Refutation of the General Single-Locus Model for the Etiology of Schizophrenia

D. H. O'Rourke,1,2 I. I. Gottesman,1,3 B. K. Suarez,1,3 J. Rice,1,4 and T. Reich1,3

SUMMARY

All published studies on the familial incidence of schizophrenia appropriate for testing the applicability of the general single-locus two-allele model are examined under the assumption of a unitary etiology for all schizophrenia. We show that the single major locus model is inadequate to predict the incidence in four classes of relatives of schizophrenic probands (parents, siblings, monozygotic, and dizygotic cotwins). In addition, the observed proportion of affected offspring from dual matings differ significantly from the model's prediction.

The lack of an overall fit between the published familial distributions and the monogenic model suggests that a single major locus is insufficient for the etiology of schizophrenia. Further efforts in examining multifactorial models, mixed models, and other transmission models may be fruitful.

INTRODUCTION

Since the pioneering family study of schizophrenia by Rüdin [1] in Germany, psychiatrists and geneticists alike have sought the etiology of this complex disorder. Although the notion of genetic involvement in the etiology of schizophrenia has been questioned (e.g., [2, 3]), the higher rate of schizophrenia observed in first-degree relatives of schizophrenic probands, the findings of adoption and cross-fostering studies [4–9], and numerous twin studies [10] conducted during the...
last half-century leave little doubt as to the relevance of genetic factors. Despite the great amount of work that has accumulated during the last 65 years, there is still considerable uncertainty about the mode of transmission of this disorder. Many models for the inheritance of the liability to schizophrenia have been suggested. Early workers in psychiatric genetics (e.g., [1, 11]) assumed that schizophrenia and other psychoses were inherited as simple recessive characters since, although familial, the parents of schizophrenic probands were themselves seldom affected. In his study of schizophrenia in a North Swedish isolate, Böök [12] suggested an incompletely penetrant dominant mode of inheritance, with as many as 80% of heterozygotes unaffected. Similarly, Slater and Slater and Cowie [13, 14] offered an algebraic argument for a partially dominant (incompletely penetrant) gene as the most parsimonious explanation for the observed rates of schizophrenia in relatives of probands (cf. [15, 16]). Carrying the case somewhat further, Heston [17] suggested a completely penetrant dominant model to account for the genetic basis of schizophrenia based on the broader concept of schizoid disease as the form of illness in nonschizophrenic heterozygotes (cf. [18] for subsequent reservations).

Polygenic models, with a large number of additive loci, have also been proposed as the genetic basis for the liability to schizophrenia. Given the dichotomous nature of diagnosis, the multifactorial threshold models developed by Crittenden [19], Falconer [20, 21], Reich et al. [22, 23], and others are most appropriate. Gottesman and Shields [10, 24] showed that such models fit the data on schizophrenia prevalence in relatives at least as well as the monogenic models. However, Elston and Campbell [25] re-analyzed the family data of Kallmann [26, 27] and concluded that a single major locus model was adequate to explain the pattern of variation observed, and, therefore, postulation of a polygenic model was unwarranted. However, upon subsequent analysis of the original pedigree data collected by Kallmann in connection with his 1946 paper, these authors [28] rejected a single-locus model.

Utilizing pooled risk estimates from several studies cited in Rosenthal [29], Kidd and Cavalli-Sforza [30] were able to fit both a single major locus model and a polygenic model. Since they were unable to reject either model, they concluded that both were acceptable and suggested that any intermediate model should also fit (also see [31, 32]). Conversely, Matthisse and Kidd [33] found that neither model could account for the observed frequency of schizophrenia in first-degree relatives of schizophrenic probands. Using frequency data on the sibs and offspring of schizophrenic probands to predict the expected values for monozygotic (MZ) twins, these workers found that neither model adequately predicted the observed incidences. Although they did not test a two-locus model, Matthisse and Kidd [33] suggested that such a model, with epistasis, might be found to fit the data better than the two models tested.

A two-locus model for schizophrenia has been entertained starting with Rüdin in 1916. Based on the familial distribution of mental illness in large Icelandic kindreds from the 18th and 19th centuries, Karlsson [34] proposed a two-locus model as the genetic mechanism predisposing to schizophrenia. In a subsequent
paper, Karlsson [35] withdrew his suggestion and opted for a single-locus explanation. In a follow-up study of a north Swedish isolate, Böök et al. [36] found a two-locus model to fit the observed family data better than either a single-locus model or a polygenic model. Similarly, Elston and Campbell [25] found a better fit to Kallmann’s [26, 27] data with a two-locus model, but rejected it because it produced an unrealistic value for the expected population prevalence.

Recently, Stewart et al. simulated 4-generation pedigrees to assess the likelihood of observing the simulated pedigrees under a variety of genetic models [37–39]. Since direct likelihood comparisons failed to distinguish the three models simulated (one-, two-, and four-locus models), the authors suggested two conclusions: (1) the etiology of schizophrenia is heterogeneous, or (2) it is nongenetic. The authors suggest, however, that since the simulated single incompletely penetrant recessive model produced a likelihood surface similar to that found in analysis of real rather than simulated data it may be "put forward as a potentially useful working hypothesis" [40].

