The Clinical Neuropsychologist, 2011, 25 (5), 799–811
http://www.psypress.com/tcn
ISSN: 1385-4046 print/1744-4144 online
DOI: 10.1080/13854046.2011.583279

Psychology Press
Taylor & Francis Group

Lorazepam Effects on Word Memory Test Performance: A Randomized, Double-Blind, Placebo-Controlled, Crossover Trial

David W. Loring[1,2], Susan E. Marino[3], Daniel L. Drane[1], David Parfitt[4], Glen R. Finney[4], and Kimford J. Meador[1,2]

[1]Department of Neurology, Emory University, Atlanta, GA, USA
[2]Department of Pediatrics, Emory University, Atlanta, GA, USA
[3]Center for Clinical and Cognitive Neuropharmacology, Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, MN, USA
[4]Department of Neurology, University of Florida, Gainesville, FL, USA

The Word Memory Test (WMT) is a common measure of symptom validity. To investigate the effects of acute benzodiazepines on WMT scores, oral lorazepam 2 mg (LOR) and placebo were administered 1 week apart in a randomized, double-blind, placebo-controlled, crossover study. A total of 28 participants completed the study and were administered the WMT during each drug condition. Within-participant comparisons of LOR vs placebo revealed significant LOR effects for Immediate Recognition ($p = .007$) and Consistency ($p = .019$), but not Delayed Recognition ($p = .085$). Significant LOR effects were present for Reaction Time Measures (Immediate Recognition RT, $p = .013$; Delayed Recognition RT, $p = .001$; Multiple Choice RT, $p = .011$) and Delayed Memory scores (Multiple Choice, $p = .007$; Paired Associates, $p = .029$; Free Recall, $p = .001$). A pattern similar to crossover results was detected for LOR vs placebo between-group differences for initial test assessment scores. When examined using publisher recommended cut scores for the principal WMT measures, there were six participants failing the WMT during initial LOR testing; all six subsequently performed in the normal range upon retesting with placebo. One participant failed WMT during placebo and obtained passing scores during LOR. These data indicate that multiple WMT measures may be affected by acute LOR dosing, and provide additional evidence that potential latent variables and their effects on both SVT performance and cognitive function should be part of the clinical decision-making process.

Keywords: Symptom validity testing; Malingering; Assessment; Benzodiazepine.

INTRODUCTION

Neuropsychological testing requires adequate task engagement from patients during task performance in order to yield valid estimates of true ability levels. Although psychological assessment measures such as the Minnesota Multiphasic Personality Inventory (MMPI) have long included special scales to indicate whether responses were likely valid, inconsistent, or systematically distorted (Meehl & Hathaway, 1946), the adoption of formal measures to assess the validity of

Address correspondence to: David W. Loring, Ph.D., ABPP-CN, Center for Neuropsychological Studies, Department of Neurology, Emory University School of Medicine, 101 Woodruff Circle, Suite 6000, Atlanta, GA 30322, USA. E-mail: dloring@emory.eduorma

Accepted for publication: March 17, 2011. First published online: June 27, 2011

neuropsychological tests has been a more recent development, beginning in earnest in the 1990s (Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009; Lezak, Howieson, & Loring, 2004). Measures that are explicitly designed to indicate the validity of neuropsychological performance are referred to as symptom validity tests (SVTs), and are an essential component of contemporary forensic neuropsychology practice (Larrabee, 2005; Sweet, 1999b).

A basic tenet of SVTs is that, although they reflect intentional distortion or incomplete task engagement, they should be largely unaffected by neurologic injury unless severe. Cutting scores for SVTs capture performances that are atypical in pattern or degree of impairment, occurring in less than 10% of patients with significant neurologic disease (Larrabee, Greiffenstein, Greve, & Bianchini, 2007). When false positives do occur in patients with genuine neurologic impairment, clinical histories are consistent with substantial neurologic disease (e.g., prolonged coma) with clear radiologic evidence of structural damage, and these patients typically require supervised care (Larrabee et al., 2007; Meyers & Volbrecht, 2003). Thus when scores are obtained that are sufficiently low that they are unlikely to result from a neurologic etiology, inferences can be generalized to the entire neuropsychological protocol that the performance levels have questionable validity.

One of the most widely used contemporary SVTs is the Word Memory Test (WMT) (Sharland & Gfeller, 2007). As noted by the publisher, the WMT is a computerized memory test with hidden measures to establish the validity of test performance (Green, 2003). The WMT has demonstrated sensitivity in neurologic populations with high risk for exaggeration of cognitive impairment (Flaro, Green, & Robertson, 2007) and in known group studies, i.e., those in whom classification of malingering/not malingering is based on accepted independent criteria (Slick, Sherman, & Iverson, 1999), although optimal cut-scores continue to be debated (Greve, Ord, Curtis, Bianchini, & Brennan, 2008; Martins & Martins, 2010). However, populations with high risk for exaggeration and known group analyses are affected by referral patterns and employ convenience samples, which introduce "spectrum" biases that influence sensitivity and specificity estimates (Frederick & Bowden, 2009; Meehl & Rosen, 1967; Willis, 2008).

