Eur Arch Psychiatry Clin Neurosci (1991) 240:182–187

European Archives of Psychiatry and Clinical Neuroscience

Sampling Strategies for Linkage Studies

Lynn R. Goldin, Maria M. Martinez, and Elliot S. Gershon

Clinical Neurogenetics Branch, National Institute of Mental Health, Bldg. 10, Rm. 3N218, 9000 Rockville Pike, Bethesda, MD 20892, USA

Summary. Broadly there are three alternative sampling strategies for linkage studies of complex disorders such as schizophrenia. These are to select either affected sib pairs (and their parents) or small-to-medium-size pedigrees, or large pedigrees. The advantages and disadvantages of each are discussed. Our studies show that, even in the presence of heterogeneity, linkage can be found in clinically realistic sample sizes of nuclear families or medium-sized pedigrees. However, relying on affected sib pair methods in less satisfactory.

Key words: Linkage – Sampling strategies – Schizophrenia

Introduction

Although there is strong evidence for heritability of major psychiatric disorders (affective disorders and schizophrenia) from twin, family, and adoption studies, it has not been possible to determine the mode of inheritance of these disorders from segregation analyses of the distribution of illness in relatives of patients (Nurnberger et al. 1986). The complex nature of these disorders (e.g., clinical phenotypes not well defined, incomplete penetrance, age and sex effects) suggest that these disorders are not simple Mendelian traits. However, the recent reports of linkage of chromosome 11 markers to affective disorder in a large Amish pedigree (Egeland et al. 1987), linkage of Xq28 markers to affective disorder in an Israeli sample (Baron et al. 1987), and linkage of chromosome 5q11–13 markers to schizophrenia in British and Icelandic families (Sherrington et al. 1988) indicates that single major loci may exist for these disorders. Unfortunately, these findings have not been sufficiently replicated in other samples to conclude confidently that disease susceptibility loci exist in these regions or to be able to determine what proportion of affected families is accounted for by these particular genes (Detera-Wadleigh et al. 1987; Berrettini et al. 1990).

Recent progress in identifying genetic polymorphisms using molecular genetic techniques has made it possible to screen the entire human genome for disease susceptibility genes. There is a growing consensus that it is desirable to carry out linkage studies in psychiatric disorders, taking advantage of the complete gene map.

There are several possible sampling strategies that can be used to carry out linkage studies. We can break these down into three classes: (1) nuclear families, including affected-sib pairs, (2) small-to-medium-sized pedigrees having an average of 15–25 individuals, and (3) very large pedigrees having on the order of 50 individuals or more. Given the fact that major psychiatric disorders are complex and are probably genetically heterogeneous, it is important to determine what the advantages and disadvantages of each of the above strategies are. It is clear that very large pedigrees in themselves are powerful enough to detect linkage of a marker and disease gene if it is present. However, there are two possible problems with this strategy. The pedigree may be from a population isolate and have a "rare" genetic cause of the disease, which will be hard to replicate in another pedigree or sample of smaller families. On the other hand, the pedigree may be segregating for more than one "common" cause of the disease, resulting in heterogeneity within the pedigree. For these reasons, it is worth considering the power and required sample size of nuclear families and small/medium pedigrees to detect linkage in the presence of complex inheritance and heterogeneity. It is also desirable to be able to confirm an initial linkage finding from a single large pedigree in a larger sample of smaller families which, if ascertained systematically, are more representative of the general population. Having several linked families also gives investigators the possibility of identifying the genetic defect shared among these families since within a single family, all ill persons share many genes in the linked region.

This paper will focus on the various approaches we have taken in order to determine the power and sample sizes for finding linkage to psychiatric disorders. In our studies, we have varied several parameters of interest. We have considered different: (1) family structures (affected sib pairs, nuclear families, small and medium pedigrees), (2) genetic models of the disease locus, (3) levels

Offprint requests to: L. R. Goldin

of heterogeneity, (4) ascertainment criteria of families or pedigrees, (5) marker information (one marker locus, map of two linked markers), and (6) methods of analysis (parameter free in affected sib pairs and likelihood approach in nuclear families and pedigrees).

