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

COLUMBIA PSYCHIATRY

McGrath, P.J.[1], Bruder, G.[1], Dillon, D.[2], Pechtel, P.[2], Adams, P.[1], Carmody, T.[3], Cooper, C.[3], Deldin, P.[5], Fava, M.[4], Kurian, B.T.[3], McInnis, M.[5], Morris, D.[3], Parsey, R.[6], Trivedi, M.[3], Weissman, M.[1], Pizzagalli, D.A.[2]

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

52nd Annual Meeting American College
of Neuropsychopharmacology
in Hollywood, Florida, December 8-12, 2013

Abstract

Background: Despite the availability of a variety of antidepressant treatments, up to 50% 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 (Establishing Moderators/Mediators for a Biosignature 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 responsiveness – have shown promise in discriminating antidepressant responders and nonresponders. Their reproducibility and test-retest reliability remains, however, largely unknown. The goal of the current analyses 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 test (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. Average reaction time was faster in the second session (449ms, SD=93) than the baseline session (488ms, SD=109, p=.01). Overall test-retest reliability was excellent (r = 0.95)

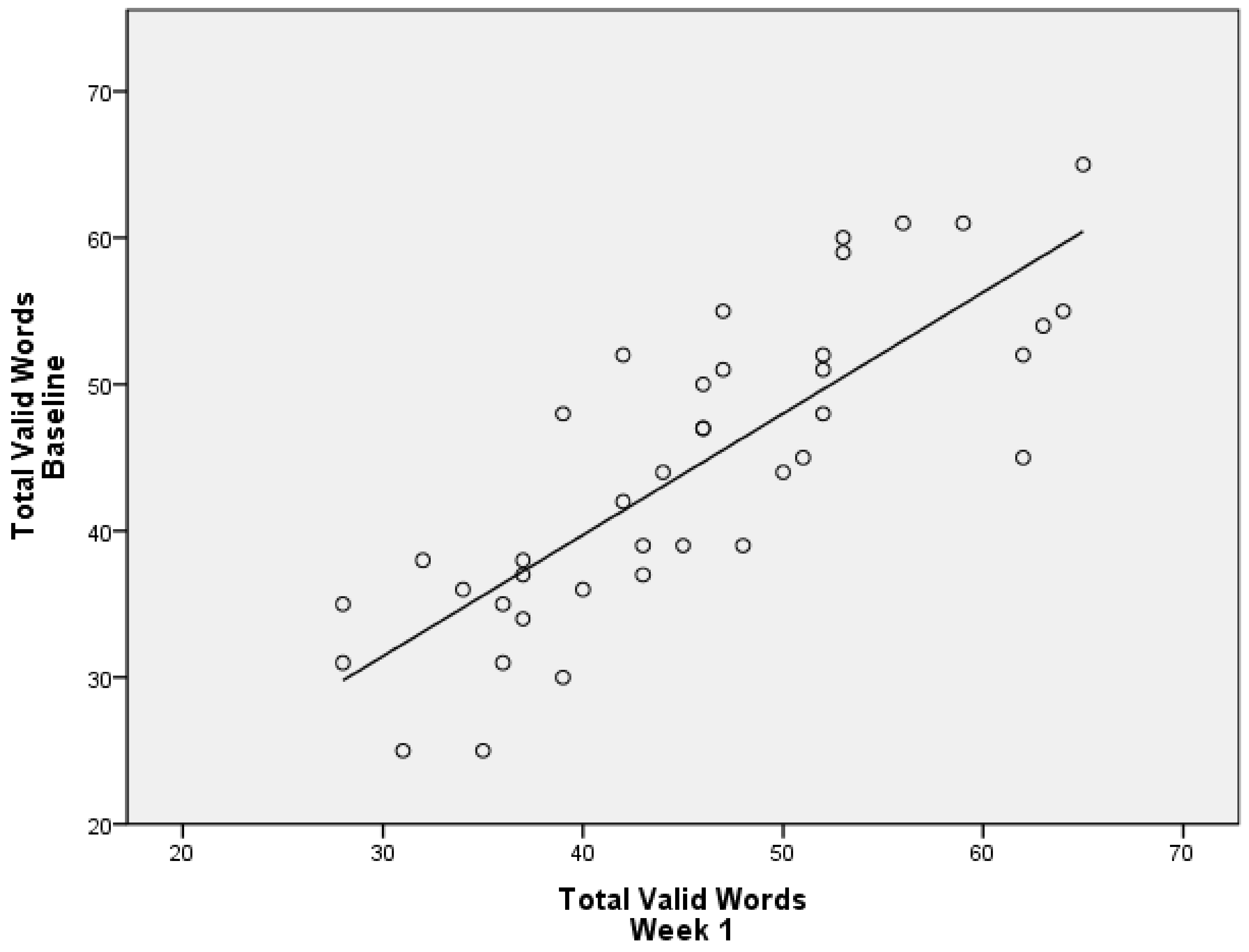
AnotB Test. There was no significant difference across sites, but average reaction time was faster in the second session (2734ms, SD=1215) than the baseline session (3228ms, SD=2121, p<.001). Overall test-retest reliability was excellent (r = 0.90).

