Comparison of Psychiatric Diagnoses From Interview Reports With Those From Best-Estimate Procedures*

KATHLEEN K. BUCHOLZ, PH.D.,† JOHN I. NURNBERGER, JR., M.D.,† JOHN R. KRAMER, PH.D.,† VICTOR M. HESSELBROCK, PH.D.,† MARC A. SCHUCKIT, M.D.,† AND LAURA J. BIERUT, M.D.

Department of Psychiatry, Washington University School of Medicine, 660 South Euclid Avenue, Campus Box 8134, St. Louis, Missouri 63110-1093

ABSTRACT. Objective: The aim of this study was to compare psychiatric diagnoses based on interview information with those based on best-estimate procedures, to evaluate information used in such procedures, and to use 5-year follow-up data to determine whether the best-estimate diagnosis is an improvement over the interview-based diagnosis. **Method:** Psychiatric diagnoses were based on interview reports from 373 probands and 2,615 relatives participating in a high-risk family study of alcoholism. The diagnosis also included clinician ratings in a best-estimate procedure of this study. **Results:** For most diagnoses, both sensitivity and specificity, using the best-estimate diagnosis (BED) as the gold standard, were excellent, in both relatives and probands. Substance abuse was an exception, with very low sensitivity, although specificity rates were excellent. For nonsubstance diagnoses, specificity was high, but sensitivity ranged from 59% to 84% across relatives and probands. In general, BED procedures led to higher prevalence estimates than those from the interview only. In the BED process, family history data were especially useful for conduct and antisocial personality disorders. Follow-up interview data supported the fact that BED procedures led to both enhancements of, as well as errors in, diagnosis. **Conclusions:** Our data attest to the utility of family history information, particularly for antisocial personality disorder and conduct disorder, and indicate that, for the phenotype of substance-dependence disorder, an interview-based diagnosis alone is adequate in classifying individuals with a minimum of error. These results should be reassuring for research studies in which costs and resources required for best-estimate procedures are not affordable. (*J. Stud. Alcohol* **67:** 157-168, 2006)

THE BEST-ESTIMATE DIAGNOSTIC PROCEDURE has been recognized as an important element in the assessment protocol of a psychiatric genetics study (Baron et al., 1990; Merikangas et al., 1989; Weissman et al., 1986). The scientific rationale for such a recommendation is that evaluation of all sources of information to produce a best-estimate diagnosis (BED) reduces diagnostic error (Merikangas et al.,1989; Weissman et al., 1986), which has a detrimental effect on linkage results (Baron et al., 1990; Martinez et al., 1989). BED is particularly common in psychiatric, as opposed to other medical, genetic studies (although, arguably, accurate diagnostic assessment is equally important) because psychiatric illnesses, unlike their medical counterparts, have no external validators, such as laboratory tests or radiological examinations. Thus, it has been typical practice in psychiatric genetics research to supplement self-report data from individuals with information from family members and from medical records. All information is evaluated by experienced clinicians who assign a "best-estimate diagnosis," which may differ from the diagnosis based solely on the individual's reported symptoms. An excellent description of this general method may be found in Leckman et al. (1982). Variations of BED procedures may be found in studies of alcoholism (Cloninger et al., 1988), affective illness (Gershon and Guroff, 1984), opiate addiction (Rounsaville et al., 1991), panic disorder (Weissman et al., 1993), depression (Klein et al., 1994, 2005), and schizophrenia (Maziade et al., 1992, 1995; Roy et al., 1997). For many studies, the BED becomes the endpoint, with relatively few investigations recording the steps along the way leading to a BED, so that comparison with the result from the interview is not always possible. One exception to this is the report of Kosten and Rounsaville (1992) regarding sensitivity and specificity of a semistructured interview, using BEDs in a family study of opiate dependence. They found that, although the best-estimate process yielded more cases, it did not result in a higher false-positive rate, which supported their conclusion that the best-estimate process improved psychiatric diagnoses rather than increased diagnostic errors.

The latter approach of evaluating the interview-based diagnosis in light of a BED is of particular interest today,

Received: July 7, 2005. Revised: September 3, 2005.

*This research was supported by National Institute on Alcohol Abuse and Alcoholism grants U10AA08401, AA11998, and AA12460 and National Institute on Drug Abuse grants DA14363 and DA14632. An early version of this work was presented at the annual meeting of the Research Society on Alcoholism, June 2002, San Francisco, CA.

†Correspondence may be sent to Kathleen K. Bucholz at the above address, or via email at: kkb@wustl.edu. John I. Nurnberger, Jr., is with the Department of Psychiatry, Institute of Psychiatric Research, Indiana University Medical Center, Indianapolis, IN. John R. Kramer is with the Department of Psychiatry, University of Iowa College of Medicine, Psychiatry Research-MEB, Iowa City, IA. Victor M. Hesselbrock is with the Department of Psychiatry, University of Connecticut Health Center, Farmington, CT. Marc A. Schuckit is with the Department of Psychiatry, University of California, San Diego, Department of Psychiatry, San Diego, CA.

not only because BED is itself an effortful process, but also because changes in ethical and regulatory research climates have had an impact on the acquisition of information used in best-estimate processes, particularly family history information and medical records. The historical practice of collecting family history reports without consent of family members has been replaced by the current approach of evaluating four conditions to determine if consent may be waived. These include the following: (1) the research involves no more than minimal risk to the person being reported on ("reportee"), (2) not obtaining consent does not adversely affect the reportee's welfare and rights, (3) the study could not be practicably carried out if consent from the reportee had to be obtained, and (4) information will be provided to the reportee (if deemed appropriate) after the study (Botvin, 2001). A waiver of consent may be granted if all four conditions are met. However, psychiatric family history has been suggested as ineligible for a waiver because it is of greater than minimal risk (Botvin, 2001). Thus, the considerable logistical, financial, and other problems associated with obtaining consent from family members pose a serious barrier to the science and conduct of psychiatric family studies, as have been detailed by others (Brunelli and Zwelling, 2001; Chen et al., 2001; Levinson et al., 2001; Nance, 2001). Although the ultimate decision to grant a waiver is left to the discretion of individual institutional review boards, researchers must provide detailed justification to substantiate their claim that the four conditions are met, yet must be prepared to follow through with the additional logistics if their claim is rejected.

Another new development in the ethical and regulatory research climate has been the stringent rules enacted under the Health Insurance Portability and Accountabilty Act (HIPAA) to protect privacy of health records, which has made acquisition of medical records for research as well as other purposes more difficult (Annas, 2003). Both of these changes have had impact on obtaining information that is needed for best-estimate diagnostic procedures. It is timely to reconsider the utility of BED and whether its contribution in today's climate merits the substantial commitment of time, clinical investment, and administrative burdens on the acquisition and management of this information.

