Working Memory in Obstructive Sleep Apnea: Construct Validity and Treatment Effects

Joshua C. Felver-Gant, B.A.; Amanda S. Bruce, Ph.D.; Molly Zimmerman, Ph.D.; Lawrence H. Sweet, Ph.D.; Richard P. Millman, M.D.; Mark S. Aloia, Ph.D.

1Department of Psychiatry and Human Behavior, Brown Medical School, Providence, RI; 2Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, Brown Medical School, Providence, RI; 3Department of Medicine, National Jewish Medical and Research Center, Denver, CO

Study Objective: To better understand the effects of obstructive sleep apnea (OSA) on working memory performance. We first examined the construct validity of a working memory task (the 2-Back task) and its relationship to other tests of cognitive functioning, and then determined the effects of positive airway pressure (PAP) treatment on measures of both working memory and its related component processes.

Methods: Fifty-six participants with OSA were administered cognitive tests of working memory and related subordinate cognitive processes prior to initiation of PAP treatment and at a 3-month follow-up visit. Objective monitors were employed to measure PAP treatment adherence. Statistical analyses were conducted to examine treatment adherence and cognitive performance over time.

Results: Performance on the 2-Back task was statistically correlated with both a second working memory task (PASAT) and all subordinate cognitive measures. Participants were separated into high and low PAP adherence groups using a median split of 4 hours of PAP use per night. Repeated measures ANOVAs demonstrated that high adherers performed better across time on both tests of working memory (2-Back: F46 = 4.73, p < 0.04; PASAT: F46 = 4.92, p < 0.04) whereas low adherers performed more poorly. There were no treatment effects for any other cognitive measure.

Conclusions: The 2-Back task demonstrated adequate construct validity as a measurement of working memory in individuals with OSA. Our treatment adherence findings suggest that the construct of working memory is more sensitive to the effects OSA treatment than any of its subordinate cognitive processes.

Keywords: Positive airway pressure, obstructive sleep apnea, working memory, n-Back, PASAT, cognition

The selected component processes for the 2-Back working memory task involve declarative memory, executive functioning, and motor speed. Declarative memory was examined using the Hopkins Verbal Learning Test - Revised (HVLT-R). Alternate forms of the HVLT-R were used at each assessment to minimize test-retest effects. The HVLT-R is a commonly used clinical measure of memory, executive, and motor speed/coordination. Declarative memory was examined using the Hopkins Verbal Learning Test—Revised (HVLT-R). Alternate forms of the HVLT-R were used at each assessment to minimize test-retest effects. The HVLT-R is a commonly used clinical mea-

WORKING MEMORY TESTING

The 2-Back working memory task was the cognitive measure under investigation in this study. The 2-Back is a verbal working memory task in which series of consonants are presented visually, one every 3000 ms (2500 ms stimulus and 500 ms interstimulus interval). At both the baseline and 3-month follow-up sessions, participants completed a practice run prior to the actual trial run to ensure that the 2-Back task was sufficiently understood. Participants were asked to make a yes/no response following each consonant, determining whether it was the same as or different from the consonant presented two earlier (e.g., f, N, b, N, B, K, b, k, N, G,…, see Figure 1). Responses were made using 2 keys on the keyboard of a laptop computer. Capitalization was randomized and each consonant block contained 33% targets. Executive coordination, phonemic buffering, and subvocal phonemic rehearsal were required to successfully perform this task.

The effects of the 2-Back condition were obtained by comparing performance to the 0-Back, a control condition. The 0-Back task included 9 consonants presented at the same rate, with 33% targets. Participants responded yes when a predetermined target consonant (“X” or “x”) appeared and no for other consonants. In total, there were four 6-minute series of four 0-Back/2-Back cycles each. The dependent variable generated for this study was the percentage of correct responses for the 2-Back condition.

