**Supplement**

**Detailed description of neurocognitive test battery**

*Cogstate*

Cogstate1 is a computerized testing software package that offers a range of 14 semi-automated assessment modules for individuals aged 4-90 years. All tasks were developed as computerized adaptations of traditional neuropsychological measures. For this study, we use five tasks in children age six and above (Detection, Identification, Continuous Paired Associate Learning, One-Card Learning, and One-Back) assessing visual learning/memory, working memory, attention, and reaction time. The entire battery takes approximately 10-15 minutes for the 4-5 year old participants to complete, and 15 to 20 minutes for those ages six and above. The primary performance outcome variables were visual learning/memory (One-Card Learning) and working memory (One-Back).

Cogstate was developed specifically for clinical trials as a sensitive assessment tool that can be administered more frequently than traditional neuropsychological measures and allows for documentation of changes in performance over time, because of a brief practice session, alternative forms, and the ability to collect more data over multiple tests, reducing susceptibility to skewed distributions and floor/ceiling effects2-3. Reliability (intra-class correlation) is 0.77 with good stability and negligible practice effects when testing intervals are greater than one week4. For healthy individuals, stability of performance is robust - with test-retest scores differing by approximately 2% on average (traditional measures differ 7-19%)5. Age-based normative data is available to quantify an individual’s performance at a single point in time, as well as to classify repeat performance as declined, stable, or improved.

*Behavior Rating Inventory of Executive Function (BRIEF)*

The BRIEF6 is a widely used behavior rating assessment of executive functioning in everyday life for children aged 3-17 years and adults aged 18-90. The parent-report version consists of 86 items, which map onto 2 broad areas: Behavioral Regulation Index (BRI) and Metacognition Index (MCI), as well as two validity scales. The BRI is further divided into three clinical subscales: Inhibit, Shift, and Emotional Control, whereas the MCI is divided into five clinical subscales: Initiate, Working Memory, Plan/Organize, Organization of Materials, and Monitor. The adult self-report is composed of 75 items, which also map onto the two indices of BRI and MCI and three validity scales. The two indices are further divided into the same clinical subscales as the parent-report for children, with the exception of the monitor domain being further subdivided into Self-Monitor and Task-Monitor. Parents of children and adult participants are asked to consider the frequency with which each item has been a problem over the last six months, responding on a 3-point Likert scale consisting of “never,” “sometimes,” and “often.” For this study, respondents were instructed to report on observations in the period of time from the previous evaluation rather than over the past six months (except for the pre-treatment assessment), as some of the follow-up assessments on the treatment trials occurred within three months of the previous assessment. Items were developed to be ecologically-valid behavioral correlates to presumed neurocognitive difficulties with executive functioning; thus, this measure was selected to provide parent and patient-reported outcomes of problems related to attention, memory, and executive function that occur in everyday life. Psychometric properties of this measure are strong using normative samples weighted to match ethnic and gender proportions in the US population. Internal consistency is high (alpha range = 0.73 to 0.90) and test-retest reliability exceeds 0.80 for both measures over intervals from 2 to 4 weeks.6-8  Scores are linear transformations of raw scores into T scores (mean = 50, SD = 10) for each of the subscales, indices, and total score; higher scores indicate greater difficulties. Normative data allows for the computation of age-adjusted scores at any single point in time, and test-retest data provides the information necessary to quantify whether changes in ratings have occurred (declines, no change, or improvements). The primary BRIEF outcome variables were MCI and BRI.

