OXFORD
UNIVERSITY PRESS

Archives
of
CLINICAL
NEUROPSYCHOLOGY

If Invalid PVT Scores Are Obtained, Can Valid Neuropsychological Profiles Be Believed?

David W. Loring[1,*] and Felicia C. Goldstein[1]

[1]*Department of Neurology, Emory University School of Medicine, 12 Executive Park Atlanta, GA 30329, USA*

*Corresponding author at: Department of Neurology, Emory Brain Health Center, 12 Executive Park Atlanta, GA 30329, USA. Tel: 404-727-4116.
E-mail address: dloring@emory.edu (D. W. Loring)

Editorial Decision 5 June 2019

Abstract

Background: Performance Validity Testing (PVT) decision-making rules may be indeterminate in patients with neurological disease in which PVT characteristics have not been adequately studied. We report a patient with multiple sclerosis (MS) who failed computerized PVT testing but had normal memory scores with a neuropsychological profile consistent with expected MS disease-related weaknesses.
Method: Neuropsychological testing was conducted on two occasions in a middle-aged woman with an established MS diagnosis to address concerns of possible memory decline. Testing was discontinued after PVT scores below recommended cut-points were obtained during the first evaluation. During the second assessment, subthreshold PVT scores on a different computerized PVT were obtained, but unlike the first assessment, the entire neuropsychological protocol was administered.
Results: Despite subthreshold computerized PVT scores, normal learning and memory performance was obtained providing objective data to answer the referral question. Other neuropsychological findings included decreased processing speed, poor working memory, and poor executive function consistent with her MS diagnosis. Embedded PVT scores were normal.
Conclusions: We speculate that poor computerized PVT scores resulted from the disease-related features of MS, although we also discuss approaches to reconcile apparently contradictory PVT versus neuropsychological results if the contributions of disease-related variables on PVTs scores are discounted. This case demonstrates the value of completing the assessment protocol despite obtaining PVT scores below publisher recommended cutoffs in clinical evaluations. If subthreshold PVT scores are considered evidence of performance invalidity, it is still necessary to have an approach for interpreting seemingly credible neuropsychological test results rather than simply dismissing them as invalid.

Keywords: Performance Validity Tests; Base rates; Test validation; Test specificity

Introduction

Performance validity testing (PVT) serves an important role in contemporary neuropsychology practice and is recommended by major professional organizations including the *American Academy of Clinical Neuropsychology* (Heilbronner et al., 2009) and the *National Academy of Neuropsychology* (Bush et al., 2005). PVTs were primarily developed for forensic neuropsychological evaluations in which high external incentives for poor performance are present, although PVTs are now commonly included in most clinical assessments. Approximately three-fourths (73.5%) of forensic neuropsychological evaluations employ multiple PVTs in their assessment protocols, contrasting with less than one-fourth (22.4%) of practices that do not see forensic patients (Sweet et al., 2015). A single stand-alone PVT is used by approximately two-fifths (42.9%) of non-forensic practices, and one or more embedded PVTs included in nearly three-fourths (73.7%) of such practices.

doi:10.1093/arclin/acz028

PVT research in patients with independently established neurologic disease is limited. Consequently, validity inferences based upon rules or algorithms derived from forensic neuropsychology may be inappropriately applied to neurologic patient samples given this incomplete understanding of disease effects on PVT performance. In addition, there is typically a lack of base rate information characterizing symptom embellishment in different neurological diseases/syndromes. Multiple studies have disproven the assumption that PVT tasks are insensitive to significant neurologic disease effects in patients with dementia, epilepsy, or acute severe traumatic brain injury (Dean et al., 2009; Grote et al., 2000; Keary et al., 2013; Loring et al., 2005, 2016; Macciocchi et al., 2006; Robinson et al., 2018). These reports not only illustrate the risks of false positive errors when using forensically determined fixed thresholds within clinical disease samples, but are also relevant when PVT performance influences the approach to assessment after obtaining subthreshold PVT scores.

We report a patient tested by two different neuropsychology services at the same academic medical center ~1 ½ years apart that employed different test continuation protocols when poor PVT scores are obtained. Computer-administered stand-alone PVT scores were below published recommended thresholds during both assessments. Testing was discontinued after obtaining subthreshold PVT scores during the first evaluation. In contrast, all neuropsychological tests were administered during the second assessment despite PVT scores in the "invalid" range. The second testing, however, yielded a neuropsychological profile consistent with the referral diagnosis of multiple sclerosis (MS). More importantly, normal performance across multiple memory measures allowed the referral question of possible MS-related memory decline to be confidently answered.

We discuss the risks of inappropriate generalization of fixed PVT thresholds to dissimilar patients with independently established neurologic disease. We also address whether neuropsychological testing should be discontinued routinely in patients referred for clinical evaluation. Different risks and benefits of PVT interpretation/misinterpretation can be expected in various clinical settings evaluating patients with different patient incentives (e.g., pursuing disability vs. pursuing treatment). We stress the importance of considering prior probability base rates of likely intentional response distortion, which are likely to be lower than compensation seeking patients even if not formally established, when making clinical inferences across neurologic diseases and conditions. There are different costs and benefits of correct/incorrect decisions when interpreting neuropsychological profiles across settings and referrals such as for patient care versus litigation or disability application.

Materials and Methods

The patient was an African American woman in her early 50s who was seen for neurologic evaluation of numbness in her fingertips and toes, with persistent left-sided paresthesia and right-sided foot-drop. She was diagnosed with MS by her neurologist ~2 years prior to neuropsychological consultation based upon CSF findings that included multiple oligoclonal bands and an elevated immunoglobulin G index. Magnetic resonance imaging (MRI) findings identified multiple non-enhancing periventricular lesions, and plaques in the bilateral temporal lobes and the corpus callosum. Additional spinal lesions were present at C2-C3, C6, and T4.