The present study investigates the effects of lorazepam (LOR), a fast-acting benzodiazepine, on multiple WMT measures. Analogous to the use of amobarbital/methohexital in Wada testing, which creates a transient pharmacological brain lesion to induce temporary neurologic impairment (Loring, Meador, Lee, & King, 1992), acute LOR administration creates a significant reversible neurologic dysfunction to examine acute neurologic effects on WMT performance. We predicted that although there would be no LOR effects on the three principal WMT measures on which performance validity is generally based, significant LOR effects would be observed on WMT measures of reaction time (Post, Chaderjian, Maddock, & Lott, 1997; Subhan, Harrison, & Hindmarch, 1986). In order to comply with contemporary reporting guidelines for clinical trials, each stage of study design and implementation for this project is reported based on CONSORT recommendations (Begg et al., 1996; Schulz, Altman, & Moher, 2010) and STROBE guidelines for neuropsychology reporting (Loring & Bowden, 2011).

METHOD

This study was conducted as part of an investigation evaluating the cognitive effects of 2 mg oral LOR in two age groups. CNS-Vitals is a computerized cognitive assessment battery (Gualtieri & Johnson, 2006a, 2006b), and selected CNS-Vitals subtests were used as cognitive pharmacodynamic measures of LOR effects. A dose of 2 mg was selected because of its established effect on choice reaction time (Subhan et al., 1986). We hypothesized that there would be age-related LOR effect differences in CNS-Vitals performance, with a greater drug effect present for the older participants; those results will be reported separately. The WMT was included as part of the investigational protocol with the permission of the study sponsor (GlaxoSmithKline) to establish whether acute LOR would affect WMT performance.

Participants

This study was performed at the University of Florida (UF) and was approved by the UF Institutional Review Board. Written consent from participants was obtained according to the Declaration of Helsinki prior to enrollment. A total of 28 healthy volunteers from two age groups completed the study. The younger participants were at least 18 and less than 40 years old ($n = 16$), and the older participants were between 60 and 80 years of age ($n = 12$). The average age of the young group was 21.6 years ($SD = 3.0$) and the average age of the older group was 64.2 years ($SD = 2.8$).

There were six females in the young group and six females in the older group, and the young group averaged 14.1 ($SD = 1.0$) years of education and the older group averaged 15.5 ($SD = 2.5$) years of education. Four additional participants withdrew after study enrollment. Two participants withdrew after baseline testing, and two additional participants withdrew following their initial treatment. Exclusion criteria included reported histories of clinically significant cardiovascular, endocrine, hematopoietic, hepatic, neurologic, psychiatric, or renal disease, or reported a history of drug or alcohol abuse. Females of childbearing potential could not be pregnant and underwent urine pregnancy testing prior to enrollment.

Exclusion criteria also included the use of concomitant medications known to affect LOR or the use of concomitant medications potentially affecting cognitive performance (e.g., psychostimulants, antidepressants). Additional exclusion criteria included a prior adverse reaction or hypersensitivity to LOR or related compounds, participants who received any investigational drug within the previous thirty days, participants with IQ estimates <70 as determined by the Test of Non-Verbal Intelligence (Brown, Sherbenou, & Johnsen, 1996), Mini Mental Status Exam (Brown et al., 1996) scores <26, Hopkins Verbal Learning Test (Brandt, 1991) learning score <2 standard deviations below their age norm, and scores on the Center for Epidemiological Studies Depression Scale (Radloff, 1977) >16. These criteria were adopted to ensure adequate levels of general cognitive abilities and absence of clinically relevant depression for all participants, and to screen older volunteers to prevent enrollment of patients with dementia or mild cognitive impairment. There were three screen failure participants, all from taking

concomitant medications thought to potentially alter cognitive function. Three potential participants declined participating during the process of obtaining informed consent.

Sample size

An original sample size of 32 participants (16 young, 16 older) for the larger CNS-Vitals protocol was targeted. This sample size yields a power of 0.8 with a two-sided *p*-value of 0.05 for detecting a drug effect as small as 0.725 times the within-participant placebo-LOR difference score standard deviation. Although the target enrollment of 16 healthy young participants was reached, study enrollment for the older participants was discontinued after 12 participants due to recruitment difficulty.

