0145-6008/95/1904-1018$03.00/0
ALCOHOLISM: CLINICAL AND EXPERIMENTAL RESEARCH

Vol. 19, No. 4
August 1995

# Comparison of Direct Interview and Family History Diagnoses of Alcohol Dependence

John P. Rice, Theodore Reich, Kathleen K. Bucholz, Rosalind J. Neuman, Roberta Fishman, Nanette Rochberg, Victor M. Hesselbrock, John I. Nurnberger, Jr., Marc A. Schuckit, and Henri Begleiter

**Using data from The Collaborative Study on the Genetics of Alcoholism, we compare direct interview diagnoses of alcohol dependence to those obtained by history from family members. Using a requirement of three or more positive implications by history, the specificity, sensitivity, and positive predictive values are 98%, 39%, and 45%, respectively.**

**A logistic analysis found the gender of the relative and alcoholism in the informant to be significant, but not the gender of the informant. The partial odds ratio of a diagnosis at interview associated with a positive family history diagnosis was 13.6. The relationship between the informant and relative was significant, with negative reports from an offspring or mate more influential than a negative report from a parent or second-degree relative.**

**We derived a recursive equation to combine a variable number of family history reports, wherein the probabilities associated with a single report are computed from the logistic analysis. This permits the use of family history information both as a proxy for an uninterviewed relative, as well as a second source of information to be used in the analysis of genetic family data.**

**Key Words: Family History Diagnoses, DSM-III-R Alcohol Dependence, Specificity, Genetic Analysis.**

WHEN CONDUCTING family and genetic studies, family history assessments provide a relatively simple and cost-effective way to select and extend families. Even when the goal of a study is to perform personal interviews, relatives will refuse or be unavailable because of death, unknown address, etc. These individuals may be nonrandom with respect to psychopathological outcomes of interest. Researchers view family study (i.e., personal interview) data as more valid and consider history information as a type of proxy interview. From another perspective, we may view both direct interview and family history diagnoses as dual sources of information to be used at the analysis stage.

*From the Department of Psychiatry (J.P.R., T.R., K.K.B., R.J.N., R.F., N.R.), Washington University School of Medicine, St. Louis, Missouri; Department of Psychiatry (V.M.H.), University of Connecticut Health Science Center, Farmington, Connecticut; Department of Psychiatry (J.I.N.), Indiana University, Indianapolis, Indiana; Department of Psychiatry (M.A.S.), Veterans Affairs Medical Center, University of California, San Diego, California; and Department of Psychiatry (H.B.), State University of New York, Health Sciences Center, Brooklyn, New York.*

*Received for publication August 3, 1994; accepted February 22, 1995*

*This study was supported in part by the U.S. Public Health Services Grants AA08401, AA08403, MH37685, MH49500, and MH31302.*

*Reprint requests: John P. Rice, Ph.D., Department of Psychiatry (Box 8134), Washington University Medical School, 4940 Children's Place, St. Louis, MO 63110-1081.*



This study focuses on the DSM-III-R diagnoses of alcohol dependence and abuse and examines a number of specific questions: (1) How sensitive and specific is the family history diagnosis of alcoholism? (2) Does the choice of informant (e.g., mother versus father) affect the sensitivity and specificity? and (3) How can a variable number of informant reports be combined within a family?

Although there have been numerous studies that consider the validity of self-reported alcohol consumption[1,2] using, for example, collateral informants,[3,4] less work has been done examining diagnostic assessments. Sher and colleagues[5,6] have compared reports of college-aged siblings on their fathers and Mann et al.[7] performed a study of the test-retest reliability of a family history questionnaire of drinking problems. Zimmerman et al.[8] examined the test-retest reliability of the Family History Research Diagnostic Criteria (FH-RDC) in 58 depressed patients. They found 100% agreement for the RDC diagnosis of alcoholism ($n$ = 12). Kendler et al.[9] examined parental diagnoses of alcoholism using 86 twin pair informants who were discordant for alcoholism. They noted a trend for the alcoholic twins to report more alcoholism in their parents.

Andreasen et al.[10] compared direct interview (SADS-L) and history diagnoses (FH-RDC) on 2,216 first-degree relatives of depressives. They found a sensitivity (the probability of a positive history diagnosis in someone alcoholic by interview) of 52% and a specificity (the probability of a negative history in an interviewed nonalcoholic) of 96%. In their study, the history diagnosis represented a consensus of two FH-RDC's: one from the proband and one from the best informant for the family.

