Reliability of Behavioral Phenotyping Predictors of Treatment Response in the EMBARC Study.

McGrath, P.J.1, Bruder, G.1, Dillon, D.2, Pechtel, P.3, Adams, P.1, Carmody, T.1, Cooper, C.3, Deldin, P.3, Fava, M.4, Kurian, B.T.1, McInnis, M.3, Morris, D.1, Parsey, R.5, Trivedi, M.3, Weissman, M.3, Pizzagalli, D.A.3

1Columbia University 2McLean Hospital/Harvard Medical School 3UT Southwestern Medical Center 4Massachusetts General Hospital 5University of Michigan 6Stony Brook University

Abstract

Background: Despite the availability of a variety of antidepressant treatments, up to 90% of patients fail to respond to treatment. The likelihood of remission is even lower, only one in three patients achieved remission in the nationally representative STAR*D study. Unfortunately, attempts to identify clinical or sociodemographic variables predicting antidepressant response have met with very limited success. Consequently, treatment in clinical practice often follows a trial-and-error approach. Identification of reliable predictors of antidepressant response would constitute major progress. One of the overarching goals of the EMBARC study (Embodiment Moderation/Mediation for a Reboot of Antidepressant Response in Clinical Care) is to identify mediators and moderators of treatment response in a large sample of MDD patients. A variety of neurocognitive tests— including measures of psychomotor slowing, cognitive control, working memory, and reward processing— have shown promise in discriminating antidepressant responders and nonresponders. Their reliability and test-retest reliability remains, however, largely unknown. The goal of the current analysis was to evaluate the test-retest reliability of these behavioral phenotyping measures in healthy adults at four research centers in the EMBARC study.

Methods: A neurocognitive battery that included a word fluency task (executive function and cognitive slowing), four-choice reaction time test (psychomotor slowing), AnotB task (speed of reasoning and working memory), Flanker task (executive function and cognitive control), and a probabilistic reward task (reward responsiveness) was administered to 40 healthy adults (10 at each of four EMBARC centers) with a test-retest interval of about 1 week.

Results: Word Fluency: The average number of valid words reported was 44.3 (SD=10.3) at baseline and 45.6 (SD=10.1) at week 1, which matches normative data for the FAS verbal fluency test. There was no significant difference across sessions or sites. The overall test-retest reliability was high (r = 0.81).

Choice Reaction Time: There was no significant difference across sites in the 4-choice reaction time task. Average reaction time was faster in the second session (104.9 ms, SD=10.9) than the baseline session (108.2 ms, t(51) = 0.37, p = 0.71). Overall test-retest reliability was excellent (r = 0.95).

AnotB: There was no significant difference across sites, but average reaction time was faster in the second session (254.8 ms, SD=21.1) than the baseline session (258.2 ms, t(52) = -2.45, p = 0.02). Overall test-retest reliability was excellent (r = 0.96).

Flanker Effects: As hypothesized, healthy controls were significantly slower and less accurate for incongruent relative to congruent trials (reaction time: t(46) = 4.19, p < .001; accuracy: t(46) = 2.68, p < .01). There was no significant difference across sites, but average reaction time was again faster in the second session (1343 ms, SD=109) than in the first session (Baseline: M=1458 ms, SD=109). Overall test-retest reliability was excellent (r = 0.95), with high reliability at the four sites.

Discussion

These findings demonstrate that neurocognitive measures previously found to predict antidepressant response in depression can be measured with high reliability in a multi-site study. The only exception was the probabilistic reward task, which showed poor test-retest reliability in healthy controls. Interestingly, preliminary analyses indicate that the test-retest reliability for the probabilistic reward task was significant for MDD patients. These findings provide the foundation to investigate the predictive validity of these behavioral phenotyping markers with respect to treatment outcome in major depression.

References
