Diagnosing solvent-induced chronic toxic encephalopathy : the effect of underperformance in neuropsychological testing

International Journal of Psychiatry in Clinical Practice

Original article

November 2018

Diagnosing solvent-induced chronic toxic encephalopathy: the effect of underperformance in neuropsychological testing

Running Title: underperformance in testing solvent-induced chronic toxic encephalopathy

Authors: Mestdagh Irmgard[1,2*], van Bergen Liesbeth[2,3] , Kocken Claudia[2,4], Heyvaert Vicky[5] , Cras Patrick[5], Van Den Eede Filip[1,2]

[1]University Department of Psychiatry, Campus Antwerp University Hospital (UZA), Antwerp (Edegem), Belgium

[2] Collaborative Antwerp Psychiatric Research Institute (CAPRI), Faculty of Medicine and Health Sciences, University of Antwerp (UA), Antwerp, Belgium

[3] Department of Abdominal Surgery, Antwerp University Hospital (UZA), Antwerp (Edegem), Belgium

[4] General Practice, Antwerp

[5] Department of Neurology, Institute Born Bunge, University of Antwerp, Faculty of Medicine and Health Sciences, Antwerp University Hospital (UZA), Antwerp (Edegem), Belgium

Software: Microsoft Word

Words abstract: 175

Words text: (introduction – discussion): 3724

(introduction - references): 5343

Pages: 24

Tables: 2

Figures: 0

* Corresponding Author:

Irmgard Mestdagh, MD

University Department of Psychiatry, Campus Antwerp University Hospital (UZA)

Wilrijkstraat 10, 2650 Edegem (Antwerp), Belgium, Europe

Tel: 00 32 3 821 39 38

Fax: 00 32 3 825 16 41

E-mail: Irmgard.mestdagh@uza.be

Abstract

Objective: The diagnoses of solvent-induced chronic toxic encephalopathy (CSE) can be supported by neuropsychological tests. However, since results not only reflect cognitive symptoms but also the patient's effort to perform well, this study examines to what extent underperformance impacts neuropsychological outcomes in individuals referred for suspected CSE.

Methods: A retrospective study of 48 suspected CSE patients having completed ten neuropsychological tests assessing different domains of cognition. Underperformance was identified using the Amsterdam Short-Term Memory Test and the Rey 15-item Memory Test. Multiple linear regression was applied to examine the effect of insufficient effort on test performance.

Results: A total of 54.1 % of the patients were identified as having underperformed on one or both performance validity tests. Analyses showed a significant effect of underperformance on most tests barring Letter-Number Sequencing.

Conclusion: Most of the neuropsychological tests evaluated showed significant effects of underperformance. Performance on Letter-Number Sequencing was not affected. In case of underperformance, the results of the neuropsychological assessment should be disregarded when weighing the final multi-disciplinary diagnosis, with the exception of Letter-Number Sequencing.

Keywords: Organic-solvent-induced chronic toxic encephalopathy, Underperformance, Malingering, Neuropsychological Tests, Depression, Performance Validity

Introduction

Organic-solvent-induced chronic toxic encephalopathy (CSE), also called painters' disease, is a persistent neuropsychiatric condition caused by long-term exposure to solvents such as paint, glue, ink and detergents. Inhalation is the most common way of absorption, but transdermal absorption has also been described (Geens et al. 2009). The first reports about solvent-induced CSE originated from Northern Europe (Hänninen et al. 1976; Arlien-Søborg et al. 1979). In Belgium, the disease was recognised as an occupational disease in 1998 (Federal Fund for Occupational Diseases 1999), with its incidence having declined during the last decade due to improved health-safety measures.

The three-tier classification of solvent-induced CSE defined during the 1986 Raleigh (North Carolina) consensus meeting is still used today (Hänninen et al. 1976; Arlien-Søborg, 1979), with type 1 being characterised by aspecific symptoms such as fatigue, unstable affect, and reversible deterioration of memory and concentration, and type 2a reflecting a permanent change in personality. In these two types, no permanent cognitive deterioration is present. Starting from type 2b, the memory and concentration problems are irreversible, as are the reduced learning capacities and emotional lability, while, in addition, peripheral neuropathy and cerebellar dysfunction can be observed. In type 3, symptoms of dementia appear with pronounced deterioration of memory functions and intelligence (Geens et al. 2009; Triebig & Hallermann 2001; Cranmer & Goldberg 1986).

A CSE diagnosis typically involves (partial or permanent) work disability, severely affecting the patient's self-esteem, social and financial status. As the disability allowance is based on CSE type and severity and the extent of the patient's disablement (Orbaek et al. 1987), a correct diagnosis is crucial, also to contain the burden on national care systems and society.

