



Neuropsychologists' ability to predict distorted symptom presentation

Brechje Dandachi-FitzGerald, Harald Merckelbach & Rudolf W. H. M. Ponds

To cite this article: Brechje Dandachi-FitzGerald, Harald Merckelbach & Rudolf W. H. M. Ponds (2017) Neuropsychologists' ability to predict distorted symptom presentation, Journal of Clinical and Experimental Neuropsychology, 39:3, 257-264, DOI: [10.1080/13803395.2016.1223278](https://doi.org/10.1080/13803395.2016.1223278)

To link to this article: <http://dx.doi.org/10.1080/13803395.2016.1223278>



© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 07 Sep 2016.



[Submit your article to this journal](#)



Article views: 102



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

Neuropsychologists' ability to predict distorted symptom presentation

Brechje Dandachi-FitzGerald^a, Harald Merckelbach^b and Rudolf W. H. M. Ponds^a

^aDepartment of Medical Psychology, Maastricht University Medical Centre, School for Mental Health and Neuroscience, Maastricht, The Netherlands; ^bForensic Psychology Section, Maastricht University, Maastricht, The Netherlands

ABSTRACT

Objective. We explored to what extent experienced neuropsychologists can predict distorted symptom presentation of clinically referred hospital outpatients. **Method.** Using clinical files and interview results, 31 neuropsychologists made predictions as to how 203 patients would perform on two response validity tests. Their predictions were matched against actual passing or failing two such tests, of which one measured overreporting of symptoms and the other underperformance on cognitive tests. **Results.** Clinical predictions and test outcomes agreed in 76% of the cases, with Cohen's kappa being .26, 95% confidence interval, CI [.08, .44]. Of the 152 patients for whom neuropsychologists had predicted nondistorted symptom presentations, 14 patients (9.2%) failed both response validity tests. Of the 51 patients for whom neuropsychologists had predicted problematic response validity, 35 patients (68.6%) passed both tests. **Conclusions.** Clinical prediction of distorted symptom presentation is far from perfect. Our findings show that response validity tests have incremental value in that they may correct initial clinical judgment.

ARTICLE HISTORY

Received 15 April 2015
Accepted 3 August 2016

KEYWORDS

Amsterdam Short Term Memory Test (ASTM); Clinical judgment; Hospital outpatients; Performance validity; Structured Inventory of Malingered Symptomatology (SIMS); Symptom validity.


Several studies have found that neuropsychologists are not very good at detecting distorted symptom presentations in their patients. For example, Heaton, Smith, Lehman, and Vogt (1978) provided 10 neuropsychologists with cognitive test profiles of instructed feigners and head-injured patients. The neuropsychologists had to determine whether profiles reflected genuine impairment or feigning. Their accuracy was far from perfect and ranged from chance level to 20% above chance. Similar results were reported by Faust, Hart, and Guilmette (1988) and Faust, Hart, Guilmette, and Arkes (1988). These authors found that the majority of neuropsychologists seemed to believe that the cognitive test profiles of adolescents and children instructed to feign reflected cortical dysfunction. Without forewarning on the base rate of malingering, none of the experts indicated feigning as a plausible explanation of the test results.

These older studies are, however, not without limitations (see for a critical review, Garb & Schramke, 1996). They were carried out in an era

when distorted symptom presentation was under-researched, and clinicians had only few tools to screen for it (e.g., Merten et al., 2013). Since the early nineties, there has been a steady increase in empirical and meta-analytic studies evaluating the diagnostic accuracy of tools designed to identify distorted symptom presentation (e.g., Sweet & Guidotti Breting, 2013). In keeping with the terminology of other authors (Larrabee, 2012; Van Dyke, Millis, Axelrod, & Hanks, 2013), we use response validity tests as an overarching term for these tools. Broadly speaking, there are two types of response validity tests: (a) self-report symptom validity tests (SVTs) that intend to measure overreporting of symptoms, and (b) performance validity tests (PVTs) that assay underperformance on cognitive tests (Greve, Bianchini, & Brewer, 2013).