Nuclear Families

Affected Sib Pairs

Sib-pair methods have long been used as a nonparametric approach to linkage analysis (Penrose 1935, 1953). During the last several years, these methods have been further refined to use information in sibling groups more efficiently. The strategy of sampling affected sib pairs (ASPs) is recognized for its advantage in testing for linkage in diseases with incomplete penetrance. The expectation is that if a disease is linked to a marker locus, ASPs should be alike at the marker locus more often than expected by chance alone. This method has been applied largely to test for linkage of diseases to the HLA region. Since HLA is so polymorphic, it is usually possible to determine if an affected sib pair share 2, 1, or 0 haplotypes identical by descent (IBD). Given no linkage, the probabilities that IBD = 2, 1, and 0, respectively, are 25%, 50%, and 25%. Thus, a distortion of these probabilities (i.e., more than 25% sharing 2 haplotypes IBD) indicates linkage. This method has been successful in detecting linkage of juvenile diabetes to HLA [reviewed in Thomson (1988)]. This method does not assume any particular mode of transmission of the disease locus. However, assuming linkage of a marker to a disease locus, it is possible to infer the parameters of the disease locus from the IBD distribution in a sample of ASPs (Suarez et al. 1978; Thomson and Bodmer 1977). The ASP method would appear to offer an advantage for studying linkage of markers to major psychiatric disorders since the mode of inheritance is unknown and the pathology of these disorders often makes large, informative families difficult to locate and study. Chakravarti et al. (1987) have looked at the power of ASPs to detect linkage and heterogeneity assuming simple inheritance. We (Goldin and Gershon 1988) have determined the required sample sizes of ASPs to find linkage under various assumptions about the inheritance of the disease locus and level of heterogeneity.

Methods. We have used the method of Blackwelder and Elston (1985) to calculate the number of ASPs needed to have 80% power to find linkage given the parameters of the disease locus, recombination fraction, and proportion of families linked. These calculations assume that the IBD status is known exactly for each affected sib pair, which implies that the marker locus is highly polymorphic (like HLA) and the marker types of the parents are also known.

Results. Table 1 shows the number of ASPs needed to have 80% power (type I error = 5%) to find linkage according to the population prevalence of the disease, recombination fraction, and proportion of families linked.

It can be seen that linkage can be detected in a feasible sample size if only 50% of families are linked. The prevalence of 1% corresponds approximately to that for schizophrenia or bipolar I disorder whereas major affective disorder (including bipolar and unipolar) has a prevalence of approximately 7% (Nurnberger et al. 1986). As is evident, linkage is somewhat easier to detect for a rarer disorder. This is expected since, as the prevalence increases, an unaffected parent is more likely to carry a disease gene, causing the IBD distribution to be less skewed. If the proportion of linked families is only 25%, then considerably larger sample sizes are required. For example, 200 ASPs are needed for the 1% prevalence case at a recombination fraction of 1%. It must also be remembered that these calculations assume that the parents' marker phenotypes are also known, which effectively doubles the sample size of individuals requiring laboratory typing. In addition, these calculations assume that the marker is completely informative (i.e., all parents have unique haplotypes), which is only true for a small number of marker loci at present. It can be concluded that the strategy of collecting ASPs is clinically feasible as long as there is a gene accounting for 50% of cases in the population.

Table 1. Number of affected sib pairs needed to have 80% power to find linkage (dominant disease with 50% penetrance)

Population prevalence	Recombination fraction	100% linked	50% linked
0.01	0.01	7	50
	0.05	15	72
	0.10	27	120
0.07	0.01	19	68
	0.05	28	100
	0.10	48	162

Since ASPs and their parents actually constitute a nuclear family, we decided to examine the more general case of the power of nuclear families under similar conditions but utilizing the traditional likelihood method (similar to the standard lod score) of testing of linkage.

General Nuclear Families

We have used the method of expected likelihood ratios in order to compute the average sample size needed to detect linkage and heterogeneity in nuclear families of different sibship sizes. For simple Mendelian diseases, the number of nuclear families needed to detect linkage and heterogeneity has been determined by Cavalli-Sforza and King (1986). Lander and Botstein (1986) carried out similar calculations to compare the power of one marker versus a map of two linked markers for detecting linkage and heterogeneity. We have compared the power of one versus a map of two marker loci for "complex" inheritance.

