Evaluating cognitive effort in a randomized controlled trial

TRAVIS H. TURNER,^{1,2,3} JENNA B. RENFROE,³ KRISTEN MORELLA,⁴ BERNADETTE P. MARRIOTT^{2,5}
& THE BRAVO WORKING GROUP*

1 Ralph H. Johnson VAMC, Mental Health Service, Charleston, SC USA

2 Department of Psychiatry and Behavioral Sciences, Charleston, SC USA

3 Department of Neurology/Neurosciences, Movement Disorders Program, Medical University of South Carolina, Charleston, SC USA

4 Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC USA

5 Nutrition Section, Division of Gastroenterology and Hepatology, Department of Medicine, Medical University of South Carolina, Charleston, SC USA

Key words

clinical trials, malingering, effort, neurocognition, suicide

Correspondence

Travis H. Turner, PhD, RHJ
VAMC (MH 116), 109 Bee Street,
Charleston, SC 29401, USA.
Telephone (+1) 843-577-5011,
ext. 5192
Email: turnertr@musc.edu

Received 6 November 2015;
accepted 8 February 2016;
revised 20 January 2016

Abstract

Many randomized controlled trials (RCTs) of neuropsychiatric conditions involve cognitive outcome measures; however, validity of cognitive data relies on adequate effort during testing, and such screening is seldom performed. Given well-established rates of 10 to 30% poor effort in clinical settings, this is not a trivial concern. This preliminary study evaluated effort during cognitive testing in an RCT of omega-3 supplementation to reduce suicidality in a high-risk psychiatric population. An interim analysis of sustained attentions measures from the Connors Performance Test (CPT-2) at baseline for the first 60 participants was conducted. Previously validated cut points to detect insufficient effort on the CPT-2 were applied. At baseline, 12% (7) were identified as giving poor effort. Follow-up analyses indicated less psychiatric distress and suicidality among those who gave poor effort. Results suggest comparable likelihood of a poor effort on cognitive testing in clinical and RCT participation. Reduced psychiatric distress in the poor effort group raises concern regarding interpretation of other measures. The importance of screening cognitive data for effort in RCTs is highlighted. Future studies will examine effort at follow-up visits, and explore relationships to attrition, adherence, and response to treatment. Copyright © 2016 John Wiley & Sons, Ltd.

*The BRAVO Group: Ron Acierno, Bashar W. Badran, Andrea Boan, Alice Bova, Jeffrey J. Borckardt, Christopher DeLeon, Mark DeSantis, Christopher J. Fink, Ashley Galloway, Mark S. George, Courtney Harrington, Mark Hamner, Joseph R. Hibbeln, Kelly Holes-Lewis, Thomas C. Hulsey, Therese K. Killeen, Fu-Shing Lee, Kathryn M. Magruder, Robert Malcolm, Alicia Marzolf, Donald Myrick, Ann Z. Poole, Marcie Pregulman, Matthew J. Roden, Ariane C. Shokri, Bryan K. Tolliver.

Introduction

A major challenge in designing randomized controlled trials (RCTs) with cognitive outcomes is selecting appropriate measures. Beyond issues such as practice effects and measurement error, collection of valid data

requires that participants put forth optimal effort during testing. This is not a trivial issue, as 10–30% of adults demonstrate insufficient effort (i.e. symptom exaggeration and/or malingering) on cognitive testing in clinical examinations to direct individual treatment (Mittenberg *et al.*, 2002); a survey of 33,531 cases by the American Board of Clinical Neuropsychology estimated a rate of 50% in medical-legal settings, where there is a financial incentive (e.g. settlement or benefits) for examinees to perform poorly (Larrabee *et al.*, 2009). Thus, it cannot be assumed that participants in a RCT will give optimal effort on cognitive tests that determine effectiveness of the intervention. Nevertheless, while the majority of practicing neuropsychologists routinely administer effort tests as part of the assessment battery (Sharland & Gfeller 2007), test results are rarely screened for effort and validity in RCTs.

