Selection of Eating-Disorder Phenotypes for Linkage Analysis

Cynthia M. Bulik,[1] Silviu-Alin Bacanu,[2] Kelly L. Klump,[3] Manfred M. Fichter,[4] Katherine A. Halmi,[5] Pamela Keel,[6] Allan S. Kaplan,[7,8] James E. Mitchell,[9] Alessandro Rotondo,[10] Michael Strober,[11] Janet Treasure,[12] D. Blake Woodside,[8] Vibhor A. Sonpar,[2] Weiting Xie,[2] Andrew W. Bergen,[13] Wade H. Berrettini,[14] Walter H. Kaye,[2]* and Bernie Devlin[2]*

[1]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina
[2]Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania
[3]Department of Psychology, Michigan State University, East Lansing, Michigan
[4]Klinik Roseneck, Hospital for Behavioral Medicine, affiliated with the University of Munich, Prien, Germany
[5]New York Presbyterian Hospital, Weill Medical College of Cornell University, White Plains, New York
[6]Department of Psychology, University of Iowa, Iowa City, Iowa
[7]Program for Eating Disorders, Toronto General Hospital, Toronto, Ontario, Canada
[8]Department of Psychiatry, Toronto General Hospital, Toronto, Ontario, Canada
[9]Neuropsychiatric Research Institute, Fargo, North Dakota
[10]Department of Psychiatry, Neurobiology, Pharmacology and Biotechnologies, University of Pisa, Pisa, Italy
[11]Department of Psychiatry and Behavioral Science, University of California at Los Angeles, Los Angeles, California
[12]Eating Disorders Unit, Institute of Psychiatry and South London and Maudsley National Health Service Trust, United Kingdom
[13]National Cancer Institute, Division of Cancer Epidemiology & Genetics, Bethesda, Maryland
[14]Center of Neurobiology and Behavior, University of Pennsylvania, Philadelphia, Pennsylvania

Vulnerability to anorexia nervosa (AN) and bulimia nervosa (BN) arise from the interplay of genetic and environmental factors. To explore the genetic contribution, we measured over 100 psychiatric, personality, and temperament phenotypes of individuals with eating disorders from 154 multiplex families accessed through an AN proband (AN cohort) and 244 multiplex families accessed through a BN proband (BN cohort). To select a parsimonious subset of these attributes for linkage analysis, we subjected the variables to a multilayer decision process based on expert evaluation and statistical analysis. Criteria for trait choice included relevance to eating disorders pathology, published evidence for heritability, and results from our data. Based on these criteria, we chose six traits to analyze for linkage. Obsessionality, Age-at-Menarche, and a composite Anxiety measure displayed features of heritable quantitative traits, such as normal distribution and familial correlation, and thus appeared ideal for quantitative trait locus (QTL) linkage analysis. By contrast, some families showed highly concordant and extreme values for three variables—lifetime minimum Body Mass Index (lowest BMI attained during the course of illness), concern over mistakes, and food-related obsessions—whereas others did not. These distributions are consistent with a mixture of populations, and thus the variables were matched with covariate linkage analysis. Linkage results appear in a subsequent report. Our report lays out a systematic roadmap for utilizing a rich set of phenotypes for genetic analyses, including the selection of linkage methods paired to those phenotypes. © 2005 Wiley-Liss, Inc.

KEY WORDS: complex disease; endophenotype; liability; clinical judgment; covariate selection; mixture model; regression

INTRODUCTION

Most common human diseases have a complex etiology involving both genetic and environmental factors. This complexity makes the task of finding relationships between phenotypes and genotypes challenging. To achieve a substantial probability of success requires highly efficient experimental designs and analytic methods. This observation is true whether association, linkage/association or pure linkage designs are employed. Here we focus on the latter, with special reference to affected sibling/relative pair linkage designs (ASP/ARP).

For reasonable models of complex disease and reasonable effects of disease loci, the power of ASP/ARP designs is low [Risch and Merikangas, 1996]. Still, these designs are commonly used for studies to find linkage. To enhance their power, the use of covariates has been proposed, and it is common for investigators to measure numerous traits on affected individuals with this goal in mind. In response, covariate methods have been developed [review in Devlin et al., 2002a; Hauser et al., 2004] and employed [Olson et al., 2001; Devlin et al., 2002b; Godart et al., 2002; Scott et al., 2003].

Because many of these covariates are quantitative, an alternative approach would be to use the covariates or "traits"

in quantitative trait locus (QTL) linkage analysis [Almasy and Blangero, 1998; Etzel et al., 2003]. QTL methodologies for ASP/ARP designs continue to be developed [TCuenco et al., 2003; Szatkiewicz and Feingold, 2004] and are being used to hunt for genetic variation affecting liability to disease [Evans et al., 2004; Loo et al., 2004].

Insofar as we are aware, a substantial void exists between the available QTL and covariate linkage methodologies and their application to ASP/ARP designs, namely there is no roadmap for selecting amongst the potentially numerous variables or for selecting the optimal methodology to analyze the resulting data [see also Rampersaud et al., 2003]. We present a novel, structured approach to variable selection and to the teaming of variables with linkage methods, which we apply to data from two eating disorder cohorts.

For quantitative traits, a natural method to test for linkage is QTL linkage analysis. Such analyses typically assume trait values follow a normal distribution, approximately, and are substantially heritable, which implies that trait values in families are correlated (Fig. 1a); we refer to these features as "classic features of quantitative traits." Whenever a trait is associated with liability to disease, however, trait values for affected individuals can be quite different than that expected from a random sample of the population, and features of the population distribution (e.g., mean, standard deviation) are critical for interpreting the data. It is possible that trait values convey no information not already embodied in diagnosis, in which case QTL linkage analysis is not useful.

