0145-6008/03/2702-0244$03.00/0
ALCOHOLISM: CLINICAL AND EXPERIMENTAL RESEARCH

Vol. 27, No. 2
February 2003

The Validity of DSM-IV Alcohol Dependence: What Do We Know and What Do We Need to Know?

Deborah S. Hasin, Marc A. Schuckit, Christopher S. Martin, Bridget F. Grant, Kathleen K. Bucholz, and John E. Helzer

This article presents the proceedings of a symposium at the 2002 RSA Meeting in San Francisco, California. Deborah S. Hasin organized the symposium and co-chaired it with Marc Schuckit. The purpose of the symposium was to provide an overview of what is known about the validity of DSM-IV and ICD-10 alcohol dependence and abuse, with a focus on work done since 1994. Presentations included: (1) Validity of DSM-III-R alcohol dependence in adolescents, by Christopher S. Martin; (2) Reliability and validity of DSM and ICD formulations of alcohol use disorders: findings from epidemiology, by Bridget F. Grant; (3) Validity and reliability of the alcohol-dependence phenotype in the context of genetic studies, by Kathleen K. Bucholz; and (4) DSM-IV and beyond: uniting the clinical utility of categories with the precision of dimensions, by John E. Helzer. The findings supported the validity of DSM-IV alcohol dependence across numerous study designs and samples, suggested some value in a dimensional dependence measure, and raised questions about the validity of the diagnosis of alcohol abuse as currently defined. Marc Schuckit, as discussant for the symposium, placed the issues in perspective for the upcoming DSM-V.

Key Words: Alcohol Dependence, Alcohol Abuse, Validity, Reliability, DSM-IV, Diagnosis.

INTRODUCTION

Deborah S. Hasin

In contrast to earlier DSMs, DSM-IV decision-making emphasized the importance of reliability and validity evidence as a basis for decision-making. Major categories such as major depression and schizophrenia did not change markedly since DSM-III (American Psychiatric Association, 1980), partly because reliability and validity data for these categories had been accumulating since the 1970s. In contrast, the criteria for alcohol abuse and dependence changed markedly from DSM-III to DSM-III-R (American Psychiatric Association, 1987), with the Alcohol Dependence Syndrome (ADS) (Edwards, 1986; Edwards and Gross, 1976) as the basis for the change (Rounsaville, 1987). The dependence criteria in DSM-III-R, ICD-10 (World Health Organization, 1992), and DSM-IV (American Psychiatric Association, 1994) were similar, reflecting a concept of dependence combining physiologic and psychological processes. Abuse, a residual category, was defined in DSM-IV by social problems or hazardous use and, in ICD-10, by hazardous use. The changes in alcohol abuse and dependence since DSM-III stimulated many US and international reliability and validity studies, most published only after 1994. Comments concerning a lack of information about alcohol-dependence validity (Meyer, 2001) indicated a need to present the substantial and accumulating reliability and validity literature on alcohol abuse and dependence.

Reliability pertains to validity because poor reliability constrains validity. Test–retest reliability studies of alcohol abuse and dependence were conducted as part of the COGA study (Bucholz et al., 1995), in the US general population (Grant et al., 1995), in substance abuse treatment settings (Hasin et al., 1996b,1997a), in international settings (Chatterji et al., 1997), and in Puerto Rican medical patients (Canino et al., 1999). These studies uniformly showed excellent test–retest reliability for alcohol dependence but much lower reliability for abuse. When abuse was considered independently from dependence, rather than as a residual category, abuse reliability improved, suggesting that the hierarchical relationship to dependence rather than the criteria themselves contributed to the poor reliability of DSM-IV alcohol abuse.

Several longitudinal studies addressed the stability and distinctiveness of the course of alcohol abuse and dependence

From the Mailman School of Public Health (DSH), Division of Epidemiology, and College of Physicians and Surgeons, Department of Psychiatry of Columbia University and the New York State Psychiatric Institute, New York, New York; San Diego VA Healthcare System (MAS) and University of California at San Diego Medical School, San Diego, California; Western Psychiatric Institute and Clinic (CSM), University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Division of Biometry and Epidemiology (BFG), National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland; Washington University School of Medicine (KKB), St. Louis, Missouri; and the Health Behavior Research Center (JEH), University of Vermont, Burlington, Vermont.

Received for publication September 13, 2002; accepted October 2, 2002.

Portions of this research were supported by NIAAA grants AA K02 AA00151 (DSH) and the New York State Psychiatric Institute (DSH).

Reprint requests: Deborah S. Hasin, PhD, Columbia University/New York State Psychiatric Institute, 1051 Riverside Drive, Box 123, New York, NY 10032; Fax: 212-543-5386; E-mail: dsh2@columbia.edu.

DOI: 10.1097/01.ALC.0000060878.61384.ED

based on DSM-III-R and DSM-IV criteria. These include two national samples (Grant et al., 2001; Hasin et al., 1990), a community sample of heavy drinkers (Hasin et al., 1997c), a sample of male university volunteers (Schuckit et al., 2000), and subjects in the (Schuckit et al., 2001) COGA study. These studies consistently show that dependence is likely to remain chronic, while abuse is likely to remit and unlikely to progress to dependence. These findings support the validity of dependence and its distinction from abuse.

When different assessment methods agree well in identifying cases of a given condition, this supports the validity of the condition (Hasin, 2002). Methods can include classification systems whose criteria vary or instruments whose probes and coding systems vary. Studies comparing DSM-III-R, DSM-IV, and ICD-10 assessments included samples of substance abuse patients and nonpatients (Rounsaville et al., 1993), alcoholism patients and their relatives (Schuckit et al., 1994), community heavy drinkers (Hasin et al., 1996a), a US national survey (Grant, 1996), and a 9-country World Health Organization/NIH study (Hasin et al., 1997b). Cross-instrument comparisons (AUDADIS, CIDI, and SCAN) were conducted in the context of the WHO study, one using ICD-10 (Pull et al., 1997) and the other using DSM-IV (Cottler et al., 1997). All studies showed excellent cross-method agreement for alcohol dependence, much lower agreement for abuse, and better agreement for abuse when considered nonhierarchically.