Major professional organizations in neuropsychology have issued policy statements that stress the routine use of SVTs and PVTs (National Academy of Neuropsychology, Bush et al., 2005; American Academy of Neuropsychology,

CONTACT Brechje Dandachi-FitzGerald  b.fitzgerald@maastrichtuniversity.nl  School for Mental Health and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

 Supplemental data for this article can be accessed [here](#).

© 2016 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). These developments reflect neuropsychologists' increased sensitivity to the diagnostic option of distorted symptom presentation and how it can be evaluated with specific tests. Indirect support for this increased awareness comes from a study by Trueblood and Binder (1997), who provided their experts with protocols of clinical cases of feigning rather than protocols of experimental (i.e., instructed) simulators. These protocols included SVT and PVT results. The authors found that the percentage of false negatives (i.e., missed feigned test protocols) across experts varied between 0 and 25% (Trueblood & Binder, 1997). Although neuropsychologists in this study were better able to detect symptom distortion than were experts in older studies (e.g., Faust, Hart, Guilmette, & Arkes, 1988), Trueblood and Binder (1997) relied on strongly distorted test profiles (e.g., below chance performance on a PVT), when in practice less extreme cases are more common and are usually the cases of greatest concern. Accordingly the potential of false negatives might be higher than 25%. Still, even if one would accept 25% as an upper bound estimate of false negatives, there is room for considerable improvement.

Can neuropsychologists accurately predict distorted symptom reporting prior to the administration of any SVTs and/or PVTs? This is an important question whenever response validity tests are not incorporated in test batteries per default. In a recent survey among European neuropsychologists (Dandachi-FitzGerald, Merten, & Ponds, 2013), only 12% of the respondents indicated that they included response validity tests in every or nearly every clinical neuropsychological assessment. Thus, many European neuropsychologists will be regularly confronted with the question of whether or not they should add response validity tests to their test battery. Yet, can neuropsychologists reliably predict on the basis of their clinical impression whether SVTs/PVTs have incremental value? Our study addressed this issue. We anticipated that neuropsychologists would not be particularly accurate in predicting the outcome of these tests, and more specifically that they would underestimate poor response validity. After all, the limited accuracy of clinical judgment was the primary reason to develop psychometric approaches to detect distorted symptom presentation (Wedding & Faust, 1989). Thus, we examined how well experienced neuropsychologists can predict the outcome of response validity testing in a

heterogeneous group of general hospital outpatients referred for neuropsychological evaluation.

Method

Participants

This study is part of a larger research project on SVT/PVT failure in clinically referred hospital outpatients and how such failure relates to performance on standard clinical instruments and external incentives (Dandachi-FitzGerald, Van Twillert, Van De Sande, Van Os, & Ponds, 2015). In total, 31 neuropsychologists in five hospitals in the southern part of the Netherlands participated in the study. At the end of the Dandachi-FitzGerald et al. (2015) study, clinicians received an exit questionnaire (see below) that asked for additional background information. All in all, 29 (94%) experts returned the questionnaire. Table 1 gives background information about this group. On average, clinicians had 10 years of experience, and more than two thirds of them were certified clinical psychologist/neuropsychologist or certified psychologist (i.e., psychologists with two-year postgraduate clinical training).

The initial sample consisted of 469 patients (52.7% women), with a mean age of 47.7 years ($SD = 14.0$, range = 17–78), who were referred for (neuro)psychological evaluation on the basis of medical considerations. Patients with severe cognitive impairment (e.g., moderate–severe Alzheimer's disease, posttraumatic amnesia) were excluded. The main five diagnostic categories upon referral were: neurological conditions (24.1%), morbid obesity (20.5%), medically unexplained symptoms (MUS; 18.1%), psychiatric disorders (15.8%), and cognitive complaints not further specified (15.1%). A total of 203 patients (43.1%) received social security benefits (full or partial), and 77 patients (16.4%) were involved in legal proceedings at the time of the assessment. Of the initial sample, 292 patients (62.3%) were referred for neuropsychological evaluation between July 2012 and May 2013, and it is this subgroup that is considered below. All diagnostic assessments were conducted for clinical purposes. Detailed information about the patient sample is given in Dandachi-FitzGerald et al. (2015).

