

## THE GENETIC IMPLICATIONS OF AGE-DEPENDENT PENETRANCE IN MANIC-DEPRESSIVE ILLNESS

RAYMOND R. CROWE\*

Department of Psychiatry, University of Iowa, College of Medicine, Iowa City, IA 52242, U.S.A.  
and

PETER E. SMOUSE†

Department of Human Genetics, University of Michigan School of Medicine, Ann Arbor, MI 48104, U.S.A.

(Received 13 May 1975. Revised 10 June 1977)

### INTRODUCTION

THE DISCOVERY by ANGTS<sup>1,2</sup> and PERRIS<sup>3</sup> that manic-depressive illness (MDI) is genetically different from unipolar depression has led to renewed interest in the genetic predisposition to MDI. WINOKUR<sup>4-5</sup> and his collaborators proposed a sex-linked dominant (SLD) mode of inheritance, on the basis of familial transmission patterns in 61 pedigrees and on the basis of apparent linkage with X-chromosomal markers.<sup>6</sup> As expected with sex-linkage, father-son transmission was not observed, but all other forms of parent-child transmission were commonly encountered. This finding was strengthened by the observation that affected sisters of female probands outnumbered affected brothers, while the sex-ratio among sibs of male probands was essentially unity. WINOKUR<sup>7</sup> extended the series by 28 pedigrees, and obtained the same patterns as found in the initial series. TAYLOR and ABRAMS<sup>8</sup>, reporting a series of 50 pedigrees, essentially corroborated all of Winokur's findings in their 20 early onset patients (onset  $\leq 30$ ), but encountered a single case of father-son transmission in the late onset ( $> 30$ ) group. HELZER and WINKOUR<sup>9</sup> studied the pedigrees of an additional 30 male probands, and found one case of father-son transmission. These rare cases of father-son transmission are entirely compatible with the SLD hypothesis, because one would expect an occasional affected father to be married to an unaffected female carrier.

Other workers, however, have reported quite contradictory results. PERRIS<sup>10</sup> and ANGST<sup>1</sup> failed to find sex-ratios suggestive of sex-linkage, and moreover Perris encountered 13 instances of father-son transmission in 138 kindreds. GOETZL *et al.*<sup>11</sup> have reported comparable results, with four cases of father-son transmission among 39 pedigrees. JAMES and CHAPMAN<sup>12</sup> and VON GRIEFF *et al.*<sup>13</sup> have also reported several cases of father-son transmission. The frequency of father-son transmission is too high in these studies for a rare sex-linked dominant, and sex-ratios are more suggestive of an autosomal dominant (AD) than of a sex-linked dominant (SLD). MENDLEWICZ and RAINER<sup>14</sup> report sex-ratios similar to these of WINOKUR *et al.*<sup>4</sup>, but also report 10 cases of father-son transmission in a series of 134 kindreds. From this same series, however, MENDLEWICZ and FLEISS<sup>15</sup> isolated several

---

\*Supported by Foundations' Fund for Research in Psychiatry, Fellowship Grant No. 72-551.

†Supported by NSF-BMS-74-11823 and NIMH-1-R01MH28294-01.

kindreds informative for linkage analysis, and obtained strong evidence for linkage between MDI and the X-linked markers for Protanopia, Deuteranopia, and the Xg blood group. The results suggest a mixture of SLD and AD segregation. It seems clear that different workers are sampling from different populations of inference, but the cause(s) of this fact are not readily apparent.

Considerable ambiguity arises in the genetic analysis of MDI, which like many other psychiatric disorders, is plagued by variable age of onset. In order to determine the transmission mode of the illness, it is necessary that the investigator be able to gauge the proportion of relatives who possess the allele in question. Since any psychiatrically *normal* relative of an affected proband may actually carry the allele (incomplete penetrance), such unambiguous categorization of relatives is not possible with MDI. The objectives of this paper are: (1) to develop an age-dependent penetrance function for MDI, and (2) to demonstrate its utility in formal genetic analysis. The strategy is to obtain age-adjusted morbidity risks for all relatives, and to use this information in the pedigree analysis.

#### AGE-DEPENDENT PENETRANCE

Several methods have been proposed to deal with the age-of-onset problem; the basic strategy is to obtain age-corrected morbidity risks for relatives of affected probands (see SLATER and COWIE.<sup>16</sup> Perhaps the most powerful method is that devised by ELSTON,<sup>17</sup> who obtains a penetrance function, using information on every individual in the pedigree. The method is statistically optimal, but the function obtained depends upon the genetic hypothesis under consideration. We wish to avoid this feature of the analysis, and have therefore used only the information on probands to derive a penetrance function. While our approach is not statistically optimal, it is both straightforward and conservative.

We shall employ the 61 pedigrees reported in Appendix B of WINOKUR *et al.*<sup>4</sup> for the analyses described below; we are implicitly adopting Winokur's diagnostic and sampling criteria, and our genetic conclusions will reflect this fact. The 61 probands range in age-of-onset from 16 to 66 yr. We have described the cumulative proportion of those *first affected by age* ( $x$ ) in Fig. 1. Separate plots for the two sexes showed no systematic differences, and in view of the small sample size ( $N=61$ ), we have arbitrarily pooled the data from both sexes.

The general shape of the cumulative onset fraction suggests an exponential life distribution for the probability of first onset by age  $x$

$$F(x) = \begin{cases} 1 - \exp[-\lambda(X-K)] & K < X < \infty \\ 0 & 0 < X < K \end{cases} \quad (1)$$

where  $K$  is the earliest age of onset possible, and  $\lambda$  is the rate parameter.