She had a history of migraine, depression, and anxiety. Because of cognitive concerns thought to be MS-related, she was referred by her neurologist for neuropsychological evaluation. Cognitive complaints initially described to her neurologist included forgetting conversations and recalling names, decreased attention, and difficulties with word-finding, spelling, and mental arithmetic. The ability to estimate distances when reaching for items was decreased, with both underestimation and overestimation of item position. She reported that she was not on disability nor was she considering pursuing disability. She graduated college and was married to a surgeon. She previously worked as a medical coder but stopped ~5 years prior to evaluation due to making frequent errors. At the time of both evaluations, medications included zonisamide 300 mg (headaches), alprazolam 2 mg prn (anxiety), lamotrigine 100 mg (anxiety and depression), venlafaxine 75 mg (depression), baclofen 10 mg (muscle spasms), gabapentin 600 mg (paresthesia), and trazodone 50 mg (sleep). MS was treated with glatiramer acetate.

Results

Initial Neuropsychological Evaluation

During her first testing, the Victoria Symptom Validity Test (VSVT; Slick et al., 1997) was administered to characterize performance validity. The VSVT is a computer-administered forced-choice PVT in which a 5-digit number is presented for 5 s followed by presentation of two 5-digit choices (target and foil) following delays of either 5, 10, or 15 s. There are 48 trials including both easy items in which the response foil differs across all five digits and hard items in which the response foil includes two transposed digits. Performance on the 24 hard items is the primary PVT measure.

Valid performance (20/24) was obtained on the easy VSVT items, although chance performance (12/24) on the difficult VSVT items raised concern about overall performance validity. The remaining neuropsychological tests were not administered after obtaining this VSVT pattern, although a psychological assessment using the MMPI-2-RF was completed. The MMPI-2-RF was interpreted as reflecting significant emotional distress due to both her current life stressors and her medical conditions. Although high distress was considered a possible factor in her poor VSVT performance on the hard items, possible lack of engagement or disinterest could not be accurately determined. She was encouraged to seek psychotherapy or additional medical treatment, with neuropsychological testing recommended following improved emotional functioning and psychological well-being. Medical management of her depression and anxiety continued to be supervised by her primary care physician and her treatment remained the same.

Second Neuropsychological Evaluation

Despite good management of her MS and follow-up MRIs demonstrating no new or enhancing MS plaques, she continued to report memory and word-finding difficulty to her neurologist. Approximately 1 ½ years after her initial testing, she was scheduled by a different neuropsychology service due to appointment availability. Stand-alone PVT testing during the second evaluation was established using the Word Memory Test and these results are presented in Table 1 (WMT; Green, 2005). We also examined embedded PVT indices recommended by Rickards, Cranston, Touradji, and Bechtold (2018), scores used by Pearson's Advanced Clinical Solutions (Pearson, 2009), and the Rey Auditory Verbal Learning Test (AVLT) Logistic Regression equation of Davis, Millis, and Axelrod (2012), which has been previously studied in neurologic disease (Loring et al., 2016).

The WMT is a computerized PVT presenting 20 semantically related word pairs for 6 s. Two trials are administered, and performance is tested across six conditions including verbal learning, recall, response consistency, and recognition. On the WMT, her scores were immediate recall = 67.5%, delayed recall = 72.5%, consistency = 50.0%, multiple choice = 60.0%, paired associates = 65%, and free recall = 30% suggesting performance invalidity. In addition, comparison of the average of the first three WMT percentages (M = 63%) to the average of the final three WMT percentages (M = 51%) did not fit the Genuine Memory Impairment Profile/Severe Impairment Profile (Howe & Loring, 2009; Robinson et al., 2018).

Embedded PVT scores. The Reliable Digit Span (Greiffenstein et al., 1994) of 7 was in the normal range. Two Wisconsin Card Sorting Test (WCST) failures to maintain set errors were not considered to be suspicious (Greve et al., 2009). Recognition scores for both Logical Memory and Visual Reproduction from the Wechsler Memory Scale-IV were above the 25% base rate threshold (Pearson, 2009). Her estimated probability of AVLT performance invalidity using Logistic Regression was p = .11 (Davis et al., 2012). Thus, scores were in the normal range across multiple embedded PVTs without suggestion of exaggeration or lack of task engagement. Normal embedded PVT performance was obtained for Complex Figure recognition (Lu et al., 2003).

General cognitive function. General cognitive function reflected by the WAIS-IV General Ability Index was average (GAI = 97) and was considered meaningfully higher than Full Scale IQ of 89, which includes both processing speed and working memory. She displayed a Verbal Comprehension Index (VCI = 108) versus Perceptual Reasoning Index (PRI = 84) discrepancy of 24 points. Although this is a pattern often associated with poor right hemisphere function, it can also be seen with diffuse brain impairment affecting novel problem solving (Farr et al., 1986; Hawkins et al., 2002). On the National Institutes of Health (NIH) Cognitive Toolbox (Weintraub et al., 2013), the Total Composite score was SS = 96. Crystallized knowledge was high average (SS = 112), whereas Fluid performance was low average (SS = 81), a pattern comparable to the WAIS-IV VCI versus PRI discrepancy.

Language. Language was normal across naming and generative verbal fluency tasks. Despite normal performance on the Boston Naming Test in terms of total number of correct responses, her mean BNT response latency was over 2 s reflecting word retrieval inefficiency (Hamberger & Seidel, 2003). Average to high average scores were obtained across all tests reflecting verbal abilities from the WAIS-IV and NIH Cognitive Toolbox.

Visual spatial. Visual spatial and constructional performances were normal overall but were impacted by slowed speed of processing and difficulties with executive functioning. In contrast to her PRI, normal perceptual ability was indicated by Line Orientation. Consistent with impaired frontal–subcortical contributions to PRI, performance was impaired on Complex Figure copy (see Fig. 1). Block design construction was low average and affected by slowed processing speed.