Only protocols that contained data on clinicians' predictions and the outcomes of two response validity tests—the Amsterdam Short Term Memory Test (ASTM) and the Structured Inventory of Malingered

Table 1. Background of neuropsychologists

Neuropsychologists	Mean (range)
Age (years)	36.0 (27–62)
Years of work experience	10.4 (1.5–35)
Number of annual neuropsychological assessments	69.1 (5–230)
Increased sensitivity to symptom distortion due to taking part in this study (0–10)	5.9 (1–9)
	% sample
Sex (female)	79.3
Professional title	
Professional title	20.6
Certified clinical psychologist/neuropsychologist	
Clinical psychologist/neuropsychologist trainee	13.7
Certified psychologist (two-year postgraduate training)	34.5
Psychologist trainee	24.1
Psychologist (only university degree)	6.9
Conducting forensic evaluations (mean number of forensic evaluations in the past year; range)	41.4 (1.4; 0–4)
Administration of response validity tests prior to start study forensic evaluations in the past year; range)	
Standard	24.1
Only certain situations	72.4
Never	6.9
Indications for administration of response validity tests ^a	
Indications for administration of response validity tests ^a :	95.5
Symptom reports during interview	
Presentation of the patient during the interview	100
Diagnosis upon referral ^b	31.8
On request of referral party	22.7
When incentives are present (e.g.e.g., legal procedures, disability claims)	90.9
Other ^{**c}	18.2

Note. *N* = 29.

^aThere were seven missing responses for this question; percentages are based on 22 responses. ^bType of diagnosis upon referral that led to inclusion of symptom validity tests (SVTs): conversion, functional symptoms, whiplash-associated disorders, chronic fatigue syndrome, commotion cerebri, and personality disorder. ^cOther indications: inconsistency in scores noticed during testing; impression of the technician during testing; behavioral presentation during testing; stagnation of treatment.

Symptomatology (SIMS; see below)—were included in the final analyses. There were eight cases with missing data on the ASTM, eight cases with missing data on the SIMS, and five cases with missing prediction data. These 21 assessments were excluded, leaving 271 cases for the final analysis.

In patient samples seen for clinical assessment, the base rate of poor symptom validity is estimated to be around 10% (Dandachi-FitzGerald et al., 2013; Mittenberg, Patton, Canyock, & Condit, 2002). With such a relatively low frequency, aggregating results across multiple response validity indicators reduces the probability of false positives (Larrabee, 2008).¹

¹This is only true when indicators are relatively independent of one another. In the current study, the shared variance (*R*²) between SIMS and ASTM was 16% (for the total sample; *N* = 271).

²Information on neuropsychologists' prediction versus test performance for the subsample of patients who failed one of either test is given in the supplemental file.

Table 2. Characteristics of patients who passed or failed both response validity tests

Patients	Patients who failed both tests (<i>n</i> = 30)	Patients who passed both tests (<i>n</i> = 173)
	Mean (range)	Mean (range)
Age (years)	54.6 (32–71)	48.6 (17–77)
Education ^a	3.57 (1–8)	4.56 (1–8)
SIMS total score	26.7 (19.3–41)	8.85 (0–19)
ASTM total score	74.8 (60–81)	86.5 (82–90)
	% sample	% sample
Sex (female)	43.3	43.9
Main diagnostic categories		
Neurological conditions	30.0	38.7
Medically unexplained symptoms	20.0	17.0
Psychiatric disorders	23.3	11.0
Cognitive complaints not further specified	16.7	25.4
Medical conditions with cognitive complaints	3.3	4.0
Other	6.7	2.9
Missing	0	1.2

Note. *N* = 203. SIMS = Structured Inventory of Malingered Symptomatology; ASTM = Amsterdam Short-Term Memory Test.