Another validation strategy involves determining whether abuse and dependence diagnoses offer clinically useful information in general population samples of heavier than average drinkers. One study addressed this question in a community sample (Hasin et al., 1997c) and another in a national sample (Hasin and Paykin, 1999). Both showed that dependence was strongly associated with clinically relevant characteristics such as drinking level, family history of alcoholism, treatment, and suicidality, supporting its validity. Abuse was weakly and inconsistently associated with these variables.

Finally, data from the National Longitudinal Alcohol Epidemiologic Survey (NLAES) (Grant, 1997) were used to examine whether alcohol dependence can more validly be seen as a categorical or continuous measure (Hasin et al., 2002). Based on early drinking onset, family history of alcoholism, comorbid drug use, and treatment for alcohol problems, evidence failed to support any specific symptom count (including three or more) as demarcating a categorical threshold. This suggests validity for alcohol dependence as a dimensional measure, consistent with its original conceptualization (Edwards, 1986; Edwards and Gross, 1976).

VALIDITY OF DSM-IV ALCOHOL DEPENDENCE IN ADOLESCENTS

Christopher S. Martin

Patterns of alcohol use and related problems tend to differ in adults and adolescents. Compared with adults, adolescent drinking tends to involve relatively infrequent but high quantity binge drinking (Deas et al., 2000), and most adolescents with DSM-IV alcohol dependence drink, on average, only every other day (Martin et al., 1995). In addition, almost all adolescent problem drinkers are polydrug users (Martin et al., 1996a), which complicates the assessment of DSM-IV alcohol abuse and dependence symptoms.

The meaning of the DSM-IV dependence criteria may differ in adolescents and adults. For example, during adolescence, onset of tolerance is probably a normal developmental phenomenon (Martin and Winters, 1998). We found that change-based definitions of tolerance (such as the DSM-IV definition) are very poor at distinguishing adolescents with and without alcohol dependence (Chung et al., 2001). Another example is the criterion: "Using more or longer than intended." Adolescents often do not have fixed intentions regarding drinking limits, so drinking more than intended in this group often seems to be due to social reasons rather than a compulsion to drink (Chung and Martin, 2002; Cooper, 1994).

Concurrent Validity. Among teens, those with DSM-IV alcohol dependence, alcohol abuse, and no alcohol use disorder (AUD) differ in alcohol use and alcohol problem severity, as well as comorbid psychiatric disorders and psychosocial functioning, indicating concurrent validity of the diagnoses. Data from a clinical sample of adolescents studied in the Pittsburgh Adolescent Alcohol Research Center (PAARC) serve as an example. Teens in this sample with current dependence ($n = 204$), abuse ($n = 169$), and no AUD ($n = 148$) differed significantly from each other on average drinks per occasion in the past year, the past-year average frequency of drinking, the number of illicit drugs ever used, and the number of other substance use disorders (SUD). This pattern of results has been observed repeatedly with adolescent clinical and community samples. Teens with alcohol dependence also have significantly greater substance use severity, more comorbid psychopathology, and poorer psychosocial functioning than nondependent teens (Lewinsohn et al., 1996; Martin et al., 1995; Winters et al., 1999).

Despite this evidence for concurrent validity of DSM-IV alcohol dependence among adolescents, alternative algorithms can also produce robust group differences in alcohol problem severity and related functioning. For example, in the PAARC sample, grouping participants by the number of current dependence criteria (0–7) produced a clear graduated increase across all seven groups for variables indicating current drinking, without evidence for a clear cutpoint indicating the most valid threshold for dependence. A threshold of 1, 2, or 4 dependence criteria produced group differences as large as the DSM-IV threshold of 3 out of 7. This apparent arbitrariness for the diagnostic threshold suggests limits in the validity of this diagnostic category.

Predictive Validity. Predictive validity is a critical component of a useful diagnostic system, as diagnoses should

ultimately tell us about likely prognosis, clinical course, and response to treatment. Data on the clinical course of adolescent-onset AUDs, however, is only beginning to emerge. The adolescent literature suggests a degree of predictive validity of DSM-IV alcohol dependence in that those with dependence tended to have worse outcomes than other groups at follow-up. However, the degree of association is modest, and variation within diagnostic groups in outcomes is much more striking than between-group variation. In the PAARC sample, the proportion of those with no AUD at 1-year follow-up was 82% in those with no baseline AUD but similar in the baseline abusers (53%) and those with dependence at baseline (43%) (Chung and Martin, 2001). More research is needed to characterize the clinical course of adolescent AUDs and to determine the predictive validity of diagnostic systems.

Limitations. There also are a number of limitations in the validity of DSM-IV alcohol dependence when applied to adolescents. One important limitation is the definition and measurement of certain dependence symptoms. Differences between adolescent studies in how the symptoms Tolerance and Larger/Longer are operationalized produce dramatic variation in their observed prevalence. The prevalence of DSM-IV Tolerance in five adolescent community samples ranged from 1.3% to 14.5%. It is implausible that the small demographic differences between these samples can account for this difference. Across four adolescent clinical samples, the prevalence of Tolerance ranged from 27.0% to 60.8%. Although this range could be due in part to differences in severity across these clinical samples, the rank order of Tolerance among the dependence symptoms ranged from first to fifth. With regard to Larger/Longer, prevalence ranged from 1% to 9.8% across five community teen samples. In four clinical samples, symptom prevalence ranged from 32.9% to 56.0% (Chung et al., 2002).