QTL linkage analysis is not ideal in related settings as well. Imagine that the disease of interest is etiologically heterogeneous, that a quantitative trait is genetically-correlated with disease via a subset of the loci generating vulnerability, and trait values convey information only on the etiology of disease. This model for the data was formalized by using a mixture model in which a certain fraction π of families traces a portion of their liability to variation at disease locus l whereas $1-\pi$ do not [Devlin et al., 2002a]. Trait value X is informative about membership in these groups. Notice that only π families would show linkage for markers proximate to l. Empirical examples of such traits, or covariates, are age-of-onset for breast cancer and Alzheimer disease, and their distribution differ from those amenable to QTL linkage analysis (Fig. 1b).

While some subtypes of eating disorders (i.e., restricting subtype of anorexia nervosa) have homogenous presentation compared to most psychiatric disease, they still have complex etiology. Many of the traits thought to underlie vulnerability to eating disorders exhibit substantial variation in the population (e.g., perfectionism), reminiscent of classic quantitative traits. Some traits, however, show quite different distributions and instead seem to reveal a mixture of populations within the eating disorders sample [Devlin et al., 2002b]. For these traits, a covariate linkage analysis based on mixture models seems more appropriate. Our structured algorithm will, therefore, focus on two methods of analysis, QTL and mixture-model-based covariate linkage analysis.

MATERIALS AND METHODS

Samples, Diagnoses, and Assessments

Supported through funding provided by the Price Foundation, three cohorts of subjects relevant to this study have been recruited. Approximately 200 people with AN and their affected relatives with an eating disorder were recruited for the AN cohort [Kaye et al., 2000]. This sample includes psychological assessments and blood samples from 196 probands and 229 affected relatives. For this analysis we focused on 154 families with affected siblings (140 with 2 affected siblings, 9 with 3, and 5 with 4; ~6% male). Approximately 316 people
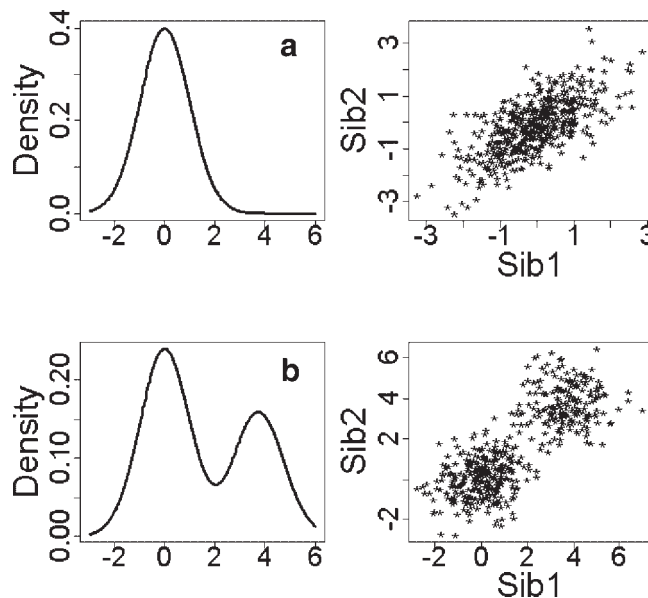

Fig. 1. Idealized distributions for a quantitative trait amenable to quantitative trait locus (QTL) linkage analysis (**a**) and a trait that identifies populations in the sample (**b**). We refer to the distribution in 1a as "classical" because the quantitative trait values are normally distributed (mean $\mu = 0$, standard deviation $\mu = 1$) and show correlation $\rho = 0.7$ between family members, in this case siblings whose trait values are randomly sampled under the specified model (a, **right**). Correlation typically would be measured as the intraclass correlation, which formally is the ratio of the between-family component of variance divided by the total variance of the trait. We describe the trait distribution in 1b as a trait that clusters families, which can be seen using simple diagnostics plotting covariate values for one sibling against the other (b, **right**). Formally the model generating these data is a mixture model, in which fraction $1-\pi = 0.6$ families are randomly sampled with trait values $\mu = 0$, $\sigma = 1$, and $\rho = 0.2$, and fraction $\pi = 0.4$ families are randomly sampled with $\mu = 3.75$, $\sigma = 1$, and $\rho = 0.2$. Diagnostics combined with formal tests are often used to determine the presence of a mixture.

with BN and their affected relatives with an eating disorder were recruited during 1996–99 for the BN cohort [Kaye et al., 2004a]. For these analyses, we focused on 244 families with affected siblings (228 with two affected siblings, 15 with 3 and 1 with 4; ~2% male). Control women (N = 697), who were screened to be free from Axis I and II pathology, were recruited from the same sites as the BN cohort. Data from measured traits of control women approximate a random sample from the population, and were used to evaluate trait distributions from the AN and BN cohorts.

The Structured Interview on Anorexia Nervosa and Bulimic Syndromes (SIAB) was used to assess lifetime history of eating disorder diagnoses. Additional diagnostic information (i.e., course, severity) was obtained from the Structured Clinical Interview for DSM IV Axis I Disorders (SCID I) [First et al., 1997]. A battery of standardized instruments (Table I) was chosen to assess potential traits related to core eating disorder symptoms, mood, temperament, and personality.