**Detailed Description of Reliable Change Methodology**

We analyzed change in outcomes over time on individual levels, using Reliable Change Index (RCI) methodology.9 RCI uses test characteristics of practice effects, score variability, and test-retest reliability to understand expected changes in test scores over time. In the absence of practice effects, the expected change score (T2 score minus T1 score) is zero, but imperfect test-retest reliability means that individuals would not be expected to obtain the exact same score twice. The standard error of measurement (SEM) of a test can be calculated using the test’s standard deviation and its test-retest reliability, and from that, the standard error of the difference (Sdiff) can be calculated to estimate the amount of variability to expect around the change score. The Sdiff is then multiplied by a specified critical value (zcrit) to construct a confidence interval (CI) around the change score. A common application is to use a zcrit of 1.64 to construct an 80% confidence interval.10 The series of formulas is as follows, where subscripts 1 and 2 refer to two testing occasions:

$$SEM=SD\*\sqrt{1-r\_{12}}$$

$SE\_{diff}$ = $\sqrt{SEM\_{1}^{}^{2}+SEM\_{2}^{}^{2}}$

$CI\_{RCI}=\pm z\_{crit}\*SE\_{diff}$, where

$z\_{crit}=$ cutoff for desired CI range ($z\_{crit} $= 1.96 for 90%, 1.64 for 80%, *etc*.),

Once the CI cutoff is established, an individual’s change score (T2 – T1) can be compared to it, and if the change score exceeds the CI cutoff, clinically significant change has occurred. In any normative sample (i.e., where no systematic changes in scores have occurred), a 90% confidence interval would identify 5% of the sample as having declined, 90% as stable, and 5% as improved. In a clinical sample, deviations from this distribution could be interpreted as clinically meaningful change having occurred in a larger or smaller proportion than expected.

For each of the four outcomes, RCIs were computed from the specific test’s normative data, and two-tailed 90% CIs were constructed. We chose two-tailed because at this stage of research with MEK inhibitors it is equally important to identify either detrimental or positive effects on functioning. For each participant, change scores were calculated between the baseline and the 6-month and baseline to 12-month follow-ups. Each change score was compared to the CI and participants were classified as “declined,” “stable,” or “improved” on each outcome for each pair of time points. We then used Chi-square to compare the frequencies of classification between the normative (5% / 90% / 5%) and clinical groups.11 This served as an evaluation of the effect of the MEK inhibitor in changing cognitive performance scores and ratings. That is, it answers the question: Does working memory or executive functioning decline or improve in a greater proportion of individuals who are given a MEK inhibitor, compared to the amount of change expected in the general population due to the passage of time?

**References**

1. Maruff P, Thomas E, Cysique L, et al: Validity of the Cogstate brief battery: Relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. Archives of Clinical Neuropsychology 24:165-178,2009
2. Ballard JC. Computerized assessment of sustained attention: Interactive effects of task demand, noise, and anxiety. Journal of Clin Exp Neuropsych. 1996; 18:864-882.
3. Betts J, McKay J, Maruff P, et al. The development of sustained attention in children: The effect of age and task load. Child Neuropsych. 2006;12:205-221.
4. Falleti MG, Maruff P, Collie A, et al. Practice effects associated with the repeated assessment of cognitive function using the Cogstate battery at 10-minute, one week, and one month test-retest intervals. J Clin Exp Neuropsych. 2006;28:1095-1112.
5. Makdissi M, Collie A, Maruff P, et al. Compterised cognitive assessment of concussed Australian Rules footballers. Brit J Sports Med. 2001;35:354-360.
6. Gioia G, Isquith P, Guy S, et al. Test Review: Brief rating inventory of executive function. Child Neuropsych. 2000;6:235-238.
7. Gioia G, Isquith P. Ecological assessment of executive function in traumatic brain injury. Dev Neuropsych. 2004;25:135-158.
8. Roth RM, Gioia GA, Isquith PK. Behavior rating inventory of executive function – adult version. Lutz, FL, Psychological Assessment Resources, Inc, 2005.
9. Chelune, G. J. Assessing reliable neuropsychological change. In Prediction in forensic and neuropsychology: Sound statistical practices (pp. 123–147). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2003.
10. Iverson, G. L. Interpreting change on repeated neuropsychological assessments of children. In E. M. S. Sherman & B. L. Brooks (Eds.), Pediatric Forensic Neuropsychology (pp. 89–112). Oxford University Press, 2012.
11. Woods, S. P., Childers, M., Ellis, R. J., Guaman, S., Grant, I., & Heaton, R. K. A battery approach for measuring neuropsychological change. Arch Clin Neuropsych. 2006;21:83–89.