The 2012 European consensus on neuropsychological assessment in CTE (van Valen et al. 2012) formulated six important recommendations: 1) A CSE diagnosis should involve the assessment of sufficient and verified exposure to neurotoxic organic solvents, while there has to be a temporal relationship

between the solvent exposure and the onset of symptoms. Since the diagnosis of CSE is made by exclusion, other potential causes for the symptoms presented need to be eliminated first. Also, a differential diagnosis should include treatment of major pathology prior to the start of the neurological assessment. 2) The European consensus stipulates that the neuropsychological assessment should test the following cognitive domains: attention (processing speed and complex attention), memory (immediate recall, delayed recall and recognition), concept formation and reasoning (verbal and non-verbal), fine motor performance (motor speed and dexterity), and construction. 3) The tests additionally need to provide an indication of the severity of any cognitive impairment. 4) The assessment of symptoms other than cognitive deficiencies needs to be performed separately. 5) The neuropsychological testing should include at least one test to assess performance validity. When underperformance is suspected, the diagnostic process should be considered inconclusive (van Valen et al. 2012). 6) If no definitive diagnosis can be made, reassessment should be conducted after a minimum of 12 months.

The recommendations of the European consensus underline the importance of an objective assessment of performance validity before abnormal neuropsychological test scores are interpreted as being reflective of cognitive deficits. Performance validity refers to 'the extent to which a person's test performance is or is not an accurate reflection of their actual level of ability' (Larrabee 2012). Indeed, sometimes patients underperform (or show poor performance validity), either intentionally or unintentionally, complicating the interpretation of the results. It is a clinical challenge to distinguish true underperformance from deliberate types of underperformance such as feigning and intentional exaggeration of cognitive deficits (simulation or malingering) as both are crucial aspects in insurance medicine and financial compensation (Iverson and Binder 2000).

To date, only a few studies have attempted to delineate underperformance during neuropsychological testing in patients with suspected CSE. Van Hout et al. (2003) demonstrated that test scores were valid in as little as 54% of the patients tested. In their follow-up study van Hout et al. (2006)

found that 23% of the patients that were assessed had underperformed and that this group was significantly more frequently involved in litigation procedures. Greve et al. (2006) estimated the prevalence of cognitive malingering after toxic exposure to be around 40%.

To delineate underperformance more closely, two types of performance validity tests (PVTs) can be administered in addition to the neuropsychological battery. The first type concerns tests that are based on the indices of existing tests. The second more recent category comprises tests that are specifically designed to assess mental effort independent of cognitive functioning usually by means of a 'forced-choice recognition design' (Rudman et al. 2011) such as the Amsterdam Short-Term Memory Test (ASTM) (Schmand and Lindeboom 2005). In the studies by van Hout et al. (2003, 2006) underperformance was gauged using multiple PVTs including the ASMT and the test of Memory Malingering (TOMM), while in the 2006 study participants also completed the Warrington's Recognition Memory Test for Faces (RMT).

Still, the influence of underperformance on widely used neuropsychological tests is mostly unclear, rendering the interpretation of neuropsychological exams as a whole insecure. Currently, it is assumed that neuropsychological test results are not representative when a patient achieves a positive score on the ASTM (van Hout et al. 2006). With the current study we seek to gain more clarity in how underperformance impacts the neuropsychological test outcomes of patients with a suspected CSE diagnosis referred for assessment.

Material and Methods

Procedure

Between 2002 and 2010, the Belgian Fund for Occupational Diseases referred 57 patients with suspected CSE for a second independent opinion to the Department of Psychiatry of the Antwerp University Hospital, Belgium, where for each patient a systematic, standardised record was compiled based on information obtained from the patient's clinical history and the results on neuropsychological tests. For our current study, a database was created retrospectively that included the patients' medical and social data, and test scores, among which are pertinent subtests of the Wechsler Adult Intelligence Scale-III (WAIS-III). Of the 57 patients tested, the datasets of nine were excluded because of incomplete data, leaving 48 datasets for analysis.

Instruments and variables

Independent variable: performance validity tests

Amsterdam Short-Term Memory Test (ASTM) delineates underperformance using 30 items, allowing a maximum score of 90, with a cut-off score of 85 reflecting under. We refer to the manual (Schmand and Lindeboom 2005) for more detailed information on the testing and scoring procedures. Test scores were entered as dichotomous variables (Schagen et al. 1997), with underperformance being coded as 1 and 'normal' performance as 0. The ASTM has been shown to be a sensitive and valid instrument to assess underperformance (Schagen et al. 1997; Bolan et al. 2002). Comparing simulators with neurological patients with severe epilepsy, multiple sclerosis or cerebral concussion, validity studies indicated that, at a cut-off score of 85, the ASTM has an 84% sensitivity and a 90% specificity. More experimental trials contrasting healthy simulators with controls achieved a sensitivity and specificity above 90% (Dandachi-FitzGerald et al. 2011).