Recently, an opportunity to investigate this issue arose in the Collaborative Study on the Genetics of Alcoholism (COGA), a large multisite family study whose main goal is to identify susceptibility genes for alcoholism (Begleiter et al., 1995; Hesselbrock et al., 2001; Reich et al., 1998). A Best-Estimate Diagnostic procedure was instituted for a sample of individuals who had been genotyped. In a recent report summarizing the familial co-aggregation of psychiatric disorders in COGA, the false-positive and false-negative rates derived from the BED results were used to correct the interview-based diagnoses of individuals who did not undergo best-estimate procedures. (Nurnberger et al., 2004). Readers are directed to Nurnberger et al. (2004) for a more complete description of this method. Prevalence estimates changed when corrections were applied (particularly for substance abuse and some dependence diagnoses and antisocial personality disorder [ASPD]). However, even though the magnitude of the association decreased when BED-corrected diagnoses were used, the overall inferences about patterns of co-aggregation of disorders in relatives remained largely unchanged.

We expanded on that study by focusing on a comparison of specific diagnoses based on the interview information with those based on best-estimate procedures. We computed sensitivity and specificity of the interview, using the BED as the gold standard. These data provided one answer to the question of the usefulness of the BED, as to whether it provided substantially better information over and above that of the interview alone. To address the issue of collection of other types of information, such as family histories and medical records, we evaluated the sources of information that were cited by the clinicians as influencing their best-estimate decision when the BED for a particular disorder differed from that of the interview. This assessment is possible because the information from each clinical review, and not just the summary data, was entered into the database. These analyses permitted evaluation of which information was used in the decisions when the BED was different from the interview diagnosis. Finally, we examined data from the 5-year follow-up in a first effort to evaluate whether the best-estimate refined diagnoses from the interview—including not only the stable negatives and positives but also the interview positive/best-estimate negative ("false-positive") and interview-negative/best-estimate positive ("false-negatives")—are an improvement over the diagnosis based on interview information only. These analyses, taken together, permitted a summary of the merits of best-estimate procedures.

Method

Data for the current study were collected as part of the first wave of assessments in COGA. COGA is a high-risk family study of alcoholism that is being conducted at six university medical centers in the United States. The first of its multistage ascertainment protocol involved recruiting from consecutive admissions to inpatient and outpatient substance-use disorder treatment facilities in the catchment areas of the COGA sites probands who met lifetime criteria for alcohol dependence (AD) as defined by Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM-III-R; American Psychiatric Association, 1987), and by the Feighner criteria for definite alcoholism (Feighner et al., 1972). Exclusion criteria included habitual intravenous drug use, any life-threatening illness that was unrelated to alcoholism, inability to speak English, known

positive status for human immunodeficiency virus (HIV), or not having four first-degree relatives available for interview (two of whom had to reside in one of the COGA centers' catchment areas, thus being available for in-person interviews). All eligible probands, their spouses, and all first-degree biological relatives over the age of 6 years were invited to participate in an interview, including a structured family history assessment (Rice et al., 1995) and completion of several personality and temperament questionnaires.

Qualifications for the second stage included having two first-degree relatives of the index proband who also met criteria for COGA AD. In these densely affected pedigrees, individuals were studied with a more extensive protocol than that used in Stage 1, including not only a semistructured psychiatric interview but also neurophysiological and neuropsychological testing, biochemistry analysis, and genotyping of individuals from a subset of the most informative and densely affected pedigrees.

Control families were also ascertained from a variety of sources to provide an estimate of the prevalence of AD in the general population, but this group will not be reported in the present article because these families were not systematically targeted for the BED process.

Assessment

All subjects were administered a comprehensive psychiatric interview (the Semi-Structured Assessment for the Genetics of Alcoholism [SSAGA]), in person (or, rarely, by telephone), which elicited information for lifetime DSM-III-R diagnoses for the major psychiatric disorders. Details about the SSAGA and its features are described more fully elsewhere (Bucholz et al., 1994; Hesselbrock et al., 1999). Interviews were audiotaped when the respondent gave permission. For alcohol dependence, questions were added to cover criteria for AD based on International Classification of Diseases, 10th Revision (ICD-10; World Health Organization, 1993), Feighner (Feighner et al. 1972), and (approximated) DSM-IV (American Psychiatric Association, 1994) classifications. Test-retest data indicated that the SSAGA had high reliability, with kappa estimates for lifetime DSM-III-R alcohol and other drug dependence exceeding .70 for most substances (Bucholz et al., 1994). Its validity was also established in a study comparing SSAGA diagnoses with those obtained from a clinical interview (Hesselbrock et al., 1999). Prior to data entry, all interviews were edited by senior staff for consistency and resolution of uncertain coding issues, with callbacks to subjects as needed. Computer scoring algorithms were developed by the COGA Assessment Committee to produce diagnoses based on interview information. In addition, medical records (not just psychiatric records) were sought for individuals who provided written releases. The COGA protocol required written informed consent from all subjects prior to enrollment in the study, as approved by the institutional review board at each COGA center.

Five years after baseline interview, probands and controls, children, and other relatives were re-interviewed with an expanded version of the original interview, which added several psychiatric disorders and updated diagnostic criteria. About 70% of those targeted were re-interviewed, with overall rates ranging from 60% to 80% across COGA centers.

Best-estimate diagnosis procedure

Modeled after procedures used in earlier family studies, a BED procedure was instituted for the genotyped sample. One center did not have readily available senior clinical personnel to implement this process, which precluded its participation in the BED procedure. Senior clinicians (M.D.s or Ph.D.s) at each COGA site (but not necessarily affiliated with the COGA project) were recruited to review all available information for each subject and to assign a "best diagnosis" on the basis of that information. None of the clinicians, even those affiliated with COGA, had conducted any of the SSAGA interviews, although some may have been consulted about coding issues during protocol editing. The following information was provided to all BED clinicians: (1) a computer printout with summary diagnoses based on the information elicited during the SSAGA interview, with a detailed listing of alcohol symptoms endorsed and the age of onset for each diagnostic system; for certain nonsubstance diagnoses (depression, conduct disorder [CD], and ASPD), an abbreviated version of symptoms was provided along with the SSAGA diagnosis; for other disorders, only the SSAGA diagnosis was provided; (2) a listing of the results of all family history reports about psychiatric symptoms (and resulting diagnostic status) for that person, along with the relationship of the reporter to the person under review (e.g., parent, sibling) for DSM-III-R diagnoses of alcohol dependence, drug dependence, depression, mania, CD, and ASPD; (3) all medical records (not just psychiatric) that were obtained for that person; (4) the original coded hard-copy SSAGA interview report, with interviewer's and editor's marginal notes; (5) the audiotape of the interview session (if available); and (6) a copy of diagnostic criteria for DSM-III-R (and ICD-10, Feighner, and DSM-IV for alcohol dependence) for reference.

