Published in final edited form as:

Comput Stat Data Anal. 2009 March 15; 53(5): 1829–1842. doi:10.1016/j.csda.2008.08.030.

The power of linkage analysis of a disease-related endophenotype using asymmetrically ascertained sib pairs

Heejong Sung^{a,d}, Fei Ji^b, Deborah L. Levy^c, Steven Matthyse^c, and Nancy Role Mendell^{d,*}

^aGenometric Section, IDRB, NHGRI, NIH, Baltimore, MD, USA

^bLaboratory of Statistical Genetics, Rockefeller University, New York, NY, USA

^cPsychology Research Laboratory, McLean Hospital, Belmont, MA, USA

^dDepartment of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY, 11794, USA

Abstract

A linkage study of a qualitative disease endophenotype in a sample of sib pairs, consisting of one disease affected proband and one sibling is considered. The linkage statistic compares marker allele sharing with the proband in siblings with an abnormal endophenotype to siblings with the normal endophenotype. Expressions for the distribution of this linkage statistic, in terms of the recombination fraction are derived and (1) the genetic parameter values (allele frequency and endophenotype and disease penetrance) and (2) the abnormal endophenotype rates in the population and in classes of relatives of disease affected probands. It is then shown that when either the disease or the abnormal endophenotype has additive penetrance, the expressions simplify to a monotonic function of the *difference* between abnormal endophenotype rates in siblings and in the population. Thought disorder is considered as a putative schizophrenia endophenotype. Forty sets of genetic parameter values that correspond to the known prevalence values for thought disorder in schizophrenic patients, siblings of schizophrenics and the general population are evaluated. For these genetic parameter values, numerical results show that the test statistic has >70% power ($\alpha = 0.0001$) in general with a sample of 200 or more proband-sibling pairs to detect the linkage between a marker ($\theta = 0.01$), and a locus pleiotropic for schizophrenia and thought disorder.

1. Introduction

Complex diseases are thought to be caused by several genes, each having a relatively small effect on the clinical phenotype. A complex disease phenotype is an imperfect indicator of genotype, however, due in part to genetic heterogeneity and reduced penetrance, both of which increase the risk of false negatives in linkage analyses. Therefore, the power to detect linkage on the basis of the disease phenotype alone is limited. Genetic analysis of disease

*Corresponding author. Tel.: +1 631 632 8373; fax: +1 631 632 8490. E-mail addresses: heejongsung@gmail.com (H. Sung), nmendell@notes.cc.sunysb.edu (N.R. Mendell).

Uncited references Casella and Berger, 2002, Matthyse et al., 1986 and Torgersen et al., 2002.

endophenotypes with a simpler genetic etiology than that of the disease can increase power substantially (Matthysse et al., 2004).

Gottesman and Gould (2003) define an endophenotype (EDP) as a heritable trait, that co-segregates with the disease and that is state-independent. The EDP should show a significantly higher recurrence risk in unaffected relatives of individuals with the disease than in the general population. Of particular importance, the recurrence risk for the EDP should be substantially higher than risk for the disease in first-degree relatives of affected individuals. EDPs are usually first detected as qualitative abnormalities or abnormal response values (in the case of a quantitative endophenotype) that appear in a large proportion of individuals with the disease. Genetic epidemiologists usually establish that a trait is an endophenotype, by first comparing the distribution of the trait in affected individuals with that of controls. If a large effect size is observed, first-degree relatives are then compared with controls. Assessment of the endophenotype should be done blind with respect to group membership (i.e., whether the individual is a case, a relative of a case, or a control). The statistical analysis assessing the difference in the distributions of the trait in the groups must incorporate all other covariates that may be associated with the candidate endophenotype such as age, gender, income, diet, and drug history.

Incorporating EDPs into linkage analyses of complex diseases has been useful in improving power to detect linked loci that were either not detected at all, or that showed much weaker linkage when the clinical disorder was considered alone (Arolt et al., 1999; Freedman et al., 1997; Gasperoni et al., 2003; Ghosh et al., 2003; Greenberg et al., 2000; Hallmayer et al., 2005; Jones et al., 2004; Keating et al., 1991; Keating and Sanguinetti, 2001; Marlow et al., 2003; Matthysse et al., 2004; Spence et al., 2006; Wang et al., 2006; Wilcox et al., 2002). Most of the cited studies considered quantitative EDPs. However these quantitative EDPs are frequently dichotomized to values above and below a threshold. In a related approach, the definition of “affectedness” is broadened to include other clinical disorders that appear at an increased frequency in family members of probands. These disorders may be alternative expressions of the same genes that contribute to risk for the primary disease of interest (Choi et al., 2006; Houy et al., 2004; Winawer et al., 2005) and hence can also be seen as EDPs. Additionally, there are many circumstances in which the EDP is, in fact, a qualitative trait. For these reasons we focus on a qualitative EDP.

Several gene model-based methods for analysis have been proposed, that incorporate information on both disease and EDP into the analysis. The most commonly used gene model-based method, involves broadening the definition of “affected” to include all individuals who have abnormal EDP, i.e. $E+$, and/or the disease and increasing the penetrance values used in the analysis. Matthysse and Parnas (1992) proposed a methodology for applying latent trait models to linkage analysis that was recently successfully applied to a linkage analysis of the joint distribution of eye tracking dysfunction and schizophrenia (Matthysse et al., 2004). Ji (2004) developed extensions of the gene model-based linkage analysis for a dichotomous phenotype (classical “LOD” score likelihood based linkage analysis) to 4 disease/EDP phenotypes, or equivalently, 4 liability classes for a dichotomous phenotype. Ji (2004) also noted that the power of this gene model-

based “bivariate” approach greatly depends on accuracy of the 7 (or more) genetic parameter values used in the analysis.

Gene model-free methods have been proposed as an alternative to gene model-based methods described above (Amos, 1994; Fiengold, 2001; Weeks and Lange, 1988). These methods evaluate the relationship between the proportion of shared identical alleles by descent (IBD), and phenotype identity in relative (usually sib) pairs. These methods are appealing, because the statistical methods are more straightforward than the gene model-based analysis methods, and the conclusions from the analysis do not depend on assumptions about disease allele frequency and mode of genetic transmission. Gene model-free linkage statistics can be computed far more quickly and easily than the gene model-based linkage statistics, which are based on likelihood maximization. Thus they are more practical for linkage analyses of tens of thousands of SNPs.

Prior to doing linkage analysis, it is advantageous to evaluate the power of the study given the sample size, study design and the statistical approach to be used. Simulation of the results, using a simple major gene generating model is often used for this purpose. Such simulations are necessary if one proposes to use gene model-based linkage analysis methods. However, in the case of a simple major gene generating model, approximate power can be computed numerically.

Risch (1990) elegantly derived expressions for numerically calculating the approximate power of linkage analysis, based on analysis of alleles IBD in *disease affected sib pairs* in terms of risk of the disease in parents, siblings and monozygotic twins of affected individuals. He showed that, if one has a major gene generating model with additive disease penetrance, power of a model-free linkage analysis is a simple function of relative risk of the disease in siblings of probands to that of the general population, recombination fraction and sample size.

In this paper, we derive a method for *direct* computation of approximate power for a family-based gene model-free linkage analysis of a qualitative EDP in a sample of sib pairs acquired using an “asymmetric” ascertainment design. In this design, probands are ascertained for having a disease, but siblings are examined for recurrence on the EDP (although they can be examined for recurrence on the disease as well).

Our derivations consider a biallelic pleiotropic gene that contributes to risk for both a disease and a qualitative EDP. Although this gene is one of several risk factors for the disease, it is a *major* gene for the EDP. Thus, the allele has much higher penetrance for the disease-associated abnormal form of the EDP (E^+), than for the disease itself (D^+).

We consider a statistic that compares the average number of alleles, identical by descent, with the D^+ proband (IBDP) of EDP abnormal siblings to the average observed in the EDP normal siblings (Sung et al., 2005) and develop the mathematical expressions for numerically calculating the power of this statistic. Our first goal is to derive a general expression for calculating the power of linkage analysis for a given sample size and recombination fraction, as a function of 7 genetic parameters of the pleiotropic major gene generating model being considered. These are (1) disease/EDP pleiotropic allele frequency,

(2)–(4) the $D+$ penetrance of each of the 3 genotypes and (5)–(7), the $E+$ penetrance of each of the 3 genotypes. A second goal is to extend the methodologies developed by Risch (1990) and Suarez et al. (1978), and determine simplified expressions for power that depend only on recurrence risks in classes of relatives and population risks.

To illustrate the application of our findings, we apply these numerical methods to genetic epidemiology findings on the rates of schizophrenia-related thought disorder, a putative EDP for schizophrenia, in schizophrenics, siblings of schizophrenics, and nonpsychiatric controls to calculate power to detect linkage to the EDP where the sample sizes are 200, 400 and 500 sib pairs.

We compare our findings with those obtained, using simulations for a model-based (lod score) analysis on a sample of families that would generate an equivalent number of sib pairs.

2. Asymmetrically ascertained sib pairs (AASP): Study design and statistics

The study design we consider, involves a random sample of n independent asymmetrically ascertained sib pairs (AASP). That is, we sample n unrelated probands (individuals with the disease of interest), who have at least one sibling and then we randomly sample one sibling per proband. Each of these AASP provides information on the sibling's EDP status, E , a binary variable with outcomes $E+$ and $E-$. Here $E+$ denotes the abnormal, less frequent form of the EDP (i.e., 'affected') and $E-$ denotes the normal, more frequent form of the EDP (i.e., 'unaffected'). We consider the situation where every AASP has completely informative marker information. That is, both parents' marker genotypes are known, and both parents are heterozygous with different genotypes. Thus, the number of alleles identical by descent to the proband, IBDP in an AASP, I , equals the observed number of identical alleles in the sib pair.

These data values can be summarized by two sets of summary statistics. A first set of summary statistics would be the 3×2 table of observed joint frequencies of sibling EDP status, E , and IBDP status, I . Letting n_{ij} denote the observed number of AASP with $I = i$ for $i = 0, 1, 2$ and $j = E+, E-$ we illustrate this statistical summary in Table 1.