The aforementioned studies indicate that family history represents a valid source of information, although there may be reduced sensitivity in a single report. However, when multiple informants are available, it is unclear how to weight the number of positive and negative reports or how to take into account other covariates that would affect the degree of certainty attached to information from a particular informant. In previous work,[11,12] we discussed the use of multiple personal interviews to reduce the impact of diagnostic error in a single, cross-sectional assessment. Combining multiple reports of family history information with interview information offers another approach for phenotype definition in genetic studies.

## SUBJECTS AND METHODS

Data presented are from The Collaborative Study on the Genetics of Alcoholism* (COGA), a multisite study of alcoholic probands and their relatives. The six sites consist of the State University of New York at Brooklyn, University of Connecticut, Indiana University, University of Iowa, University of California at San Diego, and Washington University in St. Louis. Probands are identified in treatment settings and meet both DSM-III-R[13] dependence and Feighner et al.[14] definite lifetime criteria for alcoholism.

### *Diagnostic Instruments*

Family members and probands are administered the Semi-Structured Assessment for the Genetics of Alcoholism[15] (SSAGA) designed to assess the physical, psychological, and social manifestations of alcoholism and related disorders, and the Family History Assessment Module (FHAM), which makes six specific DSM-III-R diagnoses on relatives by history. Each interviewed member (age 18 or older) of the family is administered the FHAM, so that there are a variable number of multiple sources of family history diagnoses. The SSAGA and FHAM were developed as part of the COGA project.

The FHAM is a structured diagnostic instrument to assess major DSM-III-R psychiatric disorders among relatives of the informant. All interviewed adults (aged 18+) receive the FHAM. The following disorders are diagnosed: alcoholism, drug dependence, depression, mania, schizophrenia, and antisocial personality.

The FHAM is composed of a Screener, a set of 11 questions to screen relatives likely to have a specific psychiatric disorder, and Individual Assessment Modules (IAMs), which list specific symptoms of each diagnosis. The screening question for alcoholism is, "have any of your relatives ever had any family, job, school, police, or health problems because of drinking?" Similar questions are asked regarding other psychiatric disorders. Then, all relatives listed as positive in the diagnostic screening questions are assessed by the corresponding IAM(s). The screening question for alcohol problems is asked of each first-degree relative who is not identified by the initial screening question. (Copies of the FHAM are available from Dr. Bucholz.)

### *Subjects*

In this study, we consider 452 proband families and 120 control families with at least one FHAM available for analysis. Probands are ascertained from alcoholism treatment facilities and must be able to speak English, be older than 17 years, give written informed consent, and have at least two first-degree relatives living in one of the catchment areas. Families that are found to include two additional first-degree relatives with a diagnosis of alcoholism are extended through affected individuals or through "leapfrogging" over an unaffected individual if two individuals in that secondary nuclear family are affected by family history. Control probands and their families are ascertained via random consecutive sampling from HMOs, Dental Clinics, or by randomly sampling driver's license records. These families are not eliminated if they contain alcoholic members.

Interview data and diagnoses are available on 2,654 individuals in these 572 families. These individuals have been administered the SSAGA, and their interview data have been checked for consistency by an editor at each site, entered, checked by a computer-cleaning program, and included on the Masterfile (MF19) distributed to each site. The relationship of these individuals to the proband is given in Table 1.

**Table 1.** Relationship of Interviewed Relatives to Proband

| Relationship | No. |
|---|---|
| Alcoholic proband | 380 |
| Control proband | 93 |
| Parent | 502 |
| Sibling | 900 |
| Offspring | 375 |
| Mate | 170 |
| Other | 234 |
| | 2,654 |

### *Logistic Regression*

We utilized logistic regression[16] to predict an interview diagnosis of DSM-III-R alcohol dependence in relatives of the proband. The general principles used in an analysis using logistic regression are similar to those used in the more familiar linear regression. One important difference is that, in logistic regression, the outcome variable, $y$, is dichotomous with $y = 1$ corresponding to "affected" and $y = 0$ corresponding to "unaffected." Let $X = x_1, x_2, \ldots, x_k$ be a sequence of independent variables influencing the probability of being affected, that is, of having an interview diagnosis of alcohol dependence. Let $\Pr(y = 1|X)$ denote the probability of a diagnosis of alcohol dependence as a function of the independent variables (the covariates). The following function, known as the logistic function, is used to model $\Pr(y = 1|X)$:

$$\Pr(y = 1|X) = \frac{e^{\alpha + \beta_1 X_1 + \ldots + \beta_k X_k}}{1 + e^{\alpha + \beta_1 X_1 + \ldots + \beta_k X_k}}. \quad (1)$$