Each clinician provided a final diagnosis for specific substance use disorders as unaffected, abuse, dependence, or uncertain, or a final diagnosis for nonsubstance disorders as present, organic, negative, or uncertain. Available medical records were scanned for evidence of specific alcohol symptoms (withdrawal, seizures, blackouts, psychological problems, and health problems related to drinking) and for evidence of other psychiatric disorders. When their diagnosis differed from that of the interview, clinicians

indicated which sources of information (medical records, family history, interview information) figured in the decision. Decisions were recorded by clinicians on precoded data sheets and entered into a database for analysis. Diagnoses covered by the BED process included alcohol dependence (DSM-III-R, Feighner, ICD-10, and DSM-IV); cannabis, cocaine, other stimulants, opiate, and sedative dependence or abuse; and depression, mania, panic disorder, obsessive-compulsive disorder (OCD), social phobia, CD, and ASPD. Except for alcohol dependence, for which four diagnostic classification systems were covered, all diagnoses were based on DSM-III-R lifetime criteria.

After two independent evaluations, the diagnostic assignments were reviewed by a staff member who was not involved in BED to identify discrepancies, which were resolved by additional evaluations or, at some sites, consensus. Most subjects had two reviews, with only about 4% requiring a third evaluation; 0.8% required four or more reviews.

For this report, the BEDs, as well as the full set of all clinical reviews conducted in the best-estimate process, were analyzed. The BED was compared with the interview-based diagnosis for the following disorders: ICD-10, DSM-III-R, DSM-IV alcohol dependence and abuse, and DSM-III-R abuse of or dependence on cannabis, cocaine, other stimulants, opiates, and sedatives; and depression, mania, social phobia, panic disorder, OCD, CD, and ASPD. Prevalence estimates using the SSAGA diagnosis were compared with those from the consensus BED diagnosis. Using the BED as the gold standard, sensitivity and specificity were computed for each diagnosis. In addition, the entire set of clinical reviews was analyzed to understand the importance of types of information in making a BED when this was different from that of the interview. Data were analyzed separately for probands, who were ascertained from substance use disorder treatment facilities, and their relatives.

Results

The BED process involved 6,152 clinical reviews (5,386 reviews of relatives, 766 reviews of probands) conducted by 40 clinicians at five sites, on a total of 2,988 individuals on whom analyses reported here are based. Of these, 373 were probands and 2,615 were relatives of probands, reflecting 29.9% and 37.2% of all interviewed probands and relatives, respectively, in the full COGA study. The majority of relatives were women (61.2%), and men predominated among probands (76.4%). As in the full COGA sample, approximately 73.2% of probands and 76.6 % of relatives were white, with blacks comprising 18.2% of probands and 14.8% of relatives and Hispanics comprising 8.6% of both probands and relatives. On average, ages were 36.8 and 41.0 years, with median ages of 34 and 37, for probands and relatives, respectively.

Lifetime prevalence, sensitivity, and specificity

Alcohol dependence, abuse, and combined abuse/dependence. Table 1 presents the overall lifetime prevalence derived from the SSAGA and from the BED process. Because probands were in treatment at the time of ascertainment, we do not report alcohol abuse or dependence comparisons for this group as defined by DSM-III-R. Among relatives, the prevalence estimates were comparable for SSAGA diagnosis and BED for alcohol dependence defined by DSM-III-R, DSM-IV, and ICD-10 and for COGA AD. Further, the interview was found to be both highly sensitive and highly specific for relatives across the classifications (Table 2), particularly for the definition of COGA AD, with 98.0% sensitivity and 99.2% specificity. The interview was less sensitive for ICD-10 alcohol dependence, missing about 15% of true cases, but with nearly perfect specificity for relatives. For DSM-IV alcohol abuse and for ICD-10 harmful use, prevalence estimates from the BED were consistently lower than those from the interview-based diagnoses. The opposite held true for DSM-III-R alcohol abuse: The prevalence based on the interview was very low, but much higher when determined by BED. Compared with data on alcohol dependence, sensitivity was much lower for all abuse classifications, ranging from 20.5% to 84.6%. Prevalence estimates for the diagnosis of abuse and/or dependence combined were lower based on BED compared with interview for DSM-IV and ICD-10 classifications; the opposite was true for the disorders based on DSM-III-R. Sensitivity for the combined diagnoses was lower (91.1%) than that for dependence (98.2%) for DSM-III-R definition but considerably higher than that for abuse only (20.5%). Increases in sensitivity, but decreases in specificity, were observed for both DSM-IV and ICD-10 combined abuse/dependence over the dependence diagnosis.

We studied the interplay between interview and BEDs of alcohol abuse and dependence among relatives for the three classification systems. We found that most new cases of BED-determined alcohol abuse among relatives were identified from those classified as unaffected by the interview (65.8% of BED abuse), not from cases of interview-determined dependence (13.7%). Likewise, of the 29 individuals determined by interview as having DSM-III-R alcohol abuse, 5 were changed by BED: 2 to unaffected and 3 to dependence. In contrast, only 2% of interview-diagnosed cases of alcohol dependence were changed by BED: 4 to unaffected and 16 to abuse. For DSM-IV alcohol abuse, only 4.2% of new BED abuse cases were from among those determined as unaffected by interview. Further, among the 479 individuals classified by interview as meeting criteria for DSM-IV alcohol abuse, 29% were downgraded to unaffected by BED, and 11.9% were upgraded to dependence. For ICD-10 diagnoses, the BED process changed the diagnosis of harmful use in a high

TABLE 1. Lifetime prevalence of psychiatric disorder, as derived from SSAGA interview and best-estimate procedures: Probands and relatives

	Probands			Relatives		
Diagnosis	SSAGA %	BED %	*n*	SSAGA %	BED %	*n*
Substance disorders						
Alcohol dependence						
DSM-III-R	100	100	372	34.3	34.2	2,608
DSM-IV	97.8	97.8	370	25.7	26.0	2,607
ICD-10	93.8	96.2	372	19.2	22.2	2,600
COGA-AD[a]	100	100	372	31.4	31.4	2,573
Alcohol abuse						
DSM-III-R	–	–	–	1.1	4.5	2,608
DSM-IV	2.1	2.1	370	18.3	12.7	2,607
ICD-10 (harmful use)	3.8	1.6	372	8.0	2.2	2,600
Abuse or dependence						
DSM-III-R	100	100	372	35.4	38.6	2,608
DSM-IV	100	100	370	43.9	38.7	2,607
ICD-10 (harmful use)	97.6	97.9	372	27.2	24.4	2,600
DSM-III-R dependence						
Cannabis	48.4	48.7	372	13.7	14.0	2,585
Cocaine	57.0	57.5	372	10.8	11.0	2,594
Stimulants	21.8	22.4	371	6.4	6.6	2,604
Opiates	16.0	16.0	370	2.7	2.8	2,603
Sedatives	17.9	18.2	369	3.3	3.5	2,600
DSM-III-R abuse						
Cannabis	1.1	6.2	372	0.8	3.2	2,585
Cocaine	0	3.8	372	0.4	1.7	2,594
Stimulants	1.1	3.5	371	0.4	1.3	2,604
Opiates	0.5	1.6	370	0.2	0.9	2,603
Sedatives	1.4	3.5	369	0.3	1.4	2,600
DSM-III-R abuse or dependence						
Cannabis	49.5	54.8	372	14.5	17.1	2,585
Cocaine	57.0	61.3	372	11.1	12.7	2,594
Stimulants	22.9	25.9	371	6.7	7.9	2,604
Opiates	16.5	17.6	370	2.9	3.7	2,603
Sedatives	19.2	21.7	369	3.6	4.9	2,600
Nonsubstance disorders						
Major depression	11.1	13.6	360	15.9	17.9	2,559
Mania	1.9	1.9	367	0.8	1.0	2,595
Obsessive-compulsive disorder	1.9	1.6	372	0.8	0.6	2,595
Social phobia	3.2	4.0	371	2.2	2.7	2,598
Panic disorder	5.1	5.1	371	2.6	2.9	2,598
Conduct disorder	29.5	33.3	363	10.2	11.4	2,573
Antisocial personality disorder	24.6	30.8	357	5.8	7.7	2,575