The second set of summary statistics is the average values of I for the $E+$ siblings and the $E-$ siblings, $\bar{I}_{E+}$ and $\bar{I}_{E-}$ respectively. These are calculated directly using the values of the n_{ij} as

$$\bar{I}_j = \frac{2n_{2j} + n_{1j}}{n_j} \quad \text{for } j = E+, E-.$$

3. Numerical methods for calculating the power of linkage analysis: The IBDP distribution as a function of disease-related endophenotype penetrance and allele frequency

We consider a situation where we have a single biallelic pleiotropic locus, $D+ / E+$, which determines the presence or absence of both disease and abnormal endophenotype. One of the two alleles is associated with increased risk of both disease ($D+$) and abnormal

endophenotype ($E+$), and we refer to this as the $D+ /E+$ allele. We let G_j ($j = 0, 1, 2$) denote the $D+/E+$ locus genotype with j copies the $D+/E+$ allele. We then define $D+/E+$ locus parameters that determine the joint distribution of $D+$ and $E+$ in the general population and in relatives of $D+$ individuals: (1) p , the population frequency of the $D+ /E+$ allele, and (2)–(7) f_{jt} , the penetrance of $t = D+, E+$ for genotype G_j for $j = 0, 1, 2$. We will detect $D+ /E+$ through linkage analysis of a closely linked marker locus M . Because these loci are closely linked, θ , the recombination fraction, the probability of a recombination event occurring between the marker M and the $D+ /E+$ locus, is a value close to 0.0 and substantially less than 0.5.

We extend the methods for obtaining the IBD distribution at the marker locus M , in $D+$ sib pairs (Risch, 1990; Suarez et al., 1978), and derive the IBDP distribution at marker locus M conditional on the sibling EDP , and the proband *disease phenotype* as a function of p , the f_{jt} ($j = 0, 1, 2$) $t = D+, E+$ and θ (Appendix E, Supplementary materials) Specifically the distribution of I_M for an $E+$ sibling in an AASP is (equation E.1 in Appendix E, Supplementary materials)

$$P \left(I_M | X_{D+}^{pro}=1, X_{E+}^{sib}=1 \right) = \sum_{I_L=0}^2 \left(P(I_M | I_L) \cdot \frac{\sum_{d,m=0}^2 \sum_{h,k=0}^2 (f_{kE+} \cdot f_{hD+} \cdot P(I_L | G_h^{pro}, G_k^{sib}, M_{d \otimes m}) \cdot P(G_h^{pro}, G_k^{sib} | M_{d \otimes m}) \cdot P(M_{d \otimes m}))}{P(X_{E+}^{sib}=1 | X_{D+}^{pro}=1) \cdot P(X_{D+}^{pro}=1)} \right). \quad (1)$$

Here $I_M | I_L$ denotes the number of alleles IBD at the marker (disease/EDP) locus ($I_i = 0, 1, 2$ where $i = M, L$) and $X_{D+}^{pro} (X_{E+}^{sib})$ denotes the disease (EDP) affection status of the proband (sibling), $X_{D+}^{pro} (X_{E+}^{sib}) = 0$ if $D-(E)$ and 1 if $D+(E)$, and $M_{d \otimes m}$ denotes a sib pair, in which the father is genotype G_d and the mother is genotype G_m ($d, m = 0, 1, 2$).

The expressions for $P(I_M | I_L)$ for $I_M, I_L = 0, 1, 2$ have been shown to be a simple function of the recombination fraction, θ (Haseman and Elston, 1972). These are included in Appendix A.

We assume random mating with respect to $D+$ and $E+$, thus $P(M_{d \otimes m}) = P(G_d) \cdot P(G_m)$. We also assume that sibling genotypes conditional on parents' genotypes are independent:

$P(G_h^{pro}, G_k^{sib} | M_{d \otimes m}) = P(G_h^{pro} | M_{d \otimes m}) \cdot P(G_k^{sib} | M_{d \otimes m})$. We assume Hardy Weinberg Equilibrium, and obtain the values of $P(G_j)$ for $j = 0, 1, 2$ as a function of p , the frequency of the $D+ /E+$ predisposing allele, as $P(G_0) = (1 - p)^2$, $P(G_1) = 2p(1 - p)$, $P(G_2) = p^2$.

The values of $P(G_k^{sib} | M_{d \otimes m})$ and $P(G_h^{pro} | M_{d \otimes m})$, for $h, k = 0, 1, 2$, $d, m = 0, 1, 2$ are simply the Mendelian probabilities of offspring genotypes conditional on parents' genotypes. For completeness these are also given in Appendix A.

The equations for calculating $P(X_{E+}^{sib}=1|X_{D+}^{pro}=1)$, EDP recurrence risk in siblings of $D+$ probands, $K_S^{E+|D+}$, as a function of disease/EDP predisposing allele frequency (p) and the $D+$ and $E+$ penetrance values (f_{kt} , $k=0, 1, 2$ and $t=D+, E+$) are given in Appendix B (Ji, 2004). The details of computing the conditional IBDP distributions as well as the values of $P(I_L|G_k^{pro}, G_h^{sib}, M_{d\otimes m})$ are in Appendix E, Supplementary materials.

4. Numerical methods for obtaining the IBDP distribution as a function of recurrence risks

The IBD distribution at the disease locus for a sib-pair conditional on the number of affected siblings (0, 1, or 2) in the pair was derived by Suarez et al. (1978) using variance component parameterization. These results were generalized by Risch (1990) to other classes of relatives, and were shown to simplify to functions of risk ratios (the recurrence risk of the disease in relatives of $D+$ probands divided by the population risk) in different classes of relatives of $D+$ probands. We here use Risch's 1990 methods to derive the IBD distribution at the disease/EDP locus conditional on the EDP status of the sibling in the AASP as a function of $E+$ recurrence risk in AASP and population risk of $E+$.

Let K^{E+} and $K_R^{E+|D+}$ denote abnormal EDP recurrence risk in the general population, and in a family member of a $D+$ proband of relationship R , respectively. Here, R may take the value of M (monozygotic twin), S (sibling), O (offspring), C (first cousin), etc. Thus,

$K_S^{E+|D+} = P(X_{E+}^{sib}=1|X_{D+}^{pro}=1)$. The ratio of abnormal EDP recurrence risks in relatives of relationship R (to the $D+$ proband) to the general population risk can be denoted by

$\lambda_R^{E+|D+} = \frac{K_R^{E+|D+}}{K^{E+}}$. These are the relative risks of the abnormal EDP in relatives of relationship R . Similar to Risch's 1990 result for $D+$ sib pairs, we can show that

$\lambda_M^{E+|D+} = 4\lambda_S^{E+|D+} - 2\lambda_O^{E+|D+} - 1$ (Eq. (C.8) and (C.9) in Appendix C).

The IBDP distribution of the disease proband/EDP sibling (Eq. (1)) can be re-expressed in terms of the recurrence risks of the EDP. The general form of the distribution of IBDP conditional on the EDP ($x=0, 1$) of the sibling of the $D+$ proband (Eq. (C.1) in Appendix C) is

$$P(I_M|X_{D+}^{pro}=1, X_{E+}^{sib}=x) = \sum_{I_L=0}^2 P(I_M|I_L) \cdot \frac{P(X_{E+}^{sib}=x|X_{D+}^{pro}=1, I_L) \cdot P(I_L)}{P(X_{E+}^{sib}=x|X_{D+}^{pro}=1)} \quad (2)$$

for $x=0, 1$. Here $E=E+$ if $x=1$, $E=E-$ if $x=0$.

Using our notation for population risks and recurrence risks given above, we note that

$P(X_{D+}=1) = K^{D+}$ and $P(X_{E+}^{sib}=x|X_{D+}^{pro}=1) = K_S^{E|D+}$. Noting that

$$P\left(X_{E+}^{sib}=x|X_{D+}^{pro}=1, I_L=0\right)=P\left(X_{E+}^{sib}=x\right)=K^E,$$

$$P\left(X_{E+}^{sib}=x|X_{D+}^{pro}=1, I_L=1\right)=K_O^{E|D+} \quad \text{and} \quad P\left(X_{E+}^{sib}=x|X_{D+}^{pro}=1, I_L=2\right)=K_M^{E|D+}$$

respectively where $E = E+$ if $x = 1$, $E = E-$ if $x = 0$, we obtain the IBDP distribution in terms of EDP recurrence risks in relatives and the recombination fraction as follows (Derivations are given in Appendix (C.1), (C.6), and (C.7) respectively):

$$P\left(I_M=0|X_{D+}^{pro}=1, X_{E+}^{sib}=x\right)=\psi^2 \frac{K^E}{4K_S^{E|D+}} + \psi(1-\psi) \frac{K_O^{E|D+}}{2K_S^{E|D+}} + (1-\psi)^2 \frac{K_M^{E|D+}}{4K_S^{E|D+}} \quad (3)$$

$$P\left(I_M=1|X_{D+}^{pro}=1, X_{E+}^{sib}=x\right)=2\psi(1-\psi) \frac{K^E}{4K_S^{E|D+}} + (1-2\psi(1-\psi)) \frac{K_O^{E|D+}}{2K_S^{E|D+}} + 2\psi(1-\psi) \frac{K_M^{E|D+}}{4K_S^{E|D+}} \quad (4)$$

and

$$P\left(I_M=2|X_{D+}^{pro}=1, X_{E+}^{sib}=x\right)=(1-\psi)^2 \frac{K^E}{4K_S^{E|D+}} + \psi(1-\psi) \frac{K_O^{E|D+}}{2K_S^{E|D+}} + \psi^2 \frac{K_M^{E|D+}}{4K_S^{E|D+}} \quad (5)$$

where $\psi = \theta^2 + (1-\theta)^2$ ($0 \leq \theta \leq 0.5$) (Appendix A) and $E = E+$ if $x = 1$, $E = E-$ if $x = 0$.

Moreover, values of the $K_R^{E|D+}$ ($R = S, M, O$), and K^E , $E = E+, E-$ in Eqs. (3)–(5) can be derived using genetic parameters of disease and EDP penetrances and allele frequency (Appendix E, Supplementary materials). However, since in reality, recurrence risks are more estimable than genetic parameters, investigators might use empirically obtained population risks and recurrence risks, instead.