The type of function described by (1) is commonly referred to as a "one-hit" curve, and is generally used to describe phenomena where the number of "hits" in a small time interval is Poisson distributed, and where a single hit is sufficient to trigger the change observed. We might view a "hit" as a physiological stimulus which shifts a susceptible genotype from the normal to the abnormal state. The significance of the threshold age ( $K$ ) may be that the individual must first be "primed", before the stimulus can be effective. The coincidence of this "threshold" with puberty suggests hormonal priming, but we have no hard evidence on this point. We shall refer to  $F(x)$  as the *age-dependent penetrance function*. We hasten to add

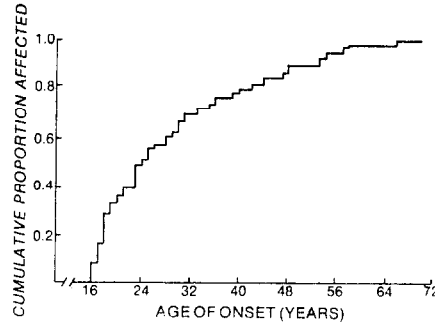


FIG. 1. Observed cumulative age of onset distribution for manic depressive illness (MDI).

that  $F(x)$  does not explain the age-of-onset pattern; rather it describes that pattern. Suffice it to say that a certain fraction of the individuals who possess the gene (whatever it is) can be expected to first manifest the illness by age ( $x$ ), and that this fraction is predictable in a regular fashion.

There are two important features of the problem not allowed for in Fig. 1, and both need to be dealt with. First, the relationship shown is a *conditional* one, since all individuals characterized have been chosen precisely because they have *already* manifested the illness; we have an ascertainment bias. These data on probands provide no information on those individuals who have the susceptible genotype, but who have yet to manifest MDI by age 66. That there exist such individuals in the population is proven by relatives of Winokur's probands who first manifest after this age. Second, we have ignored the underlying demographic attrition of the population. Older age classes represent a small fraction of the total population, and might alone account for the small number of probands who manifest at older ages. Fortunately, both of these difficulties may be circumvented.

We assume that the demographic vital rates of susceptible individuals are the same as those for the population at large, at least until onset of the illness. Given this assumption, we may describe the age-of-onset distribution by

$$g(x) = \begin{cases} f(x)c(x) & K < X < \infty \\ = 0 & 0 < X < K, \end{cases} \quad (2)$$

where  $f(x) = \lambda \exp[-\lambda(x-k)] = F'(x)$ , and where  $c(x) = b \cdot l(x) \exp[-rx]$  is the stable age-distribution of the population at large. The parameter ( $r$ ) is the intrinsic growth rate;  $l(x)$  is the survivorship function; and ( $b$ ) is a normalizing constant. The cumulative age-of-onset distribution is therefore

$$G(x) = \begin{cases} \int_0^x g(x) dx = \lambda b e^{\lambda k} \int_K^x e^{-(\lambda+r)x} l(x) dx & K < X < \infty \\ = 0 & 0 < X < K, \end{cases} \quad (3)$$

or (replacing integrals by summations)

$$G(x) = \lambda b e^{\lambda K} \sum_{k=K}^{x-1} e^{-(\lambda+r)x} \ell(x) \quad K \leq X < \infty \tag{4}$$

$$= 0 \quad 0 \leq X < K.$$

We use  $(x-1)$ , since one is 19 in the 20th year of life, etc. Equation (4) may be scaled as a fraction of the total less than some upper age ( $u$ )

$$G(x|x < u) = G(x) \div G(u) \quad K \leq X < \infty$$

$$= 0 \quad 0 \leq X < K. \tag{5}$$

The proportion of individuals with the susceptible genotype expected to manifest (for the first time) between the ages ( $x_1$ ) and ( $x_2$ ) is thus

$$Pr(x_1 < x < x_2 | x_2 < u) = G(x_2 | x_2 < u) - G(x_1 | x_1 < u). \tag{6}$$

We have employed the  $\ell(x)$  schedule and  $r$ -value ( $0.0103 \text{ yr}^{-1}$ ) from the 1965 Census for Caucasians. We were unable to obtain comparable data for the St. Louis area in 1965 (the time and location of Winokur's study), but an examination of the 1960 and 1970 Census figures shows that the U.S. Census results are quite similar to those for St. Louis. The Census estimates are so much more precise than the data used here ( $N=61$ ) that we shall treat  $\ell(x)$  and  $r$  as known without error.

The limit of age resolution in Winokur's data is about one year, and we have arbitrarily divided the onset data into seven (7) intervals (0-19, 20-24, 25-29, 30-39, 40-49, 50-59, 60-69) for statistical purposes. The intervals are coarser than the data permit, but with the limited number of observations at our disposal, finer classification is scarcely warranted. The first interval spans the no expression period, and really contains the age group  $[K-19]$ . If the probands may be considered a random sample from the population of inference, i.e. the population of affected individuals, then the logarithm of the likelihood function (except for a combinatorial constant) is given by

$$\text{Log}_e L(P|N) \propto \sum_{i=1}^7 N_i \text{Log}_e p_i, \tag{7}$$

where the  $P_i$  are the expected proportions of first onsets in the seven age-class intervals, and the  $N_i$  are the observed numbers. The probabilities  $P_i$  are determined by  $\lambda$  and  $K$ , via (6). The strategy is to maximize (7) relative to  $\lambda$  and  $K$ , and this is accomplished by simple search procedure. Since we are using one-year age intervals, we are unable to make fine distinctions in  $K$ , and have considered all integer values between 0 and 19. In Table 1, we list the estimated values of  $\lambda$  and the  $P_i$ , and the corresponding values of (7) for the range of  $K$ -values considered.