Notes: SSAGA = Semi-Structured Assessment for the Genetics of Alcoholism; BED = best-estimate diagnosis; DSM-III-R = Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised; DSM-IV = DSM, Fourth Edition; ICD-10 = International Classification of Diseases, 10th Revision; COGA-AD = Collaborative Study on the Genetics of Alcoholism alcohol dependence. [a]COGA-AD is a combination of those who are classified with both DSM-III-R alcohol dependence and Feighner definite alcoholism.

percentage of cases; only 15% of interview-based harmful use cases retained that diagnosis under BED, 65% became unaffected under BED, and 18% became dependent.

DSM-III-R lifetime dependence on and abuse of other drugs

For probands and their relatives, prevalence estimates for dependence on other drugs were similar across all classes of drugs, from both the interview and from BED. (Note: Only DSM-III-R criteria were assessed for drugs other than alcohol.) For dependence, the interview was highly specific and sensitive. In relatives, the overwhelming proportion of those who were determined to be dependent by interview retained that diagnosis under best-estimate procedure across all classes of drugs. In terms of drug abuse, as was similar to data for DSM-III-R alcohol abuse, more cases of abuse were identified by BED procedures, with most of the BED-defined abuse cases coming from among

TABLE 2. Sensitivity and specificity, by diagnosis, using best-estimate diagnosis (BED) as the true diagnosis: Probands and relatives

	Probands			Relatives		
Variable	*n*	Sens. %	Spec. %	*n*	Sens. %	Spec. %
Substance disorders						
Alcohol dependence						
DSM-III-R	–	–	–	2,608	98.2	98.8
DSM-IV	–	–	–	2,607	90.6	97.2
ICD-10	–	–	–	2,600	85.8	99.9
COGA-AD[a]	–	–	–	2,573	98.0	99.2
Alcohol abuse						
DSM-III-R	–	–	–	2,608	20.5	99.8
DSM-IV	–	–	–	2,607	84.6	91.4
ICD-10 (Harmful use)	–	–	–	2,600	57.1	93.0
Abuse or dependence						
DSM-III-R	–	–	–	2,608	91.1	99.6
DSM-IV	–	–	–	2,607	98.0	90.2
ICD-10	–	–	–	2,600	89.4	92.9
DSM-III-R dependence						
Cannabis	372	98.9	99.4	2,585	95.6	99.6
Cocaine	372	99.1	100	2,594	97.2	99.9
Stimulants	371	96.4	99.6	2,604	96.5	99.9
Opiates	370	98.3	99.7	2,603	94.4	99.8
Sedatives	369	95.5	99.3	2,600	90.2	99.9
DSM-III-R abuse						
Cannabis	372	13.0	99.7	2,585	18.3	99.8
Cocaine	372	0	100	2,594	20.0	100
Stimulants	371	23.1	99.7	2,604	11.8	99.8
Opiates	370	16.7	99.7	2,603	12.5	99.9
Sedatives	369	23.1	99.4	2,600	14.3	99.9
Abuse or dependence						
Cannabis	372	90.2	100	2,585	83.3	99.7
Cocaine	372	93.0	100	2,594	87.0	99.9
Stimulants	371	88.5	100	2,604	83.9	99.9
Opiates	370	93.8	100	2,603	78.1	100
Sedatives	369	87.5	99.7	2,600	73.2	99.9
Nonsubstance disorders						
Major depression	360	81.6	100	2,559	83.2	98.7
Mania	367	85.7	99.7	2,595	61.5	99.8
Obsessive-compulsive disorder	372	100	99.7	2,595	93.8	99.7
Social phobia	371	73.3	99.7	2,598	79.7	99.8
Panic disorder	371	94.7	99.7	2,598	84.2	99.8
Conduct disorder	363	86.0	98.8	2,573	82.3	99.1
Antisocial personality disorder	357	79.1	99.6	2,575	73.2	99.8

Notes: Sens. = sensitivity: interview +/all BED+; spec. = specificity: interview -/all BED-; DSM-III-R = Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised; DSM-IV = DSM, Fourth Edition; ICD-10 = International Classification of Diseases, 10th Revision; COGA-AD = Collaborative Study on the Genetics of Alcoholism alcohol dependence. [a]COGA-AD is a combination of those who are classified with both DSM-III-R alcohol dependence and Feighner definite alcoholism.

the unaffected and not from among those determined as dependent by the interview. For the combined abuse/dependence diagnoses, prevalence estimates were slightly greater for BED-based, compared with interview-based, diagnosis, reflecting the higher prevalence of BED-determined abuse for each class of drug. Sensitivity for the combined abuse/dependence diagnosis was lower than that for dependence but greater than that for abuse, and specificity remained very high.

Nonsubstance diagnoses

Data on prevalence, sensitivity, and specificity for nonsubstance diagnoses are displayed in Tables 1 and 2. For most disorders, prevalence estimates based on the interview were nearly identical to those based on BEDs, with the exception of CD and ASPD, in which more cases were identified by BED. The specificity of the interview was excellent across all diagnoses and for probands and relatives:

98% or higher for all disorders. Sensitivity was also high for most disorders (80% in most instances) except for ASPD and mania, in which about 25% of true ASPD cases (i.e., BED-defined) for probands and relatives, and 15% (probands) and 40% (relatives) of true mania cases, were missed by the interview.

Inferences were similar when nonindependent diagnoses (that is, those not resulting from direct effects of substance or physical illness) were considered (data not shown). Because ASPD is conditional with meeting criteria for CD, we investigated the impact of a change in the diagnosis of CD by the best-estimate procedures on the diagnosis of ASPD. Data indicated that a high percentage of BED-diagnosed CD cases that were not identified by interview also received a BED diagnosis of ASPD: 39 of 46 relatives and 15 of 15 probands. The prevalence of mania was very low (~1%), and the interview was not very sensitive, as noted previously.