Flanker effects. As hypothesized, healthy controls were significantly slower and less accurate for incongruent relative to congruent trials (reaction time: P < .001; accuracy: P < .001). There were no differences across sites or sessions. The overall test-retest reliability was high (RT: r = 0.84; accuracy: r = 0.75).

Probabilistic Reward Task. As hypothesized, healthy participants showed a preference for the more frequently rewarded stimulus. There were no differences across sites or sessions. Contrary to our hypotheses and unlike prior studies, the test-retest reliability for response bias was poor (r = 0.27).

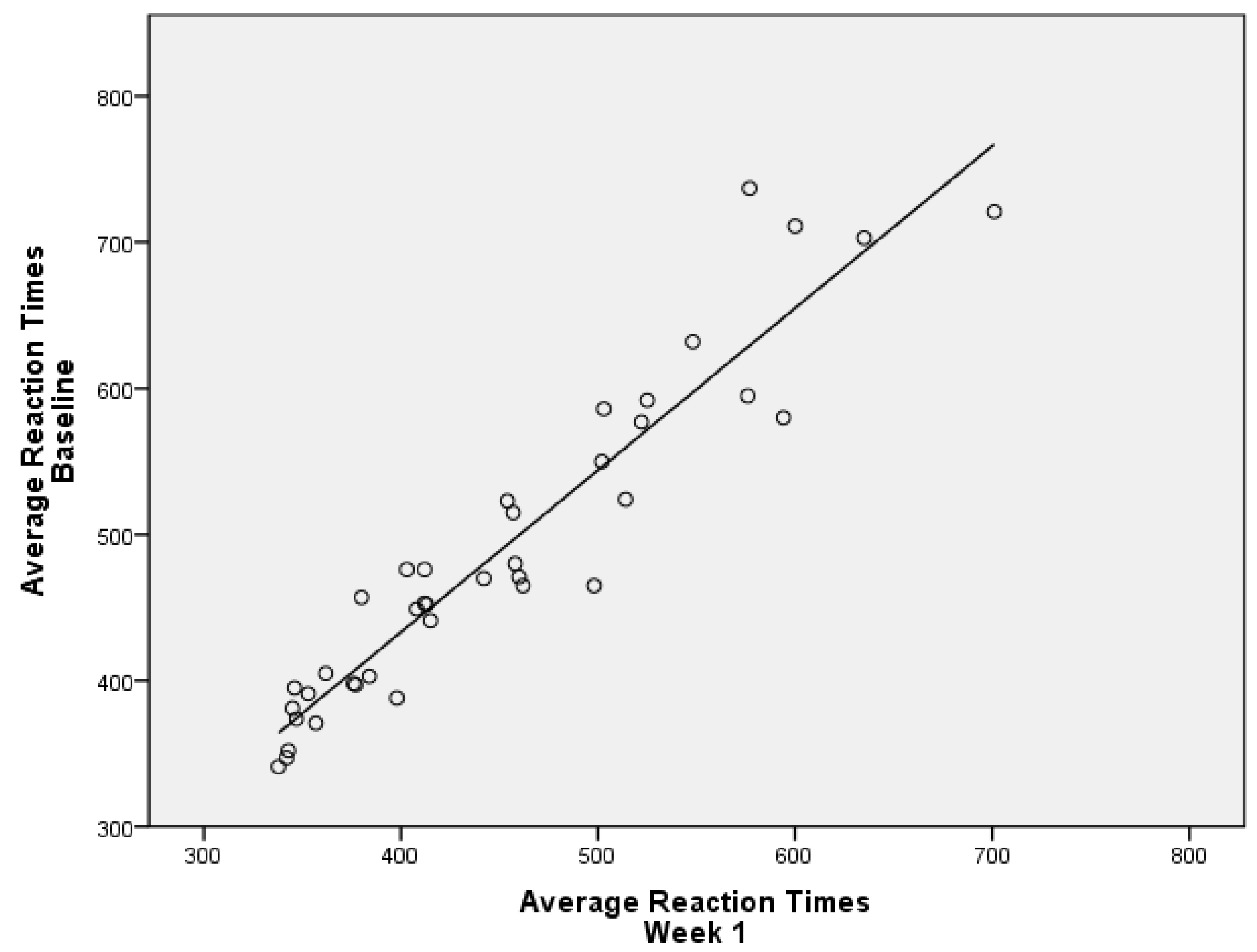
Discussion: These findings demonstrate that neurocognitive measures previously found to predict antidepressant response in depression can be measured with high reliability in a multi-site study. The only exception was the probabilistic reward task, which showed poor test-retest reliability in healthy 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.