Here we will assess the hypothesis that a generalized single-locus, two-allele model is responsible for the observed variation in the incidence of schizophrenia in relatives of schizophrenic probands. It is assumed that all schizophrenia is the result of a single etiology.

MATERIALS AND METHODS

In previous attempts to determine the genetic basis of schizophrenia, researchers either used the data of a single investigator or pooled several data sets. Unfortunately, individual studies are too small to be definitive (see [25, 41, 42]), while the grouped data obscure important variation in risk due to differences in diagnostic criteria, ascertainment, age-of-onset distributions, and population prevalence (e.g., see [30]). To avoid these pitfalls, our paper is based on all published reports of family or twin studies of schizophrenia that meet the following criteria: (1) Morbidity risks for siblings and parents or offspring of schizophrenic probands and probandwise twin concordance rates must be given or can be calculated from the published report. (2) The original data were collected in such a way as to allow some assessment of the method of ascertainment, and, therefore, potential ascertainment bias. (3) All family data must be age-corrected. In most cases, this was done by the abridged method of Weinberg. (4) Where the data used in earlier studies have been incorporated into more recent follow-up studies, the latter reports have been selected for inclusion. (5) Studies in which schizophrenic probands were selected due to the presence of another major disorder (e.g., alcoholism, mental retardation) were excluded.

The correlation between concordance and severity as well as between ages of onset in MZ but not, necessarily, dizygotic (DZ) twins results in spuriously high concordance rates when twin series are age-corrected by methods used for other classes of relatives [42, 43]. Consequently, we prefer not to use age-correlated concordance rates until adequate age-correction methods for twins are developed [10]. The studies included in this survey (cf. [44, 45]), along with the rates of schizophrenia in relatives of schizophrenic probands, are given in tables 1 and 2.

THE MODEL

In attempting to "fit" the general single major locus model to morbid risk data such as those in tables 1 and 2, a parameter problem is encountered. The two-allele (A and a) genetic model is defined by four parameters: the a gene frequency (q = 1 – p) and the three penetrances f1, f2, and f3 (for the genotypes AA, Aa, and aa, respectively), where the pene-
TABLE 1
PUBLISHED MORBIDITY RISKS IN RELATIVES OF SCHIZOPHRENIC PROBANDS UTILIZED IN THIS REPORT—CERTAIN + PROBABLE

<table>
<thead>
<tr>
<th>Risk Period</th>
<th>Siblings</th>
<th></th>
<th>Parents</th>
<th></th>
<th>Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BZ</td>
<td>$K_R \pm SE$</td>
<td>BZ</td>
<td>$K_R \pm SE$</td>
<td>BZ</td>
</tr>
<tr>
<td>Family studies:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Bleuler [46]</td>
<td>20-40</td>
<td>183.5</td>
<td>.065</td>
<td>.018</td>
<td>196.0</td>
</tr>
<tr>
<td>2. Schultz [47]</td>
<td>15-40</td>
<td>1959.5</td>
<td>.083</td>
<td>.006</td>
<td>1277.5</td>
</tr>
<tr>
<td>3. Smith [48]</td>
<td>15-40</td>
<td>611.5</td>
<td>.041</td>
<td>.008</td>
<td>400.0</td>
</tr>
<tr>
<td>4. Kallmann [26]</td>
<td>15-40</td>
<td>1996.5</td>
<td>.115</td>
<td>.007</td>
<td>1963.0</td>
</tr>
<tr>
<td>5. Bleuler [49]</td>
<td>20-40</td>
<td>257.0</td>
<td>.128</td>
<td>.021</td>
<td>195.5</td>
</tr>
<tr>
<td>6. Garrone [50]</td>
<td>15-70</td>
<td>452.5</td>
<td>.086</td>
<td>.013</td>
<td>454.0</td>
</tr>
<tr>
<td>7. Lindelius [75]</td>
<td>Life table</td>
<td>800.0</td>
<td>.068</td>
<td>.009</td>
<td>512.0</td>
</tr>
<tr>
<td>8. Reed et al. [76]</td>
<td>Life table</td>
<td>30.0</td>
<td>.083</td>
<td>.050</td>
<td>16.0</td>
</tr>
<tr>
<td>9. Tsuang et al. [77]</td>
<td>15-40</td>
<td>121.5</td>
<td>.099</td>
<td>.027</td>
<td>117.5</td>
</tr>
<tr>
<td>Family studies from twin index cases:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Luxenburger [51]</td>
<td>15-40</td>
<td>278.0</td>
<td>.115</td>
<td>.019</td>
<td>256.0</td>
</tr>
<tr>
<td>12. Slater [43]</td>
<td>15-40</td>
<td>481.0</td>
<td>.054</td>
<td>.010</td>
<td>292.0</td>
</tr>
<tr>
<td>13. Kringlen [65]</td>
<td>35-65</td>
<td>251.0</td>
<td>.068</td>
<td>.016</td>
<td>110.0</td>
</tr>
<tr>
<td>15. Fischer [66]</td>
<td>Modified life table</td>
<td>290.0</td>
<td>.107</td>
<td>.018</td>
<td>140.0</td>
</tr>
</tbody>
</table>

| Probandwise Concordance Rates† | | | | | | |
| No. pairs | MZ ± SE | No. pairs | DZ ± SE | | | |
| Twin studies: | | | | | | |
| 16. Slater [43] | 37 | .68 | .078 | 112 | .11 | .027 |
| 17. Kringlen [65] | 55 | .45 | .058 | 90 | .15 | .027 |
| 18. Tienari [68] | 17 | .35 | .104 | 20 | .13 | .067 |
| 19. Pollin et al. [52] | 95 | .43 | .046 | 125 | .09 | .019 |