As a special case, we consider the situation where $\lambda_S^{E+|D+} = \lambda_O^{E+|D+}$. This is analogous to the assumption of $\lambda_S^{D+|D+} = \lambda_O^{D+|D+}$ that was used by Risch (1990) in calculating power in a symmetrical ascertainment design. He noted that when $\lambda_S^{D+|D+} = \lambda_O^{D+|D+}$ and $\theta=0$, the distribution of I_M is a function of $K_S^{D+|D+}$ and K^{D+} . Similarly, for the case where $K_S^{E+|D+} = K_O^{E+|D+}$ and $\theta = 0$, Eqs. (3)–(5) above can be simplified further to

$$P\left(I_M=0|X_{D+}^{pro}=1, X_{E+}^{sib}=x\right)=\frac{K^E}{4K_S^{E|D+}} \quad (6)$$

$$P(I_M=1|X_{D+}^{pro}=1, X_{E+}^{sib}=x) = \frac{1}{2} \quad (7)$$

and

$$P(I_M=2|X_{D+}^{pro}=1, X_{E+}^{sib}=x) = \frac{1}{2} - \frac{K^E}{4K_S^{E|D+}} \quad (8)$$

for $x = 0, 1$. Here $E = E+$ if $x = 1$, $E = E-$ if $x = 0$ and

$$K^{E-} = 1 - K^{E+}, K_S^{E-|D+} = 1 - K_S^{E+|D+}.$$

Eqs. (6)–(8) also give expressions for the conditional mean and variance of I_M for the case where $\theta = 0$ as

$$E(I_M|X_{D+}^{pro}=1, X_{E+}^{sib}=x) = \frac{3}{2} - \frac{K^E}{2K_S^{E|D+}} \quad (9)$$

and

$$\text{Var}(I_M|X_{D+}^{pro}=1, X_{E+}^{sib}=x) = \frac{1}{4} + \frac{K^E}{2K_S^{E|D+}} - \left(\frac{K^E}{2K_S^{E|D+}} \right)^2 \quad (10)$$

where, as above, $E = E+$ if $x = 1$, $E = E-$ if $x = 0$ and

$$K^{E-} = 1 - K^{E+}, K_S^{E-|D+} = 1 - K_S^{E+|D+}.$$

Risch (1990) noted that the relationship $\lambda_S^{D+|D+} = \lambda_O^{D+|D+}$ occurs whenever the $D+$ allele has an additive effect on the penetrance, i.e., $f_{1D+} = (f_{0D+} + f_{2D+})/2$ (Risch, 1990).

However, we have noted (Appendix H) and (Sung, 2006) that this is the case whenever the disease/EDP allele has additive effect on $E+$ penetrance or $D+$ penetrance. That is,

$$K_S^{E+|D+} = K_O^{E+|D+} \text{ whenever } f_{1D+} = (f_{0D+} + f_{2D+})/2 \text{ or } f_{1E+} = (f_{0E+} + f_{2E+})/2.$$

5. Linkage statistic and power calculation

Eqs. (1)–(10) give expressions for calculating the distribution of I , the number of alleles IBDP conditional on whether the sibling of the proband has the abnormal EDP or the normal EDP for an arbitrary set of genetic penetrance, and recombination fraction values or for an arbitrary set of recurrence risks and recombination fraction values. In this section, we derive a method to calculate the approximate power of a linkage analysis.

We consider the power of a two-sample mean test. This test compares the average number of alleles IBDP in $E+$ siblings (E_+) with the average number of alleles IBDP in $E-$ siblings (E_-). The test statistic we consider is

$$Z = \frac{\bar{I}_{E+} - \bar{I}_{E-}}{\sqrt{V_0}} \quad \text{where} \quad V_0 = \frac{1}{2} \left(\frac{1}{n_{E+}} + \frac{1}{n_{E-}} \right). \quad (11)$$

Here, E_+ , E_- and n_{E+} , n_{E-} are as defined in Section 2.

We refer to this two-sample statistic as the AASP-Z statistic. This statistic successfully identified “pleiotropic genes” in a simulated Genetic Analysis Workshop 14 data set (Sung et al., 2005). It can be shown to be equivalent to a Cochran Armitage test on the same dataset, and thus is equivalent to -1.0 times the Haseman–Elston statistic (Haseman and Elston, 1972), in which the phenotype difference is set equal to ‘1’ in proband sib pairs with $E-$ sibling and ‘0’ in proband sib pairs with an $E+$ sibling.

The AASP-Z statistic has an asymptotically standard normal distribution under the null hypothesis of no linkage. The approximate power of this test statistic can be calculated (for more details, see Appendix F and Appendix G in the Supplementary materials) as

$$\text{power} \approx P \left(Z > \frac{z_\alpha \cdot \sqrt{V_0} - E(\bar{I}_{E+} - \bar{I}_{E-})}{\sqrt{V_1}} \right) = P \left(Z > z_\alpha \sqrt{\frac{V_0}{V_1}} - \frac{E(I_{E+}) - E(I_{E-})}{\sqrt{V_1}} \right) \quad (12)$$

where

$$V_0 = \frac{1}{2} \cdot \left(\frac{1}{E(n_{E+})} + \frac{1}{E(n_{E-})} \right) = \frac{1}{2} \cdot \left(\frac{1}{n \cdot K_S^{E+|D+}} + \frac{1}{n \cdot (1 - K_S^{E+|D+})} \right) \quad (13)$$

and

$$V_1 = \text{var} \left(\bar{I}_{E+} \right) + \text{var} \left(\bar{I}_{E-} \right) \approx \frac{\text{var}(I_{E+})}{n \cdot K_S^{E+|D+}} + \frac{\text{var}(I_{E-})}{n \cdot (1 - K_S^{E+|D+})}. \quad (14)$$

In the special case of $\lambda_S^{E+|D+} = \lambda_O^{E+|D+}$ and $\theta = 0$, expressions of the IBDP distribution of the disease/EDP allele are shown in Eqs. (6)–(8) and the conditional mean and variance of

I_M are shown in Eqs. (9) and (10) above. Power can then be expressed as a function of K^{E+} and $K_S^{E+|D+}$ where effect size is defined as follows.

$$\text{Effect size} = \frac{E(I_{E+}) - E(I_{E-})}{\delta_0} = \frac{\frac{K_S^{E+|D+} - K^{E+}}{2K_S^{E+|D+} \cdot (1 - K_S^{E+|D+})}}{\sqrt{\frac{1}{2}}} = \frac{K_S^{E+|D+} - K^{E+}}{\sqrt{2} \cdot K_S^{E+|D+} \cdot (1 - K_S^{E+|D+})}.$$

(15)

Here, δ_0 denotes the standard deviation of I_{E+} and I_{E-} under the null hypothesis of no linkage. The denominator term will increase as $K_S^{E+|D+}$ gets closer to 0.5. Thus, we conjecture that in this case the minimum ($K_S^{E+|D+}=0.5$) power depends on the term $K_S^{E+|D+} - K^{E+}$.

Using the software “R”, we created a program to calculate the conditional distribution of I , given the EDP status of the sibling and the approximate power of the AASP-Z statistic to detect linkage for a given sample size of n AASP and recombination fraction θ . Input to this program can be specified genetic parameters (recombination fraction (θ), disease/EDP allele frequency (p) and $D+$ penetrance values, $E+$ penetrance values (f) or specified epidemiology parameters, i.e., three out of the four parameters K^{E+} , $K_S^{E+|D+}$, $K_O^{E+|D+}$, and $K_M^{E+|D+}$ (since $K_M^{E+|D+} = 4K_S^{E+|D+} - 2K_O^{E+|D+} - 1$, Appendix C). This program is available upon request to the first author (sungho@mail.nih.gov).

6. Application I: Evaluation of power based on genetic epidemiology parameter values

Genetic epidemiology studies that are focused on identifying EDPs frequently provide estimates of (1) $K_S^{E+|D+}$, the $E+$ recurrence risk in siblings of $D+$ probands (2) $K_I^{E+|D+}$, the risk of $E+$ in $D+$ individuals and (3) K^{E+} , the risk of $E+$ in the general population. We consider a situation here, where we assume $K_O^{E+|D+} = K_S^{E+|D+}$ and apply Eqs. (6)–(10). Thus these result apply only in those cases where the disease/EDP allele has an additive effect on $E+$ penetrance or $D+$ penetrance i.e., $f_{1D+} = (f_{0D+} + f_{2D+})/2$ or $f_{1E+} = (f_{0E+} + f_{2E+})/2$.

We consider the 25 pairs of epidemiology values generated by setting $K^{E+} = 0.02, 0.05, 0.1, 0.15$ and 0.2 and $K_S^{E+|D+} = K_O^{E+|D+} = 0.31, 0.37, 0.42, 0.45$ and 0.48 . We use the expressions given in Eqs. (6)–(8) to calculate the means and variances of I_{E+} and I_{E-} , in order to calculate power using the AASP-Z test for linkage to a marker with $\theta = 0.0$ and Type I error rate $\alpha = 0.0001$ with $n = 200, 400$ and 500 AASP. Fig. 1 plots the power for these 25 cases.

As the figure shows, when $K_s^{E+|D+} - K^{E+}$ is close to 0.25, power to detect linkage to a locus for the EDP is <80% with a sample size of 200 AASP. As $K_s^{E+|D+} - K^{E+}$ gets close to 0.3, power increases up to 100% in a sample of 200 AASP. Thus, there is a clustering of the values at 100% for all three sample sizes for all $K_s^{E+|D+} - K^{E+} > 0.3$. This figure can be used to obtain the minimum sample size required to achieve the power value for specified values of $K_s^{E+|D+} - K^{E+}$.

Fig. 1 also shows that the power is very low, even with a sample size of 500 AASP, when the difference in risk for $E+$ between sibs of $D+$ probands and the general population is very low ($K_s^{E+|D+} - K^{E+} = 0.11$). These low power values correspond to the values of risk difference in $D+$ sib pairs that would be observed if $D+$ penetrance values were additive and the linkage analysis were restricted to the disease ($D+$); i.e., $K_s^{D+|D+} - K^{D+} = 0.11$.

