

## Moving Toward DSM-5: The Field Trials

The ongoing revision of APA's *Diagnostic and Statistical Manual of Mental Disorders* (DSM) began in 1999 and has a projected completion date of 2013. The longer production time for the 5th edition (DSM-5) is one of many changes from earlier DSM revision processes. For example, special efforts were taken to avoid elitism and conflicts of interest in the selection of members for the work groups charged with revisiting each major diagnostic category. DSM-5 will also differ from previous versions by its inclusion of dimensional measures (1, 2), meant to be sensitive to differences *between* patients and to changes *within* patients. Some dimensional measures are crosscutting to capture information provided by the patient that is not specific to any diagnosis but more generally useful for assessing mental health. While the work groups conferred in privacy, their proposals have been widely disseminated through publications, presentations at professional meetings, and web postings. The first draft of their diagnostic proposals was recently posted, and the resulting public comments are now being considered by the work groups for further revision of the proposed criteria. The next steps for the criteria are field trials. These DSM-5 field trials have features that may be of interest to potential users.

A field trial is an evaluation of a product in the context in which it will be used. The DSM-5 diagnostic criteria are the product to be used as a basis for clinical decision making and research to benefit patients with mental disorders. As was true for DSM-IV, the process for constructing DSM-5 relies on the currently available scientific and clinical evidence for the diagnostic features of mental disorders. As occurred for DSM-III and DSM-IV, field trials are now needed to assess the clinical utility of the criteria and their reliability when used by different clinicians, but now additional focus is on their test-retest reliability over time in the same patient (precision) and on criterion validity, the extent to which the application of the criteria matches expert consensus diagnosis (accuracy).

---

*“Field trial sites will reflect the heterogeneity of settings in which DSM is actually used, including clinicians and patients from general medicine, psychiatric, and specialty psychiatric clinics.”*

---

Earlier field trials were often conducted by the same groups that developed the proposed criteria. Emphasis was on reliability within observers of the same interview. Thus, major sources of diagnostic error, such as variability in the use of the criteria by different in-

terviewers of the same patient and day-to-day inconsistency of response by the patient, were not available for the calculation of reliability. Moreover, reliability was sometimes estimated with patients selected because they demonstrated particular symptoms of interest and by highly invested clinicians. Now, because a central group has designed the field tests for all work group diagnoses that involve new or controversial changes, there will be uniformity in the approach to field trials. Results will be analyzed centrally and then delivered to the work groups to guide revisions. Field trial sites will reflect the heterogeneity of settings in which DSM is actually used, including clinicians and patients from general medicine, psychiatric, and specialty psychiatric clinics.

### Formal Field Trials in Large Clinical Settings

Formal field trials will involve the testing of between two and five specific diagnoses at any one site. The diagnoses tested at a site will depend on their relative frequency there. For example, major depressive disorder and complex somatic symptom disorder

can be evaluated at a general medical clinic, but autism spectrum disorders require evaluation in a specialty psychiatric clinic specializing in these disorders.

At each site, a research coordinator, trained and monitored centrally, will record each successive entry to the clinic over a specific time period to provide necessary sampling weights for that site's variance in reliability and validity. DSM-IV diagnoses obtained for clinical purposes at each site will be used to place each consenting patient into either a stratum likely to be rich in a target diagnosis at that clinic or a stratum consisting of a random sample of all other diagnoses. The goal is to recruit 50 patients per stratum per site, a total of 150 to 300 patients for each diagnosis under evaluation, to have adequate power for a site-specific determination of precision. Two DSM-5-trained clinicians who are new to the patient will be assigned to conduct independent clinical interviews of the same patient at least 4 hours, but not more than 2 weeks, apart. The attending clinician will be able to observe the interviews. The interviewing clinicians will know the target diagnoses at that site but will be blinded to the stratum to which each patient is assigned and to the attending clinician's diagnosis. The interviewing clinician at each session will be provided the patient's current crosscutting assessments, conduct a clinical interview with the patient, make one or more categorical diagnoses using DSM-5 criteria, and complete associated dimensional severity ratings.