* BZ = Bezugsziffer, age corrected total.
† Method of Strongoren [53].
‡ Modified method of Hagnell [54].
§ Only offspring of MZ twins.
†† See [55].
<table>
<thead>
<tr>
<th>FAMILY STUDIES</th>
<th>RISK PERIOD</th>
<th>SIBLINGS</th>
<th></th>
<th>PARENTS</th>
<th></th>
<th>OFFSPRING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bleuler [46]</td>
<td>20-40</td>
<td>183.5 .049 .016</td>
<td>196.0 .020 .010</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2. Schultz [47]</td>
<td>15-40</td>
<td>1959.5 .067 .006</td>
<td>1277.5 .026 .004</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>3. Smith [48]</td>
<td>15-40</td>
<td>611.5 .033 .007</td>
<td>400.0 .012 .005</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>4. Galatschjan [56]</td>
<td>15-40</td>
<td>322.0 .140 .019</td>
<td>411.0 .049 .011</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>5. Kallmann [26]</td>
<td>15-45</td>
<td>1996.5 .076 .006</td>
<td>* .051 ...</td>
<td>678.5 .139 .013</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>6. Bleuler [49]</td>
<td>15-40</td>
<td>257.0 .104 .019</td>
<td>195.5 .056 .016</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>7. Bök [12]</td>
<td>15-50</td>
<td>277.0 .097 .018</td>
<td>149.5 .120 .027</td>
<td>235.0 .072 .017</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>8. Lindelius [75]</td>
<td>Life table†</td>
<td>800.0 .056 .008</td>
<td>512.0 .006 .003</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>9. Bleuler [45]</td>
<td>15-40</td>
<td>634.0 .090 .011</td>
<td>405.5 .047 .011</td>
<td>106.5 .094 .028</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

* Bezugziffer cannot be calculated from Kallmann’s presentation. Figure given is morbid risk in parents of legitimate probands.
† Method of Strömgren [53].
trances are the probabilities of the genotypes showing the phenotype in question. Morbid risk data, however, permit the estimation of only three parameters; the population prevalence ($K_p$), the additive genetic variance ($V_A$), and the dominance variance ($V_D$), no matter how many different classes of relatives are studied [57]. These two sets of parameters are related in the following way [58]:

$$K^2 = p^2 f_1 + 2pq f_2 + q^2 f_3,$$

$$V_A = 2pq \{q(f_1 - f_2) + p(f_2 - f_1)\}^2,$$

$$V_D = p^2 q^2 \{f_1 - 2f_2 + f_3\}^2.$$

Thus, while incidence data allow estimation of $K_p$, $V_A$, and $V_D$, unique estimation of the underlying parameters is usually precluded except in the special case in which one or more of the underlying penetrances are at a bound of either 0 or 1, as is the case, for instance, when transmission is dominant or recessive [59–61]. The validity of the model rests on the following assumptions [59]: (1) The character under study is scored as present or absent. (2) The character is determined by two alleles at a single locus and is not influenced by any other locus (i.e., no epistasis). (3) Any or all of the three genotypes may be incompletely penetrant. However, there is no correlation of penetrance between related individuals beyond that due to the sharing of genotypes.

As noted by James [57], the incidence in relatives of probands is merely a function of the population prevalence ($K_p$) and the covariance between relatives (COVR) for the dichotomous character,

$$K_R = K_p + \frac{\text{COVR}}{K_p}.$$

The covariance between relatives of probands is simply the sum of weighted proportions of the additive and dominance variances,

$$\text{COVR} = uV_A + vV_D.$$

The coefficients $u$ and $v$ are the probabilities of relatives sharing one particular allele and both alleles at a locus identical by descent, respectively. For the class of relatives considered in this paper, the values of $u$ and $v$ are both 1 for MZ twins, 1/2 and 0 for parents or children, and 1/2 and 1/4 for full siblings, including DZ twins.