7. Application II: Evaluating the power of linkage analysis of a Schizophrenia Endophenotype using AASP (calculation of power based on genetic parameter values)

Thought disorder occurs in a variety of psychiatric syndromes. The kind of thought disorder that is characteristic of schizophrenia differs from that found in nonschizophrenic psychotic conditions, such as bipolar disorder (Andreasen and Grove, 1986; Holzman et al., 1986; Johnston and Holzman, 1979; Makowski et al., 1997; Spohn et al., 1986). The most characteristic feature of schizophrenic thought disorder is deviant verbalizations. Deviant verbalizations refer to peculiar or odd use of language, for example, “rectangularly speaking”, “a tree head kind of person”, “posterior pronunciations”. Deviant verbalizations are also characteristic of the thought disorder found in clinically unaffected first-degree relatives of schizophrenics (Johnston and Holzman, 1979; Shenton et al., 1989; Vaever et al., 2005; Wahlberg et al., 2000), and account for the elevated amount of thought disorder found in these relatives, compared with relatives of nonschizophrenics and controls (Arboleda and Holzman, 1985; Docherty and Gottesman, 2000; Johnston and Holzman, 1979; Kinney et al., 1997; Romney, 1984).

Thought disorder was assessed using the Thought Disorder Index (TDI) (Johnston and Holzman, 1979; Solovay et al., 1986). The TDI quantifies the amount and severity of disordered thinking, using 23 categories of thought disorder, each of which is weighted at one of four severity levels (for a comprehensive review of research using the TDI (Holzman et al., 2005). Each instance of thought slippage was identified based on verbatim transcripts of responses to the ten Rorschach cards that were scored by consensus among expert raters who were blind to group identity. The protocol was then assigned a “TDI Diagnosis” by the raters. The TDI diagnosis is a qualitative designation that characterizes whether thought disorder with “schizophrenic” features (i.e., deviant verbalizations) was ($T+$) or was not ($T-$) a predominant feature of the record.

The sample consisted of 105 individuals who met DSM IV criteria for a diagnosis of schizophrenia or schizo-affective disorder, 133 of their clinically unaffected (i.e., no psychosis, lifetime; no bipolar disorder without psychotic features, no schizotypal, paranoid or schizoid personality disorder) full siblings and 86 nonpsychiatric controls. The rates of $T+$ in schizophrenics, siblings of schizophrenics, and controls were 72.4%, 39.85%, and 11.6%, respectively, indicating that $T+$ is a promising EDP for schizophrenia. A more detailed report can be found in Levy et al. (in preparation-a).

The following estimates of $T+$ risk and $D+$ recurrence risk were used in the calculations based on the equations given in Appendix B: (1) $K_s^{T+|SZ+}=0.398$ ($SE=0.083$), i.e., 39.8% of siblings of schizophrenics ($SZ+$) have $T+$, (2) $K_I^{T+|SZ+}=0.724$ ($SE=0.086\%$), i.e., 72.4% of schizophrenics have $T+$, (3) $K^{T+}=0.157$ ($SE=6\%$), or 15.7% ($SE = 6\%$) of the general population has $T+$. The latter rate has been modified to allow for unscreened members of the general population, and is thus higher than the rate observed in screened controls (about 12%). Risk for schizophrenia (Slater, 1968) was estimated to be (4) $K^{SZ+}=0.0065$ ($SE=0.0005$), i.e. 0.65% of the population has schizophrenia, and risk for schizophrenia in siblings of schizophrenics was estimated to be (5)

$K_s^{SZ+|SZ+}=0.089$ ($SE=0.048$), or 8.9% of siblings of schizophrenics have schizophrenia based on this sample (Levy et al., in preparation-a), a value consistent with other data (Kendler et al., 1993).

In this example, we consider all generating genetic model parameter values that would result in values within 1 SE of the values of K^{T+} , $K_s^{T+|SZ+}$, $K_I^{T+|SZ+}$ and $K_s^{SZ+|SZ+}$, given above for thought disorder and schizophrenia (for K^{SZ+} a $2SE$ range was used). Software was developed to input genetic parameter values and to calculate the resulting values of

K^{T+} , $K_s^{T+|SZ+}$, $K_I^{T+|SZ+}$, K^{SZ+} , and $K_s^{SZ+|SZ+}$ using Eq. (B.1) in Appendix B. Over 3 million possible genetic parameter values were considered. The search was stopped when 40 sets of genetic parameter values were found that resulted in values of $0.097 \leq K^{T+} \leq 0.217$, $0.315 \leq K_s^{T+|SZ+} \leq 0.482$, $0.64 \leq K_I^{T+|SZ+} \leq 0.81$, $0.04 \leq K_s^{SZ+|SZ+} \leq 0.137$ and $0.0055 \leq K^{SZ+} \leq 0.0075$. These genetic parameter values are listed in Appendix D. Most of these pleiotropic genetic parameter values corresponded to dominant high penetrance values for $T+$ ($f_{2T+} > 0.65$ and $f_{1T+} > (f_{0T+} + f_{1T+})/2$ with 1 exception) and to recessive low penetrance values for $D+$ ($f_{2D+} \leq 0.30$ and $f_{1D+} < (f_{0D+} + f_{1D+})/2$). For the vast majority of the parameter values, phenocopy rates for $T+$ were high ($f_{0T+} > 0.1$). A detailed discussion of the power of asymmetric ascertainment using gene model-based linkage analysis of a bivariate trait, illustrated by these 40 models, can be found in Kinney et al. (1997) and Levy et al. (in preparation-a).

We used these genetic parameter values to calculate the approximate power of linkage analysis of the EDP, schizophrenia-related thought disorder, in 174 AASP. The results of these findings are shown in Fig. 2. There appears to be a positive correlation between $K_s^{T+|SZ+} - K^{T+}$ and power with the model-free method. However, as Fig. 2 also shows,

there is a range of power values for $K_s^{T+|SZ+} - K^{T+}$ equal to 0.25 and 0.35, illustrating graphically the expected finding that if both disease penetrance and endophenotype penetrance are nonadditive, as they are in all of these parameter values, then the risk difference does not totally determine power.

In particular, when $K_s^{T+|SZ+} - K^{T+} > 0.31$ there appear to be two ‘outliers’, i.e., two parameter values for which the power values are much less than those for three other parameter values having the same risk difference. For all of these models, the values of $K_s^{T+|SZ+} - K^{T+}$ equals about 0.32. The two parameter settings with the lower power values (0.74 and 0.84) share similar allele frequency ($\hat{p} = 0.08, 0.10$) and penetrance values for thought disorder ($f_{2T+} = 0.82, 0.81$, $f_{1T+} = 0.80, 0.77$ and $f_{0T+} = 0.05, 0.01$, respectively) but not for schizophrenia in those homozygous for the $SZ+ / T+$ allele ($f_{2SZ+} = 0.55, 0.32$; $f_{1SZ+} = 0.012, 0.015$ and $f_{0SZ+} = 0.0017, 0.0014$). Upon comparing these 2 parameter settings to the 3 settings that also had $K_s^{T+|SZ+} - K^{T+} = 0.32$, but had much greater power (0.986, 0.987, and 0.988), allele frequency values are lower ($p = 0.027, 0.014$, and 0.013) and penetrance values for schizophrenia in those heterozygous for the $SZ+ / T+$ allele are higher ($f_{1SZ+} = 0.11, 0.19$ and 0.21) than the corresponding values observed in the 2 settings with lower power. Thus, the two low power settings occur when the $SZ+ / T+$ allele is recessive for $SZ+$ and dominant for $T+$. In contrast, in the three settings with high power, the $SZ+ / T+$ allele is dominant for $SZ+$, and dominant for $T+$. These values resulted in computed values of $K_o^{T+|D+} \neq K_s^{T+|D+}$ for the two settings with the lower power ($K_s^{T+|D+} = 0.48$, and 0.47; $K_o^{T+|D+} = 0.57$ and 0.53), and values of $K_s^{T+|D+} = K_o^{T+|D+}$ for the three settings with the higher power ($K_s^{T+|D+} = K_o^{T+|D+} = 0.41, 0.41$, and 0.47). This accounts for why these two low power values are ‘outliers’ ($K_o^{T+|D+} \neq K_s^{T+|D+}$), and why the 3 other settings generate power values that increase monotonically with $K_s^{T+|SZ+} - K^{T+}$. We also observe lower values of $K_M^{T+|D+} = 0.65, 0.67$ for the two parameter values with lower power (as compared to $K_M^{T+|D+} = 0.72, 0.73$ and 0.79).

Fig. 2 also shows that the empirical power obtained using gene model-based analysis is generally more powerful than that for the model-free analyses. However, for values of risk difference of 0.3 or greater, there are many situations where there is little difference in power. These gene model-based results were obtained by doing $N = 1000$ simulations of 100 nuclear families, using the family structures of our schizophrenia family data base for each of the genetic parameter values given in Table D.1. of Appendix D (see also Ji (2004)). In model-based lod score analysis, we considered both disease and EDP and assumed independence of $E+$ and $D+$ conditional on genotype. The lod score was then calculated for each sample of 100 families, using the generating model parameter values (and hence maximizing the power of the gene model-based analyses). The power to detect linkage in 174 sib pairs is generally quite good for most models whose generating parameters are consistent with our observed data on thought disorder recurrence risk.

We also calculated the power of the two-sample AASP-Z test as a function for $n = 400$ for these 40 parameter sets, setting $\theta = 0.01$. With $n = 400$ AASP sib pairs power was well over 90% for all 40 genetic parameter values.

8. Discussion and conclusions

Our numerical calculations show that the risk difference $K_s^{E+|D+} - K^{E+}$, not the risk ratio ($K_s^{E+|D+} / K^{E+}$), is the driving genetic epidemiology parameter in determining the power of a linkage analysis of an EDP. This finding is due to the fact that the linkage statistic we are investigating compares two samples of sib pairs, $E+$ and $E-$ siblings. The predictive power of the risk ratio (Risch, 1990), in contrast, applies to a single sample of sib pairs, all of whom are uniformly concordant (or discordant) for some trait. The positive correlation between $K_s^{E+|D+} - K^{E+}$ and power indicates that power is likely to be greater for linkage analyses of an EDP, than for the disease itself so long as the EDP risk difference is greater than the disease risk difference, $K_s^{E+|D+} - K^{E+} > K_s^{D+|D+} - K^{D+}$. These findings also indicate that modest sample sizes on the order of $n = 200$ sib pairs have excellent power, as long as $K_s^{E+|D+} - K^{E+}$ is 0.30.