Structured Approach to Variable Selection

Variable reduction by clinical criteria. Over 100 variables were available from the AN and BN cohorts, creating a large, multidimensional space for analysis, on the order of the number of families participating. For this reason, we used our knowledge of eating disorders and related phenotypes to reduce this pool. Our choices were guided by several criteria: must be consistently related to eating pathology; must be

TABLE I. Results From Cluster Diagnostics/Testing and Estimates of Intraclass Correlation for
All Variables Chosen at Stage 1

Variable	Correlation		Cluster	
	AN	BN	AN	BN
BDI[a] score	—	0.00	—	C
Maximum BMI	22.82	34.59	UN	C
Minimum BMI (during illness)	0.00	20.87	UN[k]	C
Cooperativeness[b]	25.32	28.02	NC	C
Concern over mistakes[c]	16.94	12.45	C	C
Doubts about actions[c]	12.27	4.24	NC	UN
EDI Bulimia[d]	16.21	—	UN	—
Body dissatisfaction[d]	15.89	—	C	—
Drive for thinness[d]	14.86	—	C[l]	—
SIAB[e] depression	26.45	—	NC	—
SIAB[e] compulsion/anxiety	9.80	—	NC	—
SIAB[e] bulimia	13.39	—	NC	—
Harm avoidance[b]	23.98	19.99	C	UN
Harm avoidance 1[f]	3.76	13.92	UN	UN
Harm avoidance 2[f]	14.11	14.56	C	C
Agreeableness domain[g]	—	18.90	—	NC
Conscientiousness domain[g]	—	1.93	—	NC
Neuroticism domain[g]	—	7.55	—	NC
Novelty seeking[b]	20.34	0.00	NC	NC
Organization[c]	20.92	0.00	C	C
Persistence[b]	6.85	11.49	UN	UN
Age at menarche	23.16	22.56	NC	NC
Personal standards[c]	19.20	8.96	NC	C
Reward dependence[b]	9.38	15.56	NC	C
Self directedness[b]	13.58	7.32	NC	NC
Self transcendence[b]	24.21	33.77	NC	NC
Anxiety[h]	12.30	11.81	NC	NC
Food obsession[i]	7.31	0.00	C	C
YBOC compulsion[j]	0.00	21.33	UN	UN
YBOC obsession[j]	21.22	31.39	UN	NC

Under clustering, C indicates variable clusters in families, NC indicates no clustering, UN indicates unknown, '—'
indicates unmeasured.
[a]Beck Depression Inventory (BDI) [see Kaye et al., 2000, 2004a for reference for this and other scales].
[b]Temperament and Character Inventory (TCI).
[c]Multidimensional Perfectionism Scale (MPS).
[d]Eating Disorder Inventory 2 (EDI-2), bulimia scale.
[e]Structured Interview of Anorexia Nervosa and Bulimic Syndromes.
[f]TCI, HA1 Anticipatory worry and pessimism versus uninhibited optimism; HA2, fear of uncertainty.
[g]Revised NEO Personality Inventory.
[h]State-Trait Anxiety Inventory (Form Y-1), Trait Anxiety.
[i]Yale-Brown-Cornell Eating Disorder Scale, worst lifetime total.
[j]Yale-Brown Obsessive Compulsive Scale, worst lifetime total.
[k]By definition, individuals diagnosed with AN show low BMI. Therefore clustering is difficult to judge, but
reasonable to expect.
[l]Because Drive-for-Thinness was used in a previous linkage analysis, and because it was not measured in BN
families, we will not consider it for further analysis.

heritable (or at least show correlation or clustering in our
families); and must be indicators of severity of illness, or
enduring traits rather than states resulting from the illness.

Dimension reduction by multivariate analysis. Us-
ing the pool of selected variables, we studied their multivariate
distribution and determined which ones can be aggregated to
produce a reduced set of variables. We used clustering methods
for these multivariate analyses [Kaufman and Rousseeuw,
1990], specifically the "hclust" statistical procedure in the
statistical package R [Dalgaard, 2002].

Familial analysis of selected variables. By these
analyses, we attempted to distinguish whether variables
cluster families (matching covariate linkage analysis) or
whether they show attributes more typical of quantitative
traits (matching QTL linkage analysis). Because age is known
to impact many of these variables, we regressed out the effect
of age and performed analyses on the residuals. For brevity, we
refer to these transformed variables as variables or traits.

The approach we took for clustering was similar to that
described previously [Devlin et al., 2002a], namely to use
simple graphical diagnostics to determine whether a selected
variable clustered families. (If the clustering is readily visible,
then we assume the variable could be biologically meaningful).
Clustering was performed using a per-family summary
statistic, which was judged to be the most informative
summary for that variable. For example, if young age-of-onset
were thought to be the most informative for genetic analysis
(e.g., breast cancer), the maximum value of age-of-onset for
each family would be used in the analysis, because the
maximum extracts more information about the within-family
values of the variable than does the mean or sum.

To evaluate clustering formally, we used the "mclust"
procedure [Banfield and Raftery, 1993] in the statistical
package R [Dalgaard, 2002]. Mclust assumes the observed
data come from one or more populations, and the data from each
population follows a normal distribution. Mclust estimates the

number of populations K from the data, as well as the mean and variance associated with each population, by means of hierarchical, mixture model-based clustering. For a given K, the estimated mean and variance of each population maximizes the likelihood of the data (approximately). This is a non-standard statistical problem for which it is well known that increasing K always increases the likelihood. Therefore K is chosen on the basis of the Bayesian Information Criterion, which is a penalized likelihood procedure that favors parsimony (smaller K).

To evaluate the "within-family" correlation of variables, the intraclass correlation was estimated by using the NESTED procedure of SAS (version 8.1). Specifying family as a class produces an estimate of the variance attributable to family, which can then be used with the estimate of the total variance to obtain the interclass correlation. To contrast the distribution of variables in the AN/BN cohorts to that for control women, we use regression models and Generalized Estimating Equations methods [Liang and Zeger, 1986], which account for relatedness by adjusting the variance of parameter estimates.

