

The Initial Field Trials of DSM-5: New Blooms and Old Thorns

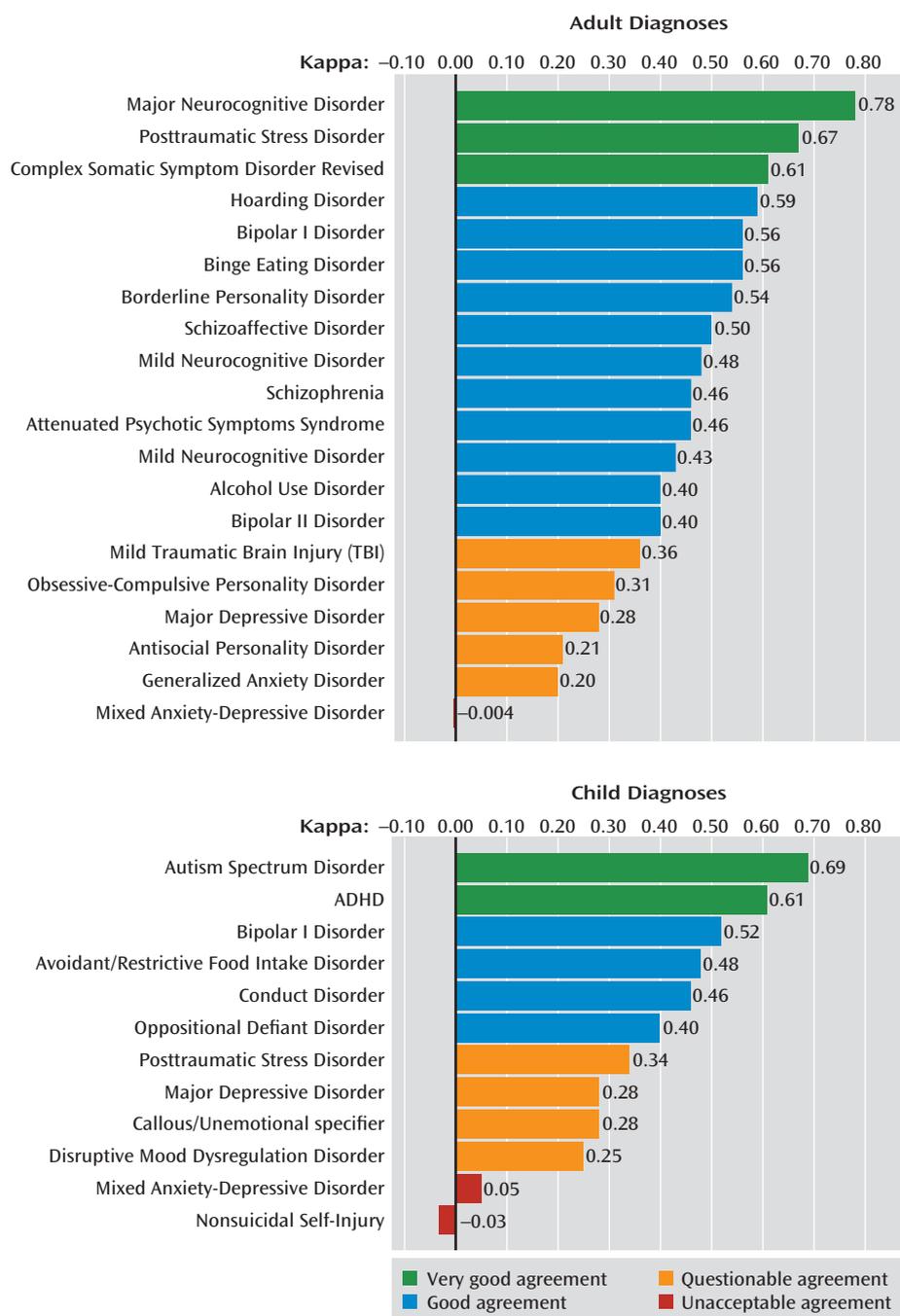
“A rose is a rose is a rose” (1). For psychiatric diagnosis, we still interpret this line as Robins and Guze did for their Research Diagnostic Criteria—that reliability is the first test of validity for diagnosis (2). To develop an evidence-based psychiatry, the Robins and Guze strategy (i.e., empirically validated criteria for the recognizable signs and symptoms of illness) was adopted by DSM-III and DSM-IV. The initial reliability results from the DSM-5 Field Trials are now reported in three articles in this issue (3–5). As for all previous DSM editions, the methods used to assess reliability reflect current standards for psychiatric investigation (3). Independent interviews by two different clinicians trained in the diagnoses, each prompted by a computerized checklist, assessment of agreement across different academic centers, and a pre-established statistical plan are now employed for the first time in the DSM Field Trials. As for most new endeavors, the end results are mixed, with both positive and disappointing findings.