Table 1. Neuropsychological test findings

Test	Score	Percentile/interpretation
Test of Premorbid Function (standard score)	112	79
WAIS-IV Full Scale IQ (standard score)	89	23
WAIS-IV General Ability Index (standard score)	97	42
WAIS-IV Verbal Comprehension Index (standard score)	108	70
Vocabulary (scaled score)	14	91
Information (scaled score)	10	50
Similarities (scaled score)	11	63
WAIS-IV Perceptual Reasoning Index (standard score)	84	14
Block design (scaled score)	6	9
Matrix reasoning (scaled score)	7	16
Visual puzzles (scaled score)	9	37
WAIS-IV Working Memory Index (standard score)	80	9
Digit span (scaled score)	5	5
Digits forward (scaled score)	6	9
Digits backward (scaled score)	6	9
Digit sequencing (scaled score)	7	16
Arithmetic (scaled score)	8	25
WAIS-IV Processing Speed Index (standard score)	84	14
Symbol search (scaled score)	8	25
Coding (scaled score)	6	9
Boston Naming Test		
Total (raw)	56/60	68
Mean response latency	2.9 s	Prolonged
Recognition	60/60	WNL
Controlled Oral Word Association (raw)	33	18
Animal Fluency (raw)	20	58
Line Orientation (raw corrected)	25/30	56
Rey–Osterrieth Complex Figure		
Copy (raw)	27/36	1
Immediate (raw)	10.5/36	3
Delay (raw)	13.5/36	10
Recognition (raw)	22 (11 + 11)/24	79
Atypical recognition errors	0	WNL
ROCFT effort equation	60	WNL
Rey Auditory Verbal Learning Test		
Total (trials 1–5; raw)	46/75	42
Delay (raw)	12/15	74
Recognition hits (raw)	15/15	WNL
Recognition false positive (raw)	0	WNL
Davis et al. (2012) Logistic Regression	$p = .11$	WNL
WMS-IV Logical Memory		
Immediate recall (raw)	27/50	63
Delayed recall (raw)	19/50	37
Delayed recognition (raw)	28/30	>75
WMS-IV Visual Reproduction		
Immediate recall (raw)	35/43	10
Delayed recall (raw)	25/42	10
Delayed recognition (raw)	6/7	51–75
Trail Making Test		
Part A time	72 s	3
Part A errors	1	WNL
Part B time	300 s	$<1^{st}$
Part B errors	2	borderline
Wisconsin Card Sorting Test (computer)		
Categories (raw)	4/6	14
Perseverative responses (raw)	25	16
Perseverative errors (raw)	23	14
Failure to maintain set	2	16
Total errors (raw)	51	6

(Continued)

Table 1. Continued

Test	Score	Percentile/interpretation
NIH Cognitive Toolbox		
Fluid composite (standard score)	81	10
Crystallized composite (standard score)	112	79
Total composite (standard score)	96	39
Picture vocabulary (standard score)	113	81
Oral reading recognition (standard score)	110	75
List sorting (standard score)	90	25
Picture sequence memory (standard score)	127	96
Flanker inhibitory task (standard score)	78	7
Dimensional card sort (standard score)	78	7
Pattern comparison (standard score)	66	1
PROMIS		
Physical functioning (*T* score)	42	21
Anxiety (*T* score)	56	73
Depression (*T* score)	50	50
Fatigue (*T* score)	53	62
Sleep disturbance (*T* score)	62	55
Ability to participate in social activities (T score)	40	16
Pain interference (*T* score)	64	92

Notes: Normative values for Boston Naming Test, COWA, Animal Fluency, and Trail Making are from Heaton and colleagues (2004). Normative values for the Rey Auditory Verbal Learning Test are from the Schmidt (1996) meta-norms. Normative values for the Rey–Osterrieth are from Meyers and Meyers (1995).

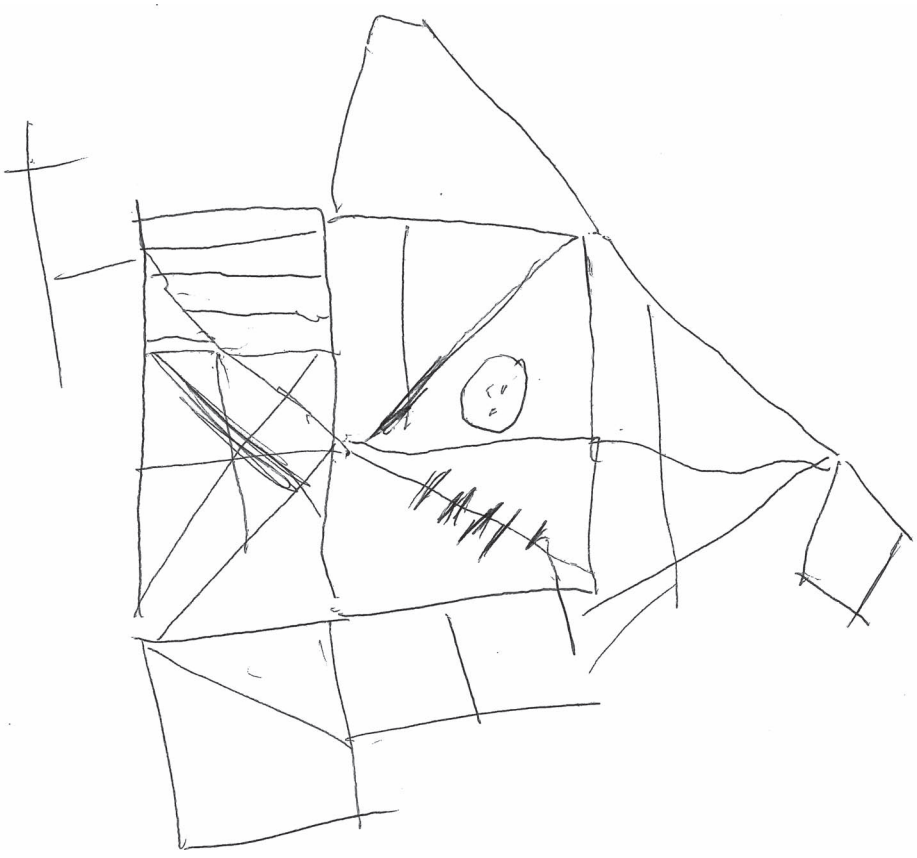

Fig. 1. Rey-Osterrieth Complex Figure copy.