RESULTS

To select a parsimonious subset of variables for linkage analysis, as well as choose the type of linkage analysis to be applied, we employed a structured approach to variable selection (Fig. 2). In Stage 1, clinical criteria were used to winnow the number of variables. To be chosen, the variable must be related to eating pathology, be heritable based on published data or at least familial in our data, and be relatively insensitive to state of illness. By these criteria, 26 variables were selected for the AN cohort and 24 for the BN cohort (Table I). There was substantial overlap between the variables selected, which was purposeful because we hoped the overlapping variables would prove useful for linkage analysis of the combined AN/BN cohorts. All of these selected variables (Table I) had some degree of support from the literature in terms of heritability and insensitivity to state of illness (i.e., trait-like qualities), as well as support from our clinical experience and from the contrast of women with eating disorders and control women (data not shown). Examples of variables excluded from consideration include Maturity Fears (associated with eating disorders, but not known to be heritable) and Agreeableness (no specific relationship to eating disorders).

In Stage 2, we sought to determine the degree of independence of the variables selected in Stage 1. Most were largely independent for both the AN and BN cohorts (Figs. 3 and 4). Several variables displayed moderate correlation (≥ 0.63). For the AN cohort, Obsessions were substantially correlated with Compulsions, Self-Directedness with Anxiety, Drive-For-Thinness with Body Dissatisfaction, and Concern Over Mistakes with Doubts About Actions; finally SIAB Compulsions (Table I) showed moderate correlation with Obsessions and Compulsions (Fig. 3). Similar results accrued for the BN cohort, with the exception that Neuroticism was substantially correlated with both Anxiety and Self-Directedness. (Neuroticism as not measured in the AN cohort).

Results from Stage 2 showed that certain variables contribute redundant information for individuals with eating disorders. Deliberations in Stage 3 focused on whether to combine these variables into composite variables by multivariate analysis, or select one of them for further analysis. Without missing data, composite variables would be preferred because they extract information for two or more variables. Missing data for either variable on a particular individual generates missing data for the composite variable for that individual (without imputation). Thus missing data were typically greater for the composite variable than for any of
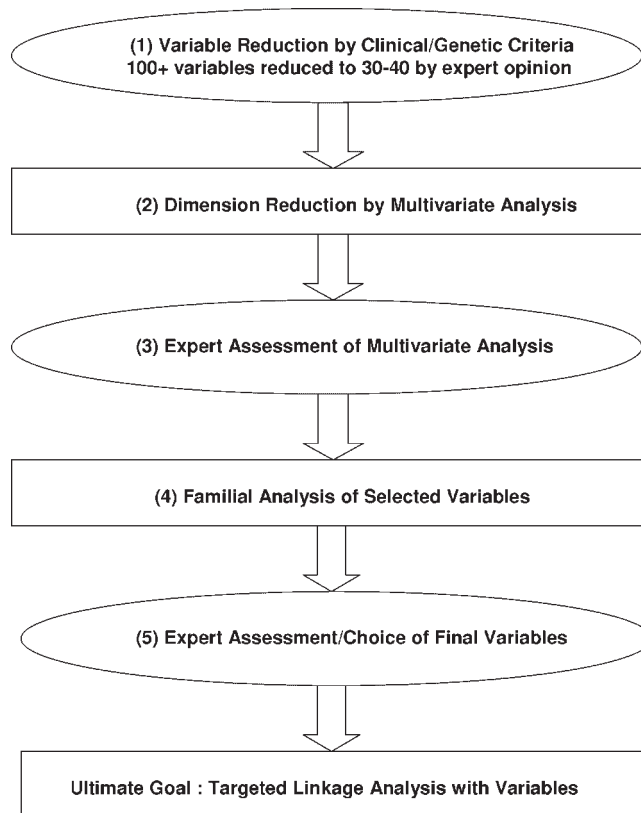

Fig. 2. Flow chart for structured approach to variable selection.

the variables to be combined. This problem was worsened for families, the unit of interest. Therefore, in most cases, we chose to target one of the variables; the exception is described below.

The group of eating-disorder experts believed that a fundamental underlying feature of the pathology of eating disorders is anxiety and that eating disorder pathology can serve an anxiolytic function. They doubted Anxiety, as measured (Table I), would capture that feature. The first subscale of Harm Avoidance, anticipatory worry, captures a key feature of anxiety seen in individuals with eating disorders [Fassino et al., 2004; Klump et al., 2004]. Therefore a composite variable was derived, consisting of the first principal component of Anxiety and the first subscale of Harm Avoidance (PC-Anxiety).

Stage 4 consisted of analysis of familiality, measured by either the magnitude of the correlation of trait values within families or whether the traits clustered families into distinct and meaningful groups (Fig. 1). Variables showing strong intraclass correlation (≥ 0.20) in both AN and BN cohorts include maximum BMI, Cooperativeness, Age at Menarche, Self Transcendence, and Obsessionality (Table I). Harm Avoidance also shows substantial intraclass correlations, missing the cutoff by 0.01 for the BN cohort only. Other variables show strong intraclass correlations in one sample only (Table I).