Of interest, also, is the expected proportion of the affected offspring when both parents are affected. Suarez et al. [60] report this expected dual mating proportion to be

$$K^2 = K_p + V_A/2K_p + V_D/2K_p^2,$$  

where $\sigma_D = \pm (V_D)^{1/2}$. From family data, however, only $V_A$ and $V_D$ are obtainable so that the sign of $\sigma_D$ is unknown. This is unimportant unless the absolute value of $(V_D)^{1/2}$ is between 0 and $K_p$ since this is the only region of ambiguity concerning the sign of $\sigma_D$. If the value of $\sigma_D$ falls within this range, it may be recovered as follows [62]:

$$\sigma_D = \frac{[K_p(K_1 - 2K_1 + K_0)]/(K_1 - K_p)}{K_p - \sigma_0 + \sigma_A(1 - 2q)/(2pq)^{1/2}}.$$  

It is clear from equations (1–5) that the region over which the single major locus may be mathematically defined can be framed either in terms of the two genetic variances or directly by the reported $K_p$s, so long as the population prevalence ($K_p$) is known. For convenience, the latter method is presented here. Noting that from equations (2–5), $(V_A)^{1/2}$ and $(V_D)^{1/2}$ are linear in the $f$'s [63]:

$$f_1 = K_p - \{\sigma_A(2pq)^{1/2}\}/p + \sigma_D q/p,$$

$$f_2 = K_p - \sigma_0 + \sigma_A(1 - 2q)/(2pq)^{1/2},$$

$$f_3 = K_p + \{\sigma_A(2pq)^{1/2}\}/q + \sigma_D p/q,$$

the range of $\sigma_D$ for...
given \( a_A, K_p, \) and \( q \) may be obtained directly [59]. This observation allows the mathematical definition of a perimeter in the bivariate \( a_A - K_p \) plane that contains all points compatible with the single major locus model at a given population prevalence and gene frequency. This procedure may be generalized to obtain a perimeter, or envelope, that defines the limits of the single major locus model for all gene frequencies. Thus, single-locus models of any penetrance vector, and even sporadics (i.e., \( f_i \neq 0 \)) are subsumed within the envelope ([59], for complete discussion of the method). The transformation of limits of the model derived in the \( a_A, a_D \) parameter space to the \( K_p \) parameter space is straightforward from equations (1-5).

Mapping the limits of this model is achieved by constructing an “envelope” in the bivariate \( K_p \) space that is defined by the limits of \( a_A \) and \( a_D \) for a particular \( K_p \) [59]. If the true proportions of affected relatives are consistent with the general single-locus model, the point of intersection of the \( \Gamma_{R(SIB)} \) and \( \Gamma_{R(P-O)} \) values must lie within the envelope or directly on its borders. If this is not the case, the data are theoretically incompatible with this model, although they may not be statistically incompatible. The discrimination between models is aided by including MZ twin concordances. According to the general single major locus model, when all three \( K_p \)s are available (\( K_{R(MZ)}, K_{R(SIB)}, \) and \( K_{R(P-O)} \)), all three must fall on a single point to be theoretically consistent with the single major locus model. More precisely, the joint observation on sibs and parents (or offspring) must fall precisely on a line transecting the envelope defined by the observed MZ twin concordance (cf. [64], see figs. 3 and 4).

RESULTS AND DISCUSSION

With the available data, the most powerful test of compatibility of individual data sets with the general single-locus model is to examine the familial incidence of schizophrenia in siblings and parents/offspring of schizophrenic probands as well as the concordance rate in MZ twins relative to each other within the confines of the limits to the model at the appropriate population prevalence. Table 1 shows that only four studies [10, 43, 65, 66] have reported all of the \( K_p \)s required.

The MZ twin concordance rate and the joint incidence of schizophrenia in the sibs and parents of the twin index cases for these four studies are plotted in figure 1 relative to the envelope defining the limit of the single major locus model at \( K_p = .01 \). All four studies use a population morbid risk very near .01. As shown in figure 1, the twin concordance rates for these four studies lie well into the upper half of the envelope while the joint incidence in sibs and parents plot very near the bottom. Although the family data from Fischer [66] plot outside the limits of the model, and hence would be judged theoretically incompatible with it, the distance from the boundary is not great so that sampling error associated with estimation of the familial incidence values may be responsible for the deviation. Of interest is the fact that neither the twin concordance rates nor the familial incidence values lie close to the points defining recessive and dominant modes of transmission. Thus, an incompletely penetrant model must be adopted if the single major locus model is to be retained.

The degree of overlap of the approximate 95% confidence intervals for familial incidence and twin concordance provide some indication of the disparity between the two morbid risks. As shown in figure 1, there is no overlap of these confidence intervals for the data of Slater [43]. The overlap of confidence intervals for the remaining three studies shown in figure 1 is small, and occurs in a region of the envelope devoid of observations as judged from the survey of world literature on
familial aggregation of schizophrenia (see figs. 3 and 4). The small proportion of overlap of these confidence intervals, as well as the location of the overlap relative to the mathematical limits of the model, strongly suggests the inapplicability of the general single-locus model to account for the incidence of schizophrenia in relatives of schizophrenic probands.