Changes in the best-estimate diagnoses from the interview diagnoses: Analysis of all clinical reviews

As noted earlier, all clinical reviews conducted as part of the best-estimate diagnostic process were entered into the database, which permitted an examination of review decisions across all clinical evaluations. When clinicians indicated that their BED differed from that of the interview, the sources of information accounting for this difference were coded on the data entry form. Coding options offered to the clinician included clinical judgment using (1) interview information, (2) family history information, (3) medical record information, or (4) some combination of these. In terms of the direction of difference, it was most often the case that a negative interview diagnosis was changed to a positive BED. This was true for both relatives and probands. The only disorder that differed from this pattern was OCD, where best-estimate prevalence was lower than that from interview.

Although medical records were available for about 25% of all clinical reviews of relatives, most records did not have evidence of the particular diagnosis. For example, among reviews of medical records of relatives, 9% contained evidence of depression, 5.1% had evidence of cocaine abuse/dependence, 3.7% had evidence of cannabis abuse/dependence, and less than 2% had evidence of drug-use disorders (such as stimulants, sedatives, opiates), ASPD and CD, manic disorder, panic disorder, social phobic disorder, and OCD. Reviews of medical records of probands revealed that these were more informative for specific disorders; for example, 22.6% had evidence of depression, 33.5% had evidence of cocaine, 19.4% had evidence of cannabis, 10.4% had evidence of other stimulants, 9.1% had evidence of ASPD, and 6.3% had evidence of CDs. However, even among reviews of probands, evidence of some disorders in medical records was negligible (e.g., 2.8% had evidence of panic disorder, 1.8% had evidence of mania, and 0.6% had evidence of social phobia). We were interested in knowing whether evidence of a given disorder in the medical record was associated with a change in interview diagnosis by the clinician. We observed a higher proportion of changes in interview diagnosis by the BED process when evidence of the disorder was present in the record (Table 3). Typically, a negative interview diagnosis was changed to positive by BED procedures. However, the overall impact was not large because few records were informative for a diagnosis.

For some diagnoses, family history information was influential in determining the BED. This was particularly true for ASPD and CD, in which family history information was cited as the reason for the difference in BED from interview diagnosis in over half of the disagreements. Also, family history information was identified as being important for drug-dependence diagnosis, particularly for cocaine and sedatives, and for the diagnosis of mania. Family history was not influential in best-estimate procedures for major depression, in which family history information was cited as influencing the BED decision in only 16% of reviews. Clinicians attributed differences between their best-estimate and interview diagnoses to information from the interview itself. Further details about the specific information was not obtained. There was no family history for panic disorder, social phobia, or OCD, precluding evaluation of its utility for those diagnoses.

Diagnostic status from 5-year follow-up interview information

We examined the diagnosis at follow-up in the 1,389 relatives with a second interview, cross-classified by their Time 1 BED and interview diagnoses. The follow-up information was divided into four groups: Group 1—positive diagnosis on both processes; Group 2—negative diagnosis by best estimate, positive by interview; Group 3—positive diagnosis by best estimate, negative by interview; and Group 4—negative diagnosis by both processes. Cell sizes were sufficient to analyze major depression, CD, and DSM-IV alcohol dependence at follow-up.

Examination (data not shown, but available from first author by request) revealed that rediagnosis among those positive by both interview and BED was highest: 50%, 58%, and 62% for depression, CD, and DSM-IV AD, respectively. There were few differences between Groups 2 and 3 in terms of the proportion who met diagnostic criteria at the 5-year follow-up interview: Approximately equal percentages of Groups 2 and 3 (40%, 20%, and 30% for depression, CD, and DSM-IV AD, respectively) met criteria at follow-up, which was lower than in Group 1. Individuals who were judged unaffected at Time 1 by both best-estimate

TABLE 3. Proportion of reviews where diagnosis from best-estimate procedures differed from interview diagnosis, cross classified by medical record availability and informativeness

	Overall change		No med. records		No evidence of dx in med. record		Evidence of dx in med. record	
Diagnosis	%	*n*	%	*n*	%	*n*	%	*n*
Relatives								
DSM-III-R dependence								
Alcohol	1.5	5,260	1.5	3,739	1.8	1,348	0.6	173
Cannabis	1.3	5,225	1.4	3,721	1.0	1,450	1.8	54
Cocaine	0.6	5,229	0.5	3,722	0.7	1,433	2.7	74
Stimulants	0.3	5,232	0.4	3,719	0.3	1,496	0	17
Opiates	0.2	5,230	0.3	3,721	0.1	1,476	0	33
Sedatives	0.5	5,222	0.4	3,716	0.5	1,484	18.2	22
Nonsubstance dx								
Major depression	4.5	5,196	4.0	3,706	4.4	1,357	18.8	133
Mania	0.6	5,207	0.6	3,703	0.5	1,487	11.8	17
Obsessive-compulsive disorder	0.3	5,214	0.4	3,707	0.1	1,502	40.0	5
Panic disorder	0.7	5,114	0.6	3,659	0.9	1,449	33.3	6
Social phobia	0.8	5,222	0.6	3,715	1.2	1,504	66.7	3
Conduct disorder	2.9	5,207	2.6	3,707	3.2	1,472	25.0	28
Antisocial personality disorder	2.2	5,199	2.1	3,701	2.2	1,484	35.7	14
Probands								
DSM-III-R dependence								
Cannabis	0.9	748	0.8	238	0.5	411	3.0	99
Cocaine	0.5	744	0.4	239	0.9	336	0	169
Stimulants	1.2	746	1.7	236	1.1	457	0	53
Opiates	0.7	743	0.8	238	0.6	487	0	18
Sedatives	1.6	742	0.8	237	1.9	481	4.2	24
Nonsubstance dx								
Major depression	2.8	740	1.2	242	2.1	389	9.2	109
Mania	0.7	741	0	236	1.0	498	0	7
Obsessive-compulsive disorder	0.3	748	0	239	0.4	508	0	1
Panic disorder	0.8	711	1.3	230	0.4	473	12.5	18
Social phobia	1.5	747	1.7	239	1.4	505	0	3
Conduct disorder	6.0	744	4.1	246	6.4	466	15.6	32
Antisocial personality disorder	7.1	732	4.1	244	6.5	445	30.2	43

Notes: The percentage reflects the within-category percent change. For example, for depression in relatives, overall 4.5% of reviews resulted in a change in the interview diagnosis, but among reviews where medical records contained evidence of depression, change occurred in 18.8%. Med. = medical; dx = diagnosis; DSM-III-R = Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised.

and interview diagnoses showed the greatest stability at follow-up, with most remaining negative at Time 2 interview. Those converting to positive diagnosis at Time 2 likely reflected new onsets of disorder in the 5-year follow-up interval (11% [depression], 6% [CD], and 4% [DSM-IV AD]).

Discussion

In this comparison of psychiatric diagnoses that were based on interview self-report with those that were determined by expert reviews of all sources of information, we found that both sensitivity and specificity, for most diagnoses and using the BED as the gold standard, were excellent, in both relatives and probands. This was particularly true for substance-dependence disorders in relatives, in which there was very little evidence of false-positive error rates across the range of substances assessed. Specificity for alcohol dependence ranged from 97.2% to 99.1%, and for other drug-dependence disorders, greater than 99% across all classes of drugs. This finding gives credence to the use of interview-based diagnoses of substance dependence in genetic linkage studies, reducing much concern that diagnostic error will impact substantially on the findings.