Rey 15-Item Memory Test (FIT): Patients are shown a grid containing 15 items for 10 seconds and asked to recall as many items as possible. A score of less than 7 is considered to indicate underperformance. Test scores were entered as dichotomous variables, with underperformance being coded as 1 and 'normal' performance as 0. The FIT has a very low sensitivity (10%) and high specificity (95%) (Reznek 2005).

Dependent variables

Trail Making Test A (TMT-A): Part A of the TMT requires the patient to connect 25 numbers in sequential order on a sheet of paper by drawing lines as fast as possible, with performance being scored as the time in seconds needed to complete the test successfully. Scores are interpreted using standardised tables. The test assesses the speed of information processing, visual scanning, visual-motor speed, coordination, attention and visual-spatial skills (Reitan 1992). Part B of the TMT had to be omitted from our analyses because of incomplete data.

Letter-Number Sequencing (WAIS-III): The task involves listening to and remembering a string of digits and letters read aloud at a speed of one per second, then recalling the information by repeating the numbers in chronological order, followed by the letters in alphabetical order (Klinkenberg and Kooij 2005). Three series are presented, allowing a maximum score of 21. The test gauges visualisation, sequencing abilities, and short-term auditory memory and processing speed.

Coding (WAIS-III): Using a key (boxes containing a numeral in the top line and a symbol in the bottom line) patients have to solve as many of the 133 items as possible within two minutes. The test measures attention, concentration, effort, and accuracy (Klinkenberg et al. 2005).

Working Memory Index (WMI-WAIS-III): This index is calculated on the basis of three subtests: Arithmetic, Digit Span, Letter-Number Sequencing (Klinkenberg et al. 2005).

Rey Auditory-Verbal Learning Test (RAVLT): Patients are asked to memorise as many words as possible from a list of 15 unrelated words read to them, of which they need to reiterate as many as possible. The procedure is repeated four times using the same list, while after 30 minutes a recall trial and recognition test are presented. The RAVLT gauges various components of memory (Rey 1958).

The Rey Complex Figure Test (Rey CFT): In this test patients are asked to reproduce a complicated line drawing (see Osterrieth, 1944), first by copying and then from memory. In this study we used the results from the 30 minute-recall condition. The Rey CFT evaluates visuospatial abilities and visual memory by non-intentional learning (Osterrieth 1944).

Perceptual Organization Index: This index is calculated on the basis of three subtests of the WAIS-III: 'Picture Completion', 'Block Design,' and 'Matrix Reasoning' (Klinkenberg et al. 2005).

Stroop Colour Word Test: In this study we applied an alternative interference score, which was derived by subtracting the Colour score (card II) from the Colour-Word score (card III) (Klinkenberg et al. 2005, Trenerry et al. 1989). The obtained score was interpreted using standardised tables.

Verbal Comprehension Index: This index is calculated on the basis of three subtests of the WAIS-III: 'Vocabulary', 'Similarities,' and 'Information' (Klinkenberg et al. 2005).

Processing Speed Index: This index is calculated on the basis of two subtests of the WAIS-III: 'Digit Symbol Coding' and 'Symbol Search' (Klinkenberg et al. 2005).

Depression

Beck Depression Inventory (BDI-II-NL): the Dutch (NL) version of the second edition of the BDI is a standardised 21-item self-report questionnaire measuring depression severity. Items are rated on a 4-point scale (0-3), with total scores ranging between 0 and 13 indicating minimal depression. Scores between 14 and 19 are indicative of mild, scores of 20-28 of moderate, and scores of 29-63 of severe depression, where scores >20 are labelled as clinical depression (Beck et al. 1961; Beck et al. 1996).

Statistics

SPSS Statistics (version 19) was used for all analyses and a significance threshold of $p<0.05$ was adopted.

For our regression analyses, the patients' ages, exposure duration and test scores were examined for normal distribution using the Shapiro-Wilk test and by interpreting the histograms, Q-Q plots, and detrended Q-Q plots. When data were not normally distributed, they were transformed to their logarithms, which were subsequently tested for normal distribution.

Number and percentages of ASTM and FIT scores were calculated. Because of the dichotomous nature of the FIT, the Fisher's exact test was used to compare the number of underperformers in the two performance validity tests.

Multiple linear regression analyses were performed to examine to what extent underperformance had affected the neuropsychological test results. To this end, we composed three performer groups based on the scores recorded for the performance validity tests (PVTs, i.e. the ASTM and the FIT), with the first group having passed the two tests (full pass PVTs), the second having passed one of the two tests only (fail

one PVT), and the third group having failed both tests (fail two PVTs). The performer groups were entered as independent variables and the test scores as the dependent variables. For each of the linear regression analyses the normal distribution of the residuals was tested.

