Computer Aids for the Diagnosis of Anxiety and Depression

The publication of DSM-5 marked many examples of progress in psychiatric diagnosis, but two diagnoses, major depressive disorder and generalized anxiety disorder, the core dysfunctions that psychiatry addresses, did not change from DSM-IV to DSM-5 (1). Yet, these two diagnoses had questionable test-retest reliability in the field tests, although paradoxically, high reliability for patients' self-rating (2). In this issue of the *Journal*, Gibbons et al. (3) report on the development and initial testing of *computerized adaptive testing* to assess patients' self-perception of their anxiety and depression.

In computerized adaptive testing, patients are first asked general questions, and then, based on their initial answers, additional questions are selected to increase the precision of assessment. A good clinician does the same, beginning with general questions and then, based on the answers to those questions, asking more specific questions until the diagnosis is reached. Use of the computer allows for many possible questions and for rapid selection of those most likely to be informative for a given patient. Similar techniques are used by giant online retailers to suggest additional items to buy after an initial purchase is made.

To assess anxiety, Gibbons et al. developed an inventory of 431 questions from initial tests in 798 patients. They developed algorithms to select questions that would produce good agreement between answers to a limited set of questions and the answers to all the questions. The algorithms were tested in 816 patients. Each patient was tested until reaching a criterion for agreement, an accuracy of ±0.3. The computer program adapts to the patient not only by selecting questions but also by continuing until the criterion for accuracy is reached. An average of 12 questions per patient produced a correlation of 0.94 with the answers to the full test. The procedure was tested in 387 patients who also received a standard DSM-IV diagnosis, and the results showed a strong association between the continuous measure of anxiety and the categorical DSM-IV diagnosis.

The Gibbons et al. approach is a truly outstanding contribution to measurement in medicine (particularly in psychiatry): it is novel and exciting, and it promises to improve the accuracy and cost-effectiveness of diagnosis both in clinical practice and in research.

In recent years, emphasis has shifted from a long-standing one-size-fits-all approach in clinical research to one of personalized medicine (4). Increasingly recognized are the facts that no one treatment is equally effective for all and that no one risk factor profile is equally useful for all. Also, no one instrument for measurement or diagnosis is equally accurate and precise for all. Computerized adaptive testing offers one solution to that problem. In this approach, answers to the first few items are used to select which of a large number of items would be most informative for each individual patient, thus coming closer to selection of the best possible instrument for each individual.

This article is featured in this month's AJP Audio

Moreover, with computerized adaptive testing, the test length and duration are shorter, and with automated test scoring, there are savings in the time and costs of processing results as well—and this comes with an *increase* in precision. The increase in precision can in turn lead to a reduction in error rates in clinical decision making, to increased power and precision, and thus to efficiency and cost-effectiveness both in clinical decision making and in clinical research.

A computerized adaptive testing measure can be used for screening or diagnosis, but its value is likely to be greatest for patient follow-up, either by a treating clinician or in a clinical trial. A common problem with the use of measures for this purpose is measure "burnout." When a patient is repeatedly faced with the same long list of items, he or she may eventually simply refuse to cooperate, leading to missing-data problems, or begin to perseverate in response, resulting in decreasing sensitivity to changes in the patient's condition. With computerized adaptive testing measures, as the patient changes, the list of questions will also change, preventing such burnout.

Ten years ago, computerized adaptive testing would hardly have been feasible, for it is not an approach amenable to use with pencil and paper. The ready and easy access to computers, and patients' increasing comfort in using computers, now make this possible. Applications might easily be developed that could be used by patients in the waiting room, probably requiring less than 10 minutes of the patient's time and none of clinician's, producing a score that could be used in its dimensional form (the actual score and its measure of precision) or in categorical form (by selection of an appropriate cut point). Indeed, it might easily be possible to set different cut points, one for screening in a community sample to optimize sensitivity, another for use in imaging or genetic studies to optimize specificity, and others for use in clinical decision making to identify which specific treatments are best indicated for each individual patient.

> *With computerized adaptive testing, the test length and duration are shorter, and with automated test scoring, there are savings in the time and costs of processing results as well—and this comes with an* increase *in precision.*

That the computerized adaptive testing approach seems successful for both anxiety and depression (5) suggests that the approach can be used for many disorders. However, the demonstration of feasibility specifically for anxiety and depression is particularly relevant. One major reason for the "questionable" reliability of major depressive disorder and generalized anxiety disorder is that the patients' expression of these disorders fluctuates from day to day, while their diagnostic status remains unchanged. Such fluctuations would have a minor impact on the reliability of a dimensional measure scale, but the same fluctuations across the diagnostic boundary for a categorical diagnosis would undermine its reliability.

With all this in mind, should the Computerized Adaptive Testing–Anxiety Inventory (CAT-ANX) and the Computerized Adaptive Testing–Depression Inventory (CAT-DI) be immediately adopted? Thus far, evaluations of these instruments have been in the hands of the developers of the method. It is important that independent researchers and clinicians at different sites try these methods out and see whether they work in other milieus. The sample on which the development was based is not representative of any specific population in which the measures might

be used. It may be that the measures work very differently in an epidemiological sample than in a clinic sample, and differently in a general psychiatry clinic than in a specialized clinic. For whom do CAT-ANX and CAT-DI work and for whom do they not work? Does the fact that the items presented to a particular patient may vary from one administration of the test to the next, an advantage in preventing burnout, impair test-retest reliability? The process by which items were selected for the CAT-ANX supports construct validity, and use of the DSM-IV categorical diagnosis as a validity criterion supports convergent validity. However, it would be helpful to assess the validity against other criteria to identify the strengths and limitations of these measures. In short, computerized adaptive testing measures might well be adopted into all current research studies addressing depression and anxiety, first because they might add to the detection of crucial signals in such studies, and second because their use would document where and how best to use these measures in clinical practice.

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 2013; 170:59–70
2. 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
3. Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, Kupfer DJ: Development of the CAT-ANX: a computerized adaptive test for anxiety. Am J Psychiatry 2014; 171:187–194
4. Lesko LJ: Personalized medicine: elusive dream or imminent reality? Clin Pharmacol Ther 2007; 81:807–816
5. Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, Kupfer DJ: Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012; 69:1104–1112

HELENA CHMURA KRAEMER, Ph.D.
ROBERT FREEDMAN, M.D.

From the Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, Calif.; the Department of Psychiatry, University of Pittsburgh, Pittsburgh; and the Department of Psychiatry, University of Colorado Denver, Aurora. (doi: 10.1176/appi.ajp.2013.13111458).