A primary reason why logistic regression is considered such a powerful analytic tool is that the coefficient, $\beta_i$, can be interpreted as the natural logarithm of the odds ratio for its corresponding variable, $x_i$. Therefore, the odds ratio $= e^{\beta i}$. For example, if the covariate $x_i$ represents gender (coded 1 for males or 2 for females), then $e^{\beta i} = 0.5$ indicates that a positive diagnosis ($y = 1$) occurs half as often among females as males when all other covariates have the same values for males and females.

An individual may be classified as alcoholic depending on our choice of the cut-off value for $\Pr(y = 1|X)$; for example, we may choose $\Pr(y = 1|X) = 0.6$ as the point of demarcation, declaring anyone with a predicted probability >0.6 as alcohol-dependent.

We performed logistic regression on all first-degree relatives of the proband who were the subject of at least one FHAM; that is, at least one additional family member was interviewed with the FHAM. Among the independent variables used in the logistic regression are two family history variables: Family History-Certain (FHC) coded as 1 if and only if the FHAM diagnosis is "certain" for alcohol dependence, and Family History-Questionable (FHQ) coded as 1 for a "questionable" diagnosis; if the FHAM diagnosis is "none," then both of these variables are coded as 0. Additional covariates are the sex of the informant, sex of the relative, the alcohol diagnosis of the informant, and a set of four variables indicating the relationship between informant and relative.

### *Statistical Model for Multiple Informants*

An analysis using logistic regression estimates the probability of being affected by interview given information from a single family history report. We also wish to analyze the ability of multiple family histories to predict the interview diagnosis. Let $y$ denote the variable "affected by interview" with values $y = 1$ for affected and $y = 0$ for unaffected. Therefore, we need to compute $\Pr(y = 1|\mathrm{FH}_1, \mathrm{FH}_2, \ldots, \mathrm{FH}_n)$ for $n$ family history reports, $\mathrm{FH}_1, \ldots, \mathrm{FH}_n$, where each event $\mathrm{FH}_k$, $k = 1, \ldots, n$ represents whether the FHAM report is positive or negative, as well as all the other covariates described previously: genders of the informant and relative, relationship

* *The Collaborative Study on the Genetics of Alcoholism (H. Begleiter, State University of New York, Health Sciences Center at Brooklyn, principal investigator; T. Reich, Washington University, co-principal investigator) includes six different centers wherein data collection takes place. The six sites and principal investigator and co-investigators are: Indiana University (J. Nurnberger, Jr., P. M. Conneally); University of Iowa (R. Crowe, S. Kuperman); University of California at San Diego and Scripps Institute (M. Schuckit, F. Bloom); University of Connecticut (V. Hesselbrock); State University of New York, Health Sciences Center at Brooklyn (H. Begleiter, B. Porjesz); and Washington University in St. Louis (T. Reich, C. R. Cloninger).*

**Table 2.** Interview Versus Family History Diagnoses in Relatives with at Least Three Chances to Be Implicated

| No. of positive FHAM implications | Males ($n$ = 716) | | | Females ($n$ = 887) | | |
|---|---|---|---|---|---|---|
| | Dependence | Abuse | None | Dependence | Abuse | None |
| 0 | 28.5* | 61.1 | 85.0 | 46.5 | 63.6 | 90.1 |
| 1 | 13.7 | 5.6 | 7.8 | 16.5 | 27.3 | 5.8 |
| 2 | 12.0 | 11.1 | 4.6 | 9.0 | 9.1 | 3.0 |
| 3 | 9.4 | 5.6 | 1.4 | 11.5 | 0.0 | 0.4 |
| 4 | 12.8 | 5.7 | 0.9 | 6.0 | 0.0 | 0.0 |
| 5 | 10.8 | 11.1 | 0.3 | 4.5 | 0.0 | 0.4 |
| 6 | 6.0 | 0.0 | 0.0 | 3.5 | 0.0 | 0.0 |
| 7 | 1.7 | 0.0 | 0.0 | 1.5 | 0.0 | 0.2 |
| 8 | 1.1 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 |
| 9 | 1.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 |
| ≥10 | 2.5 | 0.0 | 0.0 | 0.5 | 0.0 | 0.0 |
| Total | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $n$ | 351 | 18 | 347 | 200 | 11 | 676 |

* Column percents.