It is important to note that power calculations based on differences in risk of the abnormal phenotypes ($E+$ risk in siblings of $D+$ probands minus the $E+$ risk in the population), and relative risk apply only when at least one of the two pleiotropic traits (e.g., disease or EDP) has additive penetrance. Power calculations based on relative risk ratios for either trait are subject to the same caveat.

We first observed this as a requirement in doing the numerical calculations, and have since proven it algebraically. We considered 15 sets of genetic parameter values in which only the EDP had additive penetrance, 12 in which only the disease had additive penetrance, 3 sets in which both the abnormal EDP and the disease had additive penetrance, as well as 16 sets of genetic parameter values in which both EDP and disease had non additive penetrance values (i.e. both dominant, both recessive or 1 dominant and 1 recessive). In the situations where either the penetrances of $D+$ or $E+$ were additive, we computed abnormal EDP risks in the sibling of $D+$ proband equal to abnormal EDP risks in the offspring of $D+$ parent. In the remaining situations (non-additive penetrance in the heterozygote for both disease and EDP), these two risks were not equal. These models with their corresponding abnormal EDP risks, their distributions of the alleles IBDP and approximate power for a sample of $n = 200$ sib pairs, are given in the Supplementary material (Appendix H).

When the disease is recessive and the EDP is dominant, any single risk difference can result in a range of power values, some of which can be greater or less than those obtained based on the simplified Eqs. (6)–(10). This is an important caveat, because most of the models that are consistent with our data on the joint distribution of schizophrenia ($D+$) and thought disorder ($E+$), are recessive for $D+$ and dominant for $E+$ (see Table D.1, Appendix D).

The genetic parameter values we considered for evaluating power to detect linkage to a thought disorder/schizophrenia locus were obtained through an unbiased search for values

compatible with our recurrence risks estimates. In reality, investigators would probably consider a range of values for each of these seven parameters. They would then need to establish that some values result in reasonable power for their sample size, and that these values correspond to realistic values of endophenotype risk in diseased individuals, siblings of diseased individuals and the general population.

Levy et al. (2005, in preparation-b) developed a method to estimate power of a test, comparing the frequency of $D+$ and $E+$ conditioned on IBDP status in AASP. Their approximation applies to diseases with low gene frequency, for which they showed that power was a direct function of $E+$ recurrence risk in siblings of $D+$ probands. Our numerical methods for evaluating power apply to a more general case, but consider an algebraically equivalent test statistic and the same asymmetrical ascertainment design.

It is reassuring that the AASP-Z statistic is only slightly less powerful than the model-based lod score analysis for the situations we consider. Implementation of model-free linkage analysis is more practical than implementation of model-based linkage analysis, especially in the context of genome-wide scans, where thousands of candidate marker loci are considered. Moreover, our simulation results for the lod score analysis are higher than what would realistically be obtained in practice, since they are based on analyses that specify the correct genetic parameter values, which are usually not known.

An alternative to the asymmetrical ascertainment design, which conditions ascertainment of siblings on the presence of disease in the proband, would be to ascertain sib pairs for having at least one individual with the abnormal EDP. In theory, such a design might have at least as good, if not better, power to detect linkage to a major gene for the EDP. In the case of EDPs that are also common risk factors and/or deleterious in their own right, such sib pairs would be medically worthwhile and feasible to ascertain. Hypertension, a risk factor for heart disease, is an example of such an EDP. However, where a candidate EDP is not dysfunctional in its own right, for example, intermittently odd use of language, a sample of sib pairs ascertained for an abnormal EDP in at least one member may not be feasible or scientifically compelling. Additionally, if the $E+$ “probands” are likely to be phenocopies, recurrence risk in their siblings is likely to be low and of meager value for genetic studies. Conditioning on the disease in the proband, on the other hand, is particularly efficient in that it maximizes selection for a gene that has a higher probability of being detected through its effect on an EDP, than through its effect on the disease phenotype. The presence of the disease in the proband thus selects for genes for the EDP in relatives, and allele sharing for a closely linked marker would increase the rate of the EDP in first-degree relatives.

Given the joint distribution of the EDP and the number of alleles IBDP, one can consider other related statistical tests. We also developed numerical expressions for the approximate power of the two degrees of freedom Pearson chi-square test of association, comparing the distribution of the number of alleles IBDP in $E+$ siblings of $D+$ probands (I_{E+}) to the distribution of the alleles IBDP in $E-$ siblings of the $D+$ probands (I_{E-}). We present here, our findings on the power of the AASP-Z test, because the power results of the Pearson chi-square test indicated slightly lower power for the range of parameter values we considered.

For example, in a sample size of $n = 200$, $\theta = 0.0$ and assuming $\lambda_s^{E+|D+} = \lambda_o^{E+|D+}$, the power

of the AASP-Z test for $K_s^{E+|D+} - K^{E+} = 0.25$ is 87%, whereas the power of the Pearson chi-square test is 80%. The AASP-Z test, as we describe it, is equivalent to the Cochran Armitage test and the Haseman–Elston test. The latter has been shown by Blackwelder and Elston (1985) to result in greater power than the Pearson chi-square test in general, consistent with our results.

Another related statistic is the one sample mean test. This statistic simply compares the observed mean number of alleles IBDP in $E+$ siblings to the expected value of 1.0. One can see from Eqs. (9) and (10) that this statistic, like the one sample statistic considered by Risch (1990), would be a direct function of the relative risk of the abnormal EDP in the siblings of $D+$ individuals, in those cases where either $D+$ or $E+$ penetrance is additive.

The power calculations we obtain are large sample approximations. Our numerical calculations for the distribution of the number of alleles IBDP in a sibling, conditional on his EDP, are exact. The accuracy of these power approximations depends on whether or not a sample of asymmetrically ascertained sib pairs is large enough for the Central Limit Theorem to apply. We evaluated this in a modest way, through $N = 500$ simulations of samples of $n = 400$ sib pairs using the generating model parameter values $p = 0.009$, $f_{2D+} = 0.8337$, $f_{1D+} = 0.3068$, $f_{0D+} = 0.0012$, $f_{2E+} = 0.8665$, $f_{1E+} = 0.8159$ and $f_{0E+} = 0.1973$. For this sample size, the average number of $E+$ siblings per sample equals 186 ($SE = 10.11$), which is well within the expected value based on our numerical calculation of sibling recurrence risk of 0.4626 (95% prediction interval = [0.4137, 0.5115]). The observed power was 100% for a 0.0001 level test. The predicted asymptotic power equals 99.97%, with a prediction interval of approximately 99.80% to 100%. With $N = 1000$ samples of $n = 200$ the observed power was 93.1%, well within the 95% prediction interval for this value ([84.9%, 93.5%]). Thus, we might conclude that the large sample approximation to the power appears to work well for $n = 200$ sib pairs.

As expected, the conditional IBDP distribution in the $n = 200,000$ simulated sib pairs was equal to that predicted by our Eq. (1) to within 2 decimal points. For example, the predicted average values of the number of alleles IBDP given EDP status are $E(I|E+) = 1.2629$, $\text{Var}(I|E+) = 0.4305$, $E(I|E-) = 0.7737$, $\text{Var}(I|E-) = 0.4491$, and observed average values of the number of alleles IBDP given EDP status are $E(I|E+) = 1.2639$, $\text{Var}(I|E+) = 0.4313$, $E(I|E-) = 0.7717$, $\text{Var}(I|E-) = 0.4486$.

The study design includes a sample of n sib pairs. In our calculation, the n proband-sibling pairs are from different families, i.e. n unrelated proband-sibling pairs. However, in reality, investigators would like to use all of the siblings in a family for whom they have data. Since the degree of allele sharing with a proband is independent between siblings, the findings for a study using more than one proband-sibling pair per family should be equivalent, as long as all proband-sibling pairs in a given family contain the same proband (Suarez and Hodge, 1979). Thus a proband with s siblings in a given family would generate $s - 1$ sib pairs. The problem of dependence of results obtained using more than one sib pair per family arises, if all possible combinations of sib pairs are used; i.e. $s(s - 1)/2$ sib pairs per family.

The study design we consider, includes n sibling pairs, all with completely informative marker information on the parents. In reality, even if marker information is available on all parents, there will be situations (a homozygous parent or two parents of identical genotype) in which all we can be sure of is that the sib pair shares 1 or 2 alleles, identical by state. In these situations, the genetic software usually estimates the mean number of alleles identical by descent in the pair based on the number of alleles identical by state, and this value in turn is used in the analysis. In these cases, probably more reliable estimates of power can be obtained through simulating the dataset along with the distribution of marker information. However, we conjecture that a minimal power value might be obtained using our methods, by resetting the sample size to the number of sib pairs with completely informative marker information available.

The power calculations we derived, assume an ascertainment design in which the proband is ascertained for being $D+$, and siblings are ascertained at random. In fact, so long as the siblings are ascertained at random with respect to disease EDP , these calculations hold. Furthermore, we conjecture that departures from randomness are likely to (1) increase the number of observed siblings with the abnormal EDP which would (2) decrease the standard error of our estimate of the mean number of alleles IBDP in abnormal siblings, and thus (3) increase the power. This is, in fact, the typical design of genetic studies of EDPs for psychiatric disorders. In showing that even relatively modest sample sizes have sufficient power to detect linkage when either a disease or an EDP is additive, our results present a much more optimistic case for incorporating EDPs into linkage analyses than has been assumed (Faraone et al., 1995). These estimates of pow^* pertain to a linkage analysis, based simply on marker allele sharing given EDP status of the sibling and disease status of the proband. When data on the disease status of the sibling and the EDP status of the proband are also available, power may be even greater if one were to include this information in the analysis using conditional logistic regression.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Funding for this project was supported in part by the Division of Intramural Research, National Human Genome Research Institute, NIH (Z01-HG000200), by NIMH R01 44292, NIMH R01 071523, the National Alliance for Research on Schizophrenia and Depression (NARSAD), and by the Sidney R. Baer, Jr. Foundation.