Methods. We considered a single locus with allele d responsible for susceptibility to the disease. Let q be the frequency of d and $f1, f2$, f3 the penetrance of the genotypes dd, Dd, and DD, respectively. We investigated recessive and dominant models with variable pen-

Table 2. Sample size needed to detect linkage and heterogeneity when $\alpha = 0.5$: dominant model

Sampling	One marker ($\theta = 0.048$)		Two markers (10 cm map)	
	$f1 = 0.9$	$f1 = 0.5$	$f1 = 0.9$	$f1 = 0.5$
$r = 2, s = 2$				
N	92	128	76	106
Nh	$> 10^5$	$> 10^5$	84	120
$r \geq 2, s = 3$				
N	43	92	36	81
Nh	$> 10^5$	$> 10^5$	42	93
$r = 3, s = 3$				
N	39	55	32	46
Nh	$> 10^5$	$> 10^5$	38	54
$r \geq 2, s = 4$				
N	26	70	NC	NC
Nh	124	1996	NC	NC

Note: High penetrance parameters: $q = 0.01$, $f1 = f2 = 0.9$, $f3 = 0.001$; Low penetrance parameters: $q = 0.02$, $f1 = f2 = 0.5$, $f3 = 0.0001$. s = Sibship size, r = no. of affected sibs, N is the number of families needed to detect linkage (assuming heterogeneity), and *Nh* is the number of families needed to detect heterogeneity; NC, not computed

etrance; all predicted a population prevalence of 2% and a phenocopy rate of 5%.

We assumed that all matings were completely informative for the marker locus. For the case of a map of two linked marker loci, we assumed that each marker locus was completely informative and that the disease locus was midway between the two marker loci.

We assumed a model of admixture that allows for two types of families (Smith 1963). The disease locus is linked to the marker with a recombination fraction θ in a proportion α of families and it is unlinked to the marker in the remaining families 1-α. Assuming that the disease has the same mode of transmission in linked and unlinked families, the likelihood of a family is a function of the two parameters and is given by:

$$L(\alpha, \theta) = \alpha L(\theta) + (1 - \alpha) L(\theta = 0.5)$$

For each set of parameters (q, $f1$, $f2$, $f3$, θ and α), we computed for nuclear families of sibship size s, with a fixed probability of ascertainment for affected children equal to 0.001, the expectation of the *ln* likelihood ratio and the corresponding sample size needed to get a significant test. The sample sizes correspond to a type 1 error of 0.1% for tests of linkage and 5% for tests of heterogeneity. These samples are *average* sample sizes needed which corresponds to 50% power. The criterion for linkage used here is similar to that used in the classical lod score test (i.e., lod score = 3). Modified versions of the programs MLINK (Clerget-Darpoux et al. 1987) and LINKMAP of the computer package LINKAGE (Lathrop et al. 1984) were used to compute the expected in *ln* likelihood ratios.

Results. Some representative results from this study for the case of dominant inheritance are shown in Table 2. More complete results can be found in Martinez and Goldin (1989). Several general trends are evident. The power to detect linkage and heterogeneity improves considerably as the sibship size gets larger, especially as families with a larger number of *affected* sibs are ascertained. As expected, for models with reduced penetrance, linkage is more efficiently detected when all sibs are affected. Having a map of two linked marker loci improves the detection of linkage only slightly, but improves the detection of heterogeneity considerably. For dominant inheritance, it is not practical to detect heterogeneity at all except for sibship sizes larger than 3 or when a map of two marker loci is used. The fact that detection of linkage itself does not improve very much with two marker loci is not a general result, but is due to the relatively small distances considered here and the fact that we have assumed that *each* marker is completely informative. In real studies, the increased power of the map will depend on the distance between the markers and the degree of polymorphism of each one.

It is clear that linkage and heterogeneity can be detected in realistic sample sizes of nuclear families for diseases with reduced penetrance as long as 50% of the families are linked. However, the detection of linkage and heterogeneity becomes more efficient as larger sibship sizes are used. As the proportion of linked families descreases, considerably larger samples are needed. For example, for the case of 2 affected/2 sibs and penetrance = 50%, the number of families needed to detect linkage (2 marker map of 10 cM) when only 25% of families are linked is 234 and 433 for recessive and dominant inheritance, respectively. If only 10% of families are linked, then the required numbers are 1411 and 2720. Thus, it will probably not be practical in real studies to detect linkage when the proportion of linked families is much less than 25%.