We may also define  $P_i = [N_i \div N]$ , the observed proportions in the various age classes. The difference between (7), using the estimated values of  $P_i$  and (7), using the observed  $P_i$ , is a measure of the adequacy of the description. In particular,  $\Lambda = 2[\text{Log}_e L(\hat{P}/N) - \text{Log}_e L(\tilde{P}/N)]$  is asymptotically distributed as  $\chi^2$  with  $(I-3) = 4 \text{ df}$ . For the best set of estimated parameters ( $K=14, \hat{\lambda}=0.0407$ ), there is very close agreement of model to observation ( $\Lambda=1.24$ ).

Given  $\hat{\lambda}$  and  $\hat{K}$ , we may extract the *unconditional* penetrance function  $F(x)$  from (1), and this is listed in Table 2, up to age 100. The reader should bear in mind that the figures are estimates from a small data set ( $N=61$ ). We present the results to four decimal places for

TABLE 1. ESTIMATED VALUES FOR  $\lambda_i$ , ESTIMATED CLASS FREQUENCIES ( $p_i$ ), AND VALUES OF  $\log_e L$  FOR VARIOUS VALUES OF  $K$ =EARLIEST AGE OF ONSET

K	$\lambda$	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>	P <sub>6</sub>	P <sub>7</sub>	$\log_e L$
1	0.0074	0.438	0.087	0.079	0.137	0.111	0.087	0.061	-111.47
2	0.0090	0.435	0.090	0.081	0.139	0.111	0.085	0.059	-111.06
3	0.0107	0.431	0.093	0.083	0.141	0.111	0.084	0.057	-110.64
4	0.0124	0.427	0.096	0.086	0.143	0.111	0.082	0.055	-110.21
5	0.0143	0.423	0.100	0.088	0.145	0.111	0.080	0.053	-109.78
6	0.0164	0.417	0.105	0.091	0.148	0.110	0.078	0.051	-109.34
7	0.0186	0.410	0.110	0.094	0.151	0.110	0.077	0.048	-108.90
8	0.0209	0.402	0.115	0.098	0.154	0.110	0.075	0.046	-108.47
9	0.0234	0.394	0.121	0.102	0.157	0.109	0.073	0.044	-108.05
10	0.0262	0.384	0.128	0.107	0.161	0.109	0.070	0.041	-107.66
11	0.0291	0.371	0.137	0.112	0.165	0.108	0.068	0.039	-107.31
12	0.0324	0.359	0.146	0.117	0.169	0.108	0.065	0.036	-107.02
13	0.0359	0.342	0.157	0.124	0.174	0.107	0.063	0.033	-106.82
14	0.0407	0.323	0.169	0.131	0.180	0.106	0.060	0.031	-106.78
15	0.0440	0.300	0.184	0.140	0.186	0.105	0.057	0.028	-106.96
16	0.0488	0.273	0.202	0.150	0.192	0.104	0.054	0.025	-107.49
17	0.0541	0.241	0.224	0.161	0.200	0.102	0.050	0.022	-108.59
18	0.0601	0.201	0.251	0.175	0.208	0.100	0.046	0.019	-110.72
19	0.0669	0.148	0.285	0.192	0.218	0.098	0.042	0.017	-114.90
Observed	—	0.328	0.180	0.115	0.164	0.115	0.82	0.16	-106.16

computational convenience only. The figures beyond  $x=70$  are clearly derived from extrapolation, and the important point is that penetrance is still only about 90% at this age. The proper cure for imprecise estimation is a much larger data base, and we anticipate that such will be forthcoming in the not too distant future.

#### PEDIGREE ANALYSIS

We shall now demonstrate the usage of this penetrance function for purposes of formal pedigree analysis. As indicated earlier, an unambiguous mode of transmission has not been established. The sex-linked dominant (SLD) and autosomal dominant (AD) hypotheses are the most likely candidates for the present pedigrees, and we shall limit formal treatment to these two alternatives. We shall, however, briefly comment on the sex-linked recessive (SLR), autosomal recessive (AR) and general single major locus model (SML) and polygenic (PG) hypotheses.

The details of the 61 pedigrees used here are to be found in Appendix B of WINOKUR *et al.*<sup>4</sup>, and we shall not recapitulate that information here. It is convenient to divide the pedigrees into eight different classes, relative to the parentage of probands (Fig. 2). Only two sorts of pedigrees are encountered for children of probands; (a) a male proband, spouse and children, or (b) a female proband, spouse and children. All spouses of probands were normal. Most of the pedigrees are of the two-generation type, containing either proband, parents and sibs, or proband and children, but some three-generation pedigrees are also present. Relatives of the proband may be either normal or affected, and we follow the usual convention of counting unipolar relatives as affected. Since penetrance is incomplete [ $F(x) < 1$  if  $x < \infty$ ], a psychiatrically normal individual may either have the susceptible genotype or not, and we must consider both possibilities. Suppose we have two sibs of an

affected proband, one (*A*) normal and age  $x_A$ , and one (*B*) affected and age  $x_B$ . The probabilities of obtaining these two phenotypes are

$$\Pr(A \text{ normal} | x_A) = \Pr(A \text{ susceptible} | H) \cdot [1 - F(x_A)] + \Pr(A \text{ not susceptible} | H) \tag{8}$$

$$\Pr(B \text{ affected} | x_B) = \Pr(B \text{ susceptible} | H) \cdot F(x_B),$$

where (*H*) is the genetic hypothesis under consideration. The probability of a susceptible genotype depends both upon the genetic hypothesis (*H*) and upon the mating type (Fig. 2) of the parents of the proband (and sibs). The likelihood of a pair of sibs (*A, B*) is simply the product of their separate probabilities.