Although the minimal overlap of confidence intervals suggest the inapplicability of the single major locus model, it does not constitute a formal test of the fit of the observations to the expectations of the model. To perform such a test, a search of the likelihood surface for the best estimates of the additive and dominance variances was conducted using maxlik [67], based on the incidence in proband's cotwin (MZ or DZ), full sibs, and parents at a fixed population prevalence. The estimates of $V_A$ and $V_D$ obtained in this way may be used to predict new values for $K_{R(P)}$, $K_{R(SIB)}$, $K_{R(MZ)}$, and $K_{R(DZ)}$, and, therefore, predicted numbers of affected and unaffected relatives in each class. The test of the fit of the observations to expectations of the model, then, is a chi-square with 2 degrees of freedom between the observed and predicted values. The results of this procedure are summarized in table 3.

Table 3 illustrates that the incidence figures reported by Slater [43], Kringlen [65], and Fischer [66] depart significantly from expectations of the single major locus model ($P < .05$). Since the procedure used to generate the predicted values for these studies maximizes congruence with the single-locus model, the finding of significant deviations from the model demonstrates its inadequacy.
### TABLE 3

**Predicted Incidence of Schizophrenia in Three Classes of Relatives Obtained from Maximum-Likelihood Estimates of $V_a$ and $V_d$ with Chi-square Goodness-of-Fit Test on Observed and Predicted No. Affected and Unaffected Individuals**

<table>
<thead>
<tr>
<th>Author</th>
<th>$K_P$</th>
<th>$K_{R(P)}$</th>
<th>$K_{R(SIB)}$</th>
<th>$K_{R(DZ)}$</th>
<th>$K_{R(MZ)}$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slater [43]</td>
<td>.0042</td>
<td>.029(4.2)</td>
<td>.102(35.8)</td>
<td>.102(00.2)</td>
<td>.346(59.8)</td>
<td>33.897*</td>
</tr>
<tr>
<td></td>
<td>.0100</td>
<td>.029(4.2)</td>
<td>.103(35.8)</td>
<td>.103(00.2)</td>
<td>.343(59.8)</td>
<td>34.616*</td>
</tr>
<tr>
<td></td>
<td>.0240</td>
<td>.029(4.2)</td>
<td>.104(36.0)</td>
<td>.104(00.1)</td>
<td>.336(59.7)</td>
<td>36.390*</td>
</tr>
<tr>
<td>Kringlen [65]</td>
<td>.0042</td>
<td>.021(2.9)</td>
<td>.110(61.4)</td>
<td>.110(22.4)</td>
<td>.392(13.4)</td>
<td>7.223†</td>
</tr>
<tr>
<td></td>
<td>.0100</td>
<td>.021(3.0)</td>
<td>.110(62.1)</td>
<td>.110(20.8)</td>
<td>.390(14.1)</td>
<td>7.384†</td>
</tr>
<tr>
<td></td>
<td>.0240</td>
<td>.024(0.5)</td>
<td>.114(66.3)</td>
<td>.114(16.1)</td>
<td>.382(17.1)</td>
<td>7.823†</td>
</tr>
<tr>
<td>Gottesman and Shields [10]</td>
<td>.0042</td>
<td>.014(1.8)</td>
<td>.121(65.2)</td>
<td>.121(003)</td>
<td>.450(33.0)</td>
<td>5.411</td>
</tr>
<tr>
<td></td>
<td>.0100</td>
<td>.014(1.8)</td>
<td>.122(65.1)</td>
<td>.122(002)</td>
<td>.448(33.1)</td>
<td>5.547</td>
</tr>
<tr>
<td></td>
<td>.0240</td>
<td>.024(2.6)</td>
<td>.128(64.1)</td>
<td>.128(0.27)</td>
<td>.438(33.0)</td>
<td>6.625†</td>
</tr>
<tr>
<td>Fischer [66]</td>
<td>.0042</td>
<td>.019(0.2)</td>
<td>.136(24.7)</td>
<td>.136(70.8)</td>
<td>.501(04.2)</td>
<td>8.337†</td>
</tr>
<tr>
<td></td>
<td>.0100</td>
<td>.019(0.3)</td>
<td>.136(25.4)</td>
<td>.136(69.6)</td>
<td>.497(04.7)</td>
<td>8.382†</td>
</tr>
<tr>
<td></td>
<td>.0240</td>
<td>.024(0.6)</td>
<td>.139(28.1)</td>
<td>.139(64.3)</td>
<td>.482(07.0)</td>
<td>8.636†</td>
</tr>
</tbody>
</table>

**Note:** Nos. in parentheses are the percent contribution of each class of relatives to the chi-square value.

* $P < .001$.
† $P < .05$.  

O'ROURKE ET AL.
The data of Gottesman and Shields [10], on the other hand, are not significantly different from predictions of the model except at a population prevalence of .024. Although the deviation of the data from the model predictions are in the same direction as the other three studies, the magnitude is not as large.

Two further points may be mentioned regarding the information in table 3. Estimates of population prevalence seem to have little effect on the risk estimates for the four classes of relatives, generally resulting in less than a 1% change over the range examined. The $K_{R|P}$ estimates for $K_p = .024$ for all studies except Slater [43] equal the population prevalence due to the very low estimates of $V_e$ obtained from the search procedure. Finally, the pattern of deviation of the observations from the model seen in these four studies is not unique, but is shared by all the studies listed in tables 1 and 2. This will be made clear in the discussion below.