A random 20% of the interviews will be videotaped. The videotapes will be viewed by the work group responsible for the proposed diagnosis to provide an expert consensus or criterion diagnosis against which the criterion validity of the two diagnoses per patient will be assessed. The videotaped interviews may also be used to assess interobserver reliability.

The measure of *reliability* for each diagnosis at each site is the intraclass kappa coefficient, a measure of how well a first diagnosis predicts a second independent one (3). The measure of *validity* is Cohen's kappa, which measures how well a diagnosis predicts the criterion (3). The homogeneity of kappas across sites will be tested. Where there is no strong indication of heterogeneity, kappas will be pooled across sites. The validity estimate, based on 20% of the sample, will not be site specific. Confidence intervals will be provided for all estimations.

When the field trial results are transmitted to the appropriate work groups, they will have access to the accumulated data to address specific questions that might help their understanding of the results as a basis of any necessary revision. If necessary, there will be a second round of field trials to assess substantially revised criteria.

## Clinical Practice Field Trials

Many clinicians and patients for whom DSM-5 is intended are not in settings at which formal reliability and validity testing is possible (4). To include such clinicians in the field trials, a representative sample of U.S. psychiatrists and other volunteer psychiatrists, psychologists, social workers, and psychiatric nurses will be trained. Each will be instructed how to select one new and one ongoing patient in their practice for study enrollment to form a representative sample of U.S. patients. Clinicians will apply DSM-5 diagnostic criteria to this sample and assess feasibility and clinical utility by using the same assessments used in the field trials in large clinical settings.

## Summary

The purpose of successive DSM revisions is to incorporate the growing knowledge base about mental disorders into diagnosis and to bring diagnostic criteria ever closer to accurate and precise identification of corresponding disorders (5). However, this goal must be accomplished while maintaining the clinical utility of the diagnostic criteria for purposes of ready, reliable, and valid use by both clinicians and researchers for prevention, early identification, and treatment. Field trials are the first real-world, empirical test of the success of these efforts in clinical settings.

## References

1. Helzer JE, Kraemer HC, Krueger RF, Wittchen H-U, Sirovatka PJ, Regier DA (eds): Dimensional Approaches in Diagnostic Classification: Refining the Research Agenda for DSM-V. Arlington, Va, American Psychiatric Association, 2008
2. Regier DA, Narrow WE, Kuhl EA, Kupfer DJ: The conceptual development of DSM-V. *Am J Psychiatry* 2009; 166:645–650
3. Kraemer HC, Periyakoil VS, Noda A: Tutorial in biostatistics: kappa coefficients in medical research. *Stat Med* 2002; 21:2109–2129
4. West JC, Wilk JE, Olfson M, Rae DS, Marcus S, Narrow WE, Pincus HA, Regier DA: Patterns and quality of treatment for patients with schizophrenia in routine psychiatric practice. *Psychiatr Serv* 2005; 56:283–291
5. Kupfer DJ, Thase ME: Laboratory studies and validity of psychiatric diagnosis: has there been progress? In *The Validity of Psychiatric Diagnosis*. Edited by Robbins LN, Barrett JE. New York, Raven Press, 1989, pp 177–201

HELENA CHMURA KRAEMER, PH.D.  
 DAVID J. KUPFER, M.D.  
 WILLIAM E. NARROW, M.D., M.P.H.  
 DIANA E. CLARKE, PH.D.  
 DARREL A. REGIER, M.D., M.P.H.

*Address correspondence and reprint requests to Dr. Kupfer, Western Psychiatric Institute and Clinic, University of Pittsburgh School of Medicine, 3811 O'Hara St., Pittsburgh, PA 15213; kupferdj@upmc.edu (e-mail). Commentary accepted for publication July 2010 (doi: 10.1176/appi.ajp.2010.10070962).*

*Dr. Regier is Executive Director of the American Psychiatric Institute for Research and Education (APIRE) and oversees all federal and industry-sponsored research and research training grants in APIRE but receives no external salary funding or honoraria from any government or industry sources. Dr. Freedman has reviewed this editorial and found no evidence of influence from these relationships. All other authors report no financial relationships with commercial interests.*