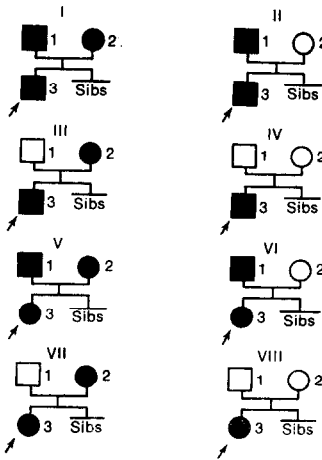


FIG. 2. Possible pedigree types for male and female probands.

The likelihood of any genetic hypothesis (*H*), given the 61 pedigrees at hand, can be written as the product of the separate likelihoods for the 61 pedigrees

$$L(H | 61 \text{ Pedigrees}) = \prod_{j=1}^{J=61} L(H | \text{Pedigree } j). \tag{9}$$

To compare the two hypotheses, we simply compute the likelihood ratio for the two hypotheses

$$\lambda = \frac{L(\text{SLD} | 61 \text{ Pedigrees})}{L(\text{AD} | 61 \text{ Pedigrees})} = \prod_{j=1}^{J=61} \frac{L(\text{SLD} | \text{Pedigree } j)}{L(\text{AD} | \text{Pedigree } j)}. \tag{10}$$

At this point in the development, the prevalence of the disorder becomes a consideration in the statistical analysis, since the likelihoods of various mating types depend on allele frequency. By way of example, pedigree type VIII of Fig. 2 implies one of the mating types of Table 3a, under the SLD hypothesis, and one of the mating types of Table 3b, under the

AD hypothesis. The relative frequencies of the different parental combinations are highly dependent on allele frequency [ $f(A)=P$  for SLD;  $f(B)=Q$  for AD].

TABLE 2. AGE DEPENDENT PENETRANCE [ $F(x) = 1 - \exp(-\lambda \{x - K\})$ ] FOR MANIC DEPRESSIVE ILLNESS

Age (x)*	F(x)	Age(x)	F(x)	Age(x)	F(x)	Age(x)	F(x)
0-13	0.0000	35	0.5916	57	0.8332	79	0.9319
14	0.0399	36	0.6078	58	0.8398	80	0.9346
15	0.0782	37	0.6235	59	0.8462	81	0.9372
16	0.1149	38	0.6385	60	0.8523	82	0.9396
17	0.1502	39	0.6529	61	0.8582	83	0.9421
18	0.1841	40	0.6668	62	0.8639	84	0.9444
19	0.2167	41	0.6800	63	0.8693	85	0.9466
20	0.2479	42	0.6928	64	0.8745	86	0.9488
21	0.2779	43	0.7051	65	0.8795	87	0.9508
22	0.3067	44	0.7168	66	0.8843	88	0.9528
23	0.3344	45	0.7281	67	0.8890	89	0.9546
24	0.3609	46	0.7390	68	0.8934	90	0.9564
25	0.3864	47	0.7494	69	0.8976	91	0.9582
26	0.4109	48	0.7594	70	0.9017	92	0.9598
27	0.4344	49	0.7690	71	0.9056	93	0.9614
28	0.4569	50	0.7782	72	0.9094	94	0.9630
29	0.4786	51	0.7870	73	0.9130	95	0.9645
30	0.4994	52	0.7955	74	0.9165	96	0.9659
31	0.5193	53	0.8037	75	0.9198	97	0.9672
32	0.5385	54	0.8115	76	0.9230	98	0.9686
33	0.5569	55	0.8190	77	0.9261	99	0.9698
34	0.5746	56	0.8262	78	0.9290	100	0.9710
			$\lambda=0.0407$	$K=14$			

\*One is 14 years of age in the 15th year of life, etc. F(14) is really  $\Pr(x < 15)$ , etc.

TABLE 3. MATING TYPES, POPULATION INCIDENCE, AND TRANSMISSION RATES FOR PEDIGREE TYPE VIII OF FIG. 3, UNDER SEX LINKED DOMINANT AND AUTOSOMAL DOMINANT INHERITANCE