One-way ANOVAs were performed to compare the three performer groups as to age, exposure duration, and BDI-II outcomes. Significant between-group differences in these variables were taken into account in the multiple linear regression analyses.

Results

Sample description

The study sample (n=48) consisted of 43 men (89.6%) and 5 women (10.4%), mean age was 49 years (± 6.69; range 35-60 yrs), mean solvent exposure time 27 years (± 9.98; range 10-45 yrs), with 35 participants being diagnosed with type-2b CTE (72.9%) and 13 (27.1%) with a milder type of CSE (type 1 or type 2a). As to occupations, 50.0% of the patients were painters and 31.3% printers. The remaining patients were joiners, carpenters, chemists, laboratory workers, or workers in the polyester industry.

Based on their PVT scores, 22 patients (45.8%) were identified as having performed at 'normal' levels on both tests, 18 (37.5%) as having passed only one PVT and eight (16.6%) as having failed both tests (Table 1).

(Table 1)

Age and exposure duration were normally distributed in the three performer groups, with the one-way ANOVA not finding any significant differences (age: F=2.247, p=.117; exposure duration: F=0.034, p=.967). BDI-II scores were not normally distributed and the Kruskal-Wallis test yielded a significant difference between the three performer groups (mean rank= 17.33 (Full pass PVT group), 29.44 (Fail one PVT group), 30.34 (Fail two PVTs group); Chi-Square= 9.840, p=.007). Consequently, BDI-II scores were included in the statistical analysis because of their confounding potential.

The effects of underperformance

Preparatory to the linear regression analyses, the distribution of the scores on the neuropsychological tests was investigated. The Coding, TMT-A and Stroop Interference scores were not normally distributed. After logarithmic transformation the distributions of the last two tests were normal, but not for Coding. However, since the Q-Q plots showed a normal distribution, we did include the test in our analyses.

The multiple linear regression analyses showed that the Letter-Number Sequencing test is not significantly predicted by underperformance, while in the other nine tests its influence was always significant (Table 2). Between-group comparisons showed that underperformance most significantly influenced the Perceptual Organization Index and Processing Speed Index in the two underperformance groups, while for the other tests (Rey 15-item memory test, Rey Complex Figure test, Interference score, Coding, Verbal Comprehension Index and Working Memory Index) this was the case in the 'fail two PVT' group failing both PVTs only.

The multiple linear regression analyses (adjusted R^2 change= 0.012, F=3.066, p=0.038) showed that in the TMT-A the significant effect of underperformance disappeared after correction for depression in the two underperformer groups (t (1vs.2)= 1.997, p=0.052; t(1vs.3)=0.849, p=400), with the BDI-II itself not significantly (t=1.258, p=0.215) influencing the TMT-A.

(Table 2)

Discussion

Underperformance on the various tests of a neuropsychological test battery widely used to assess individuals with suspected CSE was retrospectively delineated in a sample of 48 patients referred to our psychiatry department. In order to create more clarity for the clinician, we explored to which extent the outcomes were not significantly predicted by underperformance based on their scores on two performance validity tests (ASTM and the FIT). Our analyses revealed 54.1% to have underperformed, 37.5% of whom had failed only one PVT, while 16.6% had failed both PVTs. This is comparable with the results reported by van Hout et al (2003), where 46% of the CSE patients had underperformed, with 26.9% having failed one and 18.6% both PVTs.

The multiple linear regressions we ran showed that the majority of the test results was significantly predicted by underperformance, most specifically the Perceptual Organisation Index (R^2: 35.1%) and the Processing Speed Index (R^2: 27.2%). We also observed a noticeable effect (R^2: 19% to 9.5%) on the scores of the tests assessing mental flexibility, verbal comprehension, word fluency, memory and the attention tests 'Coding' and 'TMT-A'. Although they can be useful to get a general impression of the minimal cognitive capacities and the strengths and weaknesses of underperformers, these tests should be used with caution in the diagnosis of CSE (Bush et al. 2005).

At the group level, and with the exception of the perceptual organisation and processing speed indices, we found the underperformance effect to be significant in the patients having failed both PVTs only. This discrepancy may be attributable to a reduction of the false-positive and false-negative results and the resulting filtering out of 'dubious' underperformers from the full fail group. Also the relatively small sample size may have rendered the statistical power insufficient. Still, and supporting the literature, our results underscore the relevance of multiple, both embedded and stand alone, performance tests with good psychometric properties throughout CSE assessments, assuming that failure on at least two PVTs

corroborates the likelihood of intentional underperformance (Heilbronner et al., 2009; Medici 2013; Larrabee 2014), where the European consensus finds a single PVT sufficient (Van Valen et al. 2012).