^aEducation was quantified with an 8-point scale that ranges from primary school (1; fewer than six years of education) to university degree (8; 16 years of education or more) (De Bie, 1987).

Given a base rate of 10%, the posttest probability of poor response validity when both SIMS and ASTM are failed is .93. The posttest probability of valid responding when both SIMS and ASTM are passed is .99. With this in mind, two groups were selected from the pool of 271 cases: a group of patients who passed both response validity tests (*n* = 173), and a group of patients who failed both response validity tests (*n* = 30).² The characteristics of these two groups are given in Table 2. As can be seen, the majority of patients in both groups had been diagnosed upon referral with a neurological condition, medically unexplained symptoms, or a psychiatric disorder.

The research protocol was reviewed and approved by the Medical Ethical Committee of Maastricht University Medical Centre.

Measures

Symptom validity test (SVT): Structured Inventory of Malingered Symptomatology (SIMS)

Test protocols included a Dutch research version of the SIMS (Merckelbach & Smith, 2003), which is a self-report scale that consists of 75 true–false

items pertaining to rare and/or improbable symptoms. Endorsed items are summed to obtain a total SIMS score, with higher scores indicating stronger tendencies to overreport. A recent meta-analysis combined 17 samples of nonoverreporting patients ($N = 742$) and found for this group a weighted total SIMS score of 16.1 with a 95% confidence interval (CI) from 13.4 to 18.9 (Van Impelen, Merckelbach, Jelicic, & Merten, 2014). Following the recommendations of this meta-analysis, we employed a cutoff of 19. In the initial sample ($N = 469$), the rates of overreporting on the SIMS (i.e., >19) ranged between 12% and 19% for the main diagnostic subgroups, with a significantly lower failure rate of 1% in the morbid obesity subsample. However, morbidly obese patients were never seen for neuropsychological evaluation, and therefore this diagnostic category is not represented in the current study.

Performance validity test (PVT): Amsterdam Short-Term Memory Test (ASTM)

The ASTM is a measure of cognitive underperformance and involves a forced-choice word recognition procedure (Schmand & Lindeboom, 2005). The number of correct recognitions are summed (range 0–90), with lower scores reflecting poorer performance. We used a cutoff score of 82. In the original validation studies that compared 84 experimental malingerers and 206 patients suffering from neurological conditions such as cerebral contusion, advanced Parkinson disease, stroke, multiple sclerosis, and severe epilepsy, this cut score was associated with a specificity of 98% and a sensitivity of 77% (Schmand & Lindeboom, 2005). In the initial sample, 29.9% of the patients failed on the ASTM (i.e., <82). The four main diagnostic subgroups did not differ with regard to failure rates on the ASTM (range = 25–37%).

Clinician checklist

Clinicians completed a checklist³ after their interview with the patient and after they had seen the patient files, but before the test session took place. Probing symptom distortion was not part of the interview, and patient files rarely contained information directly addressing diagnostic options such as feigning or malingering. Files could contain information about previous neuropsychological assessments including SVT/PVT results. However,

for the vast majority of the neuropsychological assessments, such prior information was not available.

The checklist addressed the following patient variables: age, gender, education, referring doctor, diagnosis upon referral, medication, social situation, employment status, type of income, and current involvement in legal proceedings. The final item asked clinicians to predict on a 3-point Likert scale (i.e., unproblematic–somewhat problematic–problematic) the outcome of subsequent response validity testing. Experts endorsed the problematic option only four times, and therefore we combined the problematic and somewhat problematic categories.

Exit questionnaire for clinicians

The exit questionnaire focused on the following items: name, age, function, years of work experience, estimated number of diagnostic assessments in the past year, type of diagnostic assessment (neuropsychological, psychological or both), and experience with forensic evaluations and, if so, the number of forensic evaluations in the past year. Clinicians were queried about how often they included response validity tests in their test batteries (i.e., always, in certain situations, never). When they said they incorporated SVTs/PVTs only under certain conditions, they were asked to tick one or more options that listed potential conditions (e.g., when the patient is involved in litigation). Furthermore, clinicians indicated on a 10-point Likert scale (0 = not at all; 10 = very strongly) whether participating in this study had sensitized them to distorted symptom presentation during their interview with the patient.