These studies, by different workers in different countries using modern techniques of ascertainment, diagnosis, and analysis demonstrate the inadequacy of the single major locus model in accounting for the observed incidence of affected relatives of schizophrenic probands. Without the benefit of the MZ concordance rate, however, three of these studies would be judged compatible with a single-locus model. The twin data, then, are of critical interest.

In figure 2, MZ and DZ concordance rates are plotted relative to the envelope defined by a population morbid risk of .01. Only the concordance rates of Tienari [68] and Kringlen [65] lie within regions that are compatible with the single major locus model. The remaining studies lie well outside the envelope and would be

![Fig. 2.—MZ and DZ probandwise concordance rates relative to the limits of the general single-locus model at population prevalence of .01. Nos. refer to studies listed in table 1. Points subscripted A indicate expected values derived from $V_e$ and $V_p$ estimates obtained from incidence in sibs and parents of twin probands. (See text for discussion.)](image)
judged incompatible with a single-locus model constrained by any reasonable estimate of $K_\rho$ (e.g., .004 to .03). The observed risks in cotwins of probands may be compared to that predicted based on $V_A$ and $V_D$ estimates obtained from the incidence in sibs and parents for the four studies where these are available (subscripted $A$ in fig. 2). The fact that the observed family and twin data from Kringlen [65] are separately compatible with a single-locus model but collectively are not (see table 3) illustrates the utility and sensitivity of the method.

Essen-Möller and Fischer [69] suggested that probandwise twin concordances are artificially elevated over $KR$ values for other family members of twin index cases since the latter may never be counted twice as probands in a twin study. This bias is very small [70] and does not affect the conclusions reached here.

Further demonstration of the insufficiency of the single major locus model may be made by the use of dual-mating studies. Irrespective of the relative weight given to genetic or environmental sources of variation in schizophrenia, the offspring of two affected parents are considered a high-risk group and the proportion of affected offspring is informative.

Dual-mating data are available from three studies identified in the survey and retained for analysis. For the Kringlen [71] data, the maximum-likelihood estimates of $V_A$ and $V_D$ obtained from the search of the likelihood surface have been used. While Kringlen's [71] dual-mating study was done independently of the twin/family study, the diagnostic criteria remained the same between the studies and the probands were ascertained in similar fashion from the same hospitals. Thus, it seems appropriate to treat the two studies simultaneously without loss of information since the target population was the same in each case. For the studies by Kallmann [26] and Modrzewska [72], $V_A$ and $V_D$ estimates derived from disease incidence in sibs and parents using equations (4) and (5) were employed since data were insufficient to warrant a search of the likelihood surface. The results are presented in table 4.

It is clear that there is little, if any, agreement between the observed and expected proportions of affected offspring from dual matings. Using the observed and expected number of affected and unaffected offspring in each study, a $\chi^2$ goodness-of-fit test resulted in a significant difference between expected and observed values (see table 4).

The $\chi^2$ values should be viewed with caution since it is assumed that $V_A$ and $V_D$ estimates are in fact the true parameters. Given the magnitude of the $\chi^2$ values, however, moderate changes in the estimates of the variance components will make little difference. Furthermore, the sample sizes as well as the number of affected individuals in each sample are quite small. While this lessens somewhat the impact of the significant chi-square values, the consistent deviation from the expectations of the model are instructive. At the very least, no support for the general single major locus model is apparent here. On the contrary, the results demonstrate an incompatibility between the observations and the predictions from the monogenic model.

The negative value for $V_D$ from the Modrzewska [72] study is the result of $K_{R(P)}$ exceeding $K_{R(SIB)}$. To obtain the estimate of $K_2$ given in table 4, $V_D$ was therefore
### Table 4

**Estimated Proportion of Affected Offspring from Dual Matings Using \( V_d \) and \( V_p \) from Family Studies with Population Prevalence of 1.0%**

<table>
<thead>
<tr>
<th>Author</th>
<th>( V_d )</th>
<th>( V_p )</th>
<th>( BZ )</th>
<th>( K_1 )</th>
<th>( SE )</th>
<th>( K_2 )</th>
<th>( x^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kallmann [26]</td>
<td>0.00186</td>
<td>0.00048</td>
<td>23.5</td>
<td>0.681</td>
<td>0.096</td>
<td>0.400</td>
<td>6.60, ( p &lt; 0.05^{*} )</td>
</tr>
<tr>
<td>Kringlen [71]</td>
<td>0.00021</td>
<td>0.00359</td>
<td>17.0</td>
<td>0.290</td>
<td>0.110</td>
<td>0.094</td>
<td>5.82, ( p &lt; 0.05^{*} )</td>
</tr>
<tr>
<td>Modrzechska [72](†)</td>
<td>0.00498</td>
<td>-0.00108 (‡)</td>
<td>25.5</td>
<td>0.280</td>
<td>0.090</td>
<td>0.508</td>
<td>4.43, ( p &lt; 0.05^{*} )</td>
</tr>
</tbody>
</table>

* Chi-square values computed using Yates’ continuity correction.
† Data from 1972–1977 follow-up of North Swedish isolate by Böök et al. [36]. Although a \( K_2 \) of .01 has been used here for consistency, Böök et al. [36] report a value very near .03 with estimates of \( K_{R_{HI}} = 0.102 \) and \( K_{R_{P}} = 0.11 \). This results in estimates of \( V_d \) and \( V_p \) of .00454 and –.00088, respectively. See text for discussion.
‡ \( V_p \) set equal to zero in calculating \( K_1 \).
set equal to 0. In this case, neither the family data nor the dual-mating data are compatible with the single-locus model.