Randomization and blinding

Randomization was determined by the UF Investigational Drug Pharmacy using "Random Allocation Software" (version 1.0, May 2004) with block size = 2, and stratified by age to ensure that half of each group would be randomized to LOR as the first drug condition. Blinding was obtained by over-encapsulation of LOR with matched placebo tablets. Study medication was dispensed by the UF Investigational Drug Study on the same day to a research assistant who was present when the participant swallowed the capsule.

Study design

The study consisted of a baseline screening visit when informed consent was obtained and study eligibility determined with screening tests and history. After enrollment the randomization sequence determined the sequence of drug vs placebo administration. A 1-week interval between test sessions for the two drug conditions was targeted, although three participants in the young group had their assessments separated by 3 weeks, and two older participants exceeded the 1-week target (one with a 2-week inter-test interval, and one with a 3-week inter-test interval). Participants were paid $65 for each of the first two visits and $70 for the final visit.

Because parallel versions do not exist as part of the standard administration software, the WMT was administered only during the two blinded drug conditions and not during the pretreatment baseline. Study medication was ingested 2 hours prior to neuropsychological testing, and neuropsychological testing for each condition was performed at the same time of day.

RESULTS

Group analyses

All statistical analyses were performed with SPSS 17.0. Two approaches were employed for parametric data analysis. We first performed separate two-way mixed-design ANOVAs, with age as the between-participants factor and drug condition as

Table 1 Means (*SD*) of WMT scores for lorazepam and placebo conditions for the complete sample of participants independent of treatment order (full crossover)

	Placebo	Lorazepam	Difference	95% Confidence Interval	Partial eta squared
Immediate Recognition	97.9 (3.5)	92.1 (9.9)	5.7	1.7–10.0	0.24
Delayed Recognition	97.1 (3.4)	94.4 (8.0)	2.8	–0.4–5.9	0.11
Consistency	95.9 (4.7)	89.7 (12.6)	6.2	1.1–11.2	0.19
3 SVT Mean	97.0 (3.5)	92.1 (9.6)	4.9	1.0–8.7	0.20
Multiple Choice	91.1 (12.9)	80.0 (20.3)	11.1	3.3–18.9	0.24
Paired Associates	90.9 (13.6)	82.0 (20.4)	8.9	1.0–16.8	0.16
Free Recall	52.8 (20.0)	43.3 (20.9)	9.5	4.3–14.7	0.34
3 Memory Mean	78.2 (13.9)	68.4 (18.0)	9.8	3.9–15.7	0.30
Immediate Recognition RT	1.56 (0.56)	2.09 (0.95)	–0.53	–0.93––0.12	0.21
Delayed Recognition RT	1.27 (0.38)	1.62 (0.49)	–0.36	– 0.54––0.17	0.36
Multiple Choice RT	4.06 (2.07)	5.32 (3.14)	–1.26	–2.21––0.31	0.22

RT = Reaction time.

the within- participant factor, for the three principal WMT SVT measures (Immediate Recognition, Delayed Recognition, and Consistency). Because there were no significant effects (or trends) observed for age or the age × drug interaction, we collapsed across age for all subsequent analyses. We next performed a series of within-participant *t*-tests for all WMT measures for participants completing the study. Although we employed no formal control of Type I error rate, we considered the three principal WMT SVT scores as the variables of primary interest. Reaction Time (RT) measures were treated as secondary outcome measures and the remaining, non-timed WMT scores as observational measures. Mean performance levels, confidence intervals, and effect sizes are presented in Table 1.

For the three principal WMT measures significant LOR effects were observed for Immediate Recognition ($p < .007$) and Consistency ($p < .019$) (see Figure 1). Delayed Recognition approached statistical significance at $p = .085$. Statistically significant LOR effects were present for all of the secondary WMT RT measures (Immediate Recognition RT, $p = .013$; Delayed Recognition RT, $p = .001$; Multiple Choice RT, $p = .011$). Significant LOR effects were also observed for all three observational WMT measures (Multiple Choice, $p = .007$; Paired Associates, $p = .029$; Free Recall, $p = .001$).

Because parallel forms of the WMT are not commercially available, and because the test–retest interval was only 1 week, we also performed secondary analyses of WMT performance of study completers for only the initial WMT assessment using a between- participants approach. There were 15 participants who were administered placebo during the first treatment condition and 13 participants who received LOR during the initial assessment. Mean performance levels, confidence intervals, and effect sizes are presented in Table 2. T-tests and reported confidence intervals are based on heterogeneity of variance assumptions for all measures.