However, in contrast to the findings for substance dependence, the diagnoses of DSM-III-R abuse of alcohol and of other drugs were troublesome, as indicated by comparisons of interview results with those from best-estimate procedures. Sensitivity estimates were very low (ranging

from 11.8% to 20.5%), indicating that high percentages of cases of abuse were missed by the interview. Findings were considerably better for DSM-IV alcohol abuse, with estimated sensitivity of 84.6%, and somewhat better for ICD-10 harmful alcohol use (57.1%). Neither of these classification systems was obtained for drugs other than alcohol, so we are unable to determine whether a similar increase in sensitivity would have been observed for DSM-IV abuse of other drugs. Even so, the sensitivity estimates were markedly lower for the abuse diagnoses than for dependence.

Moreover, unlike the findings for dependence, SSAGA estimates for DSM-IV alcohol abuse (18.3%) and ICD-10 harmful use (8.0%) were higher than those from BED (12.7% and 2.2%, respectively). Operationalization of diagnostic criteria for abuse in the interview for these classification systems may have been too liberal, as suggested by the reversal of interview positive diagnoses to negative by the best-estimate decision. We also considered the possibility that changes in abuse criteria from DSM-III-R to DSM-IV might have accounted for this observation. The DSM-III-R definition of abuse (American Psychiatric Association, 1987) consisted of two criteria (both of which were criteria for dependence as well), with a duration criterion of 1 month (or repeated occurrences). In contrast, for DSM-IV abuse (American Psychiatric Association, 1994), criteria were exclusive to the abuse classification, numbered four, and required repeated occurrence over 12 months instead of 1 month. Comparing content across criteria, three of the four DSM-IV criteria (hazardous use, legal problems, continued social or interpersonal problems) were embedded within the two DSM-III-R criteria. The fourth criterion—role interference, which was a dependence criterion in DSM-III-R—was new to the abuse diagnosis. It may be that clinicians using the BED process interpreted this criterion more stringently than was the case for an interview-based diagnosis. Although reasonable, this must remain speculative because BED ratings of each diagnostic criterion were not available.

We note that it would appear that clinicians made an effort to take into account the difference between ICD-10 harmful alcohol use and other classifications, particularly that of DSM-IV alcohol abuse, judging from the low rates of ICD harmful alcohol use. Unlike the DSM system, the ICD-10 system (World Health Organization, 1993) required evidence of actual physical or psychological harm, and clinicians may have been especially rigorous in their evaluation of this criterion.

In terms of drugs other than alcohol, we note that, in contrast to BED results, very few individuals were identified as meeting criteria for the abuse diagnosis by interview. This observation may indicate a schism between the clinical definition of the diagnosis of DSM-III-R abuse as presented in the American Psychiatric Association manual and the operational definitions used in the research psychiatric interview. This possibility merits further examination because the apparent stringency in the interview may not be faithful to the nosology.

Our data add to those from others over the years documenting the poor reliability of the abuse diagnosis, in general, for all substances (Canino et al., 1999; Chatterji et al., 1997; Easton et al., 1997; Hasin, 2003); and our data now indicate that validity of abuse is similarly problematic, if the diagnosis from best-estimate procedures is taken as true. Our results suggest that the diagnosis of abuse across different classification systems merits further investigation, perhaps using longitudinal data or other characteristics of the subjects themselves to better understand the discrepancy between best-estimate and interview results. A further implication of these findings is that the phenotype of alcohol abuse may be less confidently used in genetic linkage and other biological studies, if interview data only are available to classify individuals.

In terms of other nonsubstance-use disorders, our results point to relatively high proportions of missed cases of major depression, mania, social phobia, panic disorder, CD, and ASPD (assuming that the BED is correct). For all of these disorders, sensitivity estimates were between 59% and 84% across relatives and probands. Specificity was excellent for these same disorders, indicating that false-positive error rates are very low. However, in the best-estimate procedures, as observed for substance use disorders, more cases of disorder—particularly ASPD and CD—were identified. Similar results were reported by Kosten and Rounsaville in their 1992 study of family members of opiate addicts: best-estimate procedures identified between 7% and 41% more cases of certain disorders, with ASPD being at the high end of that range. Although we also observed higher prevalence of disorder based on best-estimate procedures, the differences were not as great as those in the earlier publication.

Further, in this report we were also able to address, albeit in a preliminary way, the question posed by Kosten and Rounsaville (1992) as to whether the best-estimate process reflects an "enhancement in the accuracy of psychiatric diagnosis (or) … an increase in erroneous" diagnoses. Our analyses provide a qualified "yes" to both. It appears that best-estimate procedures enhanced psychiatric diagnosis in identification of a subgroup of individuals who "convert" to an interview diagnosis at 5-year follow-up at a higher rate than their counterparts who were negative by both interview and best-estimate procedures. However, our analyses also affirm the second part of the question, suggesting that, in some cases, BED may have erroneously overturned a positive interview diagnosis (Group 2) because these individuals were as likely as their counterparts with a positive BED diagnosis to be so determined at Time 2.

A comparison of the best-estimate procedures and the follow-up interviews raises an interesting question about which is a better "gold standard": clinician review from

multiple sources or follow-up interview confirmation diagnosis? The comparison of the "true positive" group, identified both by interview and best-estimate procedures as affected at Time 1, indicated that rediagnosis, as affected at 5-year follow-up, was somewhat lower than anticipated. A similar finding, based on the full COGA sample and using interview diagnosis only at two time points, was recently reported by Culverhouse et al. (2005). All "true positive" cases without the same diagnosis at follow-up represent an error in diagnosis either at Time 1 (false-positive) or at Time 2 (false-negative). In comparing diagnostic assignment by interview and by best estimate, clinicians seemed less likely to overturn a positive diagnosis and, instead, were more likely to "upgrade" an unaffected to affected status. It is reassuring that clinicians appear to give great weight to the positive symptoms reported by an individual because these are the basis of diagnosis for fully structured interviews.

Admittedly, the follow-up data are a first attempt to provide an answer to the intriguing question of whether best-estimate procedures enhance psychiatric diagnostic processes. The data we report here may be considered flawed in that interview information only was used as the outcome. Using other measures in the interdisciplinary COGA database, including evoked potential (ERP), neuropsychological data, and (long-term) genetic information might help to provide a less equivocal answer than the one we have suggested. Other measures, such as ERP data, might assist in differentiating whether the false negatives (those positive at Time 1, negative at Time 2) are more like those with a diagnosis at both time points than those who were unaffected at both time points.