The observed population prevalence from this isolate population is idiosyncratic in that it is roughly three times the rate reported for all other populations surveyed. If the reported population prevalence of nearly 0.03 is used in the analysis, the deviation from the model's prediction is nonsignificant ($\chi^2 = 0.744$), but is in the same direction (observed > predicted; predicted $K_r = 0.193$) as that seen for the data of Kallmann [26] and Kringlen [71].

The significant deviations from the single major locus model found in the studies considered so far are not anomalous, but typical of the results for most of the family and twin studies reported. Studies of schizophrenia in the United States and Europe suggest a population lifetime morbidity risk of about 0.01. Böök [12], however, has reported a population risk of 0.028 from a north Swedish isolate. Slater and Cowie [14] reviewed the population prevalences from a number of studies in Europe and Scandinavia conducted between 1928 and 1964 and found a range of values from 0.0042 to 0.024. These values may be taken as the range encountered in various parts of Europe and Scandinavia. The mean value is 1.17% if the very large value(s) from Switzerland are included and 0.85% if they are excluded. Using a population prevalence of 0.01 as a general referent would seem to be a reasonable approximation. Figure 3 presents the bivariate plot of the incidence values for the relatives of schizophrenic probands with respect to the envelopes defining the parameter space compatible with the general single major locus model at three population prevalences (0.0042, 0.01, and 0.024). Simultaneous presentation of these three envelopes has several advantages. First, since many authors

![Figure 3](https://example.com/figure3.png)

**Fig. 3.—**Incidence of schizophrenia in parents or offspring (denoted by $A$ and $B$) and sibs of schizophrenic probands using broad diagnosis of schizophrenia. The three envelopes define the limits of the general single-locus model at three population prevalences ($K_p$). Nos. refer to studies listed in table 1. Lines transecting the envelope for $K_p = 0.01$, and circled values associated with them refer to the expected proportion of affected MZ cotwins of probands relative to the limits of the model. (See text for discussion.)
do not report a population prevalence for their study population, it is impossible to evaluate each study relative to a specific value. It may be assumed, however, that for those studies the true value will lie within the range .0042 and .024. Second, the proximity of the three envelopes in figure 3 indicates that relatively minor changes in shape and position of the envelope occur if the population prevalence is only slightly different from .01. The horizontal lines inside the envelope for $K_p = .01$ represent the expected probandwise concordance of MZ twins.

Several authors report two incidence rates for relatives of probands; one based on a "narrow" definition of schizophrenia (table 2) and a second, somewhat broader view of the disorder, that includes "uncertain" or "probable" cases (table 1). Since this practice is not universal (some authors reporting only the proportion of affected relatives based on the broader range of criteria), the values plotted in figure 3 are those based on the incidence of "certain" plus "probable" schizophrenics. This also facilitates comparison of American and European studies [73], since the American criteria for schizophrenia are somewhat broader than those in Western Europe.

There are several points of interest in figure 3. First, it is clear that the observed values for incidences in sibs is rather restricted relative to that for the parents or offspring (denoted by subscripts $A$ and $B$, respectively). This is not entirely unexpected, since the usual methods of computation of morbid risk often results in a lower value for parents of probands than for children (cf. [74]). Since parents of schizophrenic probands are preselected for mental health [74], lack of fit to any particular model may be exaggerated. In addition, sampling bias due to recall in lieu of adequate records for diagnosis in parents and the possibility of over- or undercorrecting for age in offspring may bias the estimates of risk in these groups.

Seven studies plot below the diagonal outside the theoretical limits of the model. Of the remaining studies, three are found to plot outside the envelope drawn for population prevalence of .01. One of these [26] has long been put forth as evidence for a single-locus model, and, indeed, plots very close to the edge of the envelope.

Two studies [66, 75] lie outside the envelope in the opposite direction. The Lindelius [75] study is interesting in that the probands were old and so many relatives (particularly parents) were deceased at the time of the study. Therefore, very old hospital records were used to obtain diagnoses of these relatives. In many cases, it was clear that a relative suffered some form of psychosis but the recorded information was insufficient to determine what kind. Some of these undiagnosed psychoses were probably schizophrenia. These individuals were eliminated from the calculations used in tables 1 and 2, so the values given should be considered underestimates. If the undiagnosed psychosis group is considered to contain the same proportion of schizophrenics as would be indicated by the rest of the Lindelius data, then the incidence values become: $K_{R(SIB)} = .091; K_{R(P)} = .017; K_{R(O)} = .095$. Even this adjustment, however, does not make the data conform to the single major locus model assuming a population prevalence of .01.