To date, the only published RCT reporting assessment of effort was an investigation of cognitive rehabilitation for patients with chronic solvent-induced toxic encephalopathy by van Hout *et al.* (2008). Effort was assessed at baseline using the Test of Memory Malingering (Tombaugh, 1996) and the Warrington Recognition Memory Test for Faces (Warrington, 1984). In a sample of 109 patients, 24% gave poor effort, and effort was found to be a more important predictor of treatment success than litigation status (i.e. actively pursuing financial compensation settlement). Even more compelling, not a single participant who gave poor effort at baseline improved at follow-up, compared with improvements of 32% and 13% in the treatment and control groups, respectively.

The Conners' Continuous Performance Test (Conners, 2004: CPT-2) is a computerized test of sustained attention, vigilance, and impulsivity that is widely used for individual assessments and clinical trials. A series of letters is presented, and the examinee is instructed to press the space bar as quickly as possible for every letter except "X." The task takes 14 minutes to complete, and is monotonous by design. Primary outcome measures include errors of omission, commission, perseveration, and variability in reaction time. Two recent investigations examined whether grossly abnormal and implausibly poor CPT-2 performance (e.g. greater than four standard deviations [SDs] below mean from healthy adult normative samples on primary outcome measures) could be used to discriminate poor effort in clinical populations. Ord *et al.* (2010) examined performance in traumatic brain injury (TBI) groups who passed or failed formal malingering tests. Combined errors of omission and reaction time variability provided modest sensitivity (0.44) and excellent specificity (0.97). A similar study with mild TBI was performed by Lange *et al.* (2013). Cutoffs

previously identified by Ord *et al.* (2010) resulted in unacceptably low sensitivity (0.14). A combination of omission, commission, and perseverative errors yielded sensitivity of 0.41, with perfect specificity (1.0).

The present investigation examined participant effort based on CPT-2 performance in an intervention study with cognitive and psychological outcomes. Specifically, we examined baseline data from an initial cohort of participants in a large, ongoing RCT of omega-3 supplementation to reduce suicide risk and improve psychiatric functioning (BRAVO Study, Clinicaltrials.gov: NCT01901887).

Methods

Study design

A preliminary analysis of sustained attention measurements on the Connors' CPT-2 (Conners, 2004) was conducted at baseline for the first 60 of 300 planned participants in the RCT, regardless of intervention assignment. Given that participants had very recently or were currently experiencing severe suicidal, psychiatric symptomatology, cutoffs based on Lange *et al.*'s (2013) study with the mild traumatic brain injury (mTBI) sample were used. Classification of "poor effort" was therefore identified by any of the following: Omissions Errors ≥ 39 , Commissions Errors ≥ 32 , Perseverative Errors ≥ 8 .

Participants

Participants were adults ages 18–90 years recruited from the community who met the following eligibility criteria that defined being at risk of suicide either: (a) a suicide attempt in the last six months, or (b) a suicide attempt during the adult lifetime AND current diagnosis of an episode of depression as diagnosed on the Mini International Neuropsychiatric Interview (MINI), or (c) an inpatient admission with suicide risk in the last six months, or (d) an inpatient admission with suicide risk during the adult lifetime AND current diagnosis of an episode of depression as diagnosed on the MINI, or (e) positive suicidal behavior or ideation based on a psychiatrist-administered Columbia Suicide Severity Rating Scale (C-SSRS) and psychiatrist review of participant medical history and physical, or (f) a score of zero or greater on the Implicit Associations Test – Suicide (IAT-S: Nock *et al.*, 2010), or (g) greater than or equal to a value of nine on the Beck Hopelessness Scale (BHS) and psychiatrist review of participant medical history and physical. All participants had adequate housing to allow for shipments of juice-based omega-3 HUFA supplements

to be delivered and refrigerated. Study personnel confirmed that all participants were actively being followed by a mental health professional. Exclusion criteria included diabetes (due to caloric content of juice-based supplement), eating disorder, morbid obesity, progressive neurological disease, unstable medical condition, and history of significant behavioral instability. Aside from Major Neurocognitive Disorder and delirium, participants were not excluded for any other DSM-5 psychiatric conditions (American Psychiatric Association, 2013). While acute alcohol or other substance withdrawal was an exclusion criterion, many of the individuals recruited into this study had current alcohol and/or other substance use disorders. All participants underwent informed consent procedures in accordance with the local Institutional Review Board.