Concern over Mistakes, Harm Avoidance 2, Organization, and Obsessions over Food appear to cluster families into meaningful groups. Formal analysis for a mixture supports the visual diagnostics. Food Obsessions shows distinctive features of clustering and extreme values in individuals with eating disorders (Fig. 5). Most individuals with eating disorders are 4–6 standard deviations from the mean value for control women (Fig. 5). Moreover, when one sibling is extreme for this

Dissimilarity

| 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |

reward.dependance
cooperativeness
harm.avoidance.1
self.directedness
anxiety
factor .5
edi.bulimic
age.at.menarche
maximum.bmi
minimum.bmi
organization
personal.standards
persistance
self.transcendence
novelty.seeking
harm.avoidance.2
factor .4
yboc.obsession
yboc.compulsion
drive.for.thinness
body.dissatisfaction
concern.over.mistakes
doubts.about.actions
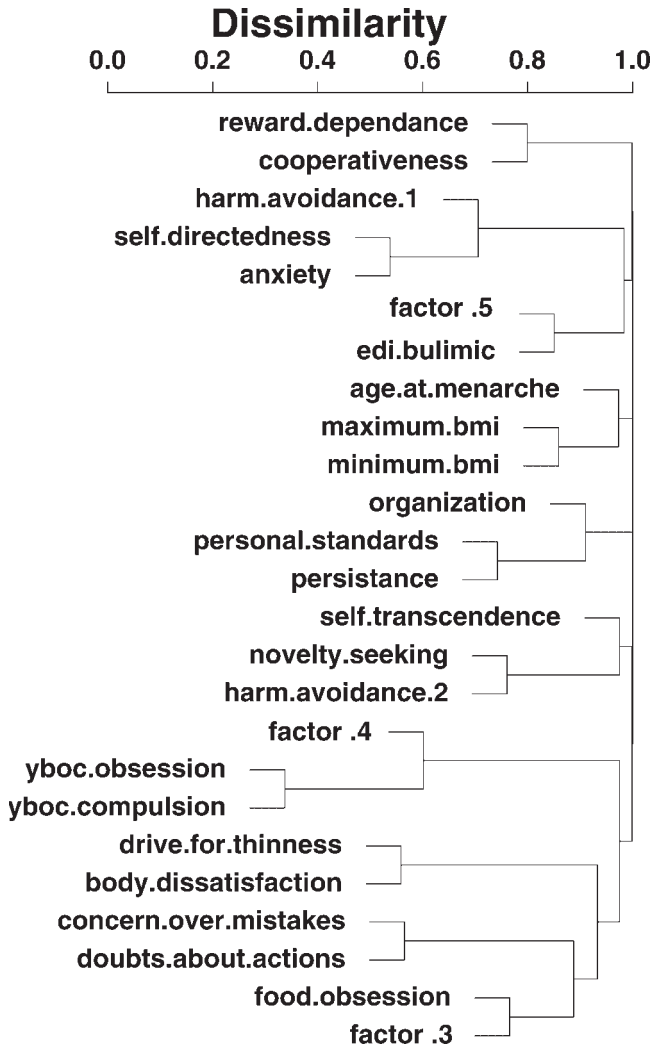food.obsession
factor .3

Fig. 3. Clustered variables for the anorexia nervosa (AN) cohort. Dissimilarities are calculated as one minus the squared correlation each pair of variables.

Dissimilarity

| 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 |

self.transcendence
novelty.seeking
harm.avoidance.2
age.at.menarche
maximum.bmi
minimum.bmi
food.obsession
yboc.obsession
yboc.compulsion
concern.over.mistakes
doubts.about.actions
depression.score
harm.avoidance.1
self.directedness
neuroticism.domain
anxiety
personal.standards
persistance
organization
conscientiousness.domain
reward.dependance
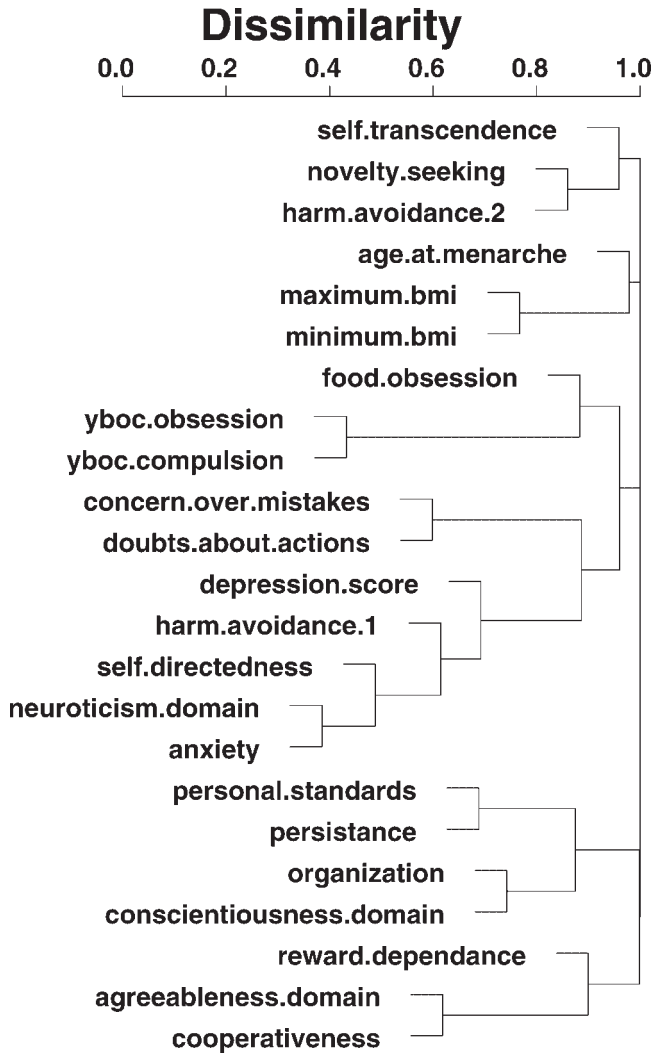agreeableness.domain
cooperativeness

Fig. 4. Clustered variables for the bulimia nervosa (BN) cohort. Dissimilarities are calculated as one minus the squared correlation each pair of variables.

trait, the other sibling tends to be extreme as well. There are exceptions, however, creating a strong cluster of ASP who are extreme and concordant for Food Obsessions, and other ASP who are dispersed in other regions of bivariate space (Fig. 5). By contrast Age at Menarche shows substantial intraclass correlation (Table I), but no evidence for clustering (Fig. 5). Minimum BMI shows clustering of families in the BN cohort, but not for families in the AN cohort. The lack of clustering in the AN cohort is largely structural, however, because by clinical definition individuals affected with AN must achieve and maintain remarkably low BMI. PC-Anxiety shows no clustering, and relatively low intraclass correlation ($\sim$0.1).

