Ott J and White R (1985) Report of the committee on methods of linkage analysis and reporting. *Cytogenetics and Cell Genetics*, **21**, 356-359.

Crowe RR, Noyes R, Pauls DL and Slyman DJ (1983) A family study of panic disorder. *Archives of General Psychiatry*, **40**, 1065-1069.

Crowe RR, Noyes R Jr, Wilson AF, Elston RC and Ward LJ (1987) A linkage study of panic disorder. *Archives of General Psychiatry*, **44**, 933-937.

Crowe RR, Noyes R Jr, Persico T, Wilson AF and Elston RC (1988) Genetic studies of panic disorder and related conditions. In: *Relatives at Risk for Mental Disorder* (Eds DL Dunner, ES Gershon and JE Barrett) pp. 73-85. Raven Press, New York.

Crowe RR, Noyes R Jr, Samuelson S, Wesner R and Wilson R (1990) Close linkage between panic disorder and α-haptoglobin excluded in 10 families. *Archives of General Psychiatry*, **47**, 377-380.

Edwards AWF (1972) *Likelihood: An Account of the Statistical Concept of Likelihood and its Application to Scientific Inference*. Cambridge University Press, Cambridge.

Greenberg DA (1981) A simple method for testing two-locus models of inheritance. *American Journal of Human Genetics*, **33**, 519-530.

Greenberg DA and Hodge SE (1989) Linkage analysis under "random" and "genetic" reduced penetrance. *Genetic Epidemiology*, **6**, 259-264.

Hopper JL, Judd FK, Derrick PL, Macaskill GT and Burrows GD (1990) A family study of panic disorder: reanalysis using a regressive logistic model that incorporates a sibship environment. *Genetic Epidemiology*, **7**, 151-161.

Kendler KS, Neale MC, Kessler RC, Heath AC and Eaves LJ (1993) Panic disorder in women: a population based twin study. *Psychological Medicine*, in press.

Mannuzza S, Fyer AJ, Martin LY *et al.* (1989) Reliability of anxiety assessment: I. Diagnostic agreement. *Archives of General Psychiatry*, **46**, 1093-1101.

Mazure C and Gershon ES (1979) Blindness and reliability in psychiatric diagnosis. *Archives of General Psychiatry*, **36**, 521-525.

Mérette C, King M-C and Ott J (1992) Heterogeneity analysis of breast cancer families by using age at onset as a covariate. *American Journal of Human Genetics*, **50**, 515-519.

Morton NE (1959) Genetic tests under incomplete ascertainment. *American Journal of Human Genetics*, **11**, 1-16.

Mutchler K, Crowe RR, Noyes R Jr and Wesner RW (1990) Exclusion of the tyrosine hydroxylase gene in 14 panic disorder pedigrees. *American Journal of Psychiatry*, **147**, 1367-1369.

Orvaschel H. Puig-Antich J, Chambers W, Tabrizi MA and Johnson R (1982) Retrospective assessment of prepubertal major depression with the Kiddie-SADS-E. *Journal of the American Academy of Child Psychiatry*, **21**, 392-397.

Pauls DL, Bucher KD, Crowe RR and Noyes R Jr (1980) A genetic study of panic disorder pedigrees. *American Journal of Human Genetics*, **32**, 639-644.

Robins LN and Regier DA (Eds) (1991) *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. The Free Press, New York.

Robins LN, Helzer JE, Weissman MM, Orvaschel H, Guenberg E, Burke JD and Regier DA (1984) Lifetime prevalence of specific psychiatric disorders in three sites. *Archives of General Psychiatry*, **41**, 949-958.

SAGE (1992) *Statistical Analysis for Genetic Epidemiology, Release 2.1*. Computer program package available from the Department of Biometry and Genetics, LSU Medical Center, New Orleans.

Torgersen S (1983) Genetic factors in anxiety disorders. *Archives of General Psychiatry*, **40**, 1085-1089.

Vieland VJ, Hodge SE and Greenberg DA (1992) Adequacy of single-locus approximations for linkage analyses of oligogenic traits. *Genetic Epidemiology*, **9**, 45-59.

Vieland VJ, Knowles J, Fyer A, Stefanovich M, Freimer NF, Lish J, Adams P, Woodley K, Rassnick H, Heiman G, White P, Das K, Klein DR, Ott J, Weissman MM and Gilliam TC (1993a) Linkage study of panic disorder: A preliminary report. In: *New Genetic Approaches to Mental Disorder* (Eds ES Gershon, CR Cloninger and JE Barrett). American Psychiatric Press, Washington, DC (in press).

Vieland VJ, Hodge SE and Greenberg DA (1993b) Adequacy of single-locus approximations for linkage analyses of oligogenic traits: Extension to arbitrary pedigree structures. *Human Heredity*, in press.

Weissman MM (1989) Evidence for comorbidity for anxiety and depression: Family and genetic studies of children. In: *Comorbidity in Anxiety and Mood Disorders* (Eds JD Maser and CR Cloninger). American Psychiatric Press, Washington, DC.

Weissman MM, Wickramaratne P, Adams P, Lish J, Horwath E, Charney D, Woods S, Leeman E and Frosch E (1992) The relationship between panic disorder and major depression: A new family study. *Archives of General Psychiatry*, in press.

(Received 9 December 1992; accepted as revised 26 February 1993)