Despite the decreased statistical power associated with a between-participants analysis of only initial WMT data, the same LOR pattern was observed. For the

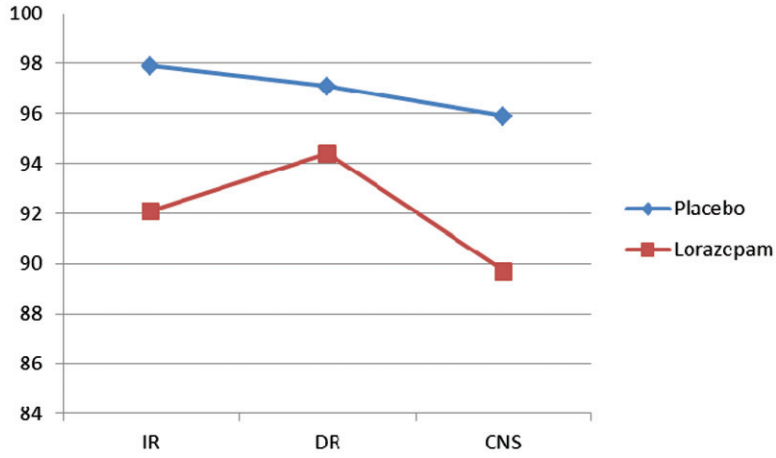

Figure 1 Mean performance levels for primary WMT scores for placebo and lorazepam conditions. IR = Immediate Recognition, DR = Delayed Recognition, CNS = Consistency.

Table 2 Means (*SD*) of WMT scores for lorazepam and placebo conditions for test scores obtained during the initial drug condition

	Placebo	Lorazepam	Difference	95% Confidence Interval	Partial eta Squared
Immediate Recognition	96.8 (4.5)	88.5 (12.9)	8.4	0.36–16.4	.17
Delayed Recognition	96.0 (3.8)	90.1 (9.9)	5.4	–0.76–1.6	.13
Consistency	94.5 (5.4)	84.4 (16.0)	10.1	.09–20.1	.17
3 SVT Mean	95.8 (4.1)	87.8 (12.2)	8.0	.36–15.6	.18
Multiple Choice	89.0 (15.6)	71.9 (24.3)	17.1	0.7–33.4	.16
Paired Associates	89.3 (15.6)	71.5 (24.9)	17.8	1.1–34.5	.17
Free Recall	51.7 (22.3)	41.4 (17.8)	10.3	–5.3–25.9	.06
3 Memory Mean	76.7 (16.1)	61.6 (20.5)	15.1	0.50–29.6	.15
Immediate Recognition RT	1.75 (0.62)	2.64 (0.94)	–0.90	–1.5—.29	.26
Delayed Recognition RT	1.34 (0.29)	1.85 (0.32)	–0.50	–.74–.26	.42
Multiple Choice RT	4.26 (2.11)	5.63 (3.23)	–1.37	–3.56–.81	.06

This is to estimate WMT performance that is not confounded by practice effects due to repeated stimuli exposure.

three principal WMT SVT measures significant drug/group effects were present for Immediate Recognition ($p = .042$) and Consistency ($p = .048$), with a trend present for Delayed Recognition ($p = .081$). For secondary WMT RT time score, significant differences were present for Immediate Recognition RT ($p = .008$) and Delayed Recognition RT ($p < .0001$), but not for Multiple Choice RT ($p = .205$). For the observational WMT measures significant drug (group) effects were present for Multiple Choice ($p = .034$), and Paired Associates ($p = .03$), but not Free Recall ($p = .193$). Thus our hypothesis that LOR RT effects would be present was supported. However, our prediction that LOR effects on principal WMT SVT

Table 3 Individual WMT scores for participants performing below recommended cut scores on any of the three principal WMT SVT measures, and scores used to calculate *Genuine Memory Impairment Profile (GMIP)* for study completers and for two participants withdrawing after single treatment session

	Group	Immediate Recognition	Delayed Recognition	Consistency	3 SVT Mean	3 Memory Mean	Difference
Completers							
LOR32	Old	57.5	70.0	47.5	58.3	16.7	***41.7***
LOR20	Young	70.0	95.0	70.0	78.3	59.2	19.2
LOR4	Young	82.5	82.5	85.0	83.3	60.0	23.3
LOR39	Old	82.5	90.0	77.5	83.3	39.2	***44.2***
LOR8	Young	85.0	72.5	62.5	73.3	45.8	27.5
LOR33	Old	85.0	97.5	82.5	88.3	55.8	***32.5***
Placebo6	Young	85.0	92.5	82.5	86.7	54.2	***32.5***
Withdrawals							
LOR22	Young	90.0	77.5	77.5	81.7	37.5	***44.2***
Placebo13	Young	95.0	82.5	77.5	85	65.8	19.2

performance would be absent was not supported, with 4 of the 6 parametric analyses (within group and between group comparisons) statistically significant, and the remaining 2/6 analyses trending toward LOR effects ($p < .10$).

Individual participant classification

Although this was not a primary goal of this investigation we examined individual classification using recommended cut scores using several approaches, since SVT results are used clinically on an individual patient level. Per recommendations in the SVT literature (Sweet, 1999a), the specific cut scores use to classify validity are not published here but are available in the test manual and from the authors.