Procedure

Prior to testing, patients received an information letter about the current study, and they gave informed consent to use their anonymized test data for this study. The information letter contained a brief description of what a psychological assessment entails. The letter explained that in order to obtain valid diagnostic data, it is important that patients exert optimal effort at the cognitive tests and fill out the psychological questionnaires as accurately as possible. Next, the letter explained that sometimes patients do not

³The checklist may be obtained from the first author upon request.

succeed in this, and that the study addressed the question of how this can be measured. When the hospital's standard test battery did not include the SIMS and/or ASTM, these tests were added to the protocol. Tests were administered either by a certified psychological technician or by trained clinical psychology doctoral students working under the supervision of a certified psychologist. All of them were familiar with the SIMS and the ASTM and had experience with administering these tests.

Results

Clinicians' prediction of response validity

For 152 cases (74.9%), clinicians predicted that the response validity would be unproblematic, while for 51 cases (25.1%), they predicted it to be (somewhat) problematic. Table 3 shows how clinicians' predictions relate to passing or failing both SIMS and ASTM. Of the 152 patients who were expected to produce unproblematic response validity test scores, 14 patients (9.2%) failed both tests. Of the 51 patients who were anticipated to produce at least somewhat problematic SVT/PVT test results, 35 patients (68.6%) passed both tests. For 16 out the 30 patients (53.3%) who failed both tests, clinicians predicted correctly that distorted symptom presentation might be an issue. Overall, clinical prediction and test outcome agreed in 76% of the cases. The corresponding Cohen's kappa was .26 ($p < .001$, 95% CI [.08, .44]), which according to widely used standards would be qualified as fair (Landis & Koch, 1977). However, given the severe consequences of misclassifications in this type of clinical decision making, and considering the fact that clinicians were only right about half of time in their prediction of the 30 cases of distorted symptom presentation, we would evaluate this level of agreement as poor.

Table 3. Neuropsychologists' prediction versus test performance.

Neuropsychologists' prediction	SIMS and ASTM		n
	Pass both	Fail both	
Unproblematic	138	14	152
(Somewhat) problematic	35	16	51
<i>n</i>	173	30	203

Note. *N* = 203. SVT = Symptom Validity Test; SIMS = Structured Inventory of Malingered Symptomatology; ASTM = Amsterdam Short-Term Memory Test.

Clinicians' use of response validity tests in clinical assessments

The majority of the clinicians (i.e., 72.4%) stated that only in certain situations would they add SVTs/PVTs to their test battery (see Table 1). The most frequently reported situations were: how symptoms are presented by patients (100%), certain types of symptoms reported during the interview (95.5%), and the presence of incentives (90.9%). Taking part in this study led to an increased alertness to the issue of distorted symptom presentation (mean score of 5.9; range = 1–9).

Discussion

We had anticipated that neuropsychologists would not be particularly accurate in their prediction of distorted symptom presentation. Agreement between neuropsychologists' predictions and actual test outcome was, indeed, far from perfect, as indicated by a relatively low Cohen's kappa. For patients who failed both SIMS and ASTM, clinicians' predictions were at chance level (i.e., hit rate of 53.3%). Furthermore, we expected that neuropsychologists would underestimate the occurrence of poor response validity. Contrary to this, however, neuropsychologists overestimated the frequency of distorted symptom reports. That is, 14.7% of the patients failed both SIMS and ASTM, whereas a (somewhat) problematic outcome was predicted for 25.1% of the assessments. This pattern deviates from those of Faust and colleagues (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988), who found that clinicians underestimated the occurrence of distorted symptom reports. It may well be that our study sensitized clinicians to the topic and that this was conducive to overprediction of symptom distortion. Indeed, on the exit questionnaire, neuropsychologists often indicated that participation in the study had made them more aware of the diagnostic option of distorted symptom presentation. When clinicians are alerted to the possibility of poor response validity, this may lead them to question genuine symptoms (e.g., Rosenhan, 1973). For example, compared with earlier studies (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988), Trueblood and Binder (1997) found relatively low levels of false negatives (i.e., missing cases of distorted symptom presentation). However, three out of 26 clinicians (11.5%) in that