For those studies in which incidence in parents and offspring are both available, the $K_{R(SIB)}, K_{R(O)}$ point is usually found to lie below the diagonal, while the $K_{R(SIB)}, K_{R(P)}$ point is contained within the envelope. The exceptions to this are the
studies of Reed et al. [76] in which both points fall below the diagonal, Tsuang et al. [77] in which the observed trend is reversed, Fischer [66] in which the sib-parent value is outside the envelope in the opposite direction, and Bleuler [45] in which both sets of observations are within the envelope. The remaining studies report joint observations on incidence of schizophrenia in relatives that are consistent with the single-locus model and, hence, plot within the envelope defined by a population prevalence of .01.

Taken as a whole, the picture presented in figure 3 neither strongly supports the single major locus model nor allows its rejection. It is not particularly surprising that recent workers, using these same incidence values, concluded that the single major locus model provided at least as good a fit as the multifactorial model. Belief in a single-locus mechanism for schizophrenia was commonly held by workers during the first half of this century, since the idea of polygenic or multifactorial inheritance was not even entertained in this area until the work on multifactorial inheritance for diseases by Falconer [20, 21]. With the exception of Ødegaard's [78] early suggestion, multifactorial models were not seriously considered until the last decade and a half [10].

The fact that nearly half of the points in figure 3 plot outside the region that may safely be assumed to cover most reasonable population prevalence values suggests that the single major locus model is inadequate, even when making provision for incomplete penetrance, but does not demonstrate it conclusively.

Since a “broad” classification of schizophrenia may contain a significant amount of noise, encompassing as it does “questionable” or “uncertain” cases, we may restrict ourselves to a narrow definition, including only certain cases. These data are presented in table 2 and plotted in the same fashion as before in figure 4. For these cases, 67% (8/12) of the joint observations are found to lie within the envelope defined by a population prevalence of .01. Three fall below the diagonal,

![Fig. 4.—Incidence of schizophrenia in relatives of schizophrenia probands using a narrow definition of schizophrenia. Symbolism used is same as in figure 3. (See text for discussion.)](image-url)
and one [75] plots well to the left, outside the range. The general pattern is the same as that seen previously. Interestingly, one study that specifically argues for an incompletely dominant form of inheritance [12] seems to be totally incompatible with the single major locus model.

The discrimination between models is aided by including twin concordances in the joint observations. The expected MZ concordances ranging from .2 to .9 are shown in figures 3 and 4. Referring to the observed MZ concordances listed in table 1, it is clear that they fall well into the upper half of the envelopes in the two figures while the sib and parent-offspring values plot entirely in the bottom half. This lack of overlap between expected and observed twin concordances strongly suggests the inadequacy of the single-locus model and demonstrates the generality of the pattern seen in the four twin/family studies discussed earlier.

The results just presented combine to form a rather convincing demonstration of the inadequacy of the general single major locus model for the etiology of schizophrenia. The published studies are a series of independent tests, where each may well fit both monogenic and multifactorial models because of small sample size, but collectively deviate consistently from the single major locus model. The error inherent in each study (not necessarily the same across studies) is overshadowed by the total pattern of nonconformance to the single major locus model. Significantly, those studies that deviate from the expectations of the monogenic model all do so in the same direction; that is, observed twin concordance rates and the proportion of affected offspring from dual matings are too high relative to the population prevalence and lifetime risks in other relatives. If the error inherent in individual studies were truly random, we would not expect to observe this consistency of deviations from predicted values.

Even though we have carried out the analysis assuming a population lifetime morbid risk of .01 for schizophrenia, which for most studies is adequate, the single-locus model treated here is generalized in the sense of subsuming all allowable penetrance vectors [59]. Thus, the general single-locus two-allele model may be rejected as inadequate to explain the observed variation in the familial distribution of schizophrenia. This does not, however, diminish the importance of genetic factors in the etiology of schizophrenia. Although the generalized single-locus two-allele model may be rejected, the possibility of involvement of major loci due to genetic heterogeneity or in conjunction with a polygenic background (i.e., a mixed model) cannot be discounted [79–81]. Other transmission models (e.g., mixed model, multifactorial, two-locus, etc.) are still viable and subject to further testing.

ACKNOWLEDGMENTS

We thank Drs. C. R. Cloninger, P. McGuffin, and D. C. Rao for their valuable suggestions.

REFERENCES

47. Schultz B: Zur Erbpathologie der Schizophrenie. Z Neurol Psychiatr 143:175–293, 1932
49. Bleuler M: Krankheitsverlauf, Persönlichkeit und Verwandtschaft Schizophrener und ihre gegenseitigen Beziehungen. Leipzig, West Germany, Thieme, 1941
51. Luxenburger H: Untersuchungen an schizophrenen Zwillingen und ihren Geschwi-


70. Rice J, Gottesman II, Suarez B, O'Rourke DH, Reich T: Ascertainment bias for non-twin relatives in twin proband studies. *Hum Hered*. In press, 1982


SCHIZOPHRENIA