Parental Combination	Population Incidence	Probability of Affected Daughter
<b>(a) Sex Linked Dominant</b>		
(1) AY × AA	$P^3[1-F(1)] [1-F(2)]$	1·F(3)
(2) AY × Aa	$2P^2(1-P) [1-F(1)] [1-F(2)]$	1·F(3)
(3) AY × aa	$P(1-P)^2[1-F(1)]$	1·F(3)
(4) aY × Aa	$2P(1-P)^2[1-F(2)]$	1/2·F(3)
(5) aY × AA	$P^2(1-P) [1-F(2)]$	1·F(3)
<b>(b) Autosomal Dominant</b>		
(1) BB × BB	$Q^4[1-F(1)] [1-F(2)]$	1·F(3)
(2) BB × Bb	$2Q^3(1-Q) [1-F(1)] [1-F(2)]$	1·F(3)
(3) Bb × BB	$2Q^3(1-Q) [1-F(1)] [1-F(2)]$	1·F(3)
(4) Bb × Bb	$4Q^2(1-Q)^2[1-F(1)] [1-F(2)]$	3/4·F(3)
(5) BB × bb	$Q^2(1-Q)^2[1-F(1)]$	1·F(3)
(6) bb × BB	$Q^2(1-Q)^2[1-F(2)]$	1·F(3)
(7) Bb × bb	$2Q(1-Q)^3[1-F(1)]$	1/2·F(3)
(8) bb × Bb	$2Q(1-Q)^3[1-F(2)]$	1/2·F(3)

To deal with the analysis in all its generality, even for  $P$  or  $Q$  known, would be a tedious task. If we must also estimate population incidence from the data, the job becomes extremely unwieldy, particularly since the estimates of  $P$  and  $Q$  will almost inevitably differ. While the task is tractable, it is unattractive, and we therefore introduce a simplification at this point.

HELGASON<sup>18</sup> presents results of an extensive Icelandic survey of the prevalence of various psychiatric disorders. Using Helgason's age-adjusted morbidity risks, we obtain a crude estimate for the prevalence of the gene for MDI of less than 1%. While other workers obtain slightly varying figures, it appears that the allele frequency is in the neighborhood of  $P \approx 0.01$ . If we accept this figure as approximately correct, then it is easily seen that mating types (a2) and (a5) are two orders of magnitude rarer than (a3) and (a4) and that mating type (a1) is two orders of magnitude rarer still. No significant error is introduced by ignoring these rarer possibilities, and we may therefore approximate the relative probabilities of (a3) and (a4) by

$$\begin{aligned} \Pr(a3) &= [\Pr(\text{father-unaffected carrier}) \times \\ &\quad \Pr(\text{transmission to daughter}) \times \Pr(\text{daughter manifests})] \div \\ &\quad [\Pr(\text{father-unaffected carrier}) \times \\ &\quad \Pr(\text{transmission to daughter}) \times \Pr(\text{daughter manifests}) \\ &+ \Pr(\text{mother-unaffected carrier}) \times \\ &\quad \Pr(\text{transmission to daughter}) \times \Pr(\text{daughter manifests})] \\ &= \frac{P(1-P)^2[1-F(1)] \cdot F(3)}{P(1-P)^2[1-F(1)+1-F(2)] \cdot F(3)} \\ &= \frac{[1-F(1)]}{[1-F(1)]+[1-F(2)]} \end{aligned} \quad (11)$$

$$P(a4) = \frac{[1-F(2)]}{[1-F(1)]+[1-F(2)]}$$

and we sidestep the necessity to estimate  $P$ . Similar arguments for  $Q \approx 0.01$  eliminate all but (b7) and (b8) from consideration, and one obtains

$$P(b7) = \frac{[1-F(1)]}{[1-F(1)]+[1-F(2)]}$$

$$P(b8) = \frac{[1-F(2)]}{[1-F(1)]+[1-F(2)]}. \quad (12)$$

A similar strategy may be used for each of the other pedigree types of Fig. 2; one assumes the most likely mating types, and ignores those orders of magnitude rarer (Table 4). The error introduced by this simplification is negligible, and is well worth the saving of computational efforts.

The individual family contributions to  $\lambda$  are listed in Table 5. There are 55 informative pedigrees (both parents assayed, along with sibs and/or children), 19 of which favor the SLD hypothesis, 10 of which favor the AD hypothesis, and 26 of which favor neither. No one pedigree is very informative but in the aggregate ( $\lambda = 88.92$ ) the SLD hypothesis is about 89 times more likely than the AD hypothesis.

We have established that the SLD hypothesis is more likely than the AD hypothesis, but we have yet to establish that it is a *good* description of the situation. This is most easily shown in another fashion. Given either hypothesis, we may compare expected and observed morbidity rates for various classes of relatives. The availability of an age specific penetrance function permits the expectations to be tailored to the ages of relatives actually encountered.



TABLE 4. ASSUMED PARENTAL GENOTYPES FOR CLASSES OF PEDIGREES SHOWN IN FIG. 2, UNDER SEX-LINKED DOMINANT (SLD) AND AUTOSOMAL DOMINANT (AD) HYPOTHESES

Type of Pedigree	Pedigree		Proband (3)	SLD		AD	
	♂	♀		Male(1)	Female(2)	Male(1)	Female(2)
I	■	●	male	AY	Aa	Bb	Bb
II	■	○	male	AY	Aa	Bb	bb
III	□	●	male	aY	Aa	bb	Bb
IV	□	○	male	aY	Aa	{ Bb bb }	{ Bb Bb }
V	■	●	female	AY	Aa	Bb	Bb
VI	■	○	female	AY	aa	Bb	bb
VII	□	●	female	aY	Aa	bb	Bb
VIII	□	○	female	AY	aa	{ Bb bb }	{ Bb Bb }

TABLE 5. LIKELIHOODS OF SEX-LINKED DOMINANT (SLD) AND AUTOSOMAL DOMINANT (AD) HYPOTHESES, WITH LIKELIHOOD RATIOS ( $\lambda$ ) FOR 61 PEDIGREES