Working memory. Working Memory was the lowest WAIS-IV domain score (WMI = 80), with a forward digit span length of 5 and a backward digit span length of 3. NIH Toolbox List Sorting was average.

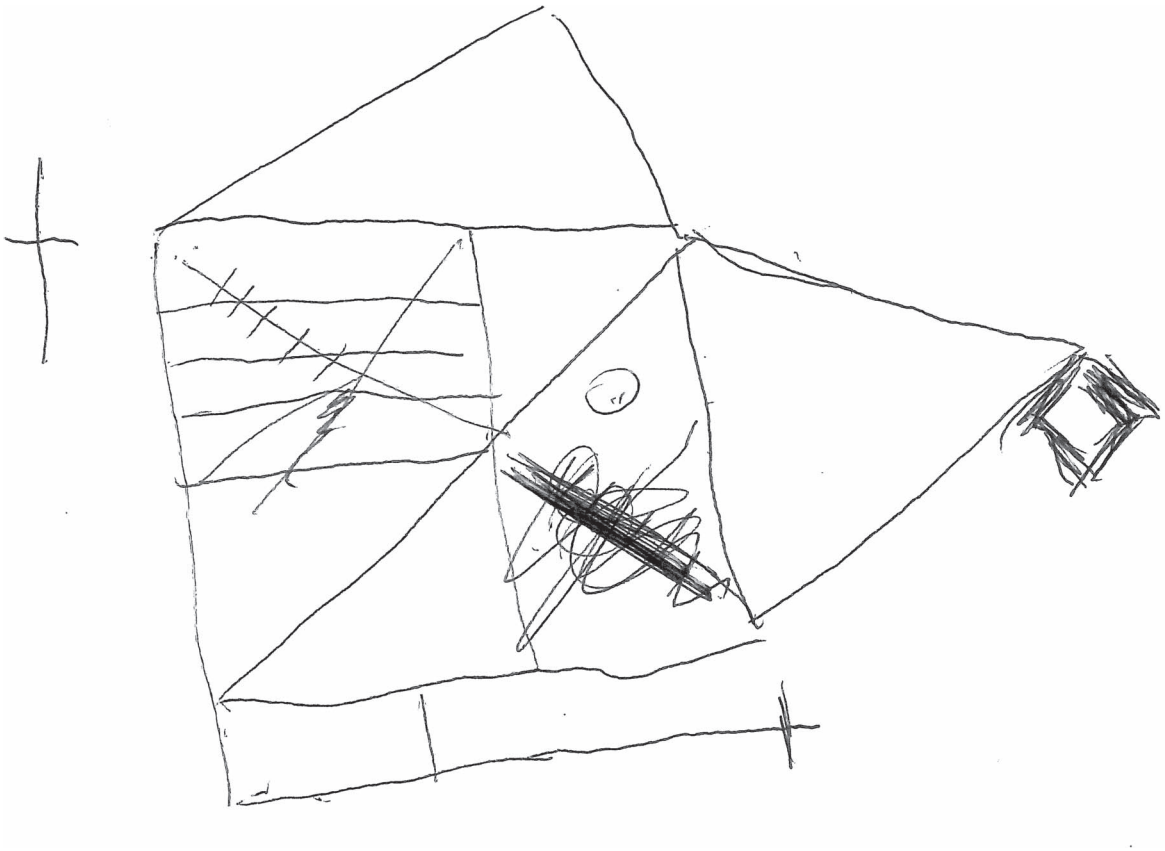

Fig. 2. Rey-Osterrieth Complex Figure immediate recall.

Learning and memory. Verbal ability to learn a list over repeated trials was normal including normal delayed free recall and recognition. Prose passage recall was average for both immediate and delayed recall conditions. Memory for simple designs was average. Although poor recall for the Complex Figure was present, this is thought to be due in part to poor organization during encoding/copy. There was no loss of information over the 30-min delay, with meaningful performance improvement when tested with recognition. NIH Toolbox Picture Sequence Learning Memory was superior.

Processing speed. Processing Speed was in the borderline range (PSI = 84). Other measures of processing speed including Trail Making Part A and NIH Toolbox Pattern Comparison were below the first percentile.

Executive function. Executive functioning was poor, a pattern consistent with white matter disease/MS (Roman & Arnett, 2016). Although generative verbal fluency to letter prompts was low average, it was meaningfully lower relative to category prompts suggesting relative impairment of frontal–subcortical/executive function. Other neuropsychological evidence of impaired executive function included poor response alternation on Trail Making Test Part B with several sequencing errors, impaired crossed response inhibition, poor visual constructional performance on the Complex Figure, poor simple design copy, and relatively poor hypothesis generation and response shifting on the WCST. Executive function measures from the NIH Cognitive Toolbox (Dimensional Card Sort, Flanker Inhibitory Task) were borderline.

