

Medical Test Development and Implementation: A Multistep Journey

A. John Rush, MD

In the context of personalized medicine, biomarkers are being developed to assist in a range of clinical tasks. These tasks include differential diagnosis, prediction of clinical prognosis, treatment selection, monitoring disease processes, monitoring treatment response, identification of those at risk who have yet to ever become symptomatic, and others.¹

The report by John Bilello and colleagues² represents an important and exciting early step toward their pursuit of a biomarker panel to assist in the identification of persons with major depressive disorder (MDD). This particular panel consists of 9 measures that, taken together, result in a single “MDDScore” (range, 1–9). In the validation set from this replication study, the test was highly accurate (91%) in differentiating patients with MDD from normal controls. The area under the curve was 0.93, the sensitivity was 96%, and the specificity was 86%. The positive likelihood ratio was 6.86, and the negative likelihood ratio was 0.047. These figures indicate that a positive test is highly indicative of MDD, with little chance of being incorrect.³

The authors are to be congratulated for specifying the 9 particular elements used in the panel, as these parameters may have pathoetiologic implications. Further, they have continued to develop the algorithm in this report to take into account the role of body mass index and gender, and they addressed the timing issue that affects the estimate of cortisol, one of the test parameters.

This report, however, is but 1 important step in a multistep journey, as the authors imply in their discussion. The first challenge is generalizability. Studies of this test panel to date have been based on rather small numbers of patients drawn from research-experienced academic sites rather than from representative practice sites, in which populations are known to be rather different.⁴ The question is whether the test performs as well in a broadly representative group of inpatients or outpatients with MDD, which we know is accompanied by a wide range of concurrent general medical and psychiatric conditions in routine care.⁵

It is also true that many depressed patients are taking a range of medications not only for their depression but also for their concurrent general medical conditions. The present study excluded patients on NSAIDs, antidepressants, antipsychotics, and anticonvulsants. It is also likely

that participants with only limited general medical and psychiatric comorbidities were included, given the sample sizes and sources. The impact of concurrent medications as well as concurrent general medical or psychiatric conditions on test performance requires further study.

Furthermore, the effect—if any—of race, ethnicity and age (eg, test performance in the elderly or children and adolescents) deserves investigation. In addition, whether the prior course of depressive illness affects test results has yet to be established. For example, is the test result abnormal early in the major depressive episode, or only after a period of time?

Given the nature of the 9 elements in the test panel, it is likely that some of these clinical circumstances and contexts will reduce test performance in ways yet to be defined. On the other hand, an encouraging aspect of the present results is the rather wide gap in the MDDScore between patients with MDD (score: 8–9) and normal controls (score: 1–3). This finding suggests that the test may still be informative and clinically useful if some of these confounds (eg, concomitant medications or medical and psychiatric conditions) only somewhat affect the total MDDScore. Further, in theory, the algorithm could be adjusted to get optimal test performance in specific clinical circumstances (eg, patients with diabetes), thereby allowing wider use of the test. These sorts of limitations are not uncommon with other biomedical tests, but they do affect when the test is clinically informative.

A second issue is the question of which depressed patients are distinguished from controls by the test. Are the results abnormal in bipolar depression? Can the test differentiate bipolar from unipolar depressive episodes? How does it perform in mixed episodes, psychotic depressions, subsyndromal depression, premenstrual dysphoric disorder, or depressions secondary to obsessive-compulsive disorder, posttraumatic stress disorder, or known medical illnesses such as Parkinson's disease or dementia? Answers to these questions will help define the specific diagnostic issues that are best addressed by the test.

A third issue is the clinical context and thus the clinical questions for which the test may be used. Will the test be used to differentiate patients with MDD from those without MDD? Or are there other diagnostic differentiations that are likely to be the basis for using the test? Test use will also be affected by cost. If the test is expensive, it will most likely be used more often in depressed patients who have already failed 1 or more treatments. If it is relatively inexpensive, it may be more widely used, potentially before any treatment is given, as might be the case in primary care.

Submitted: September 30, 2014; accepted October 6, 2014.

Corresponding author: A. John Rush, MD (c/o Zeena Akhbar), Duke-National University of Singapore, The Academia, 20 College Rd, Singapore 169856 (johnrush1960@icloud.com).

J Clin Psychiatry 2015;76(2):e220–e221 (doi:10.4088/JCP.14com09561).

© Copyright 2015 Physicians Postgraduate Press, Inc.