Importantly, this variability in the operationalization of Tolerance and Larger/Longer leads to variability in the rates of abuse and dependence diagnoses. DSM-IV's diagnostic algorithms use mutually exclusive criterion sets for the two types of disorder–1 of 4 for abuse and 3 of 7 for dependence–with the hierarchical rule that dependence precludes a diagnosis of abuse. When relatively common dependence symptoms such as Tolerance and Larger/Longer are assessed inclusively, many persons cross the 3-symptom threshold for a dependence diagnosis that would receive an abuse diagnosis or no diagnosis. This affects the rates of disorders. For example, in four adolescent community studies, the rate of alcohol abuse ranged from 0.4% to 8.2%, more than a 20-fold difference. The rate of alcohol dependence ranged from 0.6% to 4.3% in these studies, about a 7-fold difference (Chung et al., 2002).

Another limitation of DSM-IV alcohol dependence applied to teens lies in the discriminant validity of DSM-IV dependence relative to alcohol abuse. In adolescent studies, dependence and abuse symptom groups are not well distinguished from each other in symptom prevalence or ordering in the age of onset (Martin et al., 1996b; Winters et al., 1999). In addition, the PAARC data and other adolescent studies suggest that the factor structure of the 11 DSM-IV AUD symptoms does not correspond to the 2-factor model inherent in the DSM-IV classification. Instead, 1-factor models often provide a better fit to adolescent AUD symptom data (Fulkerson et al., 1999; Lynch et al., 2000).

Future Directions. There are many things that we still need to learn about the validity of DSM-IV alcohol dependence. Developmentally appropriate definitions are needed to cover such symptoms as compulsive drinking, which may manifest itself differently in adults and adolescents. Even when symptoms are well grounded conceptually and clearly defined, the details of how symptoms are queried and scored are important. Although diagnostic interviews for adolescent AUDs have shown good reliability, results across studies using different structured interviews are inconsistent.

In addition, the issue of dimensional approaches to the assessment of dependence is important. We need to know what dimensional measures validly assess alcohol dependence severity. Another issue concerns the course of adolescent-onset alcohol use disorders. Studies of the predictive validity of diagnostic criteria in this age group will be more interpretable when we know more about the natural history of the disorder.

Perhaps the most important future direction for the field is improved understanding of the construct validity of alcohol dependence. Ultimately, this involves how we understand dependence theoretically and in relation to other constructs in addiction research. At a theoretical level, what is addiction and what is alcohol dependence? Is dependence a unitary concept or does it comprise distinct addiction constructs? Many would argue that dependence comprises distinct addiction constructs, such as physical dependence, impaired control over drinking behavior, and the incentive salience of alcohol use. More research is needed on how these addiction constructs manifest themselves in adolescence and how we can measure these constructs as symptoms and in other ways.

RELIABILITY/VALIDITY OF DSM AND ICD FORMULATIONS OF ALCOHOL USE DISORDERS: FINDINGS FROM EPIDEMIOLOGY

Bridget F. Grant

This presentation delineated and defined major types of paradigms that have been used to study the validity of DSM diagnoses of alcohol abuse and dependence. These included convergent validity, population validity, construct validity, criterion-oriented validity, and validation through family history and follow-up.

For each type of validation paradigm, results are available from general population surveys conducted in the US

over the last 10 years. A number of overall trends emerge from this research.

Convergent validity. Measures of DSM alcohol abuse and dependence relate to other variables in distinct patterns predicted on theoretical or previously found empirical grounds, with stronger associations often found for dependence than abuse (Grant 1993a,b; Pull et al., 1997). These results support the convergent validity of the diagnoses.

Population validity. The relationship between DSM diagnoses and their postulated correlates generalizes across population subgroups, illustrating population validity (Harford and Grant, 1994).

Validation through family history. DSM-III-R and DSM-IV alcohol abuse and dependence were repeatedly shown to ran in families (validation through family history). Importantly, the associations for dependence with a family history were consistently stronger than those found for abuse.

Validation through follow-up. Prospective studies show that DSM-III-R and DSM-IV abuse and dependence show distinctly different longitudinal course as well as stability over time. This illustrates validation through follow-up and also indicates the validity of the distinction between the two diagnostic categories.

Construct validity through factor analysis. Factor analytic studies produce structures similar to those proposed for the DSM abuse and dependence classification (Hasin et al., 1993; Muthen et al., 1993; Nelson et al., 1999). This illustrates construct validity.

Criterion-oriented validity. Fully-structured measures of DSM abuse and dependence agreed highly with their semi-structured counterparts requiring clinical judgment.

A convergence of evidence from numerous validation studies thus suggests that the diagnostic categories are valid, although the evidence more strongly supports the validity of the dependence than the abuse category.

Future validation studies can be seen as existing in a hierarchy according to their method. Criterion-oriented validity may be least informative because the criterion measures are often less reliable than measures being validated. Convergent and construct validation paradigms can be seen as intermediate with regard to informativeness within the hierarchy. Convergent validation will continue to be informative, particularly in determining the boundaries between abuse and dependence and their delineation from other disorders. Construct validation studies will also continue to shed light on the theoretical underpinnings of the abuse and dependence categories. However, construct validation studies will need to keep abreast of state-of-the-art advances in statistical methodology. Validation through family history and follow-up were identified as the most fruitful paradigms for future research, highlighting the need for longitudinal survey designs.

Despite the validity of DSM definitions of alcohol abuse and dependence for some purposes, more refined measures may be needed to better identify neurobiological phenotypic markers for genetics studies or in predicting response to psychopharmacological treatment. A new DSM classification of alcohol use disorders may emerge that will require validation in the future. This new classification should meet the needs of the entire alcohol research community and will: (1) include diagnostic criteria informed by findings from basic research, neuroscience, genetics, developmental psychology, clinical/treatment research, epidemiology, and cultural research; (2) incorporate matrices of dimensionality and subtyping; and (3) utilize new nosological paradigms and classificatory schemes.