Interestingly, being a measure of attention and part of the Working Memory Index, Letter-Number Sequencing was the only neuropsychological test that was not significantly influenced by underperformance. Morrow et al (1991) reported that in CSE the first deficits present themselves at the level of attention and information processing, i.e. as a dysfunction in working memory. Investigating the effect of underperformance in patients with brain injury, Stulemeijer et al (2007) likewise observed a significant association between a positive test result on the ASTM and various domains of cognitive functioning (viz. attention, verbal memory, information processing, and motor speed). However, the results reported by Stulemeijer et al. (2007) and Rudman et al (2011) both partly contradict our current findings in that in our study attention outcomes, as gauged by the Letter-Number Sequencing and the TMT-A, were not significantly influenced by underperformance.

As expected, the Kruskal-Wallis test showed the BDI-II-NL scores to be significantly higher in the group of under than they were in the 'normal' performers, pointing to a higher degree of depression in the former group. Yet, a difficulty in the interpretation of the effect of depression is that the validity of self-reported depressive symptoms (BDI) has not been established. In fact, patients who underperform on neuropsychological tests may be more likely to also overreport depressive symptoms (Dandachi-FitzGerald et al. 2011, Webb et al. 2012). Also, depression can both be a component of CSE (van Valen et al. 2012) and a possible cause of underperformance (Webb et al. 2012). The causal role of depression in underperformance has as yet not been confirmed although the evidence is growing according to more recent studies (Webb et al. 2012). As the multiple linear regression showed a significant influence of depression on the TMT-A and the effect of underperformance was no longer significant after correction for depression, we can presume that poor performance does not significantly predict TMT-A outcomes. Ihrig et al (2005) observed a relationship between the duration of exposure to solvents and reduced scores

on the TMT-A and information processing speed. Our results also suggest that poor scores are indeed due to exposure to solvents. Like Letter-Number Sequencing, the TMT-A is a test that measures attention.

In addition to depressive, underperformance could be due to motivational problems, fatigue, stress, pain, fear of failure, abuse, or lack of cooperation, among other causes (Geens et al. 2009; Stulemeijer et al. 2007; Rudman et al. 2011). The need for confirmation of subjective complaints can also play a role (Schmand et al. 1998) as underperformance is more frequent in patients presenting with subjective symptoms as in chronic whiplash syndrome (Kessels et al. 2000), chronic fatigue syndrome (Werf et al. 2000), post-concussion syndrome (Binder et al. 1997) and CSE. As patients were assessed in a center for second opinions, there is a higher probability that patients show decreased performance validity. Therefore we cannot rule out that the underperformance was influenced and aggravated by compensation seeking behavior.'

Some limitations of our study warrant discussion. Because of its retrospective design, various patients had to be excluded because of incomplete data, causing the sample size to be relatively small. A potential bias can therefore not be ruled out. Furthermore, since we did not check for the validity of self-reported depressive symptoms, the BDI scores may have been affected by a response bias (i.e. symptom overreporting). Several studies also question the psychometric properties of the FIT (Reznek 2005) because of its low sensitivity, which increases the risk of false negatives. Lastly, the ASTM is generally considered an easy memory test that is independent of cognitive functioning. However, Merten et al. (2007) showed that almost all the Alzheimer patients they tested failed the ASTM and the Word Memory Test, while all control patients passed both tests. The ASTM should therefore not be used for patients suffering from CSE type 3 as this type is also characterized by dementia. In our study, none of the patients had been diagnosed with the severer type.

In the light of our study, a prospective study that includes a larger patient sample and multiple symptom validity tests, both embedded and stand alone, with good psychometric characteristics is

required. The multiple symptom validity tests then need to gauge both underperformance and symptom overreporting given that these are behavioural manifestations of symptom validity and both vulnerable to response bias. Here, we wish to point out that the European consensus does not provide explicit recommendations regarding the assessment of symptom overreporting. Additionally, the diagnostic criteria of CSE warrant further refinement and assessment considering that the current diagnosis is primarily based on neuropsychological examination. Imaging and psychomotor examination (van Valen et al. 2012) could contribute to the diagnostic process in analogy to their role in the diagnosis of major depressive disorder (Mayberg 2009; Schrijvers et al. 2009). The European consensus indeed already recommends incorporating a fine motor performance test (gauging motor speed and dexterity) in the neuropsychological assessment. Lastly, our results could be translated for the use in other clinical patient groups as underperformance is not only common in litigation procedures and forensic settings, but also in a wide range of neurological (e.g. epilepsy, toxic encephalopathy, traumatic brain injuries) and psychiatric conditions (e.g. Neurocognitive disorders, ADHD, Autism spectrum disorder) (Dandachi-FitzGerald et al. 2011).