It is not possible to compare the sample sizes needed using the IBD distribution in ASPs with the nuclear family calculations since the assumptions are different for the two studies. However, comparable calculations (not shown) demonstrate that as expected, the likelihood method is usually more powerful for detecting linkage than the non-parametric IBD method. If one uses the likelihood method but assumes the wrong genetic model for the disease, then the power to detect linkage may be decreased, depending on how large the deviation of the assumed parameters is from the true parameters. In carrying out studies, we would use likelihood methods for linkage calculations and test for linkage under a few different model assumptions.

Pedigrees

It is well known that pedigrees are more powerful than nuclear families for detecting linkage. The use of pedigrees has been a common strategy in human genetics for many years and has been the major strategy for finding linkage in psychiatric disorders. We have calculated the power of medium sized pedigrees for finding linkage under similar assumptions as we did in the nuclear family study. We have used simulation methods here since as family size becomes larger, it is not practical to use the expected likelihood ratio method, as can be done in nuclear families.

So far, studies of the power of linkage in the presence of heterogeneity have been limited to the case of identi-

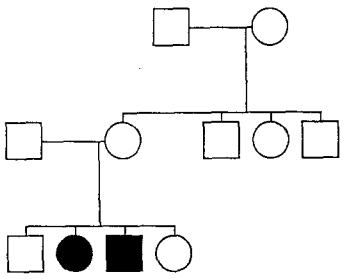

Fig. 1. Pedigree structure used for simulations

cal transmission in the two types of families. Therefore, distribution of phenotypes is assumed to be independent of the disease cause involved in a particular family. However, there is also a possibility of multiple causes of a disease, each with a different mode of transmission. Therefore, depending on the mode of sampling of families, the rate of cases due to the disease locus linked to the marker locus under study may be greater or smaller in the sample than in the population. In the presence of multiple and different disease causes, determination of sampling strategies for linkage analysis depends not only on the mode of inheritance of the illness due to the disease locus linked to the marker locus but also on the mode of transmission of the other disease causes that are independent of the marker locus. We have therefore considered the effect of different modes of sampling on the detection of linkage and heterogeneity given that a heterogeneous disorder is due to two independent and inherited causes (Martinez and Goldin 1990).

Methods. We have considered a disease determined by either one of two independent susceptibility loci, locus *A* and locus *B*, as both being diallelic. The susceptibility locus A is linked to the marker locus while the susceptibility locus B is unlinked to the marker locus. For each disease locus we define q, the frequency of the disease allele and $f1$, $f2$, $f3$ the penetrances for the disease homozygote, the heterozygote, and the normal homozygote, respectively. We have varied the heterogeneity parameter by varying the population prevalence due to each cause. We have considered a disorder having a total population prevalence of 2% and have examined several different combinations of parameters at each disease locus, including all combinations of dominant or recessive traits with high ($f = 0.90$) or low ($f = 0.50$) penetrance.

Monte Carlo methods were used to simulate both marker and disease phenotypes in three-generation families. We simulated one marker locus with 4 equally frequent alleles (polymorphism information content = 70%) linked to the disease locus *A* with a recombination fraction, θ and, unlinked to the second disease susceptibilty locus *B*.

The family data consisted of three-generation families of a given fixed structure with 11 relatives (Fig. 1). Pedigrees were ascertained according to two different sampling schemes; at least two affected children in the third generation (*S02*), or at least two affected in the third generation and at least one affected in the first or the second generation (*S12*), excluding the unrelated spouse in generation 2 (vertical transmission).

Detection of both linkage and heterogeneity of linkage have been performed using the admixture test (Smith 1963) as was used in the nuclear family study. For each generating model and for each sampling selection, 50 replicates of family samples (sizes 15, 25 or 50) were generated. For each sample, the power to detect both linkage and heterogeneity was measured by the proportion of replicates leading to a significant test. We used the same type 1 errors as before (0.1% for linkage, 5% for heterogeneity). The power of the heterogeneity test was performed conditionally on there being a significant linkage test (assuming heterogeneity).