The kappa statistic that is used for the analysis may not be familiar to most clinicians. For illustration, if an illness appears in 10% of a clinic’s patients and two colleagues agree on its diagnosis 85% of the time, the kappa statistic is 0.46, similar to the weighted composite statistic for schizophrenia in this DSM-5 Field Trial (Figure 1). Schizophrenia was radically changed in DSM-III and modified again in DSM-IV because of discrepancies worldwide in its diagnosis. Now, the problem in distinguishing schizophrenia, bipolar disorder, and schizoaffective disorder—the crux of the discrepancies—has largely resolved, and all three conditions have good kappa statistics.

The questionable reliability of major depressive disorder, unchanged from DSM-IV, is obviously a problem. Major depressive disorder has always been problematic because its criteria encompass a wide range of illness, from gravely disabled melancholic patients to many individuals in the general population who do not seek treatment. Although symptom severity on the Hamilton Depression Rating Scale distinguishes those patients who respond more specifically to pharmacotherapy, the DSM-IV criteria do not capture that distinction (6). A second problem not resolved by the DSM-IV criteria is the common co-occurrence of anxiety, which markedly diminishes the effects of antidepressant treatment (7). The DSM-5 work group decided not to change the criteria for major depressive disorder from DSM-IV and instead created other diagnoses for the mixture between anxiety and depression. However, these efforts did not improve the poor reliability of DSM-IV depression; “mixed anxiety and depression” has a kappa of 0. Clinicians often use patients’ self-rating on the Beck Depression Inventory as an indicator of severity. The dimensional cross-cutting domains in this field trial similarly rely on self-rating (5). For depression there are two domains and the intraclass correlations, which are

A single diagnostic interview, regardless of how reliable, does not capture the essence of what is happening...Accurate diagnosis must be part of the ongoing clinical dialogue with the patient.

FIGURE 1. Interrater Reliability of Diagnoses From the Initial DSM-5 Field Trials^a



^a Some of the kappa statistics did not pass the criterion of a standard error less than 0.1. They are included here for illustrative purposes. See the field trial reports for further details (3–5).

similar to the kappa statistic, for adult patients rating and rerating themselves and for parents rating their children; all exceed 0.6. Future revisions will likely need to integrate the many factors—patient self-ratings, cognitive biases, co-occurring anxiety, and vegetative symptoms—that guide treatment selection, prognosis, and assessment of suicide risk.

Experienced clinicians have severe reservations about the proposed research diagnostic scheme for personality disorder, and its applicability to clinical practice has yet to be determined (8). Most of the personality disorder diagnoses did not do

well in the field trial. Antisocial and obsessive-compulsive personality disorders had questionable or inconclusive reliability, and other types like narcissistic and schizotypal personality disorder were seen too infrequently to be assessed. The success of borderline personality disorder is nonetheless a major step forward. DSM-III relegated most personality disorders to axis II, radically severing one of psychiatry's most venerable roots. But clinicians recognized that character pathology, despite its seeming stability, was both quite disabling and amenable to treatment. Borderline personality disorder now emerges as a major diagnosis in its own right with good diagnostic reliability.