Word Fluency



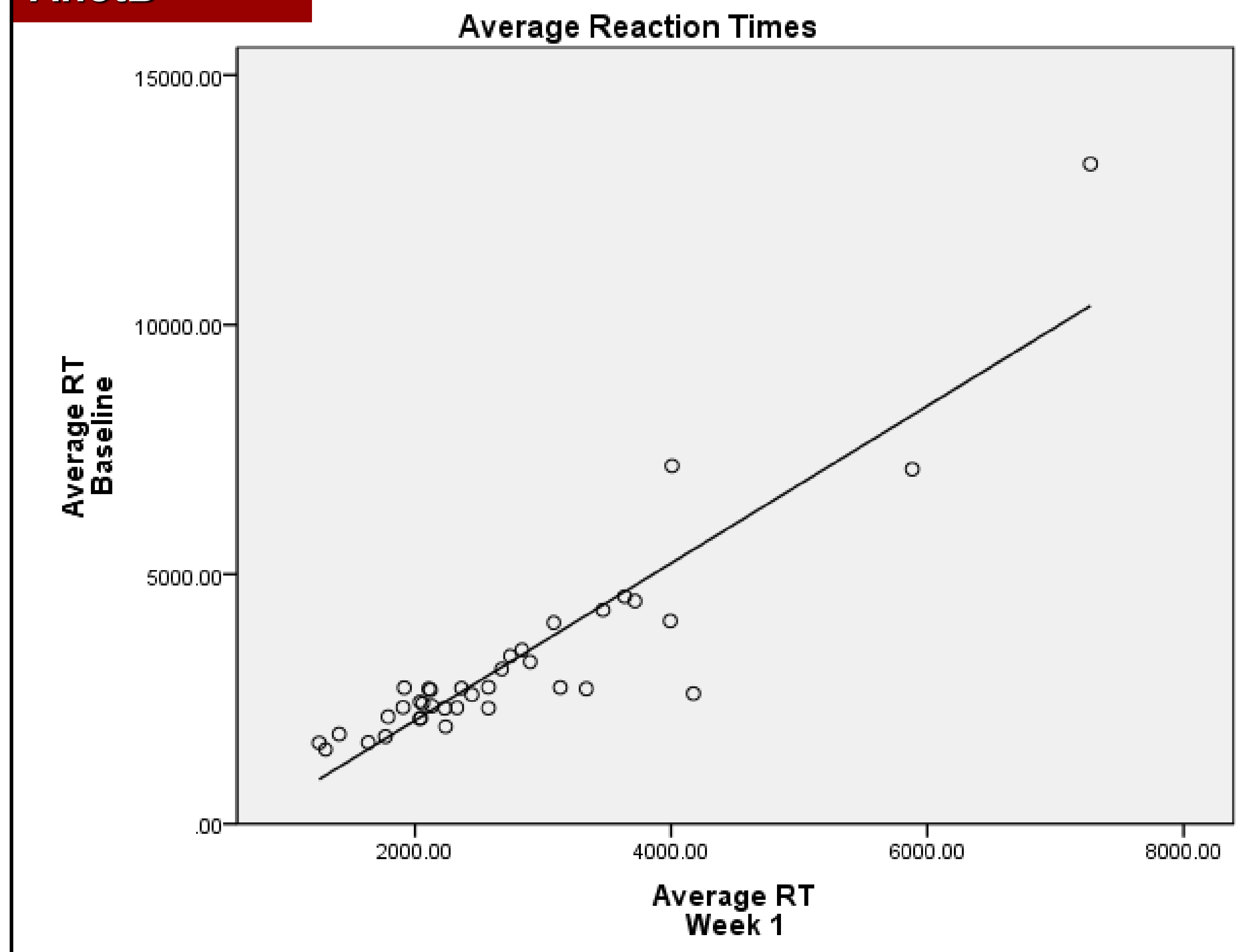
Performance on the FAS word fluency test was chosen as a predictor of SSRI response based on prior findings of Taylor et al. (2006). The average number of valid words reported was 44.3 (SD=10.3) at the baseline and 45.6 (SD=10.1) at week 1, which matches normative data for the FAS verbal fluency test. There was no significant difference in these scores across sessions or sites. The overall test-retest reliability for the 40 controls was high (r=.81). The figure above shows the scattergram giving the test-retest data the individual controls at the four sites.