A second working memory task was also administered, the Paced Auditory Serial Addition Test (PASAT). The task requires participants to add consecutive numbers as they are stated over a taped recording. The participants must add the last 2 numbers stated and answer out loud, being careful not to add their answer to the next stated number (see Table 1). This test was administered to assess working memory while minimizing the inclusion of a required motor response. This was done because previous studies have suggested that motor speed is compromised in OSA, and the authors were more interested in testing working memory than simply motor speed. This ensured that working memory findings would not be an epiphenomenon of simple psychomotor retardation.

Component Processes Testing

The selected component processes for the 2-Back working memory task involve declarative memory, executive functioning, and motor speed. Declarative memory was examined using the Hopkins Verbal Learning Test—Revised (HVLT-R). Alternate forms of the HVLT-R were used at each assessment to minimize test-retest effects. The HVLT-R is a commonly used clinical mea-

METHODS

Participants

Fifty-six participants (17 women) diagnosed with OSA were recruited from an academic sleep disorders center and followed over the course of 3 months of treatment. Participants were taken from a larger study of positive airway pressure (PAP) treatment on cognitive functioning. Eligibility for participation included the following: (a) age between 25 and 85 years; (b) diagnosis of OSA by polysomnography; (c) English-speaking; (d) no comorbid medical disorder; (e) prescribed PAP treatment for OSA; and (f) no previous treatment with PAP. Written informed consent was provided by each participant and participants were compensated for each office visit. This study was approved by Institutional Review Boards at Brown University and the Rhode Island Hospital.

Procedures

All cognitive testing was performed by trained research assistants who were blind to apnea severity and not directly involved in participant treatment. Participants were tested prior to receiving any treatment for OSA and again after 3 months of PAP treatment. Three-month visits were conducted within 2 weeks of the 3-month anniversary of the initiation of treatment. Objective adherence monitors were employed to detect the number of hours of PAP use each night over the course of the study. Adherence downloads were conducted at the follow-up visit. Cognitive testing included 2 tests of working memory as well as tests of declarative memory, executive functioning, and motor speed/coordination. The memory, executive, and motor speed tests were employed as selected component processes involved in the more global cognitive function of working memory.

Table 1—Example of the PASAT Task

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 1—Example of the 2-Back Working Memory Task

Figure 1 details the 0-Back and 2-Back conditions of the n-Back task. In both conditions, participants provide a yes/no response to each stimulus. The stimuli that should elicit a “yes” response are highlighted in bold above. In the 0-Back condition, participants were told to respond with a “yes” only if the “X or x” was presented. In the 2-Back condition, subjects were told to respond with a “yes” only if the stimulus matched one presented 2 stimuli prior, regardless of capitalization.
sure with well-developed normative data, adequate construct and content validity,35 and test-retest reliability for alternate forms.32 Executive functioning was tested using the Trail Making Test, part B (TMT-B).34 TMT-B is a putative executive test measuring the participant’s ability to rapidly alternate between letters and numbers when connecting stimuli scattered about a single page. The Grooved Pegboard test was employed as a test of psychomotor speed and coordination.34 This test required participants to rapidly grooved place pegs into holes on a board after orienting the grooves to fit properly.

**Severity Measures and Adherence**

**CLINICAL POLYSOMNOGRAPHY**

OSA was diagnosed with a full night of in-laboratory clinical polysomnography. Apneas and hypopneas were scored using the American Academy of Sleep Medicine Task Force recommended guidelines,35 with hypopneas defined as a 10-second 30%-50% drop in nasal pressure airflow associated with a 4% drop in oxygen saturation and/or American Sleep Disorders Association defined arousal from sleep. Titration was conducted during a separate full-night polysomnography in order to ascertain the pressure level that eliminated the following: snoring, obvious hypopneas and apneas, unexplained arousals, and oxygen desaturation. PAP titration determined the pressure required to reduce the apnea hypopnea index (AHI) to <5/hour and to eliminate snoring. Apnea severity measures included the AHI, an index of apneas and hypopneas per hour of sleep, and the percentage of total sleep time spent with oxygen saturation <90% (measured by pulse oximeter) during the overnight PSG (Sa90). To try to avoid missing any residual flow limitation, a pressure transducer was placed on one of the ports of the mask. In addition, a comparative pressure 2 cm above what was determined as the “best pressure” was tested to see if there was any additional improvement (i.e., a further decrease in unexplained arousals that maybe due to subtle flow limitation or to see if REM rebound was triggered). Measurements of height and weight were recorded, and body mass index (BMI) was calculated (kilograms/meters²).

