Published in final edited form as:

Am J Psychiatry. 2013 April ; 170(4): 442–443. doi:10.1176/appi.ajp.2013.13010032.

The DSM5 field trials and reliability of Alcohol Use Disorder

Deborah S. Hasin, Ph.D., Marc Auriacombe, M.D., Guilherme Borges, Ph.D., Kathleen Bucholz, Ph.D., Alan Budney, Ph.D., Thomas Crowley, M.D., Bridget F. Grant, Ph.D., Ph.D., Charles O'Brien, M.D., Ph.D., Nancy M. Petry, Ph.D., Marc Schuckit, M.D., and Melanie M. Wall, Ph.D.

DSM-5 criteria incorporate many changes to improve psychiatric diagnosis compared to DSM-IV. While test-retest reliability was not a major DSM-5 focus, valid measurement requires adequate reliability (1), so reliability should not drop precipitously. For alcohol use disorders (AUD), a 2006 review (2) indicated good to excellent reliability of current DSM-IV alcohol dependence ($kappa=.66-.82$). Reliability of DSM-IV alcohol abuse was lower, but the criteria were more reliable when analyzed independently of dependence (2). In the DSM-5 field trial, AUD reliability was much lower ($kappa=0.40$) (Regier et al., “Trials in the U.S. and Canada, Part II”; October, 2012) (1). Why the precipitous drop in reliability?

We consider three explanations. First, compared to the DSM-IV studies, DSM-5 field trial cases could have been more complex due to comorbidity that reduced reliability. This appears unlikely because the DSM-IV studies did not exclude comorbid cases and recruited from settings where comorbidity was common. Second, DSM-IV kappas could have been inflated because multi-site studies did not adjust for data pooled across sites. Making this adjustment (3), re-analysis of data from one of the studies (4) suggests this is also unlikely. Pooling data from sites with >50 patients, unadjusted and adjusted k s did not differ dramatically (Table 1).

Third, error variance could have differed between the DSM-IV studies and the DSM-5 field trials. Two main types of error variance reduce reliability (5-7): criterion variance and information variance. Criterion variance is minimized by clear guidelines (i.e., criteria) to evaluate symptoms. Information variance is minimized by standardizing (a) content, through fully or semi-structured interviews, and (b) procedures, i.e., interviewer training in use of the interviews. While never removed completely, studies using standard methodology (structured interviews, training) (7) to control information variance provide more information on criterion variance than studies that do not.

In terms of the first of these types of error variance, did the DSM-5 criteria for AUD become less clear than DSM-IV? This appears unlikely, since the biggest change from DSM-IV to DSM-5 AUD was the combination of dependence and abuse criteria into a single disorder. Except for craving (new to DSM-5) and deletion of legal problems, the AUD criteria are identical in DSM-IV and DSM-5.

Did information variance increase in the DSM-5 Field Trials compared to studies in the 2006 review? Here, the answer appears to be yes. All DSM-IV reliability studies reviewed in 2006 used standard methodology (structured interviews, interviewer training) to minimize

information variance. The DSM-5 Field Trials were very different (1, 3). After brief introduction to DSM-5 criteria, clinicians conducted unstructured diagnostic interviews, rating criteria checklists for any disorder they deemed appropriate. Because clinician training was minimal and diagnostic assessment and checklist selection unstructured, information variance in the DSM-5 Field Trial was not controlled.

Given the design differences in the DSM-IV studies and the DSM-5 Field Trial, reliability of DSM-5 AUD cannot be directly compared to DSM-IV; studies of DSM-5 AUD minimizing information variance are needed. We think such a study should be done to accurately make comparisons between DSM-IV and DSM-5 AUD.

References

1. Regier DA, Narrow WE, Clarke DE, Kraemer HC, Kuramoto SJ, Kuhl EA, Kupfer DJ. DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. *Am J Psychiatry*. 2012;1176/appi.ajp.2012.12070999
2. Hasin D, Hatzenbuehler ML, Keyes K, Ogburn E. Substance use disorders: Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) and International Classification of Diseases, tenth edition (ICD-10). *Addiction*. 2006; 101(Suppl 1):59–75. [PubMed: 16930162]
3. Clarke DE, Narrow WE, Regier DA, Kuramoto SJ, Kupfer DJ, Kuhl EA, Greiner L, Kraemer HC. DSM-5 Field Trials in the United States and Canada, Part I: Study Design, Sampling Strategy, Implementation, and Analytic Approaches. *Am J Psychiatry*. 2012;1176/appi.ajp.2012.12070998
4. Hasin D, Samet S, Nunes E, Meydan J, Matseoane K, Waxman R. Diagnosis of comorbid psychiatric disorders in substance users assessed with the Psychiatric Research Interview for Substance and Mental Disorders for DSM-IV. *Am J Psychiatry*. 2006; 163:689–696. [PubMed: 16585445]
5. Spitzer RL, Endicott J, Robins E. Clinical criteria for psychiatric diagnosis and DSM-III. *Am J Psychiatry*. 1975; 132:1187–1192. [PubMed: 1172654]
6. Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry*. 1974; 125:341–347. [PubMed: 4425771]
7. Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry*. 1978; 35:837–844. [PubMed: 678037]

Table 1

Test-retest reliability *, N=257

Kappa Adjusted for Site		
Diagnosis	No	Yes
Alcohol dependence		
Current	.82	.82
Lifetime	.76	.62
Cocaine dependence		
Current	.90	.87
Lifetime	.87	.87
Cannabis Dependence		
Current	.73	.64
Lifetime	.63	.60
Heroin Dependence		
Current	.93	.93
Lifetime	.96	.92
MDD, Primary		
Current	.77	.84
Lifetime	.72	.80
MDD, Substance-Ind		
Current	.68	.71
Lifetime	.74	.74
Antisocial PD	.73	.74
Borderline PD	.66	.66

* Data from Hasin et al., AJP 2006 (4)