VALIDITY OF THE ALCOHOL DEPENDENCE DIAGNOSIS: EVIDENCE FROM THE BEST FINAL DIAGNOSTIC PROCESS IN COGA

Kathleen K. Bucholz

A major challenge in psychiatric genetic studies is the correct identification of phenotypes. The numerous replication failures in psychiatric genetics may be due, in part, to ambiguous or imprecise phenotypes. Because diagnostic imprecision and outright errors in the form of false positives reduce the power of genetic analyses, valid phenotypes are crucial for genetics study. However, the absence of laboratory tests to definitively determine the presence of a psychiatric disorder hampers study of diagnostic validity. Indeed, in psychiatry, reliability (primarily test–retest) has been the performance measure more commonly gauged (Bucholz et al., 1994; Chatterji et al., 1997; Easton et al., 1997; Robins et al., 1981,1982). Kappa estimates for substance dependence disorders (often in the "excellent" range) have consistently exceeded those for other psychiatric disorders, regardless of the assessment instrument studied, type of population sampled, diagnostic classification system examined, and time frame of the disorder. These findings bolster confidence in the diagnoses of alcohol and other substance dependence but they are not necessarily informative about the validity of these disorders. One method of studying validity is the cleverly named LEAD standard (Spitzer, 1983). The acronym reflects the key components in the process–*L*ongitudinal *E*xpert *A*ll *D*ata. This method has been previously applied to substance disorders (Kranzler et al., 1997), with some suggestions for improvement.

Because false positives or low specificity may degrade the power of psychiatric genetic studies, family study research generally employs a "best estimate diagnosis" process to improve validity by minimizing false positives. Best estimate diagnoses capture many features of the LEAD standard (Gershon et al., 1982). The large, multisite family study of alcoholism known as the Collaborative Study on the Genetics of Alcoholism (COGA) used such a procedure. The Best Final Diagnostic (BFD) process used in COGA captures the "EAD" components in the LEAD standard, with the longitudinal component still underway. BFD data from COGA are examined below to address the

validity of alcohol use disorders and other diagnostic groupings as well.

The COGA is a multisite family study of alcoholism that has been ongoing since 1991. The sites for this study include six medical centers located in Connecticut, New York, Indiana, Missouri, Iowa, and California (a newer site in Washington, DC will not be reported here). Case families were selected in stages. In stage I, individuals in inpatient or outpatient alcoholism treatment who met lifetime criteria for both DSM-III-R alcohol dependence and Feighner definite alcoholism (COGA alcoholism) (Feighner et al., 1972) based on a semi-structured psychiatric diagnostic interview (the SSAGA) were selected as probands. All of their first-degree relatives aged 7 to 70 were recruited for interview with the SSAGA (or the child version thereof). To qualify for stage II, which included all available relatives being interviewed with the SSAGA as well as undergoing more extensive assessments such as neurophysiological and neuropsychological measurements and donating blood, two additional first-degree relatives of the proband had to meet COGA criteria for alcoholism. Stage III identified families eligible for genotyping, determined by the Ascertainment Committee. Finally, stage IV indicated those families who were sent to the laboratory for genotyping. The protocol also included control families, who were selected from a variety of sources. These control families were intended to represent the general population, so alcoholic control probands were not excluded.

The BFD procedure was applied primarily to individuals in stage IV families at five of the six COGA centers. Two experienced clinicians (primarily psychiatrists) assigned a best final diagnosis after independently evaluating the diagnostic results of the information from the SSAGA interview, which were created from application of computer algorithms to the SSAGA responses (referred to herein as SSAGA-based diagnosis), in addition to summaries of family history reports about the person; medical records, (if any); the actual hard copy of the coded SSAGA interview; and the audiotaped interview session. For ease of reference, the diagnostic criteria set forth in each classification system were also provided. If the two clinicians were not in agreement about any diagnosis, as determined by a review of BFD data sheets submitted by each clinician by a senior editor at each site, a third review by another clinician was conducted, and a fourth, and so on, as necessary.

Diagnoses covered by the BFD process included four diagnostic classifications for alcohol dependence and abuse (DSM-IIIR, DSM-IV, ICD-10, and Feighner) and DSM-III-R diagnoses for abuse and dependence on five other drug classes: major depression, mania, panic disorder, obsessive compulsive disorder, social phobia, child conduct, adult antisocial personality disorder, and adult antisocial behaviors. Analyses described here include all alcohol dependence and abuse diagnoses, DSM-III-R abuse and dependence on cannabis and cocaine, major depression, mania, and antisocial personality disorder.

A total of 6696 clinical reviews of 3081 individuals by 38 clinicians across five COGA centers were done, averaging about two per person. (Only a minority of individuals required three or more reviewers.) These individuals included 362 probands (29% of all probands), 2580 relatives (36.7% of relatives), and 139 controls (13.2% of all controls). On average, individuals were 40.3 years of age, 56% were females, and the majority were Caucasian (76.9%), with 14.3% being African American, and 8.8% Hispanic.

Comparison of prevalence estimates from interview-only information and BFD assessment yielded few differences for alcohol dependence for DSM-III-R, DSM-IV, or Feighner AD, with slightly higher estimates from the BFD-based diagnoses compared with those based on the SSAGA interview. In contrast, while SSAGA- and BFD-prevalence estimates were both very low for DSM-III-R abuse, SSAGA information alone gave much higher prevalence estimates for DSM-IV and ICD-10 alcohol abuse compared with BFD. For cannabis and cocaine dependence, rates were similar for SSAGA- and BFD-based estimates. However, prevalence estimates for marijuana and cocaine abuse were higher for SSAGA- compared with BFD-based information for all classification systems. For major depression, mania and ASPD, BFD estimates were slightly higher compared with estimates based on SSAGA information alone.