Table 3. Moderate sized pedigrees (11 persons). Power of the linkage test assuming heterogeneity and of the heterogeneity test for different sampling schemes, and *50%* of the cases in the population are due to the locus *A*, linked to the marker locus with a recombination fraction, Θ, of 0.05. Mode of transmission of the linked locus (A): DOM50

Locus B model	Sampling	Power	
		Linkage	Heterogeneity
Dom50	S02	52%	27%
	S12	76%	45%
Dom90	S02	16%	75%
	S12	42%	81%
Rec90	S02	66%	15%
	S12	100%	50%
Rec50	S02	78%	21%
	S12	98%	23%

Notes: S02: at least two affected children in the third generation; S12: at least two affected children in the third generation and at least one affected in the second or the first generation.
Dom, Dominant model; Rec, recessive model; 90, 90% penetrance; 50, 50% penetrance

Results. Generally, detection of linkage is easier when the linked disease locus A has a high penetrance, when the unlinked disease locus B has a low penetrance and, when pedigrees with multiple affected are selected. In particular, when 50% of the cases are due to a dominant locus with high penetrance ($f = 0.90$) and, with a sample size of 50 pedigrees, detection of linkage (recombination fraction = 5%) is very powerful ($\geq 94\%$), regardless of the mode of inheritance of the unliked disease locus and the sampling scheme (results not shown). Conversely, detection of linkage is hardest when the unlinked disease locus B is a dominant trait, particularly when the linked locus A is recessive.

In Table 3, we present results when the linked disease locus A is dominant with low penetrance. This table shows the power to detect both linkage and heterogeneity when the linked disease locus A accounts for 50% of the cases and is linked to the marker locus with a recombination fraction of 5%. The sample size is equal to 50 families. The values of the genetic parameters assumed for the linkage analysis were those of the disease locus A (that is, those of the linked disease locus).

As expected, selecting pedigrees showing a vertical transmission of the illness (S12) increases the power of the linkage test, particularly when the unlinked disease locus B is recessive. It appears that under this sampling scheme, pedigrees segregating for the linked dominant locus A are more likely to be selected. Furthermore, the sample is more informative for linkage, and thus the power is also improved even when the unlinked disease locus B is a highly penetrant dominant trait (DOM90).

When the disease locus A is a low penetrant recessive locus, the power to detect linkage is higher when the locus B is a recessive rather than a dominant locus (results not shown). In fact, the power to detect linkage is

Table 4. Power fo find linkage for 20 families of size 16

Proportion of families linked	$\theta = 0.01$	$\theta = 0.05$
0.50	100%	94%
0.25	80%	60%
0.10	24%	10%

very low (< 25%) in the latter case, especially when families are selected through vertical transmission.

The power to detect heterogeneity varied a great deal in our study. Because of the sampling scheme, the estimate of the rate of linked families in the sample is not an estimate of the population heterogeneity rate, unless both disease loci have the same mode of inheritance. Previous studies (Cavalli-Sforza and King 1986; Ott 1986) have shown that the power of the heterogeneity test decreases as α goes to its more extreme values: 0 or 1. On the other hand, the more homogeneous the sample, the easier the detection of linkage. Thus, because we have considered the heterogeneity test conditional on there being a significant linkage, there are no consistent trends with respect to the power to detect heterogeneity of linkage.

From our sampling criteria, linkage analysis is more powerful when the linked disease locus is dominant and highly penetrant. We have investigated this model for different pedigree sample sizes and for a rate of linked families equal to 25%.

As expected, the power of the linkage test decreases with the rate of cases due to the linked disease locus A. When the dominant-linked disease locus A accounts for 25% of the cases in the population, the power of the linkage tests is more than two times greater when selecting families having vertical transmission (S12) than when selecting through 2 affected sibs (S02), although the overall power to detect linkage is still relatively low (< 50%) except for the large sample size ($n = 50$).