Sample

The sample of 60 adults was primarily composed of men ($n = 47$, 78.3%). Twenty-one (21) participants self-identified as White/Caucasian, 32 self-identified as Black/African-American, and six self-identified as "Mixed" (one chose not to disclose). The average age in this sample was 49.1 years ($SD = 12.9$, [18, 65]), with mean education of 13.8 years ($SD = 1.8$, [12, 20]).

Study protocol

Participants completed a comprehensive baseline evaluation, totaling about six hours. In addition to structured psychiatric evaluation, formal interview for diet, alcohol, sleep, and pain symptoms, participants completed self-report measures of psychiatric functioning including the Beck Depression Scale (BDI-2; Beck *et al.*, 1996) and the Profiles of Mood State, Bipolar Form (POMS-Bi; Lorr *et al.*, 1982). Participants also completed an experimental measure of automatic processing for suicide-related words, the Suicide Stroop (Cha *et al.*, 2010). This task presents a series of 48 words categorized as neutral, positive, depressive, or suicidal in content (12 each); the task is to identify the color of the ink as quickly as possible. Reaction times are only calculated for words correctly identified by the participant, and a data screening procedure is employed to remove responses from analysis with significantly extended or reduced reaction times. Increased reaction time for identifying colors of words with suicidal content is interpreted as reflecting increased suicidal ideation and risk. Participants were randomly assigned to treatment group (placebo or control), and followed over a six-month study period, with visits at one, three, five and six month after baseline. The CPT-2,

BDI-2, POMS-Bi, and Suicide Stroop were administered at baseline, three month and six month visits.

Results

Upon application of the criteria proposed by Lange *et al.* (2013), 53 participants (88%) were found to provide consistent effort (good effort group: GEG) while seven participants (12%) were identified as providing poor effort (poor effort group: PEG) at baseline. Number of perseverative errors was the most sensitive indicator for identifying the PEG (Figure 1). While several participants also violated criteria for Omission and Commission Errors, these cutoffs did not identify any unique cases (Figures 2 and 3).

A follow-up analysis was performed to examine whether self-reported psychiatric distress differed by effort group. Though sample sizes precluded formal null-hypothesis statistical testing, 95% confidence intervals (CIs) around group means indicated less subjective distress in the PEG. With respect to depressive mood symptoms, mean total score on the BDI-2 was lower for PEG, $x = 17.1$, 95% CI: [7.9, 26.4], versus GEG, $x = 21.9$, 95% CI: [18.9, 25.0]. Responses to the POMS-Bi were similar. As shown in Figure 4, the PEG consistently reported better functioning across all subscales. Aside from the subscale for Agreeable/Hostile disposition, the 95% CIs based on the GEG did not include the mean for the PEG. Non-overlapping 95% CIs were observed for the Elated/Depressed and Energetic/Tired subscales,

Figure 1. Box plot showing raw number of perseverative errors. The cutoff is > 7 . All extreme cases except for #32 and #46 (errors = 7) were identified as having exerted insufficient effort.

Figure 2. Box plot showing raw number of omission errors. The cutoff is > 20 .

suggesting meaningful group differences in symptom reporting between the two groups.

A final analysis examined whether effort groups differed on the Suicide Stroop (Cha *et al.*, 2010). Average rate of correct responses were generally similar across groups, with GEG mean = 45.3, 95% CI: [45.0, 45.7], and PEG mean = 44.4, 95% CI: [41.5, 47.4]. However, as shown in Figure 5, participants in the PEG showed reduced reaction times to all categories of word stimuli, including those with suicide-related content; this finding suggests reduced suicidal ideation and risk.

Figure 3. Boxplot showing raw number of commission errors. The cutoff is > 31 .

Discussion

Valid data on cognitive measures is predicated on examinees providing good effort during testing. This preliminary study examined whether adults enrolled in a RCT to address suicidality and associated psychiatric disturbances provided good effort on cognitive testing administered during the study protocol. In this interim analysis, 12% (seven of 60) were identified as having *not* provided good effort. Those providing poor effort endorsed less distress on self-report and other measures of distress and suicidality.