A main objective for these analyses was using the BFD as the "true" diagnosis, thus making it possible to compute sensitivity and specificity of the SSAGA assessment for each diagnosis. Specificity for DSM-III-R and ICD-10 alcohol dependence was nearly perfect. Sensitivity for DSM-III-R alcohol dependence was similarly high and somewhat lower for DSM-IV and ICD-10. However, false positives in all systems were very rare. Investigation of the data indicated that most of the false-positive cases were borderline cases, with the minimum number of symptoms for alcohol dependence reported in the SSAGA interview. Concerning alcohol abuse, specificity was excellent across systems but sensitivity was very poor (under 30%). Sensitivity and specificity findings for marijuana and cocaine dependence were similar to those found for alcohol. Specificity was very high for dependence and abuse, while sensitivity for abuse for both marijuana and cocaine was very low. In terms of the nonsubstance diagnoses, SSAGA diagnoses of depression, mania, and ASPD had very high specificity. Sensitivity was considerably lower but still in an acceptable range.

Most discrepancies between SSAGA-based and BFD-based diagnoses were not due to the additional sources of information such as family history or medical records. Instead, the discrepancies arose due to the expert clinicians' interpretations of the information gathered in the SSAGA.

In conclusion, for alcohol dependence, the specificity findings indicated that there were few false positives across the different classification systems. This is very good news for the genetic analyses of the COGA study. Specificity also was high for dependence on cannabis and cocaine and for

the nonsubstance diagnoses. In contrast, specificity was generally lower for alcohol abuse. Sensitivity for DSM-III-R alcohol abuse also was a problem, with under-diagnosis of DSM-III-R alcohol abuse by both SSAGA and BFD processes. This problem did not apply to abuse as defined under DSM-IV and ICD-10 classifications. In general, these data suggest that abuse is more of a diagnostic problem than dependence, with under-diagnosis by the SSAGA-based information on DSM-III-R substance abuse.

Clinician attribution of differences between the diagnoses arrived at through the BFD process and those obtained from the SSAGA interview implicated re-interpretation of SSAGA information rather than incorporation of new information from other sources, such as family history reports or medical records. The results call for genetic analyses based on BFD rather than SSAGA diagnoses, incorporating longitudinal data. They also call for further investigation of the diagnosis of substance abuse, with symptom specific analyses.

DSM-V AND BEYOND: UNITING THE CLINICAL UTILITY OF CATEGORIES WITH THE PRECISION OF DIMENSIONS

John E. Helzer

The Diagnostic and Statistical Manual (DSM) published by the American Psychiatric Press (American Psychiatric Association, 1952) has served us well as a taxonomy for mental disorders. There has been a steady progression in the Manual from its initial publication. Both the first and second edition (DSM-II) (American Psychiatric Association, 1968)) gave a listing of diagnostic labels, coding numbers, and brief descriptions of the major illnesses. DSM-II expanded the diagnostic options in a few areas, such as the substance use diagnoses, and added a section on child and adolescent disorders. However, neither the original DSM nor DSM-II provided specific diagnostic criteria apart from these brief descriptions of typical features. Patterned after the Feighner diagnostic criteria (Feighner et al., 1972) that were developed mainly for research purposes, DSM-III (American Psychiatric Association, 1980) provided a set of operational definitions in which symptom requirements were carefully specified for each diagnosis. Illness boundaries were also defined, as well as a hierarchy of diagnoses for patients who met multiple definitions.

By the early 1980s, the need for precise diagnostic definitions had become widely recognized. Therefore, DSM-III found rapid acceptance not only in the United States but also internationally. The next edition, DSM-III-R (American Psychiatric Association, 1987), did not represent a conceptual advance for most diagnostic categories but served mostly as an opportunity for fine-tuning the DSM-III. The main exception was the substance use disorders. These underwent a substantial re-organization in DSM-III-R based on newer theoretical conceptualizations of substance dependence. The current edition, DSM-IV (American Psychiatric Association, 1994), provided an opportunity to use data from clinical studies to examine aspects of reliability and validity. Thus, this latest iteration of DSM was based on a more solid empirical foundation. The intent of this RSA presentation was to propose that we consider moving from a categorical classification system to a dimensional one in DSM-V, the next step in the taxonomic progression.

Classification systems, or taxonomies, in medicine can be categorical, like the DSM, or dimensional. In a categorical system, the illness definitions are typically created by a group of experts who combine their collective knowledge to construct definitions based on the typical symptoms of the various disorders. Ideally, the definitions would be mutually exclusive, i.e., any ill subject would meet criteria for only one illness. In the field of medicine, however, it is clearly possible for a patient to have more than one illness, so a categorical system sometimes defines illness hierarchies, identifying which illness is dominant when more than one can be diagnosed. In contrast, a dimensional system is constructed empirically. Typically, a questionnaire or an interview that includes a broad array of symptom questions is given to large groups of patients. Statistical analysis is used to determine which symptoms cluster together and descriptive labels can be applied to the various clusters for convenience. There is no attempt to fit patients in to one and only one category. Dimensional classification can be more complex than in a categorical system, because individual patients can receive a score on each of several symptom scales, i.e., a syndrome profile, rather than a single diagnosis.

There are many advantages of a categorical taxonomy for clinical application. Perhaps the greatest advantage is clinical convenience. A diagnostic system based on easily observable signs and reportable symptoms that results in a single diagnosis has considerable utility for communication, prognosis, and treatment. Because categorical classifications are based on clinical experience, they have face validity and are easily applied in clinical settings. Categorical classification has a long clinical tradition and has served medicine well in the endeavor to discover etiology. And, once discovered, etiology then becomes the basis for classification.