In conclusion and confirming the existing literature, underperformance (as measured by the ASTM and the FIT) is a major concern in patients suspected for CSE and has a significant influence on most neuropsychological tests. However, the scores on Letter-Number Sequencing that reflect attention were not significantly influenced by underperformance. These findings should be kept in mind when interpreting the outcomes of neuropsychological tests in CSE.

Key points

- A total of 54.1 % of patients with suspected CSE referred for neuropsychological assessment was identified as having underperformed on one or both performance validity tests.

- Underperformance has a significant effect on most neuropsychological tests with the exception of Letter-Number Sequencing assessing attention and working memory.
- In case of underperformance, the results of the neuropsychological assessment should be disregarded when weighing the final multi-disciplinary diagnosis, with the exception of Letter-Number Sequencing.

Acknowledgements

No acknowledgements

Disclosure of interests

None to declare

References

- Arlien-Søborg P, Bruhn P, Gyldensted C, Melgaard B. 1979. Chronic painters' syndrome. Chronic toxic encephalopathy in house painters. *Acta Neurologica Scandinavica*. 60:149–56.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. 1961. An inventory for measuring depression. *Archives of General Psychiatry.* 4: 561–571.
- Beck AT, Steer RA, Brown GK. 1996. *Manual for the Beck Depression Inventory, Second Edition (BDI-II-NL-II)*. San Antonio, TX: The Psychological Association.
- Binder LM, Rohling ML, Larrabee J. 1997. A review of mild head trauma. Part I: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology*. 19:421–431.
- Bolan B, Foster JK, Schmand B, Bolan S. 2002. A comparison of three tests to detect feigned amnesia: the effects of feedback and the measurement of response latency. *Journal of Clinical and Experimental Neuropsychology.* 24:154-67.
- Bush SS, Ruff RM, Tröster AI, Barth JT, Koffler SP, Pliskin NH, et al. 2005. Symptom validity assessment: Practice issues and medical necessity. NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*. 20:419–426.
- Cranmer JM, Goldberg L, editors. 1986. Human aspects of organic solvents on the central nervous system and diagnostic criteria. *Neurotoxicology*. 7: 45–56.
- Dandachi-FitzGerald B, Ponds R, Peters M, Merckelbach H. 2011. Cognitive Underperformance and Symptom Over-Reporting in a Mixed Psychiatric Sample*. The Clinical Neuropsychologist*. 25: 812-828.
- Fonds voor de Beroepsziekten (Federal Fund for Occupational Diseases) 1999. *Criteria inzake preventie en schadeloosstelling van het organisch psychosyndroom veroorzaakt door solventen.* Brussels, Belgium: FBZ.

- Geens T, Viaene M, Heyvaert V, Cosyns P. 2009. Het organisch psychosyndroom veroorzaakt door solventen: diagnostische knelpunten. *Tijdschrift voor Geneeskunde*. 65: 1038-1048.
- Greve K, Bianchini K , Black F , Heinly M , Love M , Swift D , Ciota M. 2006. The prevalence of cognitive malingering in persons reporting exposure to occupational and environmental substances. *Neurotoxicology*. 27: 940-950.
- Hageman G, van der Hoek JAF, van Hout MS, van der Laan G, Verberk MM. 2006. Enkele ontwikkelingen in de neurotoxicologie Deel A: oplosmiddelen en resultaten Solvent Teamproject 1997-2003. *Tijdschrift voor Neurologie en Neurochirurgie*. 107:127-135.
- Hänninen H, Eskelinen L, Husman K, Nurminen M. 1976. Behavioral effects of long term exposure to a mixture of organic solvents. *Scandinavian Journal of Work, Environment & Health*. 2:240–255.
- Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, Conference Participants. 2009. American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*. 23:1093-129.
- Ihrig A, Dietz MC, Bader M, Triebig G. 2005. Longitudinal study to explore chronic neuropsychological effects on solvents exposed workers. *Industrial Health*. 43:588-596.
- Iverson GL, Binder LM. 2000. Detecting exaggeration and malingering in neuropsychological assessment. *The Journal of Head Trauma Rehabilitation*. 15:829–858.
- Kessels RP, Aleman A, Verhagen WI, van Luijtelaar EL. 2000. Cognitive functioning after whiplash injury: a meta-analysis. *Journal of the International Neuropsychological Society*. 6:271–278.
- Klinkenberg E, Kooij AP. 2005. *Technische handleiding WAIS-III.* Amsterdam, Nederland: Harcourt test publishers.
- Larrabee GJ. 2008. Aggregation Across Multiple Indicators Improves the Detection of Malingering: Relationship to Likelihood Ratios. *The Clinical Neuropsychologist*. 22: 666-679.