DRAWN FROM APPENDIX B OF WINOKUR <i>et al.</i> (1969)									
Pedigree ID & Type*	Likelihood		Ratio ( $\lambda$ )	Pedigree ID & Type	Likelihood		Ratio ( $\lambda$ )		
	SLD	AD			SLD	AD			
1	-III	0.8404	0.8404	1.0000	32	-IV	0.0625	0.0294	2.1258
2	-?	ambiguous parentage		—†	33	-IV	0.0484	0.0175	2.7718
3	-VII	0.8916	0.8916	1.0000	34	-VII	0.7308	0.7308	1.0000
4	-VIII	0.8411	0.7216	1.1656	35	-IV	0.7607	0.7607	1.0000
5	-IV	1.0000	0.8387	1.1923	36	-III	0.1780	0.1780	1.0000
6	-IV	0.1200	0.1318	0.9109	37	-?	ambiguous parentage		—
7	-VII	0.0271	0.0271	1.0000	38	-III	0.6474	0.6474	1.0000
8	-VII	0.0151	0.0151	1.0000	39	-VIII	0.0074	0.0111	0.6665
9	-V	0.0603	0.0915	0.6590	40	-VIII	0.1560	0.1220	1.2787
10	-III	0.3630	0.3630	1.0000	41	-III	1.0000	1.0000	1.0000
11	-III	0.3036	0.2862	1.0610	42	-VIII	0.1093	0.0818	1.3362
12	-VII	0.3118	0.3118	1.0000	43	-V	0.2198	0.3439	0.6391
13	-III	0.2435	0.2133	1.1415	44	-VI	0.5385	0.2359	2.2830
14	-VI	0.8439	0.8439	1.0000	45	-VIII	0.4346	0.2240	1.9403
15	-?	ambiguous parentage		—	46	-VIII	0.4365	0.1581	2.7609
16	-III	not informative		—	47	-III	0.1578	0.1041	1.5152
17	-VIII	0.3274	0.2997	1.0924	48	-VII	0.2529	0.2529	1.0000
18	-III	0.5855	0.5855	1.0000	49	-IV	0.6025	0.6280	0.9593
19	-III	0.0125	0.0118	1.0609	50	-VII	0.6882	0.6882	1.0000
20	-VII	0.1804	0.1804	1.0000	51	-III	not informative		—
21	-VIII	0.3129	0.1932	1.6196	52	-VI	0.6136	0.6934	0.8850
22	-VIII	0.0007	0.0006	1.0746	53	-III	0.0888	0.0837	1.0610
23	-VII	0.4190	0.4190	1.0000	54	-VIII	0.0015	0.0020	0.7487
24	-VI	0.3068	0.4358	0.7040	55	-IV	0.9064	0.9064	1.0000
25	-IV	not informative		—	56	-III	0.5284	0.5284	1.0000
26	-VII	0.5410	0.5410	1.0000	57	-IV	1.0000	0.5236	1.9099
27	-VII	0.8916	0.8916	1.0000	58	-VII	0.1548	0.1548	1.0000
28	-VIII	0.1014	0.1506	0.6733	59	-IV	0.0509	0.0593	0.8585
29	-III	0.1828	0.1828	1.0000	60	-IV	0.5555	0.5555	1.0000
30	-VII	0.5507	0.5507	1.0000	61	-VI	0.7521	0.7180	1.0475
31	-VIII	0.8086	0.8068	1.0000					

J=61

\*Pedigree types shown in Figure 2  $\lambda = \prod_{j=1}^n \lambda_j = 88.92$

†Psychiatric status unknown.

Given a genetic hypothesis, the morbidity risks for a sib ( $S$ ) or a child ( $C$ ) are

$$\begin{aligned} \Pr(S \text{ affected}) &= \Pr(\text{mating type} | \text{proband}) \times \\ &\quad \Pr(\text{transmission to } S) \times \Pr(S \text{ manifests}) \\ \Pr(C \text{ affected}) &= \Pr(\text{transmission to } C) \times \Pr(C \text{ manifests}). \end{aligned} \quad (13)$$

By adding the morbidity risks for all relatives in a given class, we obtain the expected number of affected relatives in that class. The results for both hypotheses are shown in Table 6. Simple inspection will show close agreement of either hypothesis to observation; we attribute the close match to the efficacy of the penetrance function, and remind the reader that these relatives were not used to define the function. A likelihood ratio  $\chi^2$  comparison of expectation and observation may be computed for either hypothesis, and has 8 degrees of freedom, as indicated at the bottom of Table 6. The computed value is 3.92 for the SLD hypothesis and 12.56 for the AD hypothesis. Neither value is significant, but that for the SLD hypothesis is smaller (as expected). The SLD model is not only the better of the two, it is a *very good* description of the situation.

We turn finally to the AR, SLR, SML, and PG hypotheses. The AR hypothesis is very much less likely than either the AD or SLD hypotheses, because the number of affected parents (37) is far too high, given any reasonable value for allelic frequency. The SLR hypothesis is also orders of magnitude less likely, and cannot account for the preponderance of affected females among probands and first degree relatives. The general single major locus model (SML) described by KIDD and CAVALLI-SFORZA<sup>19</sup> involves an autosomal locus with separate penetrance relations for each of the three genotypes. In view of the closeness of expectation and observation for the SLD model, these additional refinements would seem to be unwarranted. We are left with the polygenic (PG) hypothesis, which yields heritabilities in the neighborhood of unity (or even higher)<sup>12,14</sup>. While we cannot rule out the polygenic hypothesis, there really seems to be no particular reason to invoke it. It is worth repeating here that our conclusion was largely predetermined by the choice of Winokur's material, and we make no claims relative to other studies. We submit, however, that the same basic analytic strategy may be applied elsewhere.