Our findings are consistent with estimates of 10 to 30% poor effort rate for non-litigating patients evaluated in clinical settings (Mittenberg *et al.*, 2002). Within this context, better results could not be realistically expected, as this onerous test was administered within the context of a very lengthy and intensive baseline examination. However, detection of possible effect of intervention on cognition is also contingent on valid data at three and six month follow-up visits. At present, sample sizes at these time points are inadequate to perform such analyses. Thus, rates of adequate effort across all portions of the RCT, and whether effort at baseline predicts effort at follow-up, remain unknown.

While 88% positive effort with this particular population is encouraging, anything less than 100% effort on measures of cognitive performance is concerning since poor effort can pose challenges in this or any other RCT. The risk of misinterpretation of RCT outcome analyses due to Type II errors would be most likely when effect sizes are small, or if poor effort is imbalanced across conditions. There is also inflated risk of Type I error if a disproportionately high number of participants assigned to the treatment arm give poor effort at baseline and adequate effort at follow-up (falsely suggesting improvement); a similar outcome would occur if a disproportionate number of participants in the placebo arm provided adequate effort at baseline and poor effort at follow-up (suggesting decline).

In clinical and legal settings, invalid effort on cognitive testing is most commonly associated with exaggeration and implausible symptom reporting on forensically-derived measures (Mittenberg *et al.*, 2002). In contrast, those who gave poor effort in this study tended to report less psychiatric distress as measured by the BDI-2 (Beck *et al.*, 1996) and POMS-Bi (Lorr *et al.*, 1982). Exploratory analysis of reaction times from the Suicide Stroop (Cha *et al.*, 2010), which reflects an attentional bias for suicide and depressive words, also suggested that this group was less distressed. Taken together, this pattern is inconsistent

Figure 4. Comparison of subscales from the Profiles of Mood State, Bipolar Form (POMS-Bi) questionnaire with 95% confidence intervals for group means. The *t*-scores are referenced to adults followed in outpatient psychiatric settings (mean = 50, SD = 10).

with conceptualization of poor effort on cognitive tests as representative of severe neuropsychiatric dysfunction or a “cry for help.” The most plausible explanation is that these PEG participants simply did not provide adequate effort on cognitive testing, and performance on the CPT-2 is not valid for interpretation.

When patients are determined to have provided inadequate effort on formal indices in clinical evaluations, interpretation of poor performances on other tests is made with great caution, and diagnoses based on the examination are often considered guarded or provisional. Such practice might also be considered within an RCT. In the present study going forward, an argument could be made for censoring or weighting other data from the PEG participants. One might also cross-validate if poor

effort on the CPT-2 is associated with adherence to protocol and attrition. Within this study, blood assays are able to identify level adherence to both treatment (omega-3) and placebo (macadamia nut oil). Thus, when the blind is removed it will be possible to determine whether effort on cognitive testing at baseline predicts adherence or response to treatment, as in the study by van Hout *et al.* (2008).

In addition to lack of data at follow-up, this preliminary study has several limitations. Our current sample size (60/300) precludes detailed investigation of demographic, psychiatric, psychosocial, and dietary factors associated with effort. As noted earlier, we also do not have adequate sample size at the present time to investigate effort at follow-up visits, or explore

Figure 5. Comparison of reaction times for different word stimulus categories from the Suicide Stroop (Cha *et al.*, 2010).

relationship to protocol adherence and response to treatment. Finally, the cutoffs for identifying “poor effort” from the CPT-2 measures, as well as most all other embedded and formal effort measures, strongly favor specificity over sensitivity. In the mTBI study from which the cutoffs were derived, sensitivity was 0.41 with specificity of 1.00; in other words, while only 41% of those who gave poor effort were identified with these cutoffs, there were no “false positives” cases (Lange *et al.*, 2013). In an RCT, placing greater weight on sensitivity over specificity gives individual participants the benefit of uncertainty, and allows investigators to retain more data points. That said, if treatment effect appears blunted by effort confounds, alternative cutoffs might become desirable.