study generated false positives (i.e., misclassifying genuine cognitive impairments as a form of distorted symptom presentation). Thus, clinicians' sensitivity to distorted symptom presentation may come with the price of misclassifying genuine symptom reports as not valid.

In our study, 9.2% of the patients failed on both the SIMS and the ASTM, when neuropsychologists had expected in their cases a nondistorted symptom presentation. This percentage illustrates the risk of relying too much on clinical judgment: Clinicians may decide not to include SVTs and/or PVTs in their test batteries, which means that they miss an opportunity to provide themselves with critical feedback on their incorrect initial impression.

In line with previous survey findings (Dandachi-FitzGerald et al., 2013), the majority of neuropsychologists reported that prior to participation in this study, they did not administer SVTs and/or PVTs routinely in clinical assessments. Rather, they added response validity tests to their test battery when certain features were present. These features were related to clinicians' perception of how patients present their problems, and so they were intrinsically subjective. The present results show that clinicians are well advised not to lean too much on such impressions. Relatedly, clinicians said that the presence of incentives served as an important cue for the decision to include SVTs and/or PVTs. However, patients might not wish to inform their clinicians about their anticipation of financial or legal advantages (Van Egmond, Kummeling, & Balkom, 2005). Thus, making the decision to include response validity tests contingent on information about such incentives is an imperfect rule of thumb. More generally, it is illogical to use a less accurate method (i.e., subjective judgment) to decide whether or not to employ a more accurate method.

Older studies on clinical prediction of distorted symptom presentation (Faust, Hart, & Guilmette, 1988; Faust, Hart, Guilmette, & Arkes, 1988; Heaton et al., 1978) were criticized for being too artificial (Bigler, 1990; Garb & Schramke, 1996; Schmidt, 1989). Critics pointed out that clinicians do not solely rely on test data to determine the presence of cognitive impairment or psychopathology. Rather, they would consider psychometric data along with information from the medical history, the clinical interview, and the observations. However appealing and intuitively plausible this

point might seem, there is little evidence that judgment accuracy increases as clinicians consider more sources of clinical data (e.g., Garb, 2005; Wedding & Faust, 1989). In fact, several studies have found the opposite: The more information that clinicians try to take into account, the less accurate their judgment becomes (e.g., Ægisdóttir et al., 2006; Sawyer, 1966; Wedding, 1983). In the current study, we tried to stay close to the daily practice of diagnostic assessment in a hospital setting. Thus, neuropsychologists did have all the clinical data at their disposal to inform their opinion (i.e., the medical file, their interview, observational data). Nonetheless, their predictions about distorted symptom presentation were wrong in 24% of the cases.

Several limitations of the current study deserve comment. Firstly, although they came from five different hospitals, our group of neuropsychologists was relatively small. This limits the generalizability of our findings. Also, neuropsychologists participated on a voluntary basis, and the procedure sensitized them to the issue of distorted symptom presentation. Thus, our findings may underestimate how poor clinical prediction of distorted symptom prediction really is.

Secondly, although we did include outcomes from two different response validity tests—one SVT and one PVT—our study relied on only two response validity tests. Results might have been different, had we used a whole battery of such tests. Clearly, this issue warrants further study. Thirdly, we used a 3-point Likert scale (i.e., unproblematic–somewhat problematic–problematic) to index clinicians' predictions of response validity test outcome. The problematic option, however, was only endorsed four times. Considering that clinicians were asked to make a prediction without having neuropsychological test results at their disposal, it is conceivable that they felt more comfortable to predict a “somewhat problematic” than a “problematic” outcome of response validity testing. Thus, the distinction between these two categories may reflect more the degree of confidence that clinicians placed in their suspicion of distorted symptom presentation than the number of indications that pointed in the direction of symptom distortion. In retrospect, it would have been better if we had used a dichotomous index of poor response validity (yes /no) and additionally had asked clinicians to rate the confidence they placed in their judgments.