**Table 3.** Sensitivity ($p$), Specificity ($q$), and Positive Predictive Value (PV) of History Diagnosis of Alcohol Dependence

| No. of positive FHAM implications | Males | | | Females | | | Combined | | |
|---|---|---|---|---|---|---|---|---|---|
| | $p$ | $q$ | PV | $p$ | $q$ | PV | $p$ | $q$ | PV |
| ≥1 | 0.72 | 0.84 | 0.81 | 0.54 | 0.90 | 0.60 | 0.65 | 0.88 | 0.73 |
| ≥2 | 0.58 | 0.92 | 0.87 | 0.37 | 0.96 | 0.72 | 0.50 | 0.94 | 0.82 |
| ≥3 | 0.46 | 0.96 | 0.93 | 0.28 | 0.99 | 0.88 | 0.39 | 0.98 | 0.91 |

between informant and relative, and interview alcohol diagnosis of the informant.

Under the assumption that $\Pr(FH_n|y, FH_1, \ldots, FH_{n-1}) = \Pr(FH_n|y)$, that is, the interview diagnosis is the "gold standard," the following recursive formula can be derived:

$$\Pr(y = 1|FH_1, \ldots, FH_n) = \frac{\Pr(y = 1|FH_n)\Pr(y = 1|FH_1, \ldots, FH_{n-1})\Pr(y = 0)}{\Pr(y = 0|FH_n)\Pr(y = 0|FH_1, \ldots, FH_{n-1})\Pr(y = 1) + \Pr(y = 1|FH_n)\Pr(y = 1|FH_1, \ldots, FH_{n-1})\Pr(y = 0)}. \quad (2)$$

Note that this formula depends on the "base rate" $\Pr(y = 1)$ of the interview diagnosis in the population being studied.

## RESULTS

### *Sensitivity, Specificity, and Positive Predictive Value*

We first consider first-degree relatives who had at least three chances to be implicated (i.e., three or more additional family members were interviewed with the FHAM). This data set consisted of 1,603 relatives (433 parents, 826 siblings, and 344 offspring). The cross-classification of these relatives by SSAGA diagnosis (DSM-III-R Alcohol Dependence, Abuse, none) and the number of positive FHAM implications is given in Table 2. This table displays, by sex and interview diagnosis (dependence, abuse, none), the percentage of relatives with 0, 1, ..., 10 or more positive FHAM implications. For example, the table shows that 351 males had a SSAGA diagnosis of alcohol dependence and, of these, 28.5% were not implicated by any of their relatives who were administered the FHAM.

From Table 2, we use the number of positive FHAMs to compute the sensitivity, $p$, the probability of correctly identifying a case (diagnosed by interview) and the specificity, $q$, the probability of correctly identifying a noncase. For example, Table 2 shows that $p = 71.5\%$ of the 351 male cases were identified by at least one positive FHAM. Similarly, $q = 85\%$ of the 347 male noncases had no positive FHAM implication. Table 3 also contains the positive predictive value, PV, the probability that an individual who is alcohol-dependent by history is also diagnosed dependent by personal interview. PV, $p$, and $q$ are displayed in Table 3 for three definitions of dependent by history: 1 or more, 2 or more, and 3 or more positive FHAMs.

We note from Table 3 that the specificity of a history diagnosis based on multiple positive reports can be quite good, although the sensitivity decreases with the number of implications required to make a diagnosis. Because there are variable numbers of possible implications (ranging from 3 to 27), it should be cautioned that these are not the values of sensitivity and specificity resulting from administration of a single FHAM. This will be addressed herein.

From Table 2, we note that the diagnosis of DSM-III-R alcohol abuse is low in this sample (2.5% in males and 1.2% in females), and the sensitivity for this diagnosis is below 40%, even for a cut-off of at least one implication.