In summary, this investigation demonstrates the importance of screening cognitive test data obtained within the course of RCTs for effortful performance. Our observation of 12% poor effort is consistent with estimates obtained in clinical settings. While the occurrence of poor effort does not preclude inclusion as a primary outcome in a RCT, *per se*, our finding of less than 100% “good effort” highlights the importance of screening data. Moreover, findings of reduced psychiatric symptomatology in those giving poor effort encourages

effort assessment in RCTs and other investigations with neuropsychiatric populations, even when cognition isn’t a primary outcome. Future studies will continue to investigate frequency of poor effort at follow-up visits, and relationship of effort to adherence, retention, and treatment outcome.

Acknowledgments

The BRAVO study is sponsored by award # W81XWH-13-2-0015 from the Department of Defense (DoD), US Army Medical Research and Materiel Command (USAMRMC), Congressionally Directed Medical Research Programs (CDMRP) through the US Army Medical Research Acquisition Authority (USAMRAA). The intramural program of the National Institute on Alcohol Abuse and Alcoholism also provides support for this study. The BRAVO study team wishes to acknowledge the excellent support of Samantha Wise in all aspects of the study.

Declaration of interest statement

All authors declare that they have no conflicts of interest nor do any of the authors have a financial interest in the results of the study.

References

- American Psychiatric Association (2013) Diagnostic and Statistical Manual of Mental Disorders, fifth edn, Arlington, VA: American Psychiatric Association.
- Beck A.T., Steer R.A., Ball R., Ranieri W. (1996) Comparison of Beck Depression Inventories – IA and – II in psychiatric outpatients. *Journal of Personality Assessment*, **67**(3), 588–597.
- Cha C.B., Najmi S., Park J.M., Finn C.T., Nock M. K. (2010) Attentional bias toward suicide-related stimuli predicts suicidal behavior. *Journal of Abnormal Psychology*, **119**(3), 616–622.
- Conners K.C. (2004) Conners Continuous Performance Test, second edn, Toronto: Multi-Health Systems, Inc..
- Lange R., Iverson G.L., Brickell T.A., Staver T., Pancholi S., Bhagwat A., French L.M. (2013) Clinical utility of the Conners’ Continuous Performance Test-II to detect poor effort in U.S. military personnel following traumatic brain injury. *Psychological Assessment*, **25**, 339–352.
- Larrabee G.L., Millis S.R., Meyers J.E. (2009) 40 Plus or minus 10, a new magical number: reply to Russell. *Clinical Neuropsychologist*, **23**, 841–849.
- Lorr M., McNair D.M., Fisher S.U. (1982) Evidence for bipolar mood states. *Journal of Personality Assessment*, **46**, 432–436.
- Mittenberg W., Patton C., Canyock E.M., Condit D.C. (2002) Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, **24**, 1094–1102.
- Nock M.K., Park J.M., Finn C.T., Deliberto T.L., Dour H.J., Banaji M.R. (2010) Measuring the suicidal mind: implicit cognition predicts suicidal behavior. *Psychological Science*, **21**, 511–517.
- Ord J.S., Boettcher A.C., Greve K.W., Bianchini K. J. (2010) Detection of malingering in mild traumatic brain injury with the Conners’ Continuous Performance Test-II. *Journal of Clinical and Experimental Neuropsychology*, **32**, 380–387.
- Sharland M.J., Gfeller J.D. (2007) A survey of neuropsychologists’ beliefs and practices with respect to the assessment of effort. *Archives in Clinical Neuropsychology*, **22**(2), 213–223.
- Tombaugh T.M. (1996) TOMM Test of Memory Malingering, Manual: Toronto, Multi Health Systems, Inc..
- Warrington E.K. (1984) Recognition Memory Test, Manual, Windsor, NFER-Nelson.
- van Hout M.S.E., Ellie M., Wekking E.M., Berg I.J., Deelman B.G. (2008) Psychosocial and cognitive rehabilitation of patients with solvent-induced chronic toxic encephalopathy: a randomised controlled study. *Psychotherapy and Psychosomatics*, **77**, 289–297.