In order to examine the effect of the pedigree size, we considered a pedigree with 3 sibships (a total of 16 individuals with 4 children in each sibship). Pedigrees were selected to have at least 2 affected in each sibship (i.e., total of at least 6 affected). In Table 4, we have reported the power to detect linkage under heterogeneity as a function of the degree of heterogeneity and the recombination fraction. Here, both locus A and B are dominant with 90% penetrance. Power has been calculated for a sample size of 20 families. We can see that there is very good power to detect linkage when 50% of families are linked. If only 25% of families are linked, then there is reasonably good power when $\theta = 1\%$.

A practical issue of this study concerns the general problem of mapping for "common" disorders in presence of heterogeneity in etiology. In particular, we are interested in the probability of finding a linkage to a susceptibility locus. While this study examined perhaps the worst type of genetic heterogeneity, we may relate these simulations to some other situations of heterogeneity in etiology.

For instance, the power to detect linkage to the linked disease locus A, determined from our results, is relevant to the case where the unlinked disease cause B is non-detectable. For example, it can be a representation of multiple genetic causes, each with a small effect. The case of environmental cause is also of interest. We may postulate that the detection of the susceptibility locus will be easier when the unlinked disease cause is non-genetic rather than genetic. Because of the sampling modes (multiple affected), families with the non-genetic disorder are less likely to be selected. However, if in some families the environmental correlation is very large, then these families will be similar to families with an unlinked genetic cause.

Finally, the case of a disease due to two independent, inherited causes is a realistic type of heterogeneity and may concern disorders such as bipolar illness where, as discussed earlier, previous linkage studies suggest the existence of at least two genetic causes (Egeland et al. 1987; Baron et al. 1987).

This simulation study has not only provided us with estimates of the number of pedigrees that are needed for a linkage study but has given us an indication on how the mode of selecting pedigrees affects our ability to detect linkage. It is clear that, assuming sampling methods generally used in practice, a dominant locus has a good chance to be detected (taking into account the values of θ and α), but a recessive locus with reduced penetrance has a very small chance of being detected.

Conclusions

These studies provide an estimate of the amount of resources required to find susceptibility loci for complex diseases given different sampling strategies. However, there is no one best strategy for major psychiatric disorders. As discussed earlier, very large pedigrees are powerful, assuming a single major genetic cause exists in the pedigree. If even one genetic cause can be found, then this may lead to the elucidation of a biological mechanisms. Of course, any finding in a large pedigree must be replicated in a population sample.

We have shown that linkage can be found in clinically realistic sample sizes of nuclear families or medium-sized pedigrees. For nuclear families, efficiency improves with increasing number of siblings, especially increasing the number of ill siblings. The strategy of obtaining just affected pairs of siblings (and parents) is probably not the best. From Table 2, it appears that sibship sizes of 3 are the minimal that should be collected because of the large increment in sample size savings when increasing from 2 to 3 siblings. The choice between studying nuclear families versus pedigrees will depend on the available resources. For a disorder like schizophrenia, even medium-sized pedigrees are difficult to ascertain so that sampling nuclear families may by the best choice. Conclusions obtained from either nuclear families or pedigrees will reflect the general population, assuming that they have been ascertained systematically.

Regardless of the strategy chosen, there are certain limits on the detection of linkage. For example, we have

shown that a recessive locus with reduced penetrance is unlikely to be detected in a sample of pedigrees ascertained for having at least 2 affected (minimal criteria in real studies). There are also limits on the amount of heterogeneity that can be present in order for a locus to be detected. It appears that for both nuclear families and pedigrees, the lower limit of linkage detection is when at least 25% of families are linked. At this lower limit, only close linkage can be detected so that a dense marker map would be required. If only 10% of families are linked to a given marker, this may never be detectable in realistic sample sizes using current technology.

Several questions remain to be answered with regard to the power of screening the whole genome for linkage to a disease. For example, it is worth questioning what is the optimal map density that should be used. This will depend on the polymorphism of the markers and the heterogeneity. We have done some preliminary analyses relevant to this question. For example, if $\alpha = 50\%$, there is some improvement in efficiency when going from 20 cM to 10 cM but little improvement when going from 10 cM to 5 cM. However, as α gets smaller, there will be a larger improvement in efficiency for a denser map. The decision as to whether to use a denser map or collect more families will depend on the cost and availability of clinical versus laboratory resources. The power of linkage detection may also be improved if statistical tests can be designed that take into account findings at multiple marker loci either sequentially or simultaneously (similar to a what was suggested by Lander and Botstein 1986). However, such techniques need to take into account the complex inheritance of psychiatric disorders. The optimal sampling strategies for pedigrees and sampling of individuals within pedigrees needs to be determined. Further investigation is needed on the consequence of various sampling rules for detecting linkage. In addition, within pedigrees, some individuals are more informative than others, depending on the assumed penetrance of the disease gene. Given the large amount of laboratory work required in these studies, some rules need to be designed to make the work most efficient.