Stage 5 required the final selection of variables based on the analyses of Stage 4, relevance to eating disorder pathology, and clinical experience and insight. It is possible that a trait could be familial—in that it either clusters families or shows high intraclass correlation—yet be unrelated to liability to eating disorders. If a trait were related to liability, we expect its distribution in people diagnosed with eating disorders to be displaced (e.g., trait has a different mean) relative to a sample from the population. State of illness can impact values of the traits, so displacement must be evaluated critically.

We selected three traits that seemed most appropriate for QTL linkage analysis, namely Obsessionality, Age at Menarche, and PC-Anx. The first two show substantial intraclass correlations in both the AN and BN cohorts and no evidence for a mixture of populations of multiplex families (clustering). PC-Anx also showed no clustering, but it shows fairly small intraclass correlations for both the AN and BN cohorts; nonetheless, based on the literature [Godart et al., 2002; Strober, 2004] and expert opinion, we opted to include it in the final set of variables. Total Harm Avoidance could be another candidate, but it was ruled out because of its correlation with a component of PC-Anx. Self Transcendence and Maximum BMI were ruled out because their connection to eating disorders was tenuous; there was little or no difference between the control and eating disorder samples for Self Transcendence; and the Maximum BMI, while distinct between controls and eating disorder groups, tended to be rather low in the eating disorder sample.

We also selected three traits for covariate linkage analysis, Minimum BMI, Concern over Mistakes, and Food Obsessions. All three clustered families (e.g., Fig. 5). Concern over Mistakes and Food Obsessions showed similar features in
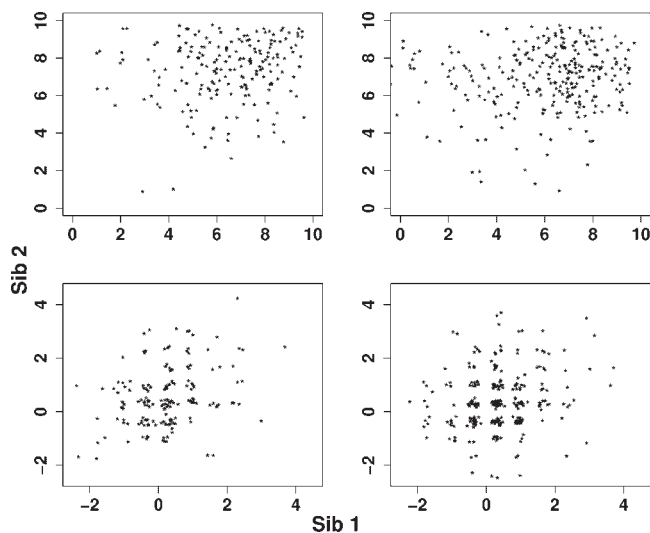
Fig. 5.   ASP values for food obsessions (**top**) and age at menarche for the AN (**left**) and BN cohorts.

both cohorts. While Minimum BMI showed no evidence of clustering in the AN cohort, that was judged unimportant because low BMI is an essential component for the diagnosis of AN. None of these variables showed substantial intraclass correlations for either data set (Table I). Organization was ruled out because its values in the eating disorder samples were only weakly differentiated from those of the control sample.

DISCUSSION

From a field of over 100 phenotypes collected from two multiplex samples of eating disorder families, as well as a sample of control women, we sought to select a parsimonious set of variables that would prove useful to linkage analysis and to match these variables to the kind of analytic method for linkage. To achieve this goal, we performed a structured analysis of the phenotypes (Fig. 2), selecting three variables for QTL linkage analysis: Obsessionality, Age at Menarche, and PC-Anx. Obsessionality is a heritable trait [Jonnal et al., 2000], is a salient personality feature of individuals with AN and BN [Halmi et al., 2000], and family studies report increased prevalence of OCD and OCPD in relatives of individuals with EDs [Lilenfeld et al., 1998]. The relation between age at menarche and eating disorders is salient yet incompletely understood. Bulimia nervosa is often associated with oligomennorhea despite the persistence of normal weight [Bulik et al., 2000], early age at menarche has been associated with the development of binge-eating in the absence of compensatory behaviors [Reichborn-Kjennerud et al., 2004], and it is heritable [Kirk et al., 2001]. Anxiety disorders are common among people with eating disorders [Kendler et al., 1995; Walters and Kendler, 1995; Godart et al., 2002], usually precede onset of AN or BN [Deep et al., 1995; Bulik et al., 1997; Kaye et al., 2004b], and are heritable, as are related traits [Hettema et al., 2001].

Some variables, such as Food Obsessions, barely show any familial correlation (Table I), yet they demonstrate strong familiality when viewed on a different scale—the ability to cluster families (Fig. 1). We use this feature in our structured analysis (Fig. 2) to select three other variables for covariate linkage analysis: Minimum BMI, Concern over Mistakes, and Food Obsessions. Covariate linkage analysis is not as straight-forward as QTL linkage analysis. We assume the covariate probabilistically identifies a cluster of families that are "linked" at a liability locus while other families are not linked [Devlin et al., 2002a,b]. Therefore, our ultimate goal for this analysis is to assign probabilities or weights of membership into the linked and unlinked groups, and biological insights must determine which group is targeted for linkage analysis.