- Larrabee GJ. 2012. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society.* 18:625-30.
- Larrabee GJ. 2014. False-Positive Rates Associated with the Use of Multiple Performance and Symptom Validity Tests. *Archives of Clinical Neuropsychology*. 29: 364–373.
- Lee G, Loring D, Martin R. 1992. Rey's 15-item visual memory test for the detection of malingering. Normative observations on patients with neurological disorders. *Psychological Assessment*. 4:43-46.
- Mayberg HS. 2009. Targeted electrode-based modulation of neural circuits for depression. *Journal of Clinical Investigation*. 119: 717-25.
- Medici R. 2013. The Use of Multiple Performance Validity Tests. *Journal of Forensic Psychology Practice*. 13: 68-78.
- Merten T, Bossink L, Schmand B. 2007. On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*. 29:308-318.
- Morrow LA, Ryan CM, Hodgson MJ, Robin N. 1991. Risk factors associated with persistence of neuropsychological deficits in persons with organic solvent exposure. *Journal of Nervous and Mental Disease*. 179:540–545.
- Orbaek, P, Lindgren M, Haeger-Aronsen B. 1987. Computed tomography and psychometric test performances in patients with solvent induced chronic toxic encephalopathy and healthy controls. *British Journal of Industrial Medicine*. 44:175-179.
- Osterrieth P. 1944. Le test de copie d'une figure complex. *Arch Psychological*. 30:206-356.
- Reitan RM. 1992. *Trailmaking test.* Tucson, Arizona, USA: Reitan neuropsychology laboratory.
- Rey A. 1958. *L'Examen clinique en psychologie.* Paris, France: Press Universitaire de France.

- Reznek L. 2005. The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*. 19: 539–543.
- Rudman N, Oyebode JR, Jones CA, Bentham P. 2011. An investigation into the validity of effort tests in a working age dementia population. *Aging & Mental Health*. 15:47-57.
- Schagen S, Schmand B, de Sterke S, Lindeboom J. 1997. Amsterdam Short-Term Memory test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology*. 19: 43–51.
- Schmand B, Lindeboom J. 2005. Amsterdam Short-Term Memory Test. Manuel. Leiden, the Netherlands: PITS.
- Schmand B, Lindeboom J, Schagen S, Heijt R, Koene T, Hamburger HL. 1998. Cognitive complaints in patients after whiplash injury: The impact of malingering. *Journal of Neurology, Neurosurgery and Psychiatry*. 64: 339–343.
- Schmand B, Ponds R. 2004. Suboptimaal presteren: simuleren, aggraveren en somatiseren. In:Deelman B, Eling P, de Haan E., van Zomeren AH, editors. Klinische neuropsychologie, Amsterdam: Boom, pp. 523-539.
- Schrijvers D, Van Den Eede F, Maas Y, Cosyns P, Hulstijn W, Sabbe BG. 2009. Psychomotor functioning in chronic fatigue syndrome and major depressive disorder: a comparative study. *Journal of Affective Disorders*. 115:46-53.
- Stulemeijer M, Andriessen TMJC, Brauer JMP, Vos PE, Van Der Werf S. 2007. Cognitive performance after Mild Traumatic Brain Injury: The impact of poor effort on test results and its relation to distress, personality and litigation. *Brain Injury*. 21:309–318.
- Trenerry M, Crosson B, DeBoe J, Leber W. 1989. *The Stroop Neuropsychological Screening Test.* New York, NY, USA: Psychological Assessment Resources.

- Triebig G, Hallermann J. 2001. Survey of solvent related chronic encephalopathy as an occupational disease in European countries. *Journal of Occupational and Environmental Medicine.* 58:575-81.
- van Hout MS, Schmand B, Wekking EM, Hageman G, Deelman BG. 2003. Suboptimal performance on neuropsychological tests in patients with suspected chronic toxic encephalopathy. *Neurotoxicology*. 24:547-551.
- van Hout MS, Schmand B, Wekking EM, Deelman BG. 2006. Cognitive functioning in patients with suspected chronic toxic encephalopathy: evidence for neuropsychological disturbances after controlling for insufficient effort. *Journal of Neurology, Neurosurgery and Psychiatry*. 77:296-303.
- van Valen E, van Thriel C, Akila R, Nilson LN, Bast-Pettersen R, Sainio M, van Dijk F, van der Laan G, Verberk M, Wekking E. 2012. Chronic solvent-induced encephalopathy: European consensus of neuropsychological characteristics, assessment, and guidelines for diagnostics. *Neurotoxicology*. 33:710-726.
- Webb JW, Batchelor J, Meares S, Taylor A, Marsh NV. 2012. Effort test failure: toward a predictive model. *The Clinical Neuropsychologist*. 26:1377-96.
- Werf SP, van der Prins JP, Jongen PJ, Meer JW, van der Bleijenberg G. 2000. Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment: a matched comparison between chronic fatigue syndrome and multiple sclerosis. *Neuropsychiatry Neuropsycholology and Behavioral Neurology*. 13:199–203.