#### DISCUSSION

It is appropriate at this juncture to examine several assumptions made in the derivation of the penetrance function. We have assumed that the age distribution of individuals with the susceptible genotype is the same as that  $c(x)$  of the general population. It is known that mortality increases and fertility decreases *after* onset of the illness, but information on these factors *prior* to onset is lacking<sup>20,21</sup>. Since we are dealing with the *distribution of the age of first onset*, we have assumed the usual  $c(x)$  schedule. To the extent that mortality is increased, the  $\ell(x)$  schedule we have used is not as steep as it should be, resulting in an inflated estimate of  $(\lambda)$ . To the extent that fertility is decreased, the  $r$ -value we have used is too large, resulting in an underestimate of  $(\lambda)$ . The net result is not entirely predictable, but would probably be partial compensation of the one bias for the other. We have also assumed an equilibrium age-distribution which applies to both sexes. An examination of the census figures for 1965 will show: (a) that the age-distributions of the two sexes are different, particularly at advanced ages, and (b) that equilibrium does not obtain.

TABLE 6. OBSERVED AND EXPECTED NUMBERS OF NORMAL AND AFFECTED RELATIVES OF PROBANDS, UNDER SEX-LINKED DOMINANT (SLD) AND AUTOSOMAL DOMINANT (AD) HYPOTHESES

Male Probands (n=26)			Female Probands (n=35)	
Normal	Affected		Normal	Affected
		<i>Brothers</i>		
20.73	8.27	Expt. (SLD)	28.34	6.66
22	7	Obs.	30	5
20.73	8.27	Expt. (AD)	25.35	9.65
		<i>Sisters</i>		
15.96	8.04	Expt. (ALD)	28.28	15.72
17	7	Obs.	30	14
15.96	8.04	Expt. (AD)	32.06	11.94
		<i>Sons</i>		
15.00	0.00	Expt. (oLD)	10.08	1.92
15	0	Obs.	10	2
13.23	1.77	Expt. (AD)	10.15	1.85
		<i>Daughters</i>		
5.66	1.34	Expt. (ALD)	9.08	1.92
5	2	Obs.	7	4
6.33	0.67	Expt. (AD)	9.10	1.90
$\Lambda(\text{SLD})=3.92 \sim \chi^2_8$ $0.75 < P < 0.90$			$\Lambda(\text{AD})=12.56 \sim \chi^2_8$ $0.10 < P < 0.25$	

We have also assumed for all the preceding that the probands were ascertained at first onset. This is not entirely justified, since some of Winokur's patients were referred during subsequent episodes. There is an upward bias (probably small) in our age of onset function due to the fact that some demographic attrition may have occurred between onset and referral. We have ignored this potential complication because the analytical alternatives involved additional assumptions we know to be seriously violated. The fact that the penetrance function (estimated from the probands) fits the relatives (sampled without bias) so closely reassures us that any such bias is small. It would probably be better to delete all but those pedigrees indexed by a proband referred at first onset, but with such limited material, we could not afford to further reduce the sample size. Until such data is available, the present penetrance function seems to provide a reasonably realistic working model.

There is also a suggestion that  $F(x)$  may differ for the two sexes<sup>16</sup>. We have conducted more elaborate analyses on separate sexes, using empiric age-distributions or separate  $l(x)$  schedules. There is some evidence that males have a larger  $K$ -value and a lower  $\lambda$ -value, but the differences are small. In view of the smaller sample sizes employed for separate sexes (35 female, 26 male), and in view of the extra complications of the more elaborate analysis, we have opted for simplicity. Moreover, the analysis described earlier provided the best description of the data at hand. A careful examination of the possible sex differential will eventually be necessary, using a much larger series of probands, but the data at hand are simply inadequate for accurate resolution. For the present, we feel justified in presenting the simplified treatment as an approximation to reality.

The decision to treat unipolar relatives of bipolar probands as "affected" is worth a final comment. It is conceivable that the unipolar manifestation of this syndrome has a different penetrance function from the bipolar manifestation. As mentioned earlier, as is the usual convention, we have assumed that this is *not* the case. The fact that the SLD model, with

penetrance function as formulated, provides an almost perfect fit to the actual morbidity risks (Table 6), is strong support for this approach. If one were to compute a separate penetrance function for the unipolar and bipolar cases, the estimation strategy we have used would have to be abandoned (there are no unipolar probands included). In view of the results, we see no reason to do so, and would only add in passing that the usual practice of lumping unipolar and bipolar relatives of bipolar probands into a single "affected" class appears to be quite reasonable.

The choice of a "one-hit" model to describe penetrance was dictated by considerations of simplicity, mathematical convenience, and a certain amount of trial and error model fitting. Any attempt to rationalize this choice in terms of current knowledge about MDI would be highly speculative. Suffice it to say that the model has been descriptive and predictive utility. It is noteworthy that equation 1 may be viewed as a special case of the more general class of penetrance functions considered by BURCH.<sup>22</sup> We have illustrated this utility for pedigree analysis, but other problems may be attacked in similar fashion; we shall briefly mention two. The penetrance function should contribute materially to the sort of linkage analysis described by MENDELWICZ and FLEISS<sup>15</sup>, since normal relatives of affected probands are genetically ambiguous. The penetrance function may also be used for purposes of genetic counseling, either to predict the probability of future morbidity for a currently normal individual, or to predict the probability that an unborn child will inherit the gene from a phenotypically normal parent who is at risk.