However, there are important disadvantages of a categorical system as well. First, the definitions are not empirically derived but rather based on expert opinion, historical precedent, and, to some extent, even political considerations. Definitions are almost certainly heterogeneous as to etiology. Alcohol dependence, for example, probably encompasses a number of disorders of varying cause. Second, given the "Chinese menu" format of DSM-IV, diagnostic reliability is highly dependent on the reliability of certain key items. In a categorical system, this is quite logical. What sense does it make to think of a patient as depressed in the absence of any evidence of mood disturbance? The tradeoff

is that diagnostic reliability falls if individual, required items are difficult to ascertain.

Another disadvantage of a categorical system is that many patients lie at the definitional boundaries. There is a highly heterogeneous mixture of anxiety, affective, psychotic, and other symptoms and a degree of arbitrariness as to where we draw the definitional margins. Further, there is meaningful variation even among those patients who fall clearly above (or below) the definitional margin. Some of this variation can be measured with rating scales. We can, for example, rate the severity of illness in a group of patients who all meet minimum criteria for major depression using a tool like the Hamilton Depression Rating Scale. However, it could be potentially useful if this variation in severity were reflected in the diagnostic nomenclature, for it is conceivable that etiologies for mild versus severe cases may differ.

A dimensional classification system overcomes the above disadvantages. To some extent, it does so at the cost of clinical convenience. It is more difficult for clinicians to communicate easily in terms of scale scores rather than a single diagnostic label. This can be mostly overcome in that each scale can be given a label corresponding to the traditional clinical diagnosis that most closely reflects the symptoms making up the scale. The most salient advantage of a dimensional system is to better position us to take advantage of our growing knowledge of the human genome in discovering the etiology of psychiatric disorders. Assuming that our current diagnostic categories are indeed highly heterogeneous, we are greatly disadvantaged in genetic epidemiology and other psychiatric research in which the goal is to match genetic variation with clinical phenotypes. Learning to speak in more precise terms about diagnostic profiles, rather than overly inclusive diagnoses, seems a small price to pay for advancing our science toward better etiologic understanding with its corresponding benefits of improved prevention and treatment.

DISCUSSION

Marc A. Schuckit

The papers presented in this symposium effectively highlight a number of issues that must be considered when the field begins to turn toward planning for DSM-V. Relevant questions were raised regarding the optimal diagnostic criteria for dependence and for abuse, and several of the presentations touched on the relative assets and liabilities of a dimensionally based label versus a categorical diagnosis. The distinctions that need to be made between research and clinically oriented diagnostic algorithms have been noted, given useful data relating to the validity of the current diagnostic criteria. The special challenges involved when generic criteria are applied to subgroups such as adolescents have also been addressed.

The impressive data related here will be useful in the process of finding the optimal diagnostic criteria. However, these efforts need to be placed in perspective. Some experiences of the DSM-IV Substance Use Disorders Workgroup might be useful for this purpose.

No diagnostic criteria for substance use disorders are likely to be perfect. The absence of highly sensitive and specific biological tests leaves the need to rely on less precise behaviorally oriented criteria sets. In addition, the substance use disorders are likely to develop from a heterogeneous set of influences that span both biological and psychosocial domains. Therefore, almost any diagnostic approach will have both assets and liabilities, and none will be perfectly accurate. Genetic research may help identify important subtypes, but the heterogeneity in biological factors that contribute to the alcoholism risk and the importance of environmental and cultural influences make it unlikely that a single genetic test will highlight a significantly improved diagnostic algorithm.

The DSM criteria must be clinically useful. Most clinicians who assign diagnoses of substance use disorders are not likely to be researchers nor experts in the field. These busy practitioners will have to consider a wide range of medical and psychiatric conditions. Therefore, the optimal criteria are those that are easy to remember and which can be evaluated without exceptional efforts. Whatever criteria are developed will have to be relatively short and easy to implement or they will be ignored in clinical settings.

The substance use disorders criteria must be relevant to at least 11 separate categories of drugs. Different types of drugs have divergent characteristics and the most accurate diagnostic criteria would take these differences into consideration. Therefore, the most sensitive and specific approach would be to modify the set of criteria to be most appropriate for each of the 11 categories of drugs of abuse. Unfortunately, such a step would make the criteria difficult to implement and jeopardize their clinical usefulness. Therefore, it is likely that any set of criteria that might be generally relevant across categories of drugs will have to be relatively generic in structure. The clinician and researcher can then modify the approach for specific settings.

A basic tenet of any DSM is the need to develop criteria that can be useful in different cultural settings. To do otherwise would mean that the DSM criteria for any specific disorder would only be useful in the United States or very similar countries. Furthermore, unless criteria are relatively generic in scope, they might require separate criteria for different age groups, the two genders, different racial backgrounds, and so on. This would create a virtually impossible task for a diagnostic manual which aims to be relevant in a wide range of situations.

While a dimensional approach has many benefits, it may not be ultimately practical. It is likely that a dimensional diagnosis (e.g., a score of 3 on a 7-point dependence scale) would be converted by busy practitioners and insurance companies to a threshold below which clinical attention or funding would not be likely to be applied. The result would

be the practical conversion of a dimensional scale to a diagnostic category.

Both clinicians and researchers have to arrive at definitions that make the generic criteria operational. In other words, the general criteria sets developed within the DSM are not of optimal use to researchers unless they define the concepts. For example, does repeated use in hazardous situations mean three or more times, five or more times, or another number as a threshold? In addition, how does the diagnostician conceptualize what "spending a great deal of time using or recovering from the effects" to mean? These concepts are stated in relatively generic terms to be guidelines for the clinician and researcher, using a general approach to optimize the commonsense modifications of the criteria that are needed in different cultural settings, for divergent subgroups of patients, and depending upon other relevant information for the patient. Researchers in the substance use disorders field might choose to develop a manual of suggested research definitions for relevant concepts, but it should still be possible for scientists to use the same generic approach as clinicians.