In each context, the test will be used to address slightly different diagnostic questions that, in turn, affect the further steps in its development. Is the objective of the test to distinguish sadness or despondency from MDD? Or might that be done most easily by assessing depressive symptom severity and daily function? In this case, one would need to know what the test adds to simpler clinical assessment procedures to address the diagnostic issue.

More often, clinicians are trying to differentiate groups of patients because that differentiation informs treatment selection. Thus, there is a clear clinical need for differentiating depressed patients from those with anxiety, obsessive-compulsive, bipolar, psychotic, or other psychiatric conditions for which the treatments differ from major depression. It is important to know how the test performs in making these differentiations.

This study also makes an important conceptual point. Results show the feasibility and promise of using a multidimensional panel with 9 items (in this case) to reliably and accurately identify members of a heterogeneous group such as those with MDD. Could the test identify subgroups of patients with MDD to inform treatment selection? Specifically, could the test panel differentiate patient groups that respond to different treatments that work by different mechanisms?

Since the test panel measures potentially etiologic processes, one might speculate that the test panel or selected tests in the panel could identify specific patients whose presentation does not conform to the definition of MDD for whom particular types of antidepressant medication or other treatments are most suited. For example, some individuals with conditions other than MDD may test positive and appropriately so if the test recommends a specific antidepressant intervention. This feature may present some interesting challenges in terms of test development and regulatory approval. Nevertheless, such a finding would have high clinical value.

An additional finding in this study deserves comment. The MDDScore total seems to be little changed during acute treatment between the symptomatic and later improved remitted state, based on a small sample ($n = 15$) treated for 8 weeks with escitalopram. This finding suggests that the test panel might be able to identify those at risk but not yet ill from depression (eg, offspring from loaded pedigrees who are not yet ill). Or it could be a trait marker that portends subsequent relapse, by identifying those who have not had their underlying pathobiology fully corrected by the treatment despite apparent symptom improvement or even

remission. Both ideas raise the possibility that this test panel could address different clinical tasks than differentiating patients with MDD from normal controls.

Finally, a caveat is in order. Psychiatry has a history of evaluating laboratory tests, such as dexamethasone suppression test or REM latency,⁶ that provide some information but that are not “diagnostic.” Hopes have risen and then faded, perhaps in part due to our own excessive expectations not being met. Most of medicine confronts the same challenges. There are few truly diagnostic tests (perhaps hemoglobin SS). We need to be clear about which clinical task(s) each test is addressing. We need to understand what each test can and cannot do in addressing that task. Clinical context, test performance, and test cost affect where, when, and how tests are used and consequently their likely clinical value or utility. It is just as important to know when the test is *not* informative as to know when it is.

That said, Bilello and colleagues’ report² raises substantial hope that the laboratory can finally come to the aid of psychiatric and mental health clinicians and patients to better accomplish the wide range of clinical tasks that we currently address with only our clinical evaluations.

Author affiliations: Duke-National University of Singapore, Singapore.

Potential conflicts of interest: Dr Rush has been a consultant to Brain Resource Ltd, Eli Lilly, Lundbeck, MedAvante, National Institute on Drug Abuse, Santium, and Takeda. He has received honoraria from the University of California at San Diego, Hershey Penn State Medical Center, and the American Society for Clinical Psychopharmacology; royalties from Guilford Publications and the University of Texas Southwestern Medical Center; a travel grant from CINP; and research support from Duke-National University of Singapore.

Funding/support: None reported.

Acknowledgment: Dr Rush acknowledges the editorial support of Jon Kilner, MS, MA (Pittsburgh, Pennsylvania).

REFERENCES

1. Robinson WH, Lindstrom TM, Cheung RK, et al. Mechanistic biomarkers for clinical decision making in rheumatic diseases. *Nat Rev Rheumatol*. 2013;9(5):267–276.
2. Bilello JA, Thurmond LM, Smith KM, et al. MDDScore: confirmation of a blood test to aid in the diagnosis of major depressive disorder. *J Clin Psychiatry*. 2015;76(2):e199–e206.
3. McGee S. Simplifying likelihood ratios. *J Gen Intern Med*. 2002;17(8):646–649.
4. Wisniewski SR, Rush AJ, Nierenberg AA, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? a STAR*D report. *Am J Psychiatry*. 2009;166(5):599–607.
5. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
6. Rush AJ, Schlessler MA, Roffwarg HP, et al. Relationships among the TRH, REM latency, and dexamethasone suppression tests: preliminary findings. *J Clin Psychiatry*. 1983;44(8 pt 2):23–29.