Appendix A. Expressions for (1) IBD distribution at the marker locus given IBD status at the disease locus (Haseman and Elston (1972)) and (2) genotype distribution in offspring given genotypes of parents

(1) The conditional probabilities for the total IBD values at marker locus $M(I_M)$, given the total IBD values at disease locus $L(I_L)$ can be obtained as follows:

$$P(I_M | I_L) = \begin{matrix} I_M=0 & I_M=1 & I_M=2 \\ \begin{matrix} I_L=0 \\ I_L=1 \\ I_L=2 \end{matrix} & \begin{pmatrix} \psi^2 & 2\psi(1-\psi) & (1-\psi)^2 \\ \psi(1-\psi) & 1-2\psi(1-\psi) & \psi(1-\psi) \\ (1-\psi)^2 & 2\psi(1-\psi) & \psi^2 \end{pmatrix} \end{matrix} \quad (\text{A.1})$$

where $\psi = \theta^2 + (1 - \theta)^2$ for $0 \leq \theta \leq 0.5$ and θ denotes the recombination fraction between marker locus and disease locus.

(2) The values of $P(G_h^{offspring} | M_{d \otimes m})$, the conditional probabilities an offspring genotype is G_h given his mother's genotype, G_m and father's genotype, G_d (h, d and $m = 0, 1, 2$), are calculated by applying Mendel's laws. Here, G_k denotes the genotype with k copies of the Allele for $k = 0, 1, 2$. Thus

$$\begin{aligned} P(G_h^{offspring} | M_{d \otimes m}) &= 1.0 && \text{for } h=d=m=0 \text{ or } 2 \text{ for } h=1 \text{ if } d=0 \text{ and } m=2 \text{ or } d=2 \text{ and } m=0; \\ &= 0.5 && \text{for } h=0, 1 \text{ if } d=1 \text{ and } m=0 \text{ or } d=0 \text{ and } m=1 \text{ and} \\ &&& \text{for } h=1, 2 \text{ if } d=1 \text{ and } m=2 \text{ or } d=2 \text{ and } m=1 \text{ and} \\ &&& \text{for } h=1 \text{ if } d=m=1; \\ &= 0.25 && \text{for } h=0, 2 \text{ and } d=m=1; \\ &= 0 && \text{otherwise.} \end{aligned}$$

(A.2)

Appendix B. Equations for calculating the endophenotype recurrence risk in a sibling of the disease affected proband as a function of penetrance and allele frequencies

The abnormal EDP recurrence risk in the sibling of the $D+$ proband, denoted by $K_s^{E+|D+}$ is obtained as

$$\begin{aligned}
K_S^{E+|D+} &= P(X_{E+}^{sib}=1|X_{D+}^{pro}=1) \\
&= \sum_{k=0}^2 P(X_{E+}^{sib}=1|G_k^{sib}) \cdot P(G_k^{sib}|X_{D+}^{pro}=1) = \sum_{k=0}^2 f_{kE} \cdot P(G_k^{sib}|X_{D+}^{pro}=1) \\
&= \sum_{k=0}^2 f_{kE+} \cdot \left[\sum_{d,m=0}^2 P(G_k^{sib}|M_{d\otimes m}) \cdot P(M_{d\otimes m}|X_{D+}^{pro}=1) \right] \\
&= \sum_{k=0}^2 f_{kE+} \cdot \left[\sum_{d,m=0}^2 P(G_k^{sib}|M_{d\otimes m}) \cdot \frac{\sum_{h=0}^2 (P(X_{D+}^{pro}=1|G_h^{pro}) \cdot P(G_h^{pro}|M_{d\otimes m}) \cdot P(M_{d\otimes m}))}{P(X_{D+}^{pro}=1)} \right] \\
&= \frac{1}{K^{D+}} \sum_{k=0}^2 f_{kE+} \cdot \left[\sum_{d,m=0}^2 P(G_k^{sib}|M_{d\otimes m}) \cdot \sum_{h=0}^2 (f_{hD+} \cdot P(G_h^{pro}|M_{d\otimes m}) \cdot P(M_{d\otimes m})) \right] \\
&= \frac{1}{K^{D+}} \sum_{d,m=0}^2 \sum_{k=0}^2 \left[f_{kE+} \cdot f_{hD+} \cdot P(G_k^{sib}|M_{d\otimes m}) \cdot P(G_h^{pro}|M_{d\otimes m}) \cdot P(M_{d\otimes m}) \right]
\end{aligned}
\tag{B.1}$$

where $P(X_{D+}^{pro}=1) = K^{D+}$, disease population risk, f_{kt} (f_{ht}) denotes the penetrance of t ($t = D+, E+$) given genotype G_k (G_h) ($h, k = 0, 1, 2$). The expressions for $P(G_k^{sib}|M_{d\otimes m})$ and $(G_h^{pro}|M_{d\otimes m})$ are given in Appendix A (A.2), assuming independent mating, where $P(M_{d\otimes m}) = P(G_d) \cdot P(G_m)$ where $d, m = 0, 1, 2$.

Appendix C. Derivation of the IBDP distribution of the disease/EDP allele as a function of recurrence risks

The IBDP distribution of the disease proband/EDP sibling can be re-expressed in terms of the recurrence risks of the EDP in various classes of relatives. For an asymmetric ascertainment study design, the IBDP distribution in EDP+ siblings of the $D+$ proband can be obtained as follows.

$$\begin{aligned}
P(I_M=0|X_{D+}^{pro}=1, X_{E+}^{sib}=1) &= \sum_{I_L=0}^2 P(I_M=0|I_L) \cdot P(I_L|X_{D+}^{pro}=1, X_{E+}^{sib}=1) \\
&= \sum_{I_L=0}^2 P(I_M=0|I_L) \cdot \frac{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L) \cdot P(X_{D+}^{pro}=1|I_L) P(I_L)}{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1) \cdot P(X_{D+}^{pro}=1)} \\
&= P(I_M=0|I_L=0) \cdot \frac{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L=0) \cdot P(I_L=0)}{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1)} \\
&\quad + P(I_M=0|I_L=1) \cdot \frac{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L=1) \cdot P(I_L=1)}{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1)} \\
&\quad + P(I_M=0|I_L=2) \cdot \frac{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L=2) \cdot P(I_L=2)}{P(X_{E+}^{sib}=1|X_{D+}^{pro}=1)} \\
&= \psi^2 \frac{K_S^{E+}}{4K_S^{E+|D+}} + \psi(1-\psi) \frac{K_O^{E+|D+}}{2K_S^{E+|D+}} + (1-\psi)^2 \frac{K_M^{E+|D+}}{4K_S^{E+|D+}}.
\end{aligned}$$

(C.1)

The values of $P(I_M=0|I_L=i)$, are obtained in (A.1) in Appendix A. The remaining values are obtained by setting $P(I_L=0) = P(I_L=2) = 0.25$ and $P(I_L=1) = 0.5$, then noting that

$$P(X_{E+}^{sib}=1|X_{D+}^{pro}=1) = K_S^{E+|D+}: E+\text{risk to sibling of } D+\text{sibling.} \quad (\text{C.2})$$

$$P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L=0) = P(X_{E+}^{sib}=1) = K^{E+} \quad (\text{C.3})$$

$$P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L=1) = K_O^{E+|D+}: E+\text{risk to offspring of a } D+\text{parent.} \quad (\text{C.4})$$

$$P(X_{E+}^{sib}=1|X_{D+}^{pro}=1, I_L=2) = K_M^{E+|D+}: E+\text{risk to monozygotic twin of } D+. \quad (\text{C.5})$$

Similarly,

$$\begin{aligned}
P(I_M=1|X_{D+}^{pro}=1, X_{E+}^{sib}=1) &= 2\psi(1-\psi) \frac{K_S^{E+}}{4K_S^{E+|D+}} + (1-2\psi(1-\psi)) \frac{K_O^{E+|D+}}{2K_S^{E+|D+}} + 2\psi(1-\psi) \frac{K_M^{E+|D+}}{4K_S^{E+|D+}} \\
&\quad + \psi^2 \frac{K_S^{E+}}{4K_S^{E+|D+}} + \psi(1-\psi) \frac{K_O^{E+|D+}}{2K_S^{E+|D+}} + (1-\psi)^2 \frac{K_M^{E+|D+}}{4K_S^{E+|D+}}.
\end{aligned}$$

(C.6)

$$P(I_M=2|X_{D+}^{pro}=1, X_{E+}^{sib}=1) = (1-\psi)^2 \frac{K^{E+}}{4K_S^{E+|D+}} + \psi(1-\psi) \frac{K_O^{E+|D+}}{2K_S^{E+|D+}} + \psi^2 \frac{K_M^{E+|D+}}{4K_S^{E+|D+}}.$$

(C.7)

The IBDP distribution in EDP- sibling ($X_{E+}^{sib}=0$) can be obtained, by replacing $E+$ by $E-$.

Here, the summation of the conditional IBDP distribution at the disease/EDP locus given E +sibling of $D+$ proband should be 1.

$$\begin{aligned} 1 &= \sum_{I_L=0}^2 P(I_L|X_D^{pro}=1, X_E^{sib}=1) \\ &= \sum_{I_L=0}^2 \frac{P(X_E^{sib}=1|X_D^{pro}=1, I_L) \cdot P(I_L)}{P(X_E^{sib}=1|X_D^{pro}=1)} = \frac{K^{E+} \cdot 0.25}{K_S^{E+|D+}} + \frac{K_O^{E+|D+} \cdot 0.5}{K_S^{E+|D+}} + \frac{K_M^{E+|D+} \cdot 0.25}{K_S^{E+|D+}} \end{aligned} \quad (C.8)$$

Therefore, we have $K_M^{E+|D+} = 4K_S^{E+|D+} - K^{E+} - 2K_O^{E+|D+}$ i.e.

$$\begin{aligned} \text{Therefore, we have } K_M^{E+|D+} &= 4K_S^{E+|D+} - K^{E+} - 2K_O^{E+|D+} \quad \text{i. e.} \\ \lambda_M^{E+|D+} &= 4\lambda_S^{E+|D+} - 2\lambda_0^{E+|D+} - 1. \end{aligned} \quad (C.9)$$

Appendix D. Table of genetic parameter values for Fig. 2

See Table D.1.