It might not be wise to change criteria unless the new approach has clear advantages over the historical criteria set. This philosophy is based on the fact that a perfect diagnostic criteria set is unlikely to appear in the foreseeable future and the fact that clinicians and researchers need to be able to use the literature and their clinical experiences to optimally treat and study a condition. Therefore, any new and different set of criteria is likely to offer some advantages and disadvantages over the prior approach, and drastic changes that limit generalizability of prior information might be best reserved for situations where data clearly indicate that the advantages far outweigh the problems.

In summary, I believe that, as both a researcher and a clinician, I can live with whatever direction DSM-V chooses to follow. The challenge for me will be to learn both the beneficial and adverse effects of whatever changes are incorporated into that manual. I hope that those who work on DSM-V will consider some of the perspectives I offer above and am grateful to the individuals who presented the excellent papers in this symposium for so clearly highlighting such important issues.

REFERENCES

American Psychiatric Association (1952) *Diagnostic and Statistical Manual of Mental Disorders (DSM)*. American Psychiatric Association, Washington, DC.

American Psychiatric Association (1968) *Diagnostic and Statistical Manual of Mental Disorders(DSM-II)*. 2nd ed. American Psychiatric Association, Washington, DC.

American Psychiatric Association (1980) *Diagnostic and Statistical Manual of Mental Disorders (DSM-III)*. 3rd ed. American Psychiatric Association, Washington, DC.

American Psychiatric Association (1987) *Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R)*. 3rd ed rev. American Psychiatric Association, Washington, DC.

American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. 4th ed. American Psychiatric Association, Washington, DC.

Bucholz KK, Cadoret R, Cloninger CR, Dinwiddie SH, Hesselbrock VM, Nurnberger JL, Reich T, Schmidt I, Schuckit MA (1994) 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.

Bucholz KK, Hesselbrock VM, Shayka JJ, Nurnberger JL, Schuckit MA, Schmidt I, Reich T (1995) Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. J Stud Alcohol 56:500–505.

Canino GJ, Bravo M, Ramírez R, Febo V, Fernández R, Hasin D, Grant BF (1999) 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.

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

Chung T, Martin C (2001) Classification and course of adolescents in addictions treatment programs. Alcohol Clin Exp Res 25:1734–1742.

Chung T, Martin C (2002) Concurrent and discriminant validity of DSM-IV symptoms of impaired control over alcohol consumption in adolescents. Alcohol Clin Exp Res 26:485–492.

Chung T, Martin C, Armstrong T, Labouvie E (2002) Prevalence of DSM-IV alcohol diagnoses and symptoms in adolescent community and clinical samples. J Am Acad Child Adolesc Psychiatry 41:546–554.

Chung T, Martin C, Winters K, Langenbucher J (2001) Assessment of alcohol tolerance in adolescents. J Stud Alcohol 62:687–695.

Cooper M (1994) Motivations for alcohol use among adolescents: development and validation of a four-factor model. Psychol Assess 6:117–128.

Cottler LB, Grant BF, Blaine J, Mavreas V, Pull CB, Hasin D, Compton WM, Rubio-Stipec M, Mager D (1997) Concordance of DSM-IV alcohol and drug use disorder criteria and diagnoses as measured by AUDADIS-ADR, CIDI and SCAN. Drug Alcohol Depend 47:195–205.

Deas D, Riggs P, Langenbucher J, Goldman M, Brown S (2000) Adolescents are not adults: developmental considerations in alcohol users. Alcohol Clin Exp Res 24:232–237.

Easton C, Meza E, Mager D, Ulüg B, Kilic C, Gögüs A, Babor TF (1997) 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.

Edwards G (1986) The Alcohol Dependence Syndrome: a concept as stimulus to enquiry. Br J Addict 81:171–183.

Edwards G, Gross M (1976) Alcohol dependence: provisional description of a clinical syndrome. Br Med J 1:1058–1061.

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

Fulkerson JA, Harrison PA, Beebe TJ (1999) DSM-IV substance abuse and dependence: are there really two dimensions of substance use disorders in adolescents? Addiction 94:495–506.

Gershon ES, Hamovit J, Guroff JJ, Dibble E, Leckman JF, Sceery W, Targum S, Nurnberger JrJI, Goldin LR, Bunney WE (1982) A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. Arch Gen Psychiatry 39:1157–1167.

Grant B (1997) Prevalence and correlates of alcohol use and DSM-IV alcohol dependence in the United States: results of the National Longitudinal Alcohol Epidemiological Survey. J Stud Alcohol 5:464–473.

Grant BF (1993a) International classification of diseases (ICD-10) and the proposed DSM-IV alcohol abuse and dependence, United States, 1988: a nosological comparison. Alcohol Clin Exp Res 17:1093–1101.

Grant BF (1993b) The relationship between ethanol intake and DSM-III-R alcohol dependence: results of a national survey. J Subst Abuse 5:257–267.

Grant BF (1996) DSM-IV, DSM-III-R, and ICD-10 alcohol and drug abuse/harmful use and dependence, United States, 1992: a nosological comparison. Alcohol Clin Exp Res 20:1481–1488.

Grant BF, Harford TC, Dawson DA, Chou PS, Pickering R (1995) The Alcohol Use Disorder and Associated Disabilities Schedule (AUDADIS): reliability of alcohol and drug modules in a general population sample. Drug Alcohol Depend 39:37–44.

Grant BF, Stinson FS, Harford TC (2001) Age at onset of alcohol use and DSM-IV alcohol abuse and dependence: a 12-year follow-up. J Subst Abuse 13:493–504.

Harford TC, Grant BF (1994) Prevalence and population validity of DSM-III-R alcohol abuse and dependence: the 1989 National Longitudinal Survey on Youth. J Subst Abuse 6:37–44.

Hasin D (2002) Classification of alcohol use disorders. Alcohol Res Health, in press.