Selection of these three variables is supported by the fact that lifetime minimum BMI is a marker for severe anorexia nervosa and is associated with poor outcome [Lowe et al., 2001]. Concern over Mistakes is a heritable component of perfectionism [Tozzi et al., 2004] and a personality feature that appears to be somewhat uniquely associated with the presence of eating disorders [Bulik et al., 2003b]. Finally, food obsessions combine the highly obsessional nature of individuals with eating disorders with a focus on food and related behaviors [Mazure et al., 1994; Halmi et al., 2000].

In the absence of clearly defined and biologically relevant endophenotypes, the selection of optimal traits for and approaches to linkage poses substantial methodological challenges in psychiatric genetics. On one level, the clinical phenotypes of anorexia nervosa and bulimia nervosa are clear. Viewed more critically, however, no measure exists that captures the essential essence of "eating disorderedness." In such instances, novel, systematic approaches to selecting and evaluating the appropriateness of traits from among a pool of many are essential. We provide a roadmap for trait selection that can be applied to genetic research on other disorders for which comprehensive phenotyping has occurred. Our algorithm offers a rational, systematic blueprint for data modeling by parsimoniously selecting traits and then matching them with linkage approach. By its nature, data modeling involves choices that depend on the properties of the data and available analytic methods, as well as the goals of the analyst. Our goal is to be as parsimonious as possible in terms of the number of tests of linkage performed. We therefore winnow a long list of possible phenotypes to a short list of traits we expect to harbor key information about liability to eating disorders. Moreover, we carry forward our parsimony principle by selecting a priori the kind of linkage method to be used with each trait. In Bacanu et al. [2005], we show that, by using our data modeling, significant and suggestive linkages arise more often than expected by chance.

ACKNOWLEDGMENTS

REFERENCES

Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet 62:1198–1211.

Bacanu S-B, Bulik CM, Klump KL, Fichter MM, Halmi KA, Kaplan AS, Strober M, Treasure J, Woodside DB, Sonpar VA, Xie W, Bergen A, Berrettini AH, Kaye WH, Devlin B. 2005. Linkage analysis of anorexia and bulimia nervosa cohorts using selected behavioral phenotypes as quantitative traits or covariates. Am J Med Genet B 139:61–68.

Banfield J, Raftery A. 1993. Model-based Gaussian and non-Gaussian clustering. Biometrics 49:803–822.

Bulik C, Sullivan P, Fear J, Joyce P. 1997. Eating disorders and antecedent anxiety disorders: A controlled study. Acta Psychiatrica Scandinavica 96:101–107.

Bulik CM, Sullivan PF, Kendler KS. 2000. An empirical study of the classification of eating disorders. Am J Psychiatry 157:886–895.

Bulik CM, Devlin B, Bacanu S-A, Thornton L, Klump K, Fichter MM, Halmi KA, Kaplan AS, Strober M, Woodside DB, Bergen AW, Ganjei JK, Crow S, Mitchell J, Rotondo A, Mauri M, Cassano G, Keel P, Berrettini WH, Kaye WH. 2003a. Significant linkage on chromosome 10p in families with bulimia nervosa. Am J Hum Genet 72:200–207.

Bulik CM, Tozzi F, Anderson C, Mazzeo SE, Aggen S, Sullivan PF. 2003b. The relation between eating disorders and components of perfectionism. Am J Psychiatry 160:366–368.

Dalgaard P. 2002. Introductory Statistics with R. New York: Springer.

Deep A, Nag L, Weltzin T, Rao R, Kaye W. 1995. Premorbid onset of psychopathology in long-term recovered anorexia nervosa. Int J Eating Disord 17:291–298.

Devlin B, Jones BL, Bacanu S-A, Roeder K. 2002a. Mixture models for linkage analysis of affected sibling pairs and covariates. Genet Epidemiol 22:52–65.

Devlin B, Bacanu S-A, Klump KL, Bulik CM, Fichter MM, Halmi KA, Kaplan AS, Strober M, Treasure J, Woodside DB, Berrettini WH, Kaye WH. 2002b. Linkage analysis of anorexia nervosa incorporating behavioral covariates. Hum Molec Genet 11:689–696.

Etzel CJ, Shete S, Beasley TM, Fernandez JR, Allison DB, Amos C. 2003. Effect of Box-Cox transformation on power of Haseman-Elston and maximum-likelihood variance components tests to detect quantitative trait loci. Hum Hered 55:108–116.

Evans DM, Zhu G, Duffy DL, Montgomery GW, Frazer IH, Martin NG. 2004. Major quantitative trait locus for eosinophil count is located on chromosome 2q. J Allergy Clin Immunol 114:826–830.

Fassino S, Amianto F, Gramaglia C, Facchini F, Abbate Daga G. 2004. Temperament and character in eating disorders: ten years of studies. Eat Weight Disord 9:81–90.

First MB, Spitzer RL, Gibbon M, Williams JBW. 1997. Structured clinical interviews for DSM-IV Axis I disorders, research version, non-patient edition (SCID-I/NP). Biometrics Research, New York State Psychiatric Institute: New York.

Godart N, Flament M, Perdereau F, Jeammet P. 2002. Comorbidity between eating disorders and anxiety disorders: a review. Int J Eat Disord 32:253–270.

Halmi KA, Sunday SR, Strober M, Kaplan A, Woodside DB, Fichter M, Treasure J, Berrettini WH, Kaye WH. 2000. Perfectionism in anorexia nervosa: variation by clinical subtype, obsessionality, and pathological eating behavior. Am J Psychiatry 157:1799–1805.

Hauser ER, Watanabe RM, Duren WL, Bass MP, Langefeld CD, Boehnke M. 2004. Ordered subset analysis in genetic linkage mapping of complex traits. Genet Epidemiol 27:53–63.

Hettema JM, Neale MC, Kendler KS. 2001. A review and meta-analysis of the genetic epidemiology of anxiety disorders. Am J Psychiatry 158:1568–1578.