References

- Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* 1994; 54(3):535–543. [PubMed: 8116623]
- Andreasen N, Grove W. Thought, language and communication in schizophrenia: Diagnosis and prognosis. *Schizophr. Bull.* 1986; 12:348–359. [PubMed: 3764356]
- Arboleda C, Holzman PS. Thought disorder in children at risk for psychosis. *Arch. Gen. Psychiatry.* 1985; 42:1004–1013. [PubMed: 4037984]
- Arolt V, Lencer R, Purmann S, Schurmann M, Muller-Myhsok B, Kreckler K, Schwinger E. Testing for linkage of eye tracking dysfunction and schizophrenia to markers on chromosomes 6, 8, 9, 20, and 22 in families multiply affected with schizophrenia. *Am. J. Med. Genet. B.* 1999; 88(6):603–606.
- Blackwelder WC, Elston RC. A comparison of sib pair linkage tests for disease susceptibility loci. *Genetic Epi.* 1985; 2:85–97.
- Casella, G., Berger, R. *Statistical Inference*. Second edition. Duxbury, California: 2002. p. 243–245.
- Choi H, Winawer MR, Kalachikov S, Pedley TA, Hauser WA, Ottman R. Classification of partial seizure symptoms in genetic studies of the epilepsies. *Neurology.* 2006; 66(11):1648–1653. [PubMed: 16769935]
- Docherty NM, Gottesman I. A twin study of communication disturbances in schizophrenia. *J. Nerv. Ment. Dis.* 2000; 188(7):395–401. [PubMed: 10919696]

- Faraone SV, Kremen WS, Lyons MJ, Pepple JR, Seidman LJ, Tsuang MT. Diagnostic accuracy and linkage analysis: How useful are schizophrenia spectrum phenotypes. *Am. J. Psychiatry.* 1995; 152:1286–1290. [PubMed: 7653682]
- Fiengold E. Methods for linkage analysis of quantitative trait loci in humans. *Theor. Popul. Biol.* 2001; 60(3):167–180. [PubMed: 11855951]
- Freedman R, et al. Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proc. Natl. Acad. Sci.* 1997; 94:587–592. [PubMed: 9012828]
- Gasparoni TL, Ekelund J, Huttunen M, Palmer GCS, Tuulio-Henriksson A, Lonnqvist J, Kaprio J, Koskenvuo M, Peltonen L, Cannon TD. Genetic linkage and association between chromosome 1q and working memory function in schizophrenia. *Am. J. Med. Genet.* 2003; 116B:8–16. [PubMed: 12497606]
- Ghosh S, Begleiter H, Porjesz B, Choelien DB, Edenberg HJ, Foroud T, Goate A, Relich T. Linkage mapping of beta 2 EEF waves via non-parametric regression. *Am. J. Med. Genet. B. Neuropsychiatry Genet.* 2003; 118 (B):66–71.
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: Etymology and strategic intentions. *Am. J. Psychiatry.* 2003; 160(4):636–645. [PubMed: 12668349]
- Greenberg DA, Durner M, Keddache M, Shinnar S, Resor SR, Moshe SL, Rosenbaum D, Cohen J, et al. Reproducibility and complications in gene searches: Linkage on chromosome 6, heterogeneity, association, and maternal inheritance in juvenile myoclonic epilepsy. *Am. J. Hum. Genet.* 2000; 66:508–516. [PubMed: 10677311]
- Hallmayer JF, Kalaydjieva L, Badcock J, Dragovic M, Howell S, Michie PT, Rock D, Vile D, Williams R, Corder EH, Hollingsworth K, Jablensky A. Genetic evidence for a distinct subtype of schizophrenia characterized by pervasive cognitive deficit. *Am. J. Hum. Genet.* 2005; 77(3):468–476. [PubMed: 16080121]
- Haseman JK, Elston RC. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* 1972; 2(1):3–19. [PubMed: 4157472]
- Holzman PS, Shenton M, Solovay M. Quality of thought disorder in differential diagnosis. *Schizophr. Bull.* 1986; 12:360–372. [PubMed: 3764357]
- Holzman, PS., Levy, DL., Johnston, MH. The use of the Rorschach technique for assessing formal thought disorder. In: Bornstein, RF., Masling, JM., editors. *Scoring the Rorschach: Seven Validated Systems.* Lawrence Erlbaum Associates; Mahwah, New Jersey: 2005. p. 55-95.
- Houy E, Raux G, Thibaut F, Belmont A, Demily C, Allio G, Haouzir S, Fouldrin G, Petit M, Frebourg T, Campion D. The promoter — 194 C polymorphism of the nicotinic alpha 7 receptor gene has a protective effect against the P50 sensory gating deficit. *Mol. Psychiatry.* 2004; 9:320–322. [PubMed: 14569275]
- Ji, F. Doctoral Dissertation. Dept. of Applied Mathematics and Statistics, Stony Brook University; Stony Brook, NY: 2004. Linkage analysis of a disease related trait using a pleiotropic model.
- Johnston, MH., Holzman, PS. *Assessing Schizophrenic Thinking.* Jossey-Bass, Inc; San Francisco: 1979.
- Jones KA, Projesz B, Almasy L, Bierut L, Goate A, Wang JC, Dick DM, Hinrichs AL, Kwon J, Rice J, Rohrbach J, Stock H, Wu W, Bauer LO, Chorlian DB, Crowe RR, Edenberg HJ, Foroud T, Hesselbrock V, Kuperman S, Nuenberger JI Jr, O'Connor SJ, Schuckit M, Stimus A, Tischfield JA, Reich T, Begleiter H. Linkage and linkage disequilibrium of evoked EEG oscillations with CHRM2 receptor gene polymorphisms: Implications for human brain dynamics and cognition. *Int. J. Psychophysiol.* 2004; 53:75–90. [PubMed: 15210286]
- Keating M, Atkinson D, Dunn C, Timothy K, Vincent M, Leppert M. Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey ras-1 gene. *Science.* 1991; 252:704–706. [PubMed: 1673802]
- Keating MT, Sanguinetti MC. Molecular and cellular mechanisms of cardiac arrhythmias. *Cell.* 2001; 104:569–580. [PubMed: 11239413]
- Kendler KS, McGuire M, Gruenberg A, O'Hare A, Spellman M, Walsh D. The Roscommon family study. I. Methods, diagnosis of probands, and risk of schizophrenia in relatives. *Arch. Gen. Psychiatry.* 1993; 50:527–540. [PubMed: 8317947]

- Kinney DK, Holzman PS, Jacobsen B, Jansson L, Faber B, Hildebrand W, Kasell E, Zimbalist ME. Thought disorder in schizophrenic and control adoptees and their relatives. *Arch. Gen. Psychiatry.* 1997; 54:475–479. [PubMed: 9152101]
- Levy DL, Coleman M, Ji F, Ulgen A, Matthyse S, Holzman PS, Mendell NR. Qualitative and quantitative phenotypes for thought disorder in linkage studies of schizophrenia 2008
- Levy DL, Ji F, Harte C.B. Krasoshevsky, Holzman PS, Mendell NR, Matthyse S. The power of asymmetric ascertainment in linkage analysis. *Schizophr. Bull.* 2005; 31:271.
- Levy DL, Ji F, Ulgen A, Krastoshevsky O, Coleman MA, Holzman PS, Mendell NR, Matthyse S. The power of asymmetric ascertainment in linkage analysis. 2008
- Makowski DG, Waternaux C, Lajonchere CM, Dicker R, Smoke N, Koplewicz H, Min D, Mendell NR, et al. Thought disorder in adolescent-onset schizophrenia. *Schizophr. Res.* 1997; 23:147–165. [PubMed: 9061811]
- Marlow AJ, Fisher SE, Franks C, MacPhie IL, Cherny SS, Richardson AJ, Talcott JB, Stein JF, Monaco AP, Cardon LR. Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am. J. Hum. Genet.* 2003; 72:561–570. [PubMed: 12587094]
- Matthyse S, Parnas J. Extending the phenotype of schizophrenia: Implications for linkage analysis. *J. Psychiatric. Res.* 1992; 26:329–344.
- Matthyse S, Holzman P, Gussella JF, Levy DL, Harte CB, Jorgensen A, Moller L, Parnas J. Linkage of eye movement dysfunction to chromosome 6p in schizophrenia: Additional Evidence. *Am. J. Med. Genet. part B. Neuro. Genet.* 2004; 128b:30–36.
- Matthyse S, Holzman P, Lange K. The genetic transmission of schizophrenia: Application of Mendelian latent structure analysis to eye tracking dysfunctions in schizophrenia and affective disorder. *J. Psychiatr. Res.* 1986; 20(1):57–67. [PubMed: 3712291]
- Risch N. Linkage strategies for genetically complex traits. I. Multilocus Models. *Am. J. Hum. Genet.* 1990; 46:222–228. [PubMed: 2301392]
- Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* 1990; 46:229–241. [PubMed: 2301393]
- Romney D. Formal thought disorder among the first-degree relatives of schizophrenics: A new look at some old data. *J. Clin. Psychol.* 1984; 40:51–52. [PubMed: 6746959]
- Shenton ME, Solovay MR, Holzman PS. Thought disorder in the relatives of psychotic patients. *Arch. Gen. Psychiatry.* 1989; 46:897–901. [PubMed: 2489936]
- Slater, E. A Review on Earlier Evidence of Genetic Factors in Schizophrenia. In: Rosenthal, D., Kety, SS., editors. *The Transmission of Schizophrenia*. Pergamon, Oxford: 1968. p. 15–26.
- Solovay M, Shenton M, Gasperetti C, Coleman M, Kestenbaum E, Carpenter J, Holzman P. Scoring manual for the thought disorder index. *Schizophr. Bull.* 1986; 12:483–496. [PubMed: 3764364]
- Spence SJ, Cantor RM, Chung L, Kim S, Geschwind DH, Alarcon M. Stratification based on language-related endophenotypes in autism: Attempt replicate reported linkage. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* 2006; 141(6):591–598.
- Spohn HE, Coyne L, Larson J, Mittleman F, Spray J, Hayes K. Episodic and residual thought pathology in chronic schizophrenia. *Schizophr. Bull.* 1986; 12:394–407. [PubMed: 2876514]
- Suarez BK, Hodge SE. A simple method to detect linkage for rare recessive diseases: An application to juvenile diabetes. *Clin. Genetics.* 1979; 15(2):126–136.
- Suarez BK, Rice J, Reich T. The generalized sib pair IBD distribution: Its use in the detection of linkage. *Ann. Hum. Genet.* 1978; 42(1):87–94. [PubMed: 686687]
- Sung H, Finch SJ, Ye KQ, Mendell NR. A gene-model free method for linkage analysis of a disease-related trait based on analysis of proband-sib pairs. *BMC Genetics.* 2005; 6(Suppl 1):S47. [PubMed: 16451658]
- Sung, H. Doctoral Dissertation. Dept. of Applied Mathematics and Statistics, Stony Brook University; Stony Brook, NY: 2006. Power calculation for linkage analysis of a disease related trait.
- Torgersen S, Edvardsen J, Oien P, Onstad S, Skre I, Lygren S, Kringlen E. Schizotypal personality disorder inside and outside the schizophrenia spectrum. *Schizophr. Res.* 2002; 54:33–38. [PubMed: 11853976]