There were seven participants completing the study who scored at or below the cut score suggested by the WMT publisher on at least one of the primary SVT scales (see Table 3). Six of these failures occurred during LOR (five were initial assessments, one was the second assessment), and one failure was during placebo and was the initial assessment (see Table 3). All six LOR participants failing the WMT produced suboptimal scores on the Immediate Recognition trial. Three of these participants were in the older group and three were in the young group. All six participants with invalid scores during LOR had valid test scores when tested during placebo. There was only a single placebo WMT failure, and this occurred in a young patient during the initial WMT assessment (1/28, 4%); normal WMT scores were obtained for this participant during LOR administration.

A second approach for individual classification relied on the 2% and 10% false positive (FP) error rate cutoffs that have been suggested as one approach to minimize FP WMT classification rates (Greve et al., 2008). Using the more conservative 2% FP criterion, a single LOR WMT failure was identified (LOR32, see Table 3), and the single placebo WMT failure identified with publisher cut-off scores (Placebo6) was no longer classified as a WMT failure. Using the 10% FP

criterion, there were three LOR WMT failures (LOR32, LOR8, LOR33) and no placebo WMT failures.

There were two participants who completed a single treatment condition (one LOR, one placebo) but who subsequently withdrew from the study. Both had principal WMT SVT scores at or below the recommended cut-score on at least one WMT measure using publisher-recommended cut-offs (see Table 3). Because neither participant completed the study they were not included in summary analyses (efficacy subset analyses). Neither participant would be classified as WMT failures using either 2% FP or 10% FP criteria.

The final approach to individual performance characterization relied on difference scores between the average of three WMT SVT scores (i.e., Immediate Recognition, Delayed Recognition, and Consistency) and average of three WMT genuine memory scores (i.e., Multiple Choice, Paired Associates, Free Recall), an approach designed to minimize false positive errors by profiles associated with dementia or significant memory impairment (Henry, Merten, Wolf, & Harth, 2010; Howe & Loring, 2009). This Genuine Memory Impairment Profile, or GMIP (also referred to as the Dementia Profile), for the WMT is operationalized as a 30 percentage points or greater difference between SVT and memory scores (Green, 2005). Of the seven participants in the completer group failing WMT SVT, four demonstrated a GMIP profile (see Table 3). Three GMIP profiles were during LOR and one GMIP was obtained during placebo. Of the two participants who participated in only a single drug condition, the LOR patient had a GMIP whereas the placebo participant did not.

Effects of SVT failure on group performance

A secondary follow-up analysis to investigate the effects of removing SVT failures from the group analysis was performed to examine whether there were individuals who were disproportionately sensitive to LOR and who were responsible for producing the group differences observed on the non SVT WMT scores (see Table 4). This approach is conceptually analogous to generating ''robust'' norms that eliminates ''contamination'' of data when elderly participants subsequently found to develop dementia are removed from the dataset (Holtzer et al., 2008). Significant drug effects were observed for Multiple Choice ($p = .043$), Free Recall ($p = .008$), 3 Memory Mean ($p = .026$), and Delayed Recognition RT ($p = .005$). Despite eliminating all participants with SVT failure, there was a drug effect trend for Immediate Recognition ($p = .065$) and a trend for drug differences for Multiple Choice RT ($p = .075$).

DISCUSSION

These data indicate that WMT SVT, memory, and RT scores are sensitive to impaired neurologic function associated with acute LOR administration. In the set of analysis of WMT SVT scores that includes practice effects there were significant LOR effects on two of the three principal WMT SVT measures. When LOR effects were examined using a between-participants analysis of only the initial assessment scores to avoid practice effects, a similar pattern was identified.

Table 4 Means (*SD*) of WMT scores for lorazepam and placebo conditions for the complete sample of participants independent of treatment order (full crossover) excluding seven participants failing SVT using publisher criteria

	Placebo	Lorazepam	Difference	95% Confidence Interval	Partial eta squared
Immediate Recognition	98.3 (2.8)	96.4 (3.9)	1.9	0.1–3.9	0.16
Delayed Recognition	97.4 (3.7)	94.5 (3.3)	–0.1	–2.0–1.8	0.00
Consistency	95.7 (4.1)	95.1 (5.2)	1.6	–1.3–4.4	0.06
3 SVT Mean	97.5 (3.1)	96.3 (3.7)	1.1	–0.9–3.1	0.06
Multiple Choice	93.6 (11.1)	88.1 (20.3)	5.5	0.2–10.8	0.20
Paired Associates	93.8 (12.3)	90.0 (11.6)	3.8	–2.8–10.4	0.09
Free Recall	55.0 (18.9)	47.7 (20.8)	7.3	2.0–12.5	0.30
3 Memory Mean	80.8 (12.8)	75.3 (12.8)	5.5	0.7–10.3	0.22
Immediate Recognition RT	1.56 (0.61)	1.92 (0.86)	–0.36	–0.84—0.13	0.10
Delayed Recognition RT	1.24 (0.38)	1.59 (0.45)	–0.35	– 0.59— 0.12	0.34
Multiple Choice RT	3.94 (1.98)	4.79 (2.52)	–0.85	–1.79—0.95	0.15

RT = Reaction time.