Self-report. Her Personality Assessment Inventory (PAI) profile did not suggest either positive or negative impression management; all PAI validity indices were $T < 50$. There were no clinical PAI scale elevations (all T scores < 70), with Depression ($T = 52$) and Anxiety ($T = 54$) both within normal limits. The single treatment elevated scale was stress ($T = 73$), although there were concerns regarding health/somatic function (SOM $T = 67$). On the PROMIS (Patient-Reported Outcomes

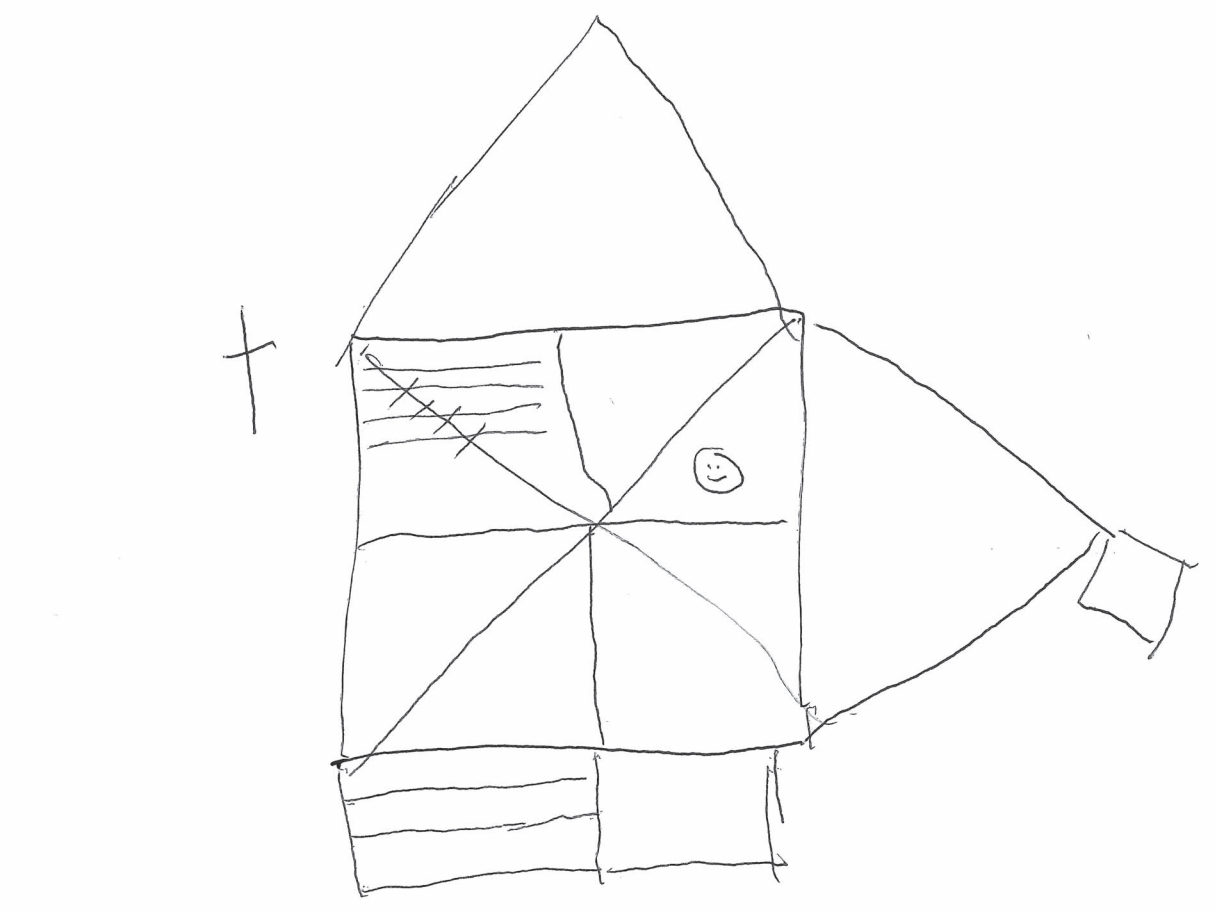

Fig. 3. Rey-Osterrieth Complex Figure delayed recall.

Measurement Information System, Cella et al., 2007), she had moderate elevations ($T > 60$) on pain interference and sleep disturbance.

Discussion

This case illustrates interpretive challenges when PVT and neuropsychological test results suggest contrasting interpretive conclusions. We discuss our approach to clinical case conceptualization in which we hypothesize that low PVT scores reflect disease-related effects of MS, but also present considerations for neuropsychologists who interpret low PVT scores to reflect invalid responding.

MSVT and WMT Score Interpretations

We do not expect most MS patients to obtain subthreshold scores on computerized PVT measures. However, we evaluated task demands and disease features to identify possible factors to account for poor computerized PVT scores on repeated occasions. The primary similarity between the VSVT and WMT is computerized presentation of target stimuli for only 5 or 6 s. Short exposure is a plausible factor because MS patients have decreased information processing speed and impaired working memory, with decreased information processing speed being a bigger contributor to cognitive inefficiency (DeLuca et al., 2004). For this patient, decreased processing speed is seen on multiple tests that include Coding, Trail Making, and Pattern Comparison. In addition, she was taking zonisamide, a drug with a high risk of cognitive impairment that includes decreased processing speed and working memory (Loring et al., 2007). Drugs with centrally acting effects such as lorazepam have robust effects on WMT performance, decreasing WMT validity performance by 8% in a double-blind, crossover study (Loring et al., 2011).

Single stand-alone PVTs are often given early during the assessment before the development of rapport associated with longer examiner interactions. During the second assessment, our patient was likely aware that the initial testing was not completed due to some aspect of her performance, which may have increased test anxiety. Thus, there are multiple factors including disease, medication, and anxiety that plausibly affected her ability to encode material presented for only 5–6 s.