We focused on clinicians' predictions of distorted symptom presentation in a hospital setting, a setting where the base rate of distorted symptom presentation will be lower than in the forensic context (Dandachi-FitzGerald et al., 2013; Mittenberg et al., 2002). Given the relatively low base rate of distorted symptom presentation, a good decisional strategy in a hospital setting would be to anticipate nondistorted symptom presentation, and to include response validity tests as a check on this assumption. Of course, this is only true to the extent that clinicians are sensitive to the psychometric information provided by response validity tests, notably their positive and negative predictive power. The issue of clinicians' prediction of response validity test outcomes warrants further research, in particular research in which purely clinical cases are compared with cases that have a forensic dimension.

Taken together, our results support the notion that it is important to routinely test for distorted symptom presentation (Bush et al., 2005; Heilbronner et al., 2009). Yet, our findings are silent about how clinicians will integrate response validity test scores in their diagnostic conclusions. What happens when their a priori predictions do not match test results? Does this, indeed, lead to a correction of their prediction or do clinicians cling to their first impressions? Clearly, this too is an important topic for future research.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., ... Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*, 341–382. doi:10.1177/0011000005285875
- Bigler, E. D. (1990). Neuropsychology and malingering: Comment on Faust, Hart, and Guilmette (1988). *Journal of Consulting and Clinical Psychology, 58*, 244–247. doi:10.1037/0022-006X.58.2.244
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., ... Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20*, 419–426. doi:10.1016/j.acn.2005.02.002
- Dandachi-FitzGerald, B., Merten, T., & Ponds, R. W. H. M. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology, 28*, 771–783. doi:10.1093/arclin/act073
- Dandachi-FitzGerald, B., Van Twillert, B., Van De Sande, P., Van Os, Y., & Ponds, R. W. H. M. (2015). Poor symptom and performance validity in regularly referred hospital outpatients: Prevalence, role of incentives, and link with standard clinical measures. *Psychiatry Research, 239*, 47–53. doi:10.1016/j.psychres.2016.02.061
- De Bie, S. E. (1987). *Standaardvragen 1987: Voorstellen voor uniformering van Vraagstellingen naar achtergrondkenmerken en interviews* [Standard questions 1987: Proposal for uniformisation of questions regarding background variables] (2nd ed.). Leiden: Leiden University Press.
- Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 56*, 578–582. doi:10.1037/0022-006X.56.4.578
- Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*, 508–515. doi:10.1037/0735-7028.19.5.508
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology, 1*, 67–89. doi:10.1146/annurev.clinpsy.1.102803.143810
- Garb, H. N., & Schramke, C. J. (1996). Judgment research and neuropsychological assessment: A narrative review and meta-analyses. *Psychological Bulletin, 120*, 140–153. doi:10.1037/0033-2909.120.1.140
- Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *The Clinical Neuropsychologist, 27*, 108–137. doi:10.1080/13854046.2012.739646
- Heaton, R. K., Smith, H. H., Lehman, R. A. W., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*, 892–900. doi:10.1037/0022-006X.46.5.892
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R.; 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*, 1093–1129. doi:10.1080/13854040903155063
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. doi:10.2307/2529310
- Larrabee, G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist, 22*, 666–679. doi:10.1080/13854040701494987

- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625–630. doi:10.1017/S1355617712000240
- Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the structured inventory of malingered symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology, 18*, 145–152. doi:10.1093/arclin/18.2.145
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmand, B. A., SantaMaría, P., & González Ordi, H. (2013). Symptom validity assessment in European countries: Development and state of the art. *Clínica Y Salud, 24*, 129–138. doi:10.1016/S1130-5274(13)70014-8
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*, 1094–1102. doi:10.1076/jcen.