- Vaever MS, Licht DM, Moller L, Perl D, Jorgensen A, Handest P, Parnas J. Thinking within the spectrum: Schizophrenic thought disorder in six Danish pedigrees. *Schizophr. Res.* 2005; 72(2-3): 137–149. [PubMed: 15560959]
- Wahlberg K-E, Wynne LC, Oja HP, Keskitalo H, Anais-Tanner P, Koistinen T, Tarvainen T, Hakko H, et al. Thought disorder index of Finnish adoptees and communication deviance of their adoptive parents. *Psychol. Med.* 2000; 30:127–136. [PubMed: 10722183]
- Wang W, Chang I, Chang C, Curb D, Ho L, Hsiung CA. Incorporating endophenotypes into allele-sharing based linkage tests. *Genet. Epidemiol.* 2006; 30:133–142.
- Weeks DE, Lange K. The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.* 1988; 42(2):315–326. [PubMed: 3422543]
- Wilcox MA, Faraone SV, Su J, Van Eerdewegh P, Tsuang MT. Genome scan of three quantitative traits in schizophrenia pedigrees. *Biol. Psychiatry.* 2002; 52:847–854. [PubMed: 12399137]
- Winawer MR, Marini C, Grinton BE, Rabinowitz D, Berkovic SF, Scheffer IE, Ottman R. Familial clustering of seizure types within the idiopathic generalized epilepsies. *Neurology.* 2005; 65(4): 523–528. [PubMed: 16116110]

+Note: $K^{E+} = 0.02, 0.05, 0.1, 0.15$ and 0.2 ; $K_S^{E+|D+} = 0.31, 0.37, 0.42, 0.45$ and 0.48 ;

EDP: endophenotype; n : number of asymmetrically ascertained sib pairs;

AASP: asymmetrically ascertained sib pairs; θ : recombination fraction;

$K_S^{E+|D+}$: Abnormal endophenotype recurrence risk in siblings of disease affected proband.

$K_O^{E+|D+}$: Abnormal endophenotype recurrence risk in offspring of disease affected parents.

K^{E+} : Abnormal endophenotype population risk.

Fig. 1.

Power of EDP linkage analysis of $n = 200, 400$ and 500 AASP as a function of

$$K_S^{E+|D+} - K^{E+}; K_S^{E+|D+} = K_O^{E+|D+}, \theta = 0.0.$$

+: Settings listed in Appendix D;
 SZ+: schizophrenia or schizo-affective disorder; T+: thought disorder present;
 n : Number of asymmetrically ascertained sib pairs; θ : recombination fraction;
 K_S^{T+SZ+} : Recurrence risk of T+ in siblings of SZ+ probands;
 K^{T+} : Thought disorder population risk;
 Model-based analysis: Bivariate Lod score analysis of SZ+ and TD+ using generating model parameter values as the analysis values;
 Model-free analysis: Analysis based on the AASP-Z statistic described in this paper.

Fig. 2.
 Power comparison of model-based methods and model-free methods of $SZ+ / T+$ analysis for the 40 genetic parameter settings⁺: $n = 174$ and $\theta = 0.01$.

Table 1Joint distribution of alleles IBD with $D+$ proband and endophenotype in AASP

Number of alleles IBD with $D+$ proband I	EDP status of sibling of $D+$ proband		Total
	$E+$	$E-$	
0	n_{0E+}	n_{0E-}	$n_{0\cdot}$
1	n_{1E+}	n_{1E-}	$n_{1\cdot}$
2	n_{2E+}	n_{2E-}	$n_{2\cdot}$
Total	n_{E+}	n_{E-}	n

n_{ij} : the number of proband-sibling pairs with $I = i(0, 1, 2)$ and $E = j(E+, E-)$.

Genetic parameter values that generate the schizophrenia ($SZ+$) and thought disorder ($T+$) rates observed in schizophrenics, siblings of schizophrenics, and the general population

Table D.1

p	f_{SZ+}	f_{1SZ+}	f_{sSZ+}	f_{2T+}	f_{1T+}	f_{0T+}	$K_S^{T+} + SZ+ - K^{T+}$	Power
0.0203	0.8656	0.1451	0.0002	0.7925	0.4202	0.0974	0.1610	30.8932
0.0069	0.2854	0.0933	0.0009	0.7891	0.7639	0.1500	0.1794	36.5568
0.0177	0.2907	0.1583	0.0013	0.9063	0.7618	0.1945	0.2243	63.6722
0.0161	0.7409	0.1335	0.0018	0.9020	0.8300	0.1822	0.2246	62.3399
0.0169	0.4875	0.1464	0.0018	0.8307	0.8173	0.1766	0.2278	65.8086
0.0107	0.6120	0.2268	0.0018	0.9241	0.8190	0.1822	0.2282	66.7958
0.0202	0.3063	0.1347	0.0009	0.8430	0.7578	0.1879	0.2365	71.5168
0.0084	0.3825	0.2888	0.0012	0.8530	0.7791	0.1775	0.2376	73.5198
0.0391	0.4680	0.0665	0.0007	0.7596	0.7079	0.1597	0.2384	67.0357
0.0185	0.5294	0.1515	0.0009	0.8407	0.7476	0.1764	0.2388	72.7582
0.0164	0.7947	0.1607	0.0012	0.8498	0.7762	0.1671	0.2460	76.6874
0.0682	0.7034	0.0227	0.0007	0.9322	0.4460	0.0837	0.2460	87.2274
0.0094	0.8132	0.2513	0.0020	0.8839	0.8446	0.1349	0.2462	80.0157
0.0732	0.5503	0.0234	0.0010	0.9742	0.5360	0.1148	0.2463	79.5659
0.0393	0.8292	0.0486	0.0016	0.9749	0.7398	0.1342	0.2485	70.4845
0.0095	0.8337	0.3068	0.0012	0.8665	0.8159	0.1973	0.2536	81.2211
0.0148	0.7041	0.1619	0.0017	0.9124	0.8796	0.1796	0.2557	81.5925
0.0251	0.5132	0.0880	0.0022	0.9700	0.9064	0.1321	0.2564	81.2580
0.0182	0.2822	0.1305	0.0015	0.9227	0.8479	0.1514	0.2570	83.9205
0.0132	0.6891	0.2011	0.0013	0.8162	0.8006	0.1348	0.2622	86.6648
0.0086	0.4826	0.3001	0.0014	0.9563	0.8374	0.1515	0.2635	87.2958
0.0245	0.8162	0.1124	0.0002	0.9453	0.7144	0.1724	0.2652	84.9703
0.0127	0.8720	0.2504	0.0005	0.9152	0.7211	0.1353	0.2673	88.8040
0.0337	0.9380	0.0577	0.0023	0.9422	0.8849	0.1192	0.2682	77.9553
0.0108	0.3976	0.2582	0.0012	0.8726	0.7889	0.1126	0.2714	91.4739
0.0147	0.7137	0.2175	0.0003	0.7814	0.6817	0.1042	0.2721	91.5193
0.0328	0.3922	0.0824	0.0013	0.8345	0.8280	0.1246	0.2723	87.9814

d	f_{SZ+}	f_{1SZ+}	f_{oSZ+}	f_{2T+}	f_{1T+}	f_{oT+}	$K_S^{T+ SZ+} - K^{T+}$	Power
0.0198	0.7715	0.1534	0.0006	0.8151	0.7245	0.1140	0.2744	90.7294
0.0214	0.9546	0.1326	0.0005	0.8830	0.7363	0.1387	0.2759	89.6381
0.0108	0.5986	0.2665	0.0006	0.8225	0.7449	0.1221	0.2769	92.5128
0.0146	0.9786	0.1634	0.0020	0.9684	0.9284	0.0946	0.2926	95.5249
0.0163	0.5172	0.1644	0.0008	0.9524	0.7743	0.0752	0.2956	97.0077
0.0119	0.8480	0.2581	0.0006	0.8715	0.8146	0.1519	0.2977	96.0985
0.0458	0.6268	0.0527	0.0011	0.8768	0.8150	0.1128	0.2990	90.8492
0.0208	0.7128	0.1616	0.0001	0.8887	0.7285	0.1037	0.3010	96.8291
0.0269	0.1358	0.1143	0.0008	0.8956	0.8050	0.0608	0.3107	98.6437
0.0140	0.6754	0.1870	0.0011	0.9032	0.8514	0.0751	0.3158	98.7827
0.0788	0.5456	0.0124	0.0017	0.8236	0.8000	0.0507	0.3172	74.4710
0.0982	0.3211	0.0148	0.0014	0.8122	0.7716	0.0067	0.3195	83.9093
0.0127	0.2518	0.2075	0.0010	0.9660	0.9141	0.1327	0.3196	98.8037

+ From Levy et al. (in preparation-b), recombination fraction $\theta = 0.01$;

p : pleiotropic (schizophrenia/thought disorder) allele frequency; $f_{SZ+}(T+)$: schizophrenia (thought disorder) penetrance, where i is the number of the pleiotropic (schizophrenia/thought disorder) alleles, $i = 0, 1, 2$; Power: power of AASP-Z test of $n = 174$ sib pairs; Genetic model parameter values satisfy the following 5 constraints:

$$0.09 \leq K^{T+} \leq 0.21, 0.31 \leq K_S^{T+|SZ+} \leq 0.47, 0.64 \leq K_I^{T+|SZ+} \leq 0.80, 0.03 \leq K_S^{SZ+|SZ+} \leq 0.13 \text{ and } 0.0055 \leq K^{SZ+}.$$