Unstable mood, a cardinal feature of borderline personality disorder in adulthood, is also the prominent feature in childhood of a new disorder, disruptive mood dysregulation disorder. This disorder has a more modest kappa statistic. Disruptive mood dysregulation disorder was more reliably assessed in the inpatient setting where it was examined, as was borderline personality disorder early in its history. Perhaps as clinical experience with this new childhood diagnosis increases, its diagnostic performance will improve. Reliability of ADHD and childhood bipolar disorder diagnoses, which had been problematic particularly when irritability was present, likely benefitted from the alternative of disruptive mood dysregulation disorder; both have good kappa statistics. The newly reorganized autism spectrum disorder, also subject of much previous debate, has a very good kappa, although the trials did not include children under 6 years old.

PTSD is another historic accomplishment, with a kappa of 0.67. The DSM series was initiated because “the ‘psychoneurotic label’ had to be applied to men reacting briefly with neurotic symptoms to considerable stress; individuals who...were not ordinarily psychoneurotic” (9). Four editions and 60 years later, PTSD is now a reliable diagnosis for a disorder that might have been dismissed as pathologizing normal behavior. Other new or redefined diagnoses have been introduced with good reliability: major neurocognitive disorder, hoarding disorder, complex somatic symptoms disorder, and binge eating disorder, in addition to those already discussed.

The field trials required that a diagnosis be reached from a single patient interview with minimal collateral information. For a general psychiatric practice, the diagnostic reliability data suggest that two-thirds of patients will receive a reliable DSM-5 principal diagnosis at the first visit. These common, reliable diagnoses are childhood ADHD, PTSD, borderline personality disorder, and alcohol use disorder. The one-third of patients with mild TBI or major depressive disorder may not have a reliable diagnosis from a single interview. Of course, this estimate—derived by combining Table 1 (sample weights in an adult outpatient setting, inserting childhood ADHD as the “other diagnosis” category) with Tables 2 and 4 (reliability of adult and childhood diagnoses [4])—will be different for each clinical setting. Robins and Guze introduced an “undiagnosed” category to urge that patients be re-examined over time when their initial symptoms do not lead to an unambiguous diagnosis. The DSM-5 Field Trials did not examine the increased reliability derived from the same treating clinician assessing the patient over time as the illness unfolds.

“A rose is a rose is a rose is a rose” had deeper meaning for Gertrude Stein, to do not only with the classification of the flower but also with its enduring essence (10). Understanding the natural course of a disorder, its response to treatment, and its impact on the life of the individual are the reasons that we strive to make reliable diagnoses, but a single diagnostic interview, regardless of how reliable, does not capture the essence of what is happening to a patient. If there are lessons for clinicians and patients and families reading these field trials, perhaps the most

important one is that accurate diagnosis must be part of the ongoing clinical dialogue with the patient.

The improvement of diagnosis is also ongoing. Future tests need to consider clinical utility in actual treatment situations and the reliability and practicality of applying the new criteria outside academic medical centers. Solo practitioners and mental health clinics may not have resources for the level of training that the field trials required. The patients were required to speak and read English, although some were bilingual. Reliability may not be the same for patients who have lower levels of education or for whom English is not their most fluent language. The findings of these field trials will be used to make further improvements, and hence the final criteria may change and require further testing after DSM-5 publication. Like its predecessors, DSM-5 does not accomplish all that it intended, but it marks continued progress for many patients for whom the benefits of diagnoses and treatment were previously unrealized.