### *Logistic Analysis*

We created a data set with all possible pairs of relatives and informants. Relatives consisted of the 1,603 interviewed first-degree relatives whose family had at least one FHAM available for analysis. We excluded probands as relatives because they were identified through treatment settings at the time of interview and might not be comparable with their alcoholic relatives. The informants were all people in that family who filled out a FHAM. The resulting

**Table 4.** Odds Ratios for Predictors of Interview Diagnosis of Alcohol Dependence

| | Odds ratio ($e^{\beta}$) | $\chi^2$ | $p$ value |
|---|---|---|---|
| Variable | | | |
| Intercept ($\alpha$) | 1.86 | 27.4 | <0.0001 |
| Family history diagnosis | | | |
| FHC | 13.6 | 1,157.1 | <0.0001 |
| FHQ | 4.3 | 65.6 | <0.0001 |
| Gender of relative | 0.3 | 606.3 | <0.0001 |
| Gender of informant | 1.1 | 3.3 | 0.07 |
| Alcohol diagnosis in informant | 1.5 | 57.3 | <0.0001 |
| Relationship | | | |
| Sibling | 0.85 | 6.4 | 0.01 |
| Child | 0.33 | 182.3 | <0.0001 |
| Mate | 0.34 | 58.6 | <0.0001 |
| Other | 1.1 | 1.9 | 0.17 |

file consisted of 12,266 observations. The relationship variable was defined as the relationship of the informant to the relative.

Using the LOGISTIC[17] procedure of SAS, we fit a logistic model using the variables listed in Table 4: the two family history diagnosis variables (described herein), the gender of both the informant and relative (males coded 1, females coded 2), interview alcohol diagnosis of informant, and four variables indicating the relationship between informant and relative. The four relationship variables consist of one each for sibling, child, mate, and other (not first degree). If the informant is a parent, all four of these variables are coded 0. Therefore, each of the four classes of relatives will be compared with parent when forming the odds ratio.

We note the highest odds ratio corresponds to a positive history diagnosis at the "certain" level. The next highest is a positive history diagnosis at the "questionable" level. This indicates that, although the precision of this latter diagnosis is lower, there is validity associated with its use. As expected, the gender of the relative is significant (the odds ratio associated with a male is 1/0.3 = 3.3), whereas the sex of the informant (controlling for relationship) is not. Positive reports from a parent, sibling, or "other" relative give comparable increases in the odds to implicate, whereas those associated with a child or mate are reduced by one-third compared with parents. Here, children had to be at least 18 years of age to be included in analysis, so this is not solely explained by youth. Finally, the odds of being implicated by FHAM were increased if the informant had a positive diagnosis of alcoholism by interview.

The odds ratios in Table 4 correspond to a single, multivariate model, and each should be interpreted as the effect of that variable controlling for all other variables. Because alcoholism is familial, an alcoholic informant is more likely to have an alcoholic relative, so univariate significance is to be expected. Herein, the odds associated with a positive report are further augmented when the informant has a positive diagnosis him/herself.

Using Eq. 1, the predicted probability of a positive interview diagnosis may be computed from the model parameters underlying Table 5. A predetermined cut-off may be used to classify individuals as affected or unaffected. The sensitivity and specificity associated with various cut-offs are given in Table 5. Notice that the decision to call affected anyone with a predicted probability of 0.70 or more only identifies 33% of those who are affected by interview; however, the specificity for this cut-off values is quite high at 97%.

**Table 5.** Sensitivity and Specificity from Logistic Analysis

| Cut-off probability | Sensitivity ($p$) | Specificity ($q$) |
|---|---|---|
| 0.1 | 0.97 | 0.19 |
| 0.2 | 0.85 | 0.44 |
| 0.3 | 0.68 | 0.74 |
| 0.4 | 0.61 | 0.82 |
| 0.5 | 0.41 | 0.94 |
| 0.6 | 0.34 | 0.97 |
| 0.7 | 0.33 | 0.97 |

### *Use of Multiple Informants*

Overall, 49% of males and 23% of females in our sample have a diagnosis of alcohol dependence. In Table 6, we display the predicted probability of being affected in relatives with 1, 2, or 3 sources of family history. We use the coefficients from Table 4 for a male informant who himself does not have a diagnosis of alcoholism, and use Eq. 2 to compute these probabilities. For example, the predicted rate in brothers with one positive FHAM implication is 88%. The predicted rate with one positive and two negative implications is 72%.