Although the phenomenon reported by Kosten and Rounsaville (1992) of best-estimate procedures resulting in a greater number of diagnosed individuals than from the interview methods was also evinced in our study, we did not observe the same weight being given to medical record information by the clinicians who participated in our study. In fact, medical record information, even when available, was not identified by clinicians as being influential in decisions about diagnoses that were different from those of the interview. Several possibilities may explain this. First, the medical records were uninformative about many of the psychiatric disorders under study. Second, those who had a medical record with evidence of a disorder may have been more severely affected, reporting more symptoms at interview, and may also have been more likely to have family members who reported the same symptoms. Therefore, medical records may not have revealed much more than what was contained in the family history and interview data. A third possibility is that the phenotype under study in COGA may not be as severe a phenotype as was reported in the Kosten and Rounsaville (1992) study (opiate addiction), for which medical records may have been more definitive. However, the same proportion of individuals had medical records in our study as in the earlier one (25%).

Finally, in the other study, there were no attributions from the clinicians themselves as to the importance given to medical record information. Individuals with medical records might have been more severely affected. It is not possible to determine whether records truly made a difference or simply were a marker for more severe cases in which the diagnosis would be less equivocal.

Our findings must be interpreted in light of several limitations. First, study probands are ascertained from substance-use disorder treatment settings and are more severely affected with alcohol dependence than a comparable group of subjects ascertained from a general population. Their family members, although not necessarily in treatment themselves, might be more severely affected with a range of substance disorders by virtue of their close biological relationship to a severely alcohol-dependent individual. Information used in the best-estimate process, while keeping clinicians blind to the status of the individual (e.g., proband, relative), did include the interview diagnosis, which may have resulted in an overstatement of agreement with all sources of information. Attribution of sources in the BED process was obtained only in instances in which the BED differed from the interview diagnosis. Although it is relevant to query types of information that clinicians believe influenced their evaluation, it would have been helpful to have had clinicians rate the importance of information for all of their diagnostic decisions, not just those in which there was a disagreement. However, it was interesting that the clinical re-evaluation and re-interpretation of the interview information was cited as an important component of the best-estimate decision. The effortful nature of collecting medical records, coupled with the heightened procedures to safeguard those records under the HIPAA legislation, would not seem to be warranted, at least in a high-risk study such as COGA. This decision might need to be reconsidered if the phenotype were major depression, in which our data did suggest an influence of medical record information on the diagnosis.

Our report affirms the value placed by clinicians on family history information in diagnostic decision making, particularly for ASPD and CD. These data were an integral component of the best-estimate procedure, and although the effort to obtain approval to collect it appears worthwhile, further analysis of its role in predicting stability of diagnosis is warranted. Finally, from the standpoint of genetic and other biological studies, our data indicate that the BED does not greatly refine the diagnosis of substance dependence, and an interview-based diagnosis is highly adequate to classify individuals. These results should be reassuring to research studies in which the costs and resources required for best-estimate procedures are not affordable.

Acknowledgments

The Collaborative Study on the Genetics of Alcoholism (COGA)—principal investigator, Henri Begleiter (State University of New York Health Sciences Center [SUNY HSC] at Brooklyn); co-principal investigators, Laura J. Bierut (Washington University in St. Louis), Howard J. Edenberg (Indiana University), Victor M. Hesselbrock (University of Connecticut), Bernice Porjesz (SUNY HSC at Brooklyn)—includes nine different centers at which data collection, analysis, and storage take place. The nine sites and principal investigators and co-investigators include the following: University of Connecticut (Victor M. Hesselbrock); Indiana University (Howard J. Edenberg, John I. Nurnberger, Jr., P. Michael Conneally, Tatiana Foroud); University of Iowa (S. Kuperman, R. Crowe); SUNY HSCB (B. Porjesz, H. Begleiter); Washington University in St. Louis (Laura J. Bierut, Alison Goate, John Rice); University of California at San Diego (Marc A. Schuckit); Howard University (R. Taylor); Rutgers University (Jay Tischfield); Southwest Foundation (Laura Almasy). Zhaoxia Ren serves as the National Institute on Alcohol Abuse and Alcoholism Staff Collaborator.

In memory of Theodore Reich, M.D., co-principal investigator of COGA since its inception, and one of the founders of modern psychiatric genetics. We acknowledge his immeasurable and fundamental scientific contributions to COGA and the field.

References

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

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), Washington, DC, 1994.

Annas, G.J. HIPAA regulations: A new era of medical-record privacy? New Eng. J. Med. **348:** 1486-1490, 2003.

Baron, M., Endicott, J., and Ott, J. Genetic linkage in mental illness: Limitations and prospects. Brit. J. Psychiat. **157:** 645-655, 1990.

Begleiter, H., Reich, T., Hesselbrock, V., Porjesz, B., Li, T.-K., Schuckit, M.A., Edenberg, H.J., and Rice, J.P. The collaborative study on the genetics of alcoholism. Alcohol Hlth Res. World **19:** 228-236, 1995.

Botvin, J. Protecting the privacy of family members in survey and pedigree research. JAMA **285:** 207-211, 2001.

Brunelli, C.A. and Zwelling, L.A. Letter to the editor. JAMA **285:** 1961, 2001.

Bucholz, K.K., Cadoret, R., Cloninger, C.R., Dinwiddie, S.H., Hesselbrock, V.M., Nurnberger, J.I., Jr., Reich, T., Schmidt, I., and Schuckit, M.A. A new, semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. J. Stud. Alcohol **55:** 149-158, 1994.

Canino, G., Bravo, M., Ramirez, R., Febo, V.E., Rubio-Stipec, M., Fernandez, R.L., and Hasin, D. The Spanish Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): Reliability and concordance with clinical diagnoses in a Hispanic population. J. Stud. Alcohol **60:** 790-799, 1999.

Chatterji, S., Saunders, J.B., Vrasti, R., Grant, B.F., Hasin, D., and Mager, D. Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule—Alcohol/Drug-Revised (AUDADIS-ADR): An international comparison. Drug Alcohol Depend. **47:** 171-185, 1997.

Chen, D.T., Worrall, B.B., and Meschia, J.F. Letter to the editor. JAMA **285:** 1961-1962, 2001.

Cloninger, C.R., Reich, T., Sigvardsson, S., von Knorring, A.-L., and Bohman, M. Effects of changes in alcohol use between generations on inheritance of alcohol abuse. In: Rose, R.M. and Barrett, J.E. (Eds.) Alcoholism: Origins and Outcome, New York: Raven Press, 1988, pp. 49-74.

Culverhouse, R., Bucholz, K.K., Crowe, R.R., Hesselbrock, V., Nurnberger, J.I., Jr., Porjesz, B., Schuckit, M.A., Reich, T., and Bierut, L.J. Long-term stability of alcohol and other substance dependence diagnoses and habitual smoking: An evaluation after 5 years. Arch. Gen. Psychiat. **62:** 753-760, 2005.

Easton, C., Meza, E., Mager, D., Ulug, B., Kilic, C., Gogus, A., and Babor, T.F. Test-retest reliability of the alcohol and drug use disorder sections of the schedules for clinical assessment in neuropsychiatry (SCAN). Drug Alcohol Depend. **47:** 187-194, 1997.