#### SUMMARY

An age dependent penetrance function was derived for manic-depressive illness, using age-of-onset data from sixty-one affected probands. The function used was a one-hit model, with earliest age-of-onset at about 14 yr, and a steadily increasing probability of manifesting the illness thereafter. The utility of the penetrance function for pedigree analysis was illustrated, using the families of the sixty-one probands. A sex-linked dominant model of inheritance was about eighty-nine times more likely than an autosomal dominant model, and both were far more likely than autosomal recessive or sex-linked recessive models. The more general single major locus model and the polygenic model cannot be ruled out, but would seem to be unnecessarily over-parameterized for the data at hand. The sex-linked dominant and autosomal dominant models were also compared, by means of the age specific morbidity risks and sibs and children. Both models provided a fairly close fit of expectation and observation, but the sex-linked model was preferable. Although the genetic conclusions cannot automatically be applied to other material, the analytical techniques should be useful elsewhere. Other uses of the penetrance function were indicated.

*Acknowledgements*—The authors wish to acknowledge the encouragement of Dr. GEORGE WINOKUR, who was kind enough to make his pedigree data available. Thanks are also due Dr. JAMES NEEL and Dr. CHARLES SING for critical comments on the manuscript. Finally, we should like to express our appreciation to an anonymous reviewer, whose diligent efforts have materially improved the paper. The sole responsibility for the treatment and results reported here remains our own.

#### REFERENCES

1. ANGST, J. *Zur Aetiologie und Nosologie Endogener Depressiver Psychosen*. Springer, Berlin, 1966.

2. ANGST, J. and PERRIS, C. Zur nosologie endogener depressionen. *Arch. Psychiat. Zschr. Neurol.* **210**, 373, 1968. Reprinted in English in *Int. J. ment. Hlth* **1**, 145, 1972.
3. PERRIS, C. A study of bipolar (manic-depressive) and unipolar recurrent depressive psychoses. *Acta psychiat. scand. Suppl.* **194**, 1966.
4. WINOKUR, G., CLAYTON, P. J. and REICH, T. *Manic Depressive Illness*. C. V. Mosby, St. Louis, 1969.
5. REICH, T., CLAYTON, P. J. and WINOKUR, G. Family history studies—V: The genetics of mania. *Am. J. Psychiat.* **125**, 1358, 1969.
6. WINOKUR, G. and TANNA, V. L. Possible role of X-linked dominant factor in manic-depressive disease. *Dis. nerv. Syst.* **30**, 89, 1969.
7. WINOKUR, G. Genetic findings and methodological considerations in manic depressive disease. *Br. J. Psychiat.* **117**, 267, 1970.
8. TAYLOR, M. and ABRAMS, R. Manic states. *Archs gen. Psychiat.* **28**, 656, 1973.
9. HELZER, J. E. and WINOKUR, G. A family interview study of male manic depressives. *Archs gen. Psychiat.* **31**, 73, 1974.
10. PERRIS, C. Genetic transmission of depressive psychoses. *Acta psychiat. scand. Suppl.* **203**, 45, 1968.
11. GOETZL, U., GREEN, R., WHYBROW, P. and JACKSON, R. X linkage revisited. *Archs gen. Psychiat.* **31**, 665, 1974.
12. JAMES, N. MCI. and CHAPMAN, C. J. A genetical study of bipolar affective disorder. *Br. J. Psychiat.* **126**, 449, 1975.
13. VON GRIEFF, H., MCHUGH, P. R. and STOKES, P. E. The familial history in sixteen males with bipolar manic-depressive illness. In: *Genetic Research in Psychiatry*, FIEVE, R. R., ROSENTHAL, D. and BRILL, H. (Editors), pp. 233–239. Johns Hopkins University Press, Baltimore, 1975.
14. MENDLEWICZ, J. and RAINER, J. D. Morbidity risk and genetic transmission in manic-depressive illness. *Am. J. hum. Genet.* **26**, 692, 1974.
15. MENDLEWICZ, J. and FLEISS, J. L. Linkage studies with X-chromosome markers in bipolar (manic-depressive) and unipolar (depressive) illnesses. *Biol. Psychiat.* **9**, 261, 1974.
16. SLATER, E. and COWIE, V. *The Genetics of Mental Disorders*. Oxford University Press, London, 1971.
17. ELSTON, R. C. Ascertainment and age of onset in pedigree analysis. *Hum. Hered.* **23**, 105, 1973.
18. HELGASON, T. Epidemiology of mental disorders in Iceland. *Acta psychiat. scand. Suppl.* **173**, 1964.
19. KIDD, K. K. and CAVALLI-SFORZA, L. L. An analysis of the genetics of schizophrenia. *Soc. Biol.* **20**, 254, 1973.
20. GUZE, S. B. and ROBINS, E. Suicide and primary affective disorders. *Br. J. Psychiat.* **117**, 437, 1970.
21. SLATER, E., HARE, E. H. and PRICE, J. S. Marriage and fertility of psychiatric patients compared with national data. *Soc. Biol.* **V. 18**, *Suppl.* **1**, p. S60, 1971.
22. BURCH, P. R. J. Twin studies in relation to the mechanism of age dependent penetrance. *Acta Med. Auxologia* **8**, 87–106, 1976.