Table 6 underscores the differences associated with relationship of the informant. A positive report from a parent carries relatively more weight than a negative report. For example, 3 of 3 positive reports predict a rate of 100% in sons or daughters, whereas all three negative reports predict rates of 24% and 10%, respectively. In contrast, 3 of 3 negative reports from offspring reporting on their parents predict 13% and 0%, respectively. This has major implications for mixtures of positive and negative reports. For example, a negative report from a sibling given two positive reports carries little weight when compared with a negative report from an offspring. The predicted rate in a brother decreases from 98% to 97%, whereas in father it decreases from 90% to 67%.

## DISCUSSION

Consistent with other reports, we find family history information for alcohol dependence to be valid and yield a high level of specificity. Using multiple sources allows the investigators to set a cut-off value for number of positive implications, or a cut-off on the logistic function, so that the specificity and predictive value approach 1.0. This is important in disease screening, in high-risk designs, or extending families through uninterviewed relatives as in COGA. Setting a cut-off to achieve high specificity is at the expense of the value of sensitivity.

**Table 6.** Predicted Rates in Relatives with Multiple Family Histories

| No. of implications | | Siblings reporting on: | | Offspring (or "other") reporting on: | | Parents Reporting on: | |
|---|---|---|---|---|---|---|---|
| Positive | Negative | Brothers | Sisters | Fathers | Mothers | Sons | Daughters |
| 0 | 0* | 49 | 23 | 49 | 23 | 49 | 23 |
| 1 | 0 | 88 | 71 | 75 | 49 | 90 | 75 |
| 0 | 1 | 36 | 15 | 18 | 6 | 40 | 18 |
| 2 | 0 | 98 | 95 | 90 | 75 | 99 | 97 |
| 1 | 1 | 82 | 60 | 40 | 18 | 86 | 68 |
| 0 | 2 | 25 | 10 | 5 | 2 | 31 | 13 |
| 3 | 0 | 100 | 99 | 96 | 91 | 100 | 100 |
| 2 | 1 | 97 | 93 | 67 | 42 | 98 | 95 |
| 1 | 2 | 72 | 48 | 13 | 5 | 81 | 61 |
| 0 | 3 | 16 | 6 | 13 | 0 | 24 | 10 |

* Rates are assumed to be 49% in males and 23% in females.

Anecdotal reports would indicate that women are better informants than men. However, in our data, the sex of the informant was not significant in predicting the presence of alcoholism. There was a significant difference based on the relationship between the relative and the informant. This is best illustrated in Table 6 by examining the impact of a negative implication. A negative report from a child (or mate) is much more influential than from a parent or "other," with a sibling being intermediate.

There are two novel features of our analysis. The first is the ability to use covariates, such as relationship and the diagnostic status of the informant, as well as the sex of the relative. These are significant predictors that increase the precision of the imputed diagnosis. The logistic model is well suited to creating a quantitative scale that takes these covariates into account.

The second novel feature is the use of Eq. 2 to combine multiple family history interviews. Although the logistic model is natural for a single history report, it would not be for a variable number of such reports. Our approach requires some model-based assumption to calculate the joint distribution of family history diagnoses. The assumption we made is that the direct interview is the "gold standard." In fact, a direct interview has an error component and its own sensitivity and specificity, although there is no diagnostic test for the "true" status of an individual. There are cases where the interview is negative, but the individual has many positive FHAM diagnoses. In these cases, it would be reasonable to assume these individuals represent false negatives by interview and use the FHAM information (at perhaps a higher cut-off than for someone not interviewed) to classify these individuals as affected.

Most statistical procedures either eliminate missing observations, or impute their values as the mean, conditional on other observed variables. In genetic analysis, wherein the phenotype of an uninterviewed parent may not be random, family history information can be used to provide information. Even when an individual is interviewed, we can combine all history reports into a single probability of affection as previously shown, and consider the bivariate phenotype based on interview and history. This should enhance diagnostic validity in genetic analysis and utilize the full set of data collected.

There are a few technical aspects of the analyses that need comment. The 12,266 "observations" are not independent. The logistic analysis requires the assumption of conditional independence of the observations. The parameter estimates will be consistent, but the reported $\chi^2$'s in Table 4 may be somewhat inflated. However, with such large sample sizes, the magnitudes of the odds ratios are of interest rather than the level of statistical significance.