Continuation Versus Termination of Testing

Agreement does not exist regarding the appropriate course of action after obtaining PVT scores suggesting performance invalidity. A survey of North American neuropsychologists reported the most common approach is to continue the evaluation, although approximately 20% of forensic evaluations and 51% of clinical evaluations are stopped early (Martin et al., 2015). The higher likelihood of test stoppage following failed PVT scores reflects the belief that further testing is unlikely to yield meaningful results. This approach was followed during the patient's first neuropsychological testing. Termination assumes that low PVT accurately reflects poor task engagement independent of the clinical condition and that additional testing will yield scores that are similarly invalid and non-clarifying. However, discontinuing testing may alert the patient to the critical tests used to characterize performance validity and will facilitate the likelihood of good performance on these measures in the future. This reasoning likely explains the discrepancy between clinical and forensic discontinuation figures reported in the survey in which more forensic evaluations are terminated following low PVT scores compared to purely clinical assessments.

Several arguments exist for continuing assessment following poor PVT scores, particularly in the context of clinical referrals. The first, illustrated by the present case, is that poor PVT scores may result from neurologic disease, from cognitive side effects of medications (zonisamide), or both. Thus, PVT scores genuinely reflect diminished cognitive performance associated clinical disease semiology, and the inference of "invalidity" is inappropriately made based upon specious data from limited clinical research. Neuropsychological profiles obtained in this context are generally considered to reflect minimal estimates of true ability levels, but to the extent that an impaired PVT reflects impaired brain function, the overall profile can be interpreted to reflect true ability levels. Despite poor PVT scores in this patient, which are thought to be related to disease-consistent white matter effects on processing speed, continuing the neuropsychological evaluation permitted the primary referral question addressing memory to be answered. If initial PVT failure is considered to reflect performance invalidity, normal embedded PVTs following failed stand-alone PVTs may be interpreted as reflecting engagement fluctuation, which is the rationale for assessing performance validity multiple times throughout an assessment (Heilbronner et al., 2009). This further supports continuing testing in clinical patients if initial PVT scores suggest performance invalidity.

Continuing the assessment provided the opportunity to directly address the referral question. Her overall neuropsychological profile is consistent with expectations for MS patients (Benedict et al., 2017; Chiaravalloti et al., 2013). Normal learning and memory were the most important neuropsychological findings addressing the specific referral question. Based upon normal performance across multiple memory measures, we concluded that despite subjective concerns of poor memory, her complaints were likely related to cognitive inefficiencies from poor frontal–subcortical/executive function.

By continuing the evaluation, the possible misinterpretation of patient motivation by other clinicians reading her electronic medical record (EMR) is minimized. Although the initial neuropsychological report cautioned that there were multiple factors that could have been contributing to subthreshold PVT scores, a rapid reading of the report may simply suggest poor volitional engagement and patient "malingering." These judgments are particularly difficult to dispel in EMRs given the tendency to auto populate fields with information from prior clinic visits (Bowman, 2013). We advocate to our neuropsychology trainees a similar dictum as followed by our physician colleagues: "*primum non nocere" (first, do no harm).* Reporting PVT scores as reflecting poor effort based upon thresholds that have not been appropriately validated in the neurological condition of interest creates an unnecessary hardship for the patient and family, potentially affecting longer term medical care.

Risks of Incorrect Inference

The optimal cutpoint for any diagnostic test not only incorporates prior probabilities of the alternative conditions, but also the benefits and costs associated with positive versus false negative diagnostic decisions (Swets, 1992). The cost–benefit tradeoff has not been adequately discussed in the PVT literature and will vary based upon assessment setting. In our patient, stating that the neuropsychological profile is invalid, even if correct, is unlikely to alter the course of treatment and care of her MS. Stating that the neuropsychological profile is valid, even if incorrect, allows inferences about normal memory function to be given to the patient. Thus, the costs associated with incorrectly inferring a valid neuropsychological evaluation in this case are not great. In general, costs and benefits related to correct/incorrect decisions should be considered when making PVT inferences in appropriate clinical evaluations when contradictory neuropsychological findings are present. This is the same approach used

when interpreting discrepant findings in our programmatic neuropsychological evaluations for possible Deep Brain Stimulation or Epilepsy Surgery.

"Ontological-Epistemological One-Worldness" and Choosing Between Apparent Contradictory Results

Faust's (2003) provocative article introduces the concept of "ontological-epistemological one-worldness," that is, the belief that "careful analysis and synthesis (of seemingly discrepant data) allows one to integrate them into meaningful or orderly results and patterns" (p. 430). PVT scores are often given interpretive priority in the interpretation of neuropsychological test findings regardless of the referral source. The finding of low PVT scores provides a simple "ontological-epistemological one-worldness" explanation to account for unexpected performance patterns by simply inferring "poor effort" while discounting normal performance variability in which cognitive healthy individuals frequently have scores in the impaired range (Binder et al., 2009; Brooks et al., 2009). Considering PVT scores as primary in profile interpretation can lead to similar assertions such as a patient being able to perform at even higher levels should some PVT scores be at or below recommended thresholds, even in the context of normal neuropsychological scores in the average range. This is, of course, speculative and we are not aware of the literature that has demonstrated this empirically.

If one accepts that PVT characteristics are presently unknown in various neurological conditions, then "a key task for the interpreter is not to synthesize the data but to decide what to use and what to disregard or throw out" (Faust, 2003, p. 431). We interpreted WMT outside of the traditional cut-scores framework because, in addition to WMT not being systematically studied in MS, we identified a plausible disease-related explanation for low WMT scores. While we acknowledge the rationale for giving PVT interpretive priority in medicolegal applications where PVT characteristics have been studied extensively, we suggest that PVT scores should be treated no differently than any other neuropsychological test measure in the absence of appropriate PVT validation when interpreting test results, some of which may be internally inconsistent.