Jonnal AH, Gardner CO, Prescott CA, Kendler KS. 2000b. Obsessive and compulsive symptoms in a general population sample of female twins. Am J Med Genet B 96:791–796.

Kaufman L, Rousseeuw J. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. New York: Wiley.

Kaye WH, Lilenfeld LR, Berrettini WH, Strober M, Devlin B, Klump KL, Goldman D, Bulik CM, Halmi KA, Fichter MM, Kaplan A, Woodside DB, Treasure J, Plotnicov KH, Pollice C, Rao R, McConaha CW. 2000. A search for susceptibility loci for anorexia nervosa: methods and sample description. Biol Psychiatry 47:794–803.

Kaye WH, Devlin B, Barbarich N, Bulik CM, Thornton L, Bacanu S-A, Fichter MM, Halmi KA, Kaplan AS, Strober M, Woodside DB, Bergen AW, Crow S, Mitchell J, Rotondo A, Mauri M, Cassano G, Keel P, Plotnicov K, Pollice C, Klump KL, Lilenfeld LR, Ganjei JK, Quadflieg N, Berrettini WH. 2004a. Genetic analysis of bulimia nervosa: methods and sample description. Int J Eat Disord 35:556–570.

Kaye WH, Bulik CM, Thornton L, Barbarich N, Masters K. 2004b. Anxiety disorders comorbid with anorexia and bulimia nervosa. Am J Psychiatry 161:2215–2221.

Kendler KS, Walters EE, Neale MC, Kessler RC, Heath AC, Eaves LJ. 1995. The structure of the genetic and environmental risk factors for six major psychiatric disorders in women: phobia, generalized anxiety disorder, panic disorder, bulimia, major depression and alcoholism. Arch Gen Psychiatry 52:374–383.

Kirk KM, Blomberg SP, Duffy DL, Heath AC, Owens IP, Martin NG. 2001. Natural selection and quantitative genetics of life-history traits in Western women: a twin study. Evolution Int J Org Evolution 55:423–435.

Klump KL, Strober M, Bulik CM, Thornton L, Johnson C, Devlin B, Fichter MM, Halmi KA, Kaplan AS, Woodside DB, Crow S, Mitchell J, Rotondo A, Keel PK, Berrettini WH, Plotnicov K, Pollice C, Lilenfeld LR, Kaye WH. 2004. Personality characteristics of women before and after recovery from an eating disorder. Psychol Med 34:1407–1418.

Liang K, Zeger S. 1986. Longitudinal data analysis using generalized linear models. Biometrika 73:13–22.

Lilenfeld LR, Kaye WH, Greeno CG, Merikangas KR, Plotnicov K, Pollice C, Rao R, Strober M, Bulik CM, Nagy L. 1998. A controlled family study of restricting anorexia and bulimia nervosa: comorbidity in probands and disorders in first-degree relatives. Arch Gen Psychiatry 55:603–610.

Loo SK, Fisher SE, Francks C, Ogdie MN, MacPhie IL, Yang M, McCracken JT, McGough JJ, Nelson SF, Monaco AP, Smalley SL. 2004. Genome-wide scan of reading ability in affected sibling pairs with attention-deficit/hyperactivity disorder: unique and shared genetic effects. Mol Psychiatry 9:485–493.

Lowe B, Zipfel S, Buchholz C, Dupont Y, Reas DL, Herzog W. 2001. Long-term outcome of anorexia nervosa in a prospective 21-year follow-up study. Psychol Med 31:881–890.

Mazure CM, Halmi KA, Sunday SR, Romano SJ, Einhorn AM. 1994. The Yale-Brown-Cornell Eating Disorder Scale: development, use, reliability and validity. J Psychiatr Res 28:425–445.

Olson JM, Goddard KAB, Dudek DM. 2001. The amyloid precursor protein locus and very-late-onset Alzheimer disease. Am J Hum Genet 69:895–899.

Rampersaud E, Allen A, Li YJ, Shao Y, Bass M, Hayne C, Ashley-Koch A, Martin ER, Schmidt S, Hauser ER. 2003. Adjusting for covariates on a slippery slope: linkage analysis of change over time. BMC Genet 4: Suppl 1:S50.

Reichborn-Kjennerud T, Bulik CM, Sullivan PF, Tambs K, Harris JR. 2004. Psychiatric and medical symptoms in binge eating in the absence of compensatory behaviors. Obes Res 12:445–454.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. Science 273:1516–1517.

Scott WK, Hauser ER, Schmechel DE, Welsh-Bohmer KA, Small GW, Roses AD, Saunders AM, Gilbert JR, Vance JM, Haines JL, Pericak-Vance MA. 2003. Ordered-subsets linkage analysis detects novel Alzheimer disease loci on chromosomes 2q34 and 15q22. Am J Hum Genet 73:1041–1051.

Strober M. 2004. Pathologic fear conditioning and anorexia nervosa: On the search for novel paradigms. Int J Eat Disord 35:504–508.

Szatkiewicz JP, Feingold E. 2004. A powerful and robust new linkage statistic for discordant sibling pairs. Am J Hum Genet 75:906–909.

Tcuenco K, Szatkiewicz JP, Feingold E. 2003. Recent advances in human quantitative-trait-locus mapping: comparison of methods for selected sibling pairs. Am J Hum Genet 73:863–873.

Tozzi F, Aggen SH, Neale BM, Anderson CB, Mazzeo SE, Neale MC, Bulik CM. 2004. The structure of perfectionism: a twin study. Behav Genet 34:483–494.

Walters EE, Kendler KS. 1995. Anorexia nervosa and anorexic-like syndromes in a population-based female twin sample. Am J Psychiatry 152:64–71.