Choice Reaction Time

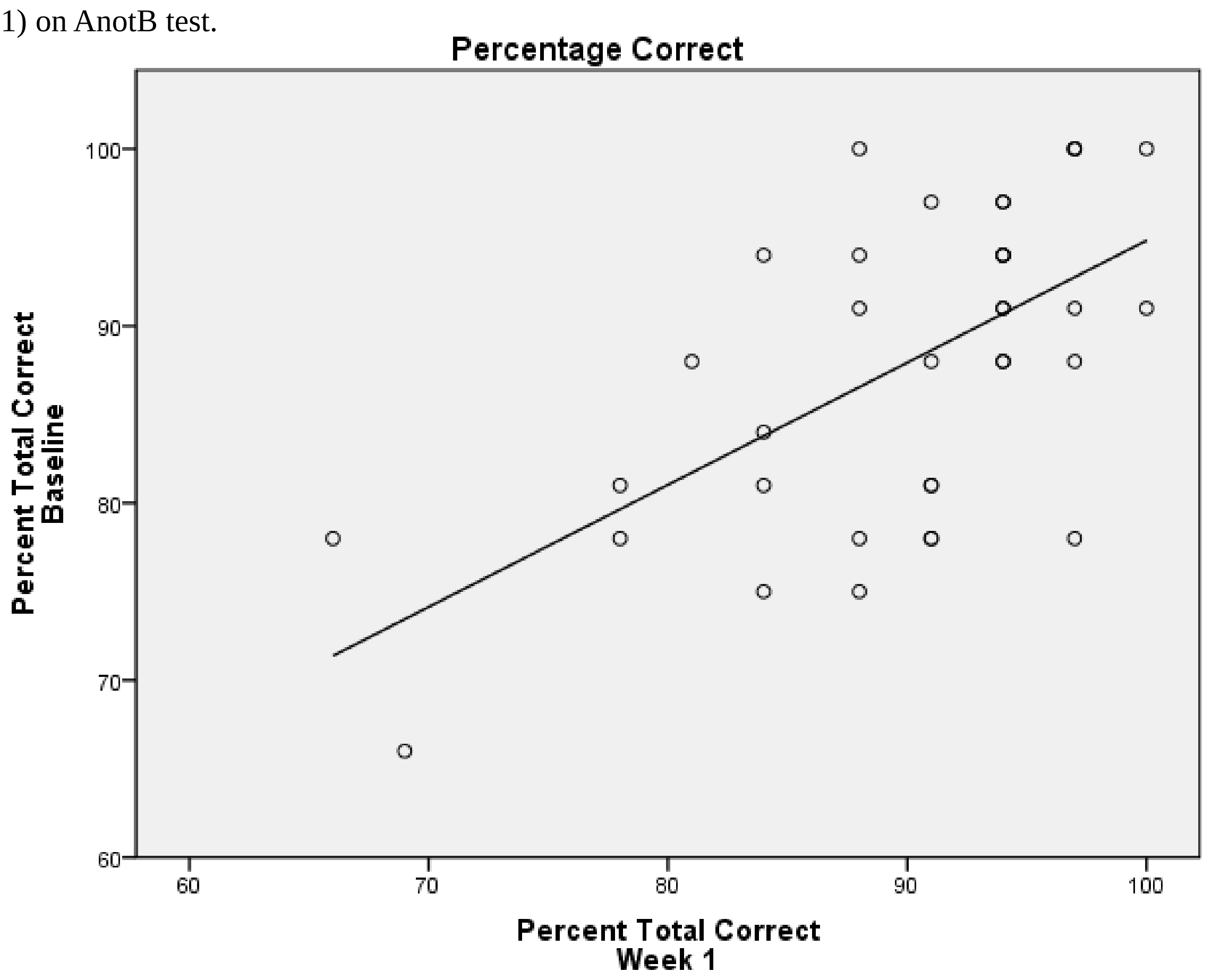


A four-choice reaction time test was used based on the hypothesis that cognitive slowing is predictive of non-response to an SSRI (Taylor et al. 2006). There was no significant difference across sites in the 4-choice reaction time. As might be expected, average reaction time was faster in the second session (week 1; M=449ms, SD=93) than in the first session (Baseline; M=488ms, SD=109). Overall test-retest reliability was excellent (r=.95), with high reliability at the four sites.