A methodological difference between our family assessment instrument (FHAM) and the commonly used FH-RDC[18] is that the FHAM does not probe in detail on each relative. One concern may be that the FH-RDC would have a lower specificity but higher sensitivity as a result. A direct comparison between our results and those reported by Andreasen et al.[10] is problematic, because the latter study used precisely two FH-RDCs and reported a consensus history diagnosis. They report a specificity of 96% and sensitivity of 52%. For a specificity of 94% in our Table 5 (based on a single FHAM), the sensitivity is 41%. Requiring two positive FHAM implications (Table 3), the sensitivity is 50% and the specificity is 94%, almost identical to those of Andreasen et al. The FHAM has the advantage that relatives more distant than first degree may be evaluated. Our results suggest that a corresponding sensitivity and specificity may be achieved by using an appropriate cut-off to classify relatives.

Our findings provide a way to combine multiple family history reports into a single probability of being alcoholic. This procedure takes into account the sex of the relative, the relationship of the informants to the relative, etc. This may be viewed as a "data reduction" technique to summarize all family history information. Future work will include using family history information along with interview data to discriminate false-positives from true-positives and false-negatives from true-negatives. In the analyses herein, an interview diagnosis was used as the "true" state. If cases could be defined using, for example, best estimate clinical consensus, then both interview and history variables could be used as predictors. Another area that merits further

work is consideration of age effects. The duration of the relative's alcohol problems, or the recency of these problems, would likely play a role in an informant's awareness of individual symptoms.

## REFERENCES

1. Sobell LC, Sobell MB: Convergent validity: An approach to increasing confidence in treatment outcome conclusions with alcohol and drug abusers, in Sobell LC, Sobell MB, Ward E (eds): Evaluating Alcohol and Drug Abuse Treatment Effectiveness. New York, Pergamon Press, 1980

2. Midanik LT: Validity of self-reported alcohol use: A literature review and assessment. Br J Addict 83:1019–1039, 1988

3. Maisto SA, Sobell MB, Sobell LC: Comparison of alcoholics' self-reports of drinking behavior with reports of collateral informants. J Consult Clin Psychol 47:106–112, 1979

4. McCrady BS, Paolino TJ, Longabaugh R: Correspondence between reports of problem drinkers and spouses on drinking behavior and impairment. J Stud Alcohol 39:1252–1257, 1978

5. Sher KJ, Discutner C: Reports of paternal alcoholism: Reliability across siblings. Addict Behav 11:25–30, 1986

6. Crews TM, Sher KJ: Using adapted short MASTs for assessing parental alcoholism: Reliability and validity. Alcohol Clin Exp Res 16: 576–584, 1992

7. Mann RE, Sobell LC, Sobell MB, Pavan D: Reliability of a family tree questionnaire for assessing family history of alcohol problems. Drug Alcohol Depend 15:61–67, 1985

8. Zimmerman M, Coryell W, Pfohl B, Stangl D: The reliability of the family history method for psychiatric diagnoses. Arch Gen Psychiatry 45:320–322, 1988

9. Kendler KS, Silberg JL, Neale MC, Kessler RC, Heath AC, et al: The family history method: Whose psychiatric history is measured? Am J Psychiatry 148:1501–1504, 1991

10. Andreasen NC, Rice J, Endicott J, Reich T, Coryell W: The family history approach to diagnosis. Arch Gen Psychiatry 43:421–429, 1986

11. Rice JP, Rochberg N, Endicott J, Lavori PW, Miller C: Stability of psychiatric diagnoses. Arch Gen Psychiatry 49:824–830, 1992

12. Rice JP: Phenotype definition for genetic studies. Psych Clin Neurosci 243:158–163, 1993

13. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R). Washington, D.C., American Psychiatric Association, 1987

14. Feighner JP, Robins E, Guze SB, Woodruff RA Jr, Winokur G, Munoz R: Diagnostic criteria for use in psychiatric research. Arch Gen Psychiatry 26:57–63, 1972

15. Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JI Jr, Reich T, Schmidt I, Schuckit MA: A new, semi-structured psychiatric interview for use in genetic linkage studies: A report of the reliability of the SSAGA. J Stud Alcohol 55:149–158, 1994

16. Cox DR: The Analysis of Binary Data. London, England, Methuen, 1970

17. SAS Institute, Inc.: SAS/STAT User's Guide Volume 2, GLM-VARCOMP, ver. 6, ed 4. Cary, NC, SAS Institute, Inc., 1990

18. Andreasen NC, Endicott J, Spitzer RL, Winokur G: The family history method using diagnostic criteria. Arch Gen Psychiatry 34:1229–1235, 1977