Faust (2003) also emphasizes the importance of incorporating base rate information when making clinical inference. When low base rates are present, most patients with positive test findings nevertheless will not have the condition of interest, even with highly sensitive and specific measures (i.e., low positive predictive power; Robinson et al., 2016). Unfortunately, the base rate of response exaggeration in treatment seeking, non-litigating MS patients is unknown. Nevertheless, we expect the base rate in this group small given the absence of external incentives to perform poorly. In contrast, the target false positive rate for PVTs is often identified as 10% (Larrabee, 2012; Victor et al., 2009), although false positive PVT failure rates are likely higher in patients with independently established neurologic disease (e.g., Dean et al., 2009; Grote et al., 2000; Keary et al., 2013; Loring et al., 2016; Macciocchi et al., 2006). If true, then by virtue of base rate differences alone, we believe the interpretation of credible performance is more likely based solely upon base rate difference. "In some circumstances, we know in advance that base rates are superior to alternative signs or combinations of signs and that we will always defer to the base rates in situations of conflict" (Faust, 2003, p. 436). Failure to consider base rates may lead to the "false positive paradox" in which even tests with high specificity are associated with more false positive errors than true positive identifications when incidence of the condition is lower than the test's false positive rate (Wolpaw et al., 2010). Unfortunately, base rates are frequently not incorporated into diagnostic decision-making, although they are important in establishing post-test (posterior) probability.

Summary

This case suggests that computerized PVTs may be affected in some MS patients by disease-related decreased information processing speed and working memory impairment. Even for those who disagree, there remains a need for viewing the apparently valid neuropsychological test profile within a consistent interpretative framework that addresses superficially incompatible findings within some logical decision-making framework. Different recommendations regarding best practice after obtaining subthreshold scores on PVT measures will vary based upon the clinical context in which the evaluations are being performed (i.e., forensic vs. medical). However, this case demonstrates that credible patterns of neuropsychological performance addressing the primary referral question may be obtained in at least some neurologic patients with subthreshold PVTs scores. Clinical neuropsychological evaluation requires judicious use of clinical judgment, and a "one-size fits all" approach to interpretation and assessment protocol when subthreshold PVT scores are obtained is insufficient to replace good clinical practice.

Conflict of Interest

None declared.

Acknowledgements

We thank Samantha John, Ph.D., Bruce Hermann, Ph.D., and our helpful reviewers for their comments on earlier versions of this report.

References

Benedict, R. H. B., DeLuca, J., Enzinger, C., Geurts, J. J. G., Krupp, L. B., & Rao, S. M. (2017). Neuropsychology of multiple sclerosis: Looking back and moving forward. *Journal of the International Neuropsychological Society*, *23* (*9–10*), 832–842 doi:10.1017/s1355617717000959.

Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24* (*1*), 31–46 doi:10.1093/arclin/acn001.

Brooks, B. L., Iverson, G. L., Feldman, H. H., & Holdnack, J. A. (2009). Minimizing misdiagnosis: Psychometric criteria for possible or probable memory impairment. *Dementia and Geriatric Cognitive Disorders*, *27* (*5*), 439–450 doi:10.1159/000215390.

Bowman, S. (2013). Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspectives in Health Information Management*, *10*, 1c Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3797550/pdf/phim0010-0001c.pdf.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H. et al. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20* (*4*), 419–426 doi:10.1016/j.acn.2005.02.002.

Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B. et al. (2007). The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Medical Care*, *45*(*5 Suppl 1*), S3–s11 doi:10.1097/01.mlr.0000258615.42478.55.

Chiaravalloti, N. D., Stojanovic-Radic, J., & DeLuca, J. (2013). The role of speed versus working memory in predicting learning new information in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *35* (*2*), 180–191 doi:10.1080/13803395.2012.760537.

Davis, J. J., Millis, S. R., & Axelrod, B. N. (2012). Derivation of an embedded Rey Auditory Verbal Learning Test performance validity indicator. *The Clinical Neuropsychologist*, *26* (*8*), 1397–1408 doi:10.1080/13854046.2012.728627.

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, *23* (*1*), 133–152 doi:10.1080/13854040701819050.

DeLuca, J., Chelune, G. J., Tulsky, D. S., Lengenfelder, J., & Chiaravalloti, N. D. (2004). Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? *Journal of Clinical and Experimental Neuropsychology*, *26* (*4*), 550–562 doi:10.1080/13803390490496641.

Farr, S. P., Greene, R. L., & Fisher-White, S. P. (1986). Disease process, onset, and course and their relationship to neuropsychological performance. In Filskov, S. B., & Boll, T. J. (Eds.), *Handbook of Clinical Neuropsychology*, (pp. 213–253). New York: Wiley.

Faust, D. (2003). Holistic thinking is not the whole story. *Assessment*, *10* (*4*), 428–441 doi:10.1177/1073191103259534.

Green, P. (2005). *Green's Word Memory Test for windows user's manual-revised*. Edmonton, Alberta: Green's Publishing.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6* (*3*), 218–224 doi:10.1037/1040-3590.6.3.218.

Greve, K. W., Heinly, M. T., Bianchini, K. J., & Love, J. M. (2009). Malingering detection with the Wisconsin Card Sorting Test in mild traumatic brain injury. *The Clinical Neuropsychologist*, *23* (*2*), 343–362 doi:10.1080/13854040802054169.

Grote, C. L., Kooker, E. K., Garron, D. C., Nyenhuis, D. L., Smith, C. A., & Mattingly, M. L. (2000). Performance of compensation seeking and non-compensation seeking samples on the Victoria Symptom Validity Test: Cross-validation and extension of a standardization study. *Journal of Clinical & Experimental Neuropsychology*, *22*, 709–719 doi:10.1076/jcen.22.6.709.958.

Hamberger, M. J., & Seidel, W. T. (2003). Auditory and visual naming tests: Normative and patient data for accuracy, response time, and tip-of-the-tongue. *Journal of the International Neuropsychological Society*, *9* (*3*), 479–489 doi:doi-org.proxy.library.emory.edu/10.1017/S1355617709090754.