References

1. Goodwin DW: Preface, in *Psychiatric Diagnosis*. By Woodruff RA, Goodwin DW, Guze SB. New York, Oxford University Press, 1974
2. Robins E, Guze SB: Establishment of diagnostic validity in psychiatric illness: its application to schizophrenia. *Am J Psychiatry* 1970; 126:983–987
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* 2013; 170:43–58
4. 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* 2013; 170:59–70
5. Narrow WE, Clarke DE, Kuramoto SJ, Kraemer HC, Kupfer DJ, Greiner L, Regier DA: DSM-5 Field Trials in the United States and Canada, part III: development and reliability testing of a cross-cutting symptom assessment for DSM-5. *Am J Psychiatry* 2013; 170:71–82
6. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, Fawcett J: Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010; 303:47–53
7. Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, Biggs MM, Zisook S, Leuchter A, Howland R, Warden D, Trivedi MH: Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am J Psychiatry* 2008; 165:342–351
8. Shedler J, Beck A, Fonagy P, Gabbard GO, Gunderson J, Kernberg O, Michels R, Westen D: Personality disorders in DSM-5. *Am J Psychiatry* 2010; 167:1026–1028
9. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC, American Psychiatric Association, 1952
10. Stein G: "Sacred Emily," in *Geography and Plays*. Madison, University of Wisconsin Press, 1922 (reissued 1993)

ROBERT FREEDMAN, M.D.
DAVID A. LEWIS, M.D.
ROBERT MICHELS, M.D.
DANIEL S. PINE, M.D.
SUSAN K. SCHULTZ, M.D.
CAROL A. TAMMINGA, M.D.
GLEN O. GABBARD, M.D.
SUSAN SHUR-FEN GAU, M.D., PH.D.
DANIEL C. JAVITT, M.D., PH.D.
MARIA A. OQUENDO, M.D., PH.D.
PATRICK E. SHROUT, PH.D.
EDUARD VIETA, M.D., PH.D.
JOEL YAGER, M.D.

From the Editors' Office of The American Journal of Psychiatry. Address correspondence to Dr. Freedman (ajp@psych.org). Editorial accepted for publication September 2012 (doi: 10.1176/appi.ajp.2012.12091189).

Authors are Editors or were invited by the Editors to collaborate in this editorial. Several have other roles in the DSM-5 process. Dr. Freedman is co-chair of the Scientific Review Committee, Dr. Pine is chair of the Child Disorders work group, Dr. Schultz is a member of the Geriatric Disorders work group and text editor, and

Dr. Yager is co-chair of the Clinical and Public Health Review Committee. Both Dr. Yager and Dr. Freedman also serve as members of the Summit Task Force, which makes final recommendations to the American Psychiatric Association Board of Trustees. Dr. Gabbard is Editor-in-Chief of the 5th edition of Treatments of Psychiatric Disorders, to be published by American Psychiatric Publishing as the initial DSM-5 treatment book. Financial disclosures of the Editor and Deputy Editors are published each year in the January issue. Dr. Gau has received speaking honoraria and travel funds from Eli Lilly; she has been an investigator in a clinical trial sponsored by Eli Lilly; and she has received speaker's honoraria from AstraZeneca and Janssen. Dr. Javitt has received research grants from Jazz Pharmaceuticals, Pfizer, and Roche; has served as a consultant to AstraZeneca, Bristol-Myers Squibb, Cypress, Lilly, Lundbeck, Merck, NPS, Pfizer, Sanofi, Schering-Plough, Sepracor, Solvay, Takeda, and Sunovion; serves on the advisory board of Promentis Pharmaceuticals; and has equity in Glytech, Inc. Dr. Oquendo has received unrestricted educational grants or lecture fees from AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Janssen, Otsuka, Pfizer, Sanofi-Aventis, and Shire; owns equity in Bristol-Myers Squibb; and receives royalty payments for eC-SSRS from ERT, Inc. Dr. Vieta has served as a consultant and speaker and received research support from AstraZeneca, Bristol-Myers Squibb, and Forest and served as a speaker for GlaxoSmithKline. The remaining authors report no financial relationships with commercial interests.