Feighner, J.P., Robins, E., Guze, S.B., Woodruff, R.A., Jr., Winokur, G., and Munoz, R. Diagnostic criteria for use in psychiatric research. Arch. Gen. Psychiat. **26:** 57-63, 1972.

Gershon, E.S. and Guroff, J.J. Information from relatives: Diagnosis of affective disorders. Arch. Gen. Psychiat. **41:** 173-180, 1984.

Hasin, D. Classification of alcohol use disorders. Alcohol Res. Hlth **27:** 5-17, 2003.

Hesselbrock, M., Easton, C., Bucholz, K.K., Schuckit, M., and Hesselbrock, V. A validity study of the SSAGA—A comparison with the SCAN. Addiction **94:** 1361-1370, 1999.

Hesselbrock, V.M., Foroud, T., Edenberg, H., Nurnberger, J.I., Jr., Reich, T., and Rice, J.P. Genetics and Alcoholism: The COGA Project. In: Agarwal, D.P. and Seitz, H.K. (Eds.) Alcohol in Health and Disease, New York: Marcel Dekker, 2001, pp. 103-124.

Klein, D.N., Lewinsohn, P.M., Rohde, P., Seeley, J.R., and Olino, T.M. Psychopathology in the adolescent and young adult offspring of a community sample of mothers and fathers with major depression. Psychol. Med. **35:** 353-365, 2005.

Klein, D.N., Ouimette, P.C., Kelly, H.S., Ferro, T., and Riso, L.P. Test-retest reliability of team consensus best-estimate diagnoses of Axis I and II disorders in a family study. Amer. J. Psychiat. **151:** 1043-1047, 1994.

Kosten, T.A. and Rounsaville, B.J. Sensitivity of psychiatric diagnosis based on the best estimate procedure. Amer. J. Psychiat. **149:** 1225-1227, 1992.

Leckman, J.F., Sholomskas, D., Thompson, W.D., Belanger, A., and Weissman, M.M. Best estimate of lifetime psychiatric diagnosis: A methodological study. Arch. Gen. Psychiat. **39:** 879-883, 1982.

Levinson, D.F., Reich, T., Weissman, M.M., Crowe, R.R., DeLisi, L., Faraone, S.V., Folstein, S., Pelias, M.K., Pulver, A.E., and Silverman, J.M. Letter to the editor. JAMA **285:** 1960-1961, 2001.

Martinez, M., Khlat, M., Leboyer, M., and Clerget-Darpoux, F. Performance of linkage analysis under misclassification error when the genetic model is unknown. Genet. Epidemiol. **6:** 253-258, 1989.

Maziade, M., Raymond, V., Cliché, D., Fournier, J.P., Caron, C., Garneau, Y., Nicole, L., Marcotte, P., Couture, C., Simard, C., Boivin, R., Rodrigue, C., Boutin, P., De Brackeleer, M., Martinez, M., and Merette, C. Linkage results on 11Q21-22 in eastern Quebec pedigrees densely affected by schizophrenia. Amer. J. Med. Genet. **60:** 522-528, 1995.

Maziade, M., Roy, M.-A., Fournier, J.-P., Clince, D., Merette, C., Caron, C., Garneau, Y., Montgrain, N., Shriqui, C., Dion, C., Nicole, L., Potvin, A., Lavallee, J.-C., Pires, A., and Raymond, V. Reliability of best-estimate diagnosis in genetic linkage studies of major psychoses: Results from the Quebec Pedigree Studies. Amer. J. Psychiat. **149:** 1674-1686, 1992.

Merikangas, K.R., Spence, M.A., and Kupfer, D.J. Linkage studies of bipolar disorder: Methodologic and analytic issues. Report of MacArthur Foundation Workshop on Linkage and Clinical Features in Affective Disorders. Arch. Gen. Psychiat. **46:** 1137-1141, 1989.

Nance, W.E. Letter to the editor. JAMA **285:** 1961, 2001.

Nurnberger, J.I., Jr., Wiegand, R., Bucholz, K., O'Connor, S., Meyer, E.T., Reich, T., Rice, J., Schuckit, M., King, L., Petti, T., Bierut, L., Hinrichs, A.L., Kuperman, S., Hesselbrock, V., and Porjesz, B. A

family study of alcohol dependence: Coaggregation of multiple disorders in relatives of alcohol-dependent probands. Arch. Gen. Psychiat. **61:** 1246-1256, 2004.

Reich, T., Edenberg, H.J., Goate, A., Williams, J.T., Rice, J.P., Van Eerdewegh, P., Foroud, T., Hesselbrock, V., Schuckit, M.A., Bucholz, K., Porjesz, B., Li, T.-K., Conneally, P.M., Nurnberger, J.I., Jr., Tischfield, J.A., Crowe, R.R., Cloninger, C.R., Wu, W., Shears, S., Carr, K., Crose, C., Willig, C., and Begleiter, H. Genome-wide search for genes affecting the risk for alcohol dependence. Amer. J. Med. Genet. **81:** 207-215, 1998.

Rice, J.P., Reich, T., Bucholz, K.K., Neuman, R.J., Fishman, R., Rochberg, N., Hesselbrock, V.M., Nurnberger, J.I., Jr., Schuckit, M.A., and Begleiter, H. Comparison of direct interview and family history diagnoses of alcohol dependence. Alcsm Clin. Exp. Res. **19:** 1018-1023, 1995.

Rounsaville, B.J., Kosten, T.R., Weissman, M.M., Prusoff, B., Pauls, D., Anton, S.F., and Merikangas, K. Psychiatric disorders in relatives of probands with opiate addiction. Arch. Gen. Psychiat. **48:** 33-42, 1991.

Roy, M.-A., Lanctot, G., Merette, C., Cliche, D., Fournier, J.-P., Boutin, P., Rodrigue, C., Charron, L., Turgeon, M., Hamel, M., Montgrain, N., Nicole, L., Pires, A., Wallot, H., Ponton, A.-M., Garneau, Y., Dion, C., Lavallee, J.-C., Potvin, A., Szatmari, P., and Maziade, M. Clinical and methodological factors related to reliability of the best-estimate diagnostic procedure. Amer. J. Psychiat. **154:** 1726-1733, 1997.

Weissman, M.M., Merikangas, K.R., John, K., Wickramaratne, P., Prusoff, B.A., and Kidd, K.K. Family-genetic studies of psychiatric disorders: Developing technologies. Arch. Gen. Psychiat. **43:** 1104-1116, 1986.

Weissman, M.M., Wickramaratne, P., Adams, P.B., Lish, J.D., Horwath, E., Charney, D., Woods, S.W., Leeman, E., and Frosch, E. The relationship between panic disorder and major depression: A new family study. Arch. Gen. Psychiat. **50:** 767-780, 1993.

World Health Organization. The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research, Geneva, Switzerland, World Health Organization, 1993.