Applying conventional criteria to characterize η^2 effect sizes (<0.01 = trivial; 0.01–0.06 = small; 0.06–0.14 = moderate; 0.14 + = large) (Ellis, 2010), the magnitude of LOR's effect on Immediate Recognition and Consistency is considered large, and is present in both the full crossover analyses and between group initial treatment comparisons.

This study was performed to investigate potential effects of LOR on WMT, not to estimate WMT SVT failure rates associated with single acute 2.0 mg LOR. A crossover design greatly decreases the required sample sizes to statistically demonstrate a treatment effect, and the sample size employed in the present study is comparable to multiple psychopharmacological reports of drug and alcohol effects (Howard et al., 2007; Mills, Spruill, Walker, & Lamson, 2009; Post et al., 1997; Subhan et al., 1986; Verster et al., 2009). However, because these results may have potential implications for clinical application of WMT, several approaches to individual performance classification were examined. There were seven participants completing the entire study with WMT SVT performance levels at or below publisher-recommended cut scores, and six of these were during LOR. Because failed WMT scores occurred almost exclusively following LOR administration, the results are due to medication effects and cannot easily be attributed to intentional response distortion. Further support for the absence of intentional performance distortion is the normal WMT SVT performance levels of these participants during placebo. Thus altered cognitive function associated with changes in neurologic status is sufficient to produce WMT SVT results in the invalid range using publisher-recommended criteria.

Although 6/28 (21%) participants obtained invalid WMT SVT during LOR but only 1/28 (4%) during placebo, other approaches to WMT SVT characterization yielded different findings. The 2% and 10% FP criteria decreased WMT failure rate, although there is no assurance that this resulted in more accurate overall performance characterization. For example, a single participant failing WMT SVT

during placebo (Placebo 6; Table 3) had a 32.5-point discrepancy between WMT SVT and memory indices, a pattern inconsistent with college enrollment. Further, this participant's normal WMT scores during LOR indicate an absence of an enduring pathological substrate to account for poor WMT SVT performance during placebo.

All six LOR participants failing the WMT SVT produced invalid scores on the WMT Immediate Recognition trial, suggesting that LOR may be interfering with some aspect of initial encoding of novel stimuli, perhaps through altered levels of attention or alertness. We have advanced a similar conceptualization of poor WMT in epilepsy surgery candidates in which high rates of WMT SVT failure (28%) have been described (Drane et al., 2006; Williamson, Drane, & Stroup, 2007). When patients having a seizure within 24 hours prior to neuropsychological testing were excluded, WMT SVT failures decreased to 8%. Because similar high failure/questionably valid rates have been reported in other epilepsy surgery series using different SVT measures but without consideration of recent seizure effects (Cragar, Berry, Fakhoury, Cibula, & Schmitt, 2006; Loring, Lee, & Meador, 2005), depleted post-ictal neuronal function is a likely etiology for many SVT failures in epilepsy surgery candidates who, as a group, are well motivated to preserve cognitive abilities following surgery. Also supporting this interpretation of SVT failure is a case study in which WMT SVT was failed when the patient was assessed during frequent left temporal epileptiform discharges, but who easily passed WMT SVT when tested during quiescent EEG (Drane, in press).

Although we hypothesized significant LOR RT effects, we were surprised to observe significant LOR effects for the principal WMT SVT scores. However, there is a striking difference in the magnitude of drug effects across different WMT measures. The largest LOR effect on the principal WMT SVT scores was present for Immediate Recognition, and was associated with a partial η^2 of .24. In contrast, the largest LOR effect on RT was present for Delayed Recognition RT with an effect size accounting for an additional 12% of the variance explained by LOR (partial $\eta^2 = .36$). Thus, although both WMT SVT and RT scores are affected by acute LOR administration, the LOR effect is greater for RT.

The major strength of this report is the use of a randomized, double-blind trial using medication to transiently create a pharmacologic lesion with a placebo comparator. To our knowledge this is the first study to employ such an approach to examine neurologic effects on WMT performances. Further, unlike non-randomized patient studies, the crossover design minimizes individual differences since all participants serve as their own control. Increasingly, biomedical journals require levels of evidence to be reported with clinical findings to facilitate the development and implementation of evidence-based practice guidelines. Evidence-based study classifications are generally rated into one of four classes (I–IV), with randomized blinded controlled trials representing the strongest (Class I) evidence (Gross & Johnston, 2009). As a study with treatment randomization, concealed allocation, double-blind assessment, exclusion/inclusion criteria clearly defined, and adequate accounting for dropouts (with at least 80% of enrolled participants completing the study), this report meets Class I evidence standards.