Finally, we have looked at the effects of some basic variables on the detection of linkage. However, psychiatric disorders are more complex clinically and there are other known environmental factors (such as age, sex, and cohort) that affect the risk of a person developing a disorder. These factors may also need to be taken into account in determining the power for different sampling strategies.

References

Baron M, Risch N, Hamburger R, Mandel B, Kushner S, Newman M, Drumer D, Belmaker RH (1987) Genetic linkage between X-chromosome markers and bipolar affective illness. Nature 326:289–292

Berrettini WH, Goldin LR, Gelernter J, Gejman PV, Gershon ES, Detera-Wadleigh S (1990) X-chromosome markers and manic depressive illness; rejection of linkage to Xq28 in nine bipolar pedigrees. Arch Gen Psychiatry 47:366–373

Blackwelder WC, Elston RC (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. Genet Epidemiol 2:85–97

Cavalli-Sforza LL, King MC (1986) Detecting linkage for genetically heterogeneous diseases and detecting heterogeneity with linkage data. Am J Hum Genet 38:599–616

Chakravarti A, Badner JA, Li CC (1987) Tests of linkage and heterogeneity in mendelian diseases using identity by descent scores. Genet Epidemiol 4:255–266

Clerget-Darpoux F, Babron MC, Bonaiti-Pellie C (1987) Power and robustness of the linkage homogeneity test in genetic analysis of common disorders. J Psychiatr Res 21:625–630

Detera-Wadleigh SD, Berrettini WH, Goldin LR, Boorman D, Anderson S, Gershon ES (1987) Close linkage of c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in three North American pedigrees. Nature 326:289–292

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

Goldin LR, Gershon ES (1988) The power of affected-sib-pair method for heterogeneous disorders. Genet Epidemiol 5:35–43

Lander ES, Botstein D (1986) Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment lengh polymorphisms. Proc Natl Acad Sci USA 83:7753–7757

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Martinez M, Goldin LR (1989) The detection of linkage and heterogeneity in nuclear families for complex disorders: one vs. two marker loci. Am J Hum Genet 44:552–559

Martinez M, Goldin LR (1990) Power of the linkage test for a heterogeneous disorder due to two independent inherited causes: A simulation study. Genet Epidemiol 7:219–230

Nurnberger JI JR, Goldin LR, Gershon ES (1986) Genetics of psychiatric disorders. In: Winokur G, Clayton P (eds) The Medical basis of psychiatry. Saunders, Philadelphia, pp 486–521

Ott J (1986) The number of families required to detect or exclude linkage heterogeneity. Am J Hum Genet 39:159–165

Penrose LS (1935) The detection of autosomal linkage in data which consists of pairs of brother and sisters of unspecified parentage. Ann Eugen 6:133–138

Penrose LS (1953) The general purpose sib-pair linkage test. Ann Eugen 18:120–124

Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dudleston K, Barraclough B, Wasmuth J, Dobbs M, Gurling H (1988) Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature 336:164–167

Smith CAB (1963) Testing for heterogeneity of recombination fraction values in human genetics. Ann Hum Genet 27:175–182

Suarez BK, Rice J, Reich T (1978) The generalized sib pair IBD distribution: its use in the detection of linkage. Ann Hum Genet 42:87–94

Thomson G, Bodmer WF (1977) The genetics of HLA and disease associations. In: Christiansen FB, Fenchal R, Barndorff-Nielson O (eds) Measuring selection in natural populations. Springer, Berlin Heidelberg New York, pp 545–564

Thomson G (1988) HLA disease associations: models for insulin dependent diabetes mellitus and the study of complex human genetic disorders. Ann Rev Genet 22:31–50