**Sleepiness**

Subjective sleepiness was measured using the Epworth Sleepiness Scale (ESS).36 This self-report scale requires participants to rate their likelihood of falling asleep under various circumstances on a 0 (no chance) to 3 (high chance) scale. Scores range from 0 to 24 with higher scores indicative of greater subjective daytime sleepiness. The clinical cutoff score for this scale is a score of ≥10. This measure has demonstrated adequate reliability and validity.37

**PAP Adherence**

All participants were prescribed Respironics REMstar Pro CPAP or C-Flex machines that were set-up and maintained through a single home healthcare company. Heated humidification was supplied for each PAP device to minimize the impact of upper airway dryness on treatment adherence.38 Adherence was covertly monitored using internal microprocessors housed within the PAP devices. Adherence was reported as the total number of hours of PAP use at the prescribed pressure per 24-hour period. Adherence data were collected at the 3-month follow-up visit and are presented in this study as the average nightly use over the 3-month period. An a priori decision was made to not inform participants that their PAP use was being monitored unless they asked for this information directly. None of the participants in this study asked for that information.

**Analyses**

All analyses were conducted using the Statistical Package for the Social Sciences (SPSS) 12.0 (SPSS Inc, Chicago, IL). Data are presented as means and standard deviations unless otherwise noted. Single-order Pearson correlation coefficients were calculated to determine the relationship between 2-Back performance and demographic and severity measures. Pearson correlation coefficients were also calculated to examine the relationship between the cognitive measures prior to treatment. Finally, we employed a Repeated Measures Analysis of Variance (RM-ANOVA) to examine change in each cognitive measure over time by adherence group.

**RESULTS**

Demographic characteristics of all participants are presented in Table 2. Data in Table 2 compare the 2 adherence groups. We have included this information as part of this study was designed to assess cognitive change with regards to a known effective treatment (PAP). We believe that treatment effects in OSA should be examined with consideration of treatment adherence. The PAP adherence data were not normally distributed, necessitating the examination of the adherence data as a categorical, rather than continuous, variable. We therefore split the sample into 2 adherence groups (“high adherence,” n=24 and “low adherence,” n=27) using the median adherence for the group as a whole (an average of 4 hours per night).

The overall sample was predominantly male, obese, well-educated, excessively sleepy, and diagnosed with severe OSA. The average nightly use of PAP for the 51 participants who returned for their assessments was 4.00 ± 2.22 over the course of the 3-month follow-up. The high and low adherence groups did not differ on any demographic variable. Education, however, tended to be higher in the participants who were strong adherers to PAP (high adherence = 16.4 ± 3.54, t99 = -1.72, p = 0.09). Therefore, education was used as a covariate for all subsequent adherence analyses where treatment was concerned.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sample (N=56)</th>
<th>High Use (N=24)</th>
<th>Low Use (N=27)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>52.8 (11.2)</td>
<td>52.3 (9.6)</td>
<td>53.6 (12.8)</td>
<td>0.41</td>
<td>0.69</td>
</tr>
<tr>
<td>BMI</td>
<td>34.1 (7.4)</td>
<td>34.8 (7.9)</td>
<td>32.9 (6.5)</td>
<td>-0.93</td>
<td>0.36</td>
</tr>
<tr>
<td>Education</td>
<td>15.5 (3.6)</td>
<td>16.4 (3.5)</td>
<td>14.8 (3.0)</td>
<td>-1.72</td>
<td>0.09</td>
</tr>
<tr>
<td>AHI (events/hr.)</td>
<td>41.4 (22.1)</td>
<td>44.7 (21.5)</td>
<td>38.6 (23.7)</td>
<td>-0.96</td>
<td>0.34</td>
</tr>
<tr>
<td>Epworth</td>
<td>11.5 (4.9)</td>
<td>12.5 (4.8)</td>
<td>10.4 (5.0)</td>
<td>-1.48</td>
<td>0.14</td>
</tr>
<tr>
<td>Sa90 (Percent time below 90% O2)</td>
<td>20.4 (24.8)</td>
<td>19.8 (22.6)</td>
<td>17.9 (23.5)</td>
<td>-0.29</td>
<td>0.78</td>
</tr>
<tr>
<td>PAP pressure (mm/H2O)</td>
<td>10.2 (2.8)</td>
<td>10.7 (2.6)</td>
<td>10.4 (5.0)</td>
<td>-0.95</td>
<td>0.35</td>
</tr>
</tbody>
</table>

*Table 2—Demographic Characteristics for the Entire Sample and for the Divided Adherence Groups.*
Correlational Analyses

Performance on the 2-Back working memory test was not significantly related to any demographic or disease severity measure (see Table 3). Correlations between 2-Back performance and other cognitive performance demonstrated significant relationships across all cognitive tests (see Table 4). The weakest 2-Back correlation was with the motor task, grooved pegboard (r = -0.27, p <0.05). The other working memory test (PASAT) was only associated with the component process of executive functioning (r = -0.51, p <0.01). The motor task correlated with both declarative memory (r = -0.35, p <0.01) and executive functioning (r = -0.51, p <0.01). It unexpectedly, however, did not correlate with the other working memory task, the PASAT.

Repeated measures ANOVA for change in 2-Back performance over time by adherence group (controlling for education) demonstrated a significant interaction between time and adherence group (F = 4.73, p <0.04), with no main effects (see Figure 2). Results demonstrated that high adherers’ (>4 hours average PAP use/night) accuracy on the 2-Back improved over time (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%). Although a control group was not included in the design of this study, consideration of our OSA sample’s performance with that of published control data acquired from the same task indicates that both high adherers’ (>4 hours average PAP use/night) accuracy increased (baseline = 82%±15%, 3-month = 86%±11%) while low adherers’ (<4 hours average PAP use/night) accuracy decreased (baseline = 82%±15%, 3-month = 86%±11%)

DISCUSSION

This study had 2 primary aims. First, we were interested in determining whether the 2-Back working memory task would demonstrate construct validity with other tests of working memory or selected component processes involved in working memory. The findings suggest that 2-Back performance in OSA patients is related to performance on another working memory test, the PASAT, and on various tests of component cognitive processes including declarative memory, executive functioning, and motor speed. These findings provide convergent evidence that the 2-Back task does test the construct of working memory in this patient sample. Also, results indicated that the high adherence group demonstrated selective improvement with PAP treatment on the 2-Back task and on the proxy measure of working memory (PASAT), and not on the component cognitive process measures. That performance on both the 2-Back and the PASAT cognitive measures responded well to treatment suggests that both measure a similar facet of cognition, further supporting the validity of the 2-Back task as a measure of working memory. Although 2-Back performance did not correlate with demographic or dis-
ease severity factors, the lack of association with disease severity measures is not entirely surprising. Other cognitive studies have demonstrated similar findings, regardless of the cognitive domain tested. More recently, a study of working memory in OSA demonstrated no significant association with severity measures. It is possible that the effects of disease severity are much more complex than measured in this study, or that there are other, yet unknown, moderators of the effect of apnea on cognition.

The second aim of the study was to determine the degree to which change in 2-Back performance with treatment mirrored treatment effects on the other component cognitive tasks. If, for example, change in 2-Back performance was similar to that of executive function change with treatment, but not declarative memory, we could argue that the working memory component most closely linked to treatment was executive functioning. Our findings suggest no main effect for time with treatment on any cognitive measure. This is not completely surprising, as many patients fail to adhere to treatment, which may limit main effects for any group as a whole. We did, however, find 2 interactions between time and PAP adherence group, with high adherers performing better with time on 2 working memory tasks, the 2-Back and the PASAT. No interactions were seen for declarative memory, executive functioning, or motor speed. This suggests that the operationalized construct of working memory in our study is more sensitive to the effects of treatment in OSA than are any selected component cognitive tasks.

It seems intuitive that OSA patients with good adherence would demonstrate an improvement over time on working memory tasks. It is a noteworthy finding, however, as these 2 tasks are quite different from each other. The 2-Back working memory task involves a motor response, while the PASAT requires a verbal response. The PASAT requires simple arithmetic, which can elicit anxiety in some, while the 2-Back task requires no special calculations. The PASAT is completely verbal, while the 2-Back has both verbal and visual components. The demonstrated convergence between these two tasks suggest that the construct of working memory is truly the common denominator that shows a change with effective treatment for OSA. This is supported by the lack of change seen in selected tests of subordinate component processes.

Although the OSA patients in this study who demonstrated high adherence to PAP showed small improvements in working memory measures over time, it is important to note that the low adherence group evidenced a decline in their scores over time. One possible explanation for this finding is that the low adherence group did not receive the benefits of PAP treatment, and thus demonstrated a relative decline in their cognitive functioning between the baseline and follow-up assessments. This finding further highlights the importance of an examination of cognitive function in OSA over time. A second possible explanation for the low adherers’ cognitive decline in working memory may be that PAP treatment itself is not a completely innocuous form of therapy. It has been assumed that PAP provides primarily beneficial effects for individuals with OSA, but perhaps there are unknown deleterious effects of the treatment, especially when used only partially. Further exploration is clearly needed to investigate these findings.

This study has several limitations. We did not include several measures for each subordinate cognitive process of working memory. We employed tests used in previous studies of cognitive functioning in OSA and attempted to maintain a very focussed assessment. Future studies might employ larger cognitive batteries and use factor analysis to verify construct validity. Our study also used a block design in the 2-Back task to assess working memory. These short blocks are typically not used during cognitive assessments of sleepiness, as the breaks between blocks can have an alerting effect by allowing subject to rest between tasks. Future studies examining working memory in OSA patients may wish to use longer, continuous testing blocks, as this could potentially eliminate any associated alerting effect. Emerging outcome studies that employ the 2-Back task, such as the Apnea Positive Pressure Long-Term Efficacy Study (APPLES), may be able to address this issue, in addition to elucidating other long term neurocognitive outcomes (including working memory impairment) associated with OSA and PAP treatment. We also split our sample into 2 groups based on PAP adherence due to a non-normal distribution of PAP data and the restrictions of our sample size. Clearly, treatment adherence is a complicated issue that may not be ideally represented by dichotomous groupings. Future studies with larger samples may wish to examine more discrete groupings of PAP adherence to further examine their relation with cognitive function.

Despite these limitations, we believe that this study provides useful information for both the clinician and the clinical researcher. Our findings suggest that the construct of working memory might be particularly sensitive to the effects of PAP treatment in the patient with OSA. Our study also demonstrates construct validity of the 2-Back task in OSA. This is a common task in cognitive neuroscience that has been extensively used in functional neuroimaging studies. We believe that inclusion of this specific task is important given its known neurofunctional mechanisms.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by a grant to author (MSA) from the National Heart, Lung, and Blood Institute (R01 HL67209).

REFERENCES


JCSM Journal of Clinical Sleep Medicine, Vol. 3, No. 6, 2007