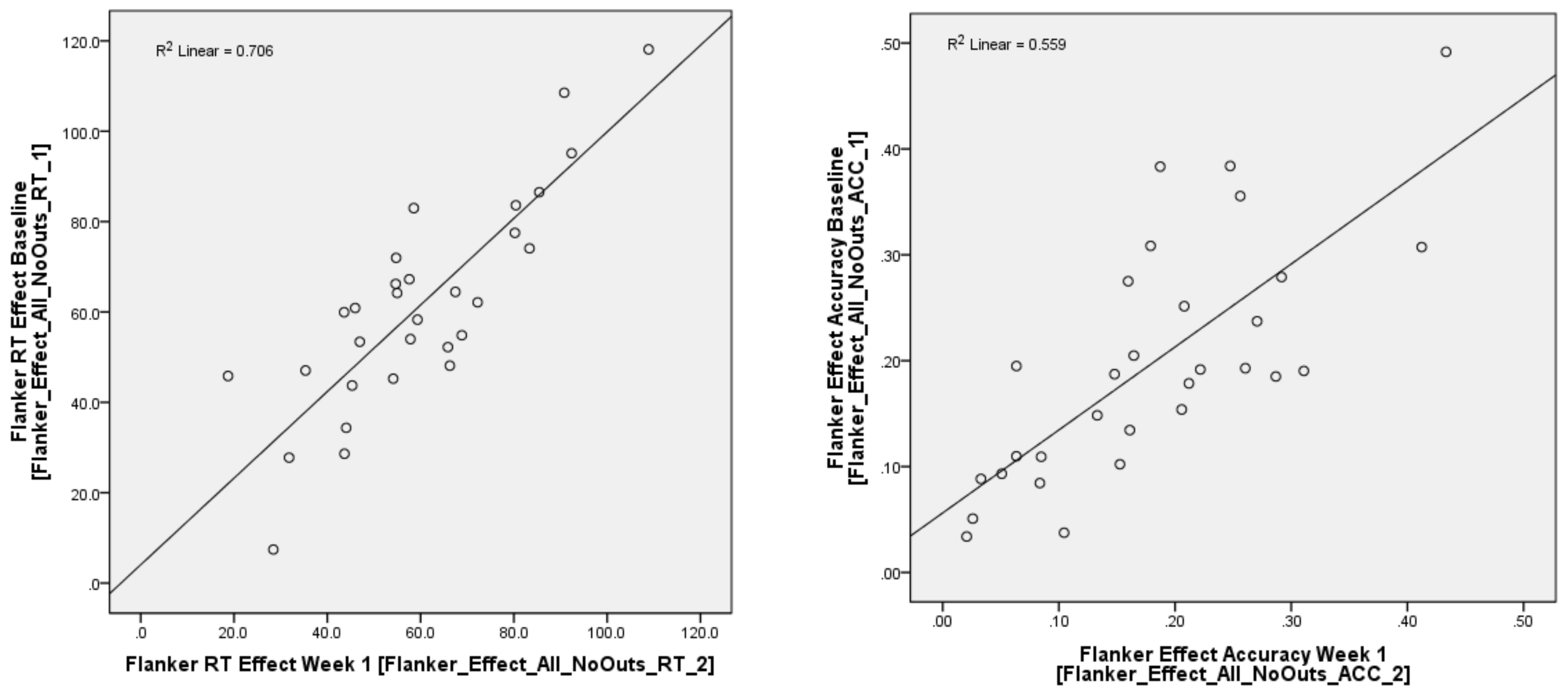
AnotB



Speed of reasoning and working memory on the AnotB test was chosen as a main predictor based on prior findings predicting response to SSRI antidepressants (Gorlyn et al. 2008). There was no significant difference in AnotB reaction time across sites, but average reaction time was again faster in the second session (week 1; M=2743ms, SD=1215) than the first session (Baseline; M=3228, SD=2121). Overall test-retest reliability was excellent for reaction time (r=.90), but was somewhat lower for accuracy scores (r=.61) on AnotB test.