Figure and Table legends

- Table 1
  Crosstabulation of the patients suspected of CSE having obtained low and normal scores on the Rey 15-item memory test and the ASTM

- Table 2
- Mean scores (SD) on the neuropsychological tests for the three performer groups; results of multiple regressions. PVT, performance validity test.

Tables

Table 1: Crosstabulation of the number of patients suspected of CSE having obtained low and normal scores on the Rey 15-item memory test and the ASTM

		ASTM		Total
		0	1	
Rey 15-item memory test	0	22	14	36
	1	4	8	12
Total		26	22	48

The Fisher's exact test showed no significant difference between the number of patients who had failed one or both performance validity tests (p=.094). ASTM, Amsterdam Short-Term Memory Test

Table 2: Mean scores (SD) on the neuropsychological tests for the three performer groups; results of multiple regressions.
PVT, performance validity test.

Dependent variable	Group 1 PASS two PVT	Group 2 FAIL one PVT	Group 3 FAIL two PVT	R^2	F (P-value)	B	S.E	Group comparison t (p-value)
Trail Making Test A	48.00 (19.021)	68.21 (28.849)	59.62 (18.236)	0.143	3.760 (0.031)	Group2: 20.211 Group3: 11.625	Group2: 7.395 Group3: 9.703	1 vs.2: 2.733 (0.009) 1 vs.3: 1.198 (0.237)
Letter-Number Sequencing	7.48 (2.159)	7.26 (3.364)	5.50 (3.117)	0.061	1.437 (0.270)	Group 2: -0.213 Group3: -1.976	Group2: 0.902 Group3: 1.183	1 vs.2:-0.263 (0.814) 1 vs.3: -1.670 (0.102)
Rey 15-item memory test	39.76 (10.104)	37.42 (7.113)	29.88 (7.699)	0.144	3.795 (0.030)	Group2: -2.341 Group3: -9.887	Group2: 2.739 Group3: 3.594	1 vs.2: -0.855 (0.397) 1 vs.3: -2.751 (0.009)
The Rey Complex Figure Test	16.33 (4.004)	13.68 (5.334)	10.86 (4.140)	0.158	4.138 (0.023)	Group2: -2.649 Group3: -5.476	Group2: 1.460 Group3: 2.013	1 vs.2: -1.814 (0.076) 1vs.3: -2.721 (0.009)
Interference Score	49.76 (25.608)	68.47 (42.355)	87.12 (44.228)	0.131	3.388 (0.043)	Group2: 18.712 Group3: 37.363	Group2:11.474 Group3:15.057	1 vs.2: 1.631 (0.110) 1 vs.3: 2.481 (0.017)
Perceptual Organization Index	98.57 (11.461)	81.21 (17.008)	74.50 (10.730)	0.351	12.177 (0.000)	Group2: -17.361 Group3: -24.071	Group2:4.387 Group3: 5.757	1 vs.2: -3.957 (0.000) 1 vs.3: -4.181 (0.000)
Coding	7.00 (2.294)	5.42 (2.950)	3.75 (1.753)	0.190	5.153 (0.010)	Group2: -1.579 Group3: -3.250	Group2: 0.805 Group3: 1.052	1 vs.2: -1.960 (0.056) 1 vs.3: -3.090 (0.003)
Verbal Comprehension Index	87.86 (9.936)	85.11 (14.212)	74.75 (10.110)	0.137	3.570 (0.036)	Group2: -2.752 Group3: -13.107	Group2: 3.754 Group3: 4.926	1 vs.2: -0.733 (0.467) 1 vs.3: -2.661 (0.011)
Processing Speed Index	86.29 (10.565)	77.89 (14.869)	65.75 (9.004)	0.272	8.391 (0.001)	Group2: -8.391 Group3: -20.536	Group2: 3.886 Group3: 5.100	1 vs.2: -2.159 (0.036) 1 vs.3: -4.027 (0.000)
Working Memory Index	85.62 (9.373)	81.63 (12.239)	73.62 (11.759)	0.095	3.481 (0.039)	Group2: -3.987 Group3: -11.994	Group2: 3.475 Group3: 4.560	1 vs.2:-1.147 (0.257) 1 vs.3:-2.630 (0.012)