Hasin D, Carpenter KM, McCloud S, Smith M, Grant BF (1997a) The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): reliability of alcohol and drug modules in a clinical sample. Drug Alcohol Depend 44:133–141.

Hasin D, Grant BF, Cottler L, Blaine J, Towle L, Üstün B, Sartorius N (1997b) Nosological comparisons of alcohol and drug diagnoses: a multisite, multi-instrument international study. Drug Alcohol Depend 47:217–226.

Hasin D, Grant BF, Endicott J (1990) The natural history of alcohol abuse; implications for definitions of alcohol use disorders. Am J Psychiatry 147:1537–1541.

Hasin D, Li Q, McCloud S, Endicott J (1996a) Agreement between DSM-III, DSM-III-R, DSM-IV and ICD-10 alcohol diagnoses in US community-sample heavy drinkers. Addiction 91:1517–1527.

Hasin D, Paykin A (1999) Alcohol dependence and abuse diagnoses: concurrent validity in a nationally representative sample. Alcohol Clin Exp Res 23:144–150.

Hasin D, Trautman K, Miele G, Samet S, Smith M, Endicott J (1996b) Psychiatric Research Interview for Substance and Mental Disorders (PRISM): reliability for substance abusers. Am J Psychiatry 153:1195–1201.

Hasin D, Van Rossem R, McCloud S, Endicott J (1997c) Alcohol dependence and abuse diagnoses: validity in community sample heavy drinkers. Alcohol Clin Exp Res 21:213–219.

Hasin DS, Muthen B, Grant BF (1993) The dimensionality of DSM-IV alcohol abuse and dependence: factor analysis in a clinical sample. Drug Alcohol Depend 88:1079–1090.

Kranzler HR, Tennen H, Babor TF, Kadden RM, Rounsaville BJ (1997) Validity of the longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. Drug Alcohol Depend 45:93–104.

Lewinsohn P, Rodhe P, Seeley J (1996) Alcohol consumption in high school students: frequency of use and dimensional structure of related problems. Addiction 91:375–390.

Lynch K, Martin C, Chung T (2000) Factor structure of DSM-IV alcohol use disorder symptoms in a clinical sample of adolescents. Alcohol Clin Exp Res 24S:S75A.

Martin C, Kaczynski N, Maisto S, Bukstein O, Moss H (1995) Patterns of DSM-IV alcohol abuse and dependence symptoms in adolescent drinkers. J Stud Alcohol 56:672–680.

Martin C, Kaczynski N, Maisto S, Tarter R (1996a) Polydrug use in adolescent drinkers with and without DSM-IV alcohol abuse and dependence. Alcohol Clin Exp Res 20:1099–1108.

Martin C, Langenbucher J, Kaczynski N, Chung T (1996b) Staging in the onset of DSM-IV alcohol abuse and dependence symptoms in adolescents: survival/hazard analyses. J Stud Alcohol 57:549–558.

Martin C, Winters K (1998) Diagnosis and assessment of alcohol use disorders among adolescents. Alcohol Health Res World 22:95–105.

Meyer RE (2001) Finding paradigms for the future of alcoholism research: an interdisciplinary perspective. Alcohol Clin Exp Res 25:1393–1406.

Muthen B, Grant B, Hasin D (1993) The dimensionality of alcohol abuse and dependence: factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. Addiction 88: 1079–1090.

Nelson CB, Rehm J, Üstün B, Grant BF, Chatterji S (1999) Factor structure for DSM-IV substance disorder criteria endorsed by alcohol, cannabis, cocaine and opiate users: results from the World Health Organization Reliability and Validity Study. Addiction 94:843–855.

Pull CB, Saunders JB, Mavreas V, Cottler LB, Grant BF, Hasin DS, Blaine J, Mager D, Üstün B (1997) Concordance between ICD-10 alcohol and drug use disorder criteria and diagnoses as measured by the AUDADIS-ADR, CIDI and SCAN: results of a cross-national study. Drug Alcohol Depend 47:207–216.

Robins LN, Helzer JE, Croughan J, Ratcliff K (1981) National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. Arch Gen Psychiatry 38:381–389.

Robins LN, Helzer JE, Ratcliff KS, Seyfried W (1982) Validity of the Diagnostic Interview Schedule, version II: DSM-III diagnoses. Psychol Med 12:855–870.

Rounsaville BJ (1987) An evaluation of the DSM-III substance-use disorders, in *Diagnosis and Classification in Psychiatry* (Tischler G ed), pp 175–194, Cambridge University Press, New York.

Rounsaville BJ, Bryant K, Babor TF, Kranzler H, Kadden R (1993) Cross system agreement for substance use disorders: DSM-III-R, DSM-IV and ICD-10. Addiction 88:337–348.

Schuckit MA, Hesselbrock V, Tipp JE, Anthenelli R, Bucholz KK, Radziminski S (1994) A comparison of DSM-III-R, DSM-IV and ICD-10 substance use disorders diagnoses in 1922 men and women subjects in the COGA study. Addiction 89:1629–1638.

Schuckit MA, Smith TL, Danko GP, Bucholz KK, Reich T, Bierut L (2001) Five-year clinical course associated with DSM-IV alcohol abuse or dependence in a large group of men and women. Am J Psychiatry 158:1084–1090.

Schuckit MA, Smith TL, Landi NA (2000) The 5-year clinical course of high-functioning men with DSM-IV alcohol abuse or dependence. Am J Psychiatry 157:2028–2035.

Spitzer RL (1983) Psychiatric diagnosis: are clinicians still necessary? Compr Psychiatry 24:399–411.

Winters K, Latimer W, Stinchfield R (1999) The DSM-IV criteria for adolescent alcohol and cannabis use disorders. J Stud Alcohol 60:337–344.

World Health Organization (1992) *The ICD-10 Classification of Mental and Behavioral Disorders*. World Health Organization, Geneva.