Hawkins, K. A., Plehn, K., & Borgaro, S. (2002). Verbal IQ-performance IQ differentials in traumatic brain injury samples. *Archives of Clinical Neuropsychology*, *17* (*1*), 49–56 doi:dx.doi.org/10.1016/S0887-6177(00)00099-8.

Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults*. Lutz, FL: Psychological Assessment Resources.

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 doi:10.1080/13854040903155063.

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 doi:10.1080/13854040801945060.

Keary, T. A., Frazier, T. W., Belzile, C. J., Chapin, J. S., Naugle, R. I., Najm, I. M. et al. (2013). Working memory and intelligence are associated with Victoria Symptom Validity Test hard item performance in patients with intractable epilepsy. *Journal of the International Neuropsychological Society*, *19* (*03*), 314–323 doi:10.1017/S1355617712001397.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 625–631 doi:10.1017/S1355617712000240.

Loring, D. W., Goldstein, F. C., Chen, C., Drane, D. L., Lah, J. J., Zhao, L. et al. (2016). False-positive error rates for reliable digit span and Auditory Verbal Learning Test performance validity measures in amnestic mild cognitive impairment and early Alzheimer disease. *Archives of Clinical Neuropsychology*, *31* (*4*), 313–331 doi:10.1093/arclin/acw014.

Loring, D. W., Lee, G. P., & Meador, K. J. (2005). Victoria Symptom Validity Test performance in non-litigating epilepsy surgery candidates. *Journal of Clinical and Experimental Neuropsychology*, *27* (*5*), 610–617 doi:10.1080/13803390490918471.

Loring, D. W., Marino, S., & Meador, K. J. (2007). Neuropsychological and behavioral effects of antiepilepsy drugs. *Neuropsychology Review*, *17* (*4*), 413–425 doi:10.1007/s11065-007-9043-9.

Loring, D. W., Marino, S. E., Drane, D. L., Parfitt, D., Finney, G. R., & Meador, K. J. (2011). Lorazepam effects on Word Memory Test performance: A randomized, double-blind, placebo-controlled, crossover trial. *The Clinical Neuropsychologist*, *25* (*5*), 799–811 doi:10.1080/13854046.2011.583279.

Lu, P. H., Boone, K. B., Cozolino, L., & Mitchell, C. (2003). Effectiveness of the Rey–Osterrieth Complex Figure Test and the Meyers and Meyers recognition trial in the detection of suspect effort. *The Clinical Neuropsychologist*, *17* (*3*), 426–440 doi:10.1076/clin.17.3.426.18083.

Macciocchi, S. N., Seel, R. T., Alderson, A., & Godsall, R. (2006). Victoria Symptom Validity Test performance in acute severe traumatic brain injury: Implications for test interpretation. *Archives of Clinical Neuropsychology*, *21* (*5*), 395–404 doi:10.1016/j.acn.2006.06.003.

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, *29* (*6*), 741–776 doi:10.1080/13854046.2015.1087597.

Meyers, J. E., & Meyers, K. R. (1995). *Rey Complex Figure Test and recognition trial*. Odessa, FL: Psychological Assessment Resources.

Pearson (2009). *Advanced clinical solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual*. San Antonio, TX: Author.

Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2018). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology. Adult*, *25* (*3*), 219–230 doi:10.1080/23279095.2017.1278602.

Robinson, J. S., Collins, R. L., Miller, B. I., Pacheco, V. H., & Wisdom, N. M. (2018). The severe impairment profile: A conceptual shift. *Archives of Clinical Neuropsychology*, *33* (*2*), 238–246 doi:10.1093/arclin/acx069.

Robinson, M., Boissoneault, J., Sevel, L., Letzen, J., & Staud, R. (2016). The effect of base rate on the predictive value of brain biomarkers. *The Journal of Pain*, *17* (*6*), 637–641 doi:10.1016/j.jpain.2016.01.476.

Roman, C. A., & Arnett, P. A. (2016). Structural brain indices and executive functioning in multiple sclerosis: A review. *Journal of Clinical and Experimental Neuropsychology*, *38* (*3*), 261–274 doi:10.1080/13803395.2015.1105199.

Schmidt, M. (1996). *Rey Auditory and Verbal Learning Test: A handbook*. Los Angeles: Western Psychological Services.

Slick, D. J., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *Manual for the Victoria Symptom Validity Test*. Odessa, FL, Psychological Assessment Resources.

Swets, J. A. (1992). The science of choosing the right decision *47* (*4*), 522–532. doi:dx.doi.org/10.1037/0003-066X.47.4.522.

Sweet, J. J., Benson, L. M., Nelson, N. W., & Moberg, P. J. (2015). The American Academy of Clinical Neuropsychology, National Academy of Neuropsychology, and Society for Clinical Neuropsychology (APA division 40) 2015 TCN professional practice and 'Salary Survey': Professional practices, beliefs, and incomes of U.S. neuropsychologists. *The Cllinical Neuropsychologist*, *29* (*8*), 1069–1162 doi:10.1080/13854046.2016.1140228.

Victor, T. L., Boone, K. B., Serpa, J. G., Buehler, J., & Ziegler, E. A. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23* (*2*), 297–313 doi:10.1080/13854040802232682.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J. et al. (2013). Cognition assessment using the NIH toolbox. *Neurology*, *80*(*11*Supplement 3), S54–S64 doi:10.1212/WNL.0b013e3182872ded.

Wolpaw, B. J., Mathews, C., Chopra, M., Hardie, D., de Azevedo, V., Jennings, K. et al. (2010). The failure of routine rapid HIV testing: A case study of improving low sensitivity in the field. *BMC Health Services Research*, *10*, 73 doi:10.1186/1472-6963-10-73.