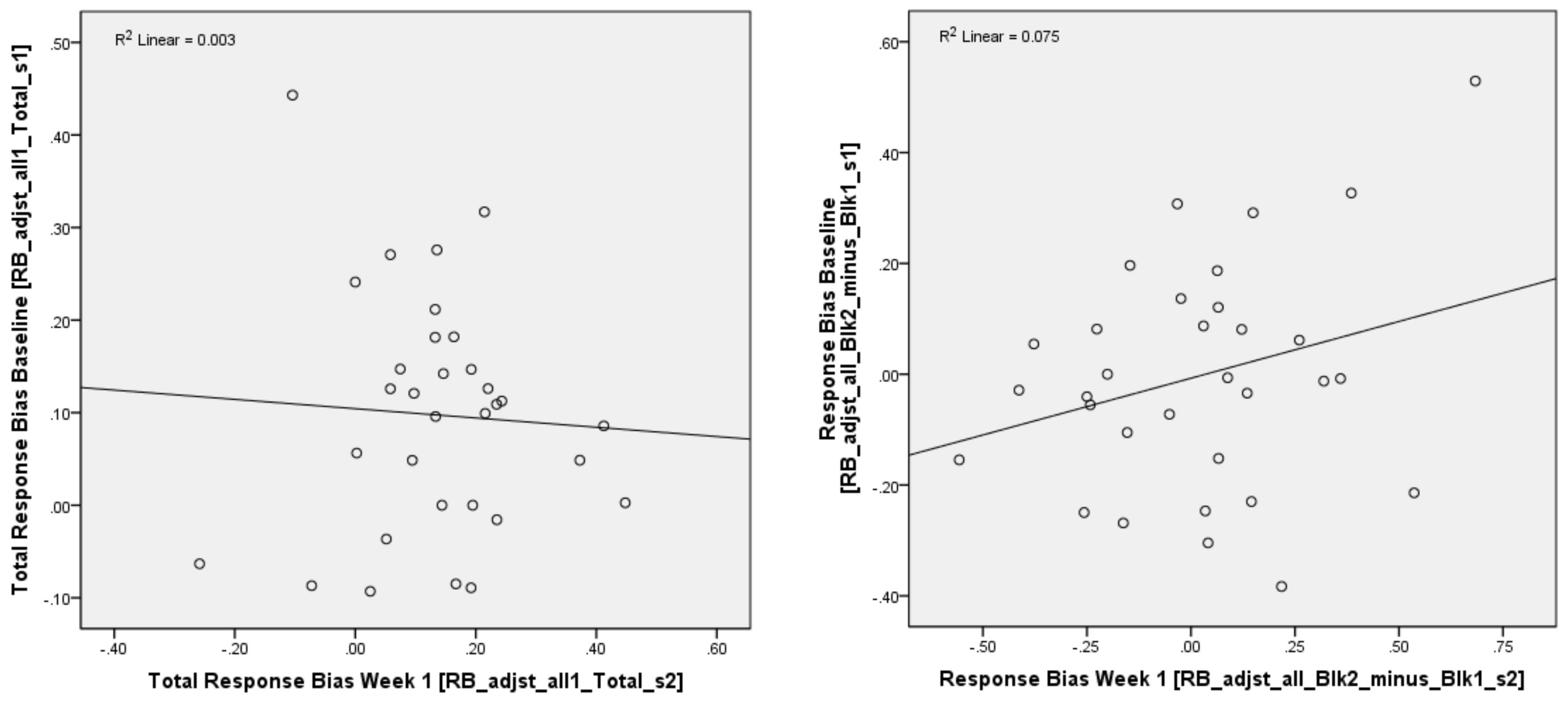
Flanker



Baseline Session
Flanker effects (n = 36). As hypothesized, healthy controls were significantly slower and less accurate for incongruent relative to congruent trials (reaction time: mean±SD = 418.76±53.65 vs. 361.59±42.95 ms, t(35) = 14.41, P < .001; accuracy: 0.76±0.13 vs. 0.99±0.01, t(35) = -10.65, P < .001). Thus, the Flanker effect was robust and highly significant for both RT and accuracy scores.

Test-Retest
Flanker effects (n = 30). Pearson correlations assessing test-retest reliability were highly significant for both RT (r = 0.84, p < .001) and accuracy (r = 0.75, p < .001)

Probabilistic Reward



Test-retest correlation for Total Response Bias and Reward Learning [DRB = Response Bias (Block 2) – Response Bias (Block 1)] among Healthy Reliability subjects (n = 32)

Baseline Week:
Response Bias (n = 35). Healthy subjects showed a response bias toward the more frequently rewarded stimulus throughout the task: in both Block 1 (mean±SD, 0.098±0.148) and Block 2 (0.108±0.172), their response bias was significantly higher than zero (both ts(34)=3.70, ps < 0.001). This indicates that healthy participants quickly learned to modulate their behavior based on reward feedback, and sustained this modulation throughout the task.

Test-Retest:
Test-retest reliability of the total response bias across the two blocks was not significant for healthy control subjects (r= -0.06, p > 0.76, n = 32) (left panel). Similarly, reward learning scores [operationalized as DRB = Response Bias (Block 2) – Response Bias (Block 1)] at the Baseline and Week 1 session were not significantly correlated (r = 0.27, p = 0.13) (right panel). This is contrary to prior findings of a significant correlation (r=.57, p<.004) of total response bias across two sessions separated by 39 days (Santesso et al. 2008).

Discussion

These findings demonstrate that neurocognive 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

Gorlyn M, Keilp JF, Grunebaum MF, Taylor BP, Oquendo MA, Bruder GE, Stewart JW, Zalsman G, Mann JJ. (2008). Neuropsychological characteristics as predictors of SSRI treatment response in depressed subjects. J Neural Transm, 115(8): 1213-1219.

Santesso, DL., Dillon DG., Birk JL., Holmes AJ., Goetz E., Bogdan R., Pizzagalli DA., (2008). Individual differences in reinforcement learning; behavioral, electrophysiological, and neuroimaging correlates. Neuroimage, 42: 807-816.

Taylor BP, Bruder GE, Stewart JW, McGrath PJ, Halperin J, Ehrlichman H, Quitkin, FM (2006). Psychomotor slowing as a predictor of fluoxetine nonreponse in depressed outpatients. Am J Psychiatry 163(1):73-78.