Unlike group comparisons based on clinical samples, results of the present study are not biased by criterion-group and clinical control samples with unknown

base rates and potentially inflated pretest probabilities. In particular, an approach with pharmacologic lesions in healthy volunteers avoids referral biases that affect many clinical studies, namely those that are not based on population-representative sampling or random allocation. These biases are associated with convenience sampling, even in careful, consecutive patient series, are sometimes referred to as "spectrum" biases, and are well known to potentially distort base rates and the associated traditional classification statistics such as sensitivity, specificity, and likelihood characterization (Frederick & Bowden, 2009; Meehl & Rosen, 1967; Willis, 2008). A limitation of this report is that testing was separated by only a single week, increasing the likelihood of significant practice or carry-over effects. However, significant practice effects would bias the results in favor of the null hypothesis. Further, a similar pattern of LOR effects was associated with the less statistically powerful group level of analyses, examining LOR vs placebo WMT scores during the initial assessment only. Because 5/6 LOR SVT failures occurred during the initial assessment, there is a suggestion that the magnitude of the practice effect associated with repeat testing offsets the magnitude of the drug effect.

Our findings provide additional evidence that potential latent variables and their effects on both SVT performance and cognitive function should be part of the clinical decision-making process. While the current study examined an acute dose of LOR administered to naïve participants, it demonstrates that, under certain circumstances, medications can impact SVT results with poor SVT scores unrelated to apparent intentional attempts at response distortion. Attributing cause for test invalidity is more complex than a dichotomous option of whether or not a participant is actively avoiding full task engagement since multiple potential causes of suboptimal task engagement exist (Donders & Boonstra, 2007). As with all neuropsychological measures, SVT performances should not be interpreted in isolation. In addition to employing multiple measures of performance validity (Larrabee et al., 2007), SVTs should be interpreted within the entire clinical context of patient presentation including history, laboratory findings, and other behavioral features.

ACKNOWLEDGMENT

Funding for this project was provided by GlaxoSmithKline. Drs. Loring, Finney, and Meador, and Mr. Parfitt, have received research support from GlaxoSmithKline. We thank our reviewers for their helpful comments.

REFERENCES

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reporting of randomized controlled trials. *The CONSORT statement*. *JAMA*, *276*(8), 637–639.

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new verbal memory est with six equivalent forms. *The Clinical Neuropsychologist*, *5*, 125–142.

Brown, L., Sherbenou, R. J., & Johnsen, S. K. (1996). *Test of Non-verbal Intelligence* (3rd ed., TONI-3). Austin, TX: Pro-Ed.

Cragar, D. E., Berry, D. T., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2006). Performance of patients with epilepsy or psychogenic non-epileptic seizures on four measures of effort. *The Clinical Neuropsychologist*, *20*(3), 552–566.

Donders, J., & Boonstra, T. (2007). Correlates of invalid neuropsychological test performance after traumatic brain injury. *Brain Injury*, *21*(3), 319–326.

Drane, D. L. (in press). Neuropsychological evaluation of the epilepsy surgery candidate. In C. Morrison & W. B. Barr (Eds.), *Handbook of the neuropsychology of epilepsy*. New York, NY: Springer.

Drane, D. L., Williamson, D. J., Stroup, E. S., Holmes, M. D., Jung, M., Koerner, E., et al. (2006). Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia*, *47*(11), 1879–1886.

Ellis, P. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. New York, NY: Cambridge University Press.

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury*, *21*(4), 373–383.

Frederick, R. I., & Bowden, S. C. (2009). The test validation summary. *Assessment*, *16*(3), 215–236.

Green, P. (2003). *Green's Word Memory Test: User's manual*. Edmonton, Canada: Green's Publishing Inc.

Green, P. (2005). *Green's Word Memory Test for Windows User's manual–revised*. Edmonton, Canada: Green's Publishing.

Greve, K. W., Ord, J., Curtis, K. L., Bianchini, K. J., & Brennan, A. (2008). Detecting malingering in traumatic brain injury and chronic pain: A comparison of three forced-choice symptom validity tests. *The Clinical Neuropsychologist*, *22*(5), 896–918.

Gross, R. A., & Johnston, K. C. (2009). Levels of evidence: Taking Neurology® to the next level. *Neurology*, *72*(1), 8–10.

Gualtieri, C. T., & Johnson, L. G. (2006a). Antidepressant side effects in children and adolescents. *Journal of Child and Adolescent Psychopharmacology*, *16*(1–2), 147–157.

Gualtieri, C. T., & Johnson, L. G. (2006b). Reliability and validity of a computerized neurocognitive test battery, CNS Vital Signs. *Archives of Clinical Neuropsychology*, *21*(7), 623–643.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129.

Henry, M., Merten, T., Wolf, S. A., & Harth, S. (2010). Nonverbal Medical Symptom Validity Test performance of elderly healthy adults and clinical neurology patients. *Journal of Clinical and Experimental Neuropsychology*, *32*(1), 19–27.

Holtzer, R., Goldin, Y., Zimmerman, M., Katz, M., Buschke, H., & Lipton, R. B. (2008). Robust norms for selected neuropsychological tests in older adults. *Archives of Clinical Neuropsychology*, *23*(5), 531–541.

Howard, M. E., Jackson, M. L., Kennedy, G. A., Swann, P., Barnes, M., & Pierce, R. J. (2007). The interactive effects of extended wakefulness and low-dose alcohol on simulated driving and vigilance. *Sleep*, *30*(10), 1334–1340.

Howe, L. L., & Loring, D. W. (2009). Classification accuracy and predictive ability of the medical symptom validity test's dementia profile and general memory impairment profile. *The Clinical Neuropsychologist*, *23*(2), 329–342.

Larrabee, G. J. (Ed.). (2005). *Forensic neuropsychology: A scientific approach*. New York, NY: Oxford University Press.

Larrabee, G. J., Greiffenstein, M. F., Greve, K. W., & Bianchini, K. J. (2007). Refining diagnostic criteria for malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits*. New York, NY: Oxford University Press.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment* (4th ed.). New York, NY: Oxford University Press.

Loring, D. W., & Bowden, S. C. (2011). *Developing evidence based neuropsychology practice: More than a flicker of hope using STROBE*. Manuscript submitted for publication.

Loring, D. W., Lee, G. P., & Meador, K. J. (2005). Victoria Symptom Validity Test performance in non-litigating epilepsy surgery candidates. *Journal of Clinical & Experimental Neuropsychology*, *27*(5), 610–617.

Loring, D. W., Meador, K. J., Lee, G. P., & King, D. W. (1992). *Amobarbital effects and lateralized brain function: The Wada Test*. New York, NY: Springer-Verlag.

Martins, M., & Martins, I. P. (2010). Memory malingering: Evaluating WMT criteria. *Applied Neuropsychology*, *17*(3), 177–182.

Meehl, P. E., & Hathaway, S. R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*(5), 525–564.

Meehl, P. E., & Rosen, A. (1967). Antecedent probability and the efficiency of psychometric signs, patterns, on cutting scores. In D. N. Jackson & S. Messick (Eds.), *Problems in human assessment*. New York: McGraw-Hill.

Meyers, J. E., & Volbrecht, M. E. (2003). A validation of multiple malingering detection methods in a large clinical sample. *Archives of Clinical Neuropsychology*, *18*(3), 261–276.

Mills, K. C., Spruill, S. E., Walker, J. M., & Lamson, M. (2009). A clinical trial demonstration of a web-based test for alcohol and drug effects. *Journal of Studies on Alcohol and Drugs*, *70*(2), 308–313.

Post, R. B., Chaderjian, M. R., Maddock, R. J., & Lott, L. A. (1997). Effects of lorazepam on the distribution of spatial attention. *Experimental and Clinical Psychopharmacology*, *5*(2), 143–149.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*(340), c332.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology*, *22*(2), 213–223.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(4), 545–561.

Subhan, Z., Harrison, C., & Hindmarch, I. (1986). Alprazolam and lorazepam single and multiple-dose effects on psychomotor skills and sleep. *European Journal of Clinical Pharmacology*, *29*(6), 709–712.

Sweet, J. J. (1999a). Malingering: Differential diagnosis. In J. J. Sweet (Ed.), *Forensic neuropsychology: Fundamentals and practice*. Exton, PA: Swets & Zeitlinger.

Sweet, J. J. (Ed.). (1999b). *Forensic neuropsychology: Fundamentals and practice*. Exton, PA: Swets & Zeitlinger.

Verster, J. C., Wester, A. E., Goorden, M., van Wieringen, J. P., Olivier, B., & Volkerts, E. R. (2009). Novice drivers' performance after different alcohol dosages and placebo in the divided-attention steering simulator (DASS). *Psychopharmacology (Berl)*, *204*(1), 127–133.

Williamson, D. J., Drane, D. L., & Stroup, E. S. (2007). Symptom validity tests in the epilepsy clinic. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment*. New York, NY: Guilford Publications.

Willis, B. H. (2008). Spectrum bias–why clinicians need to be cautious when applying diagnostic test studies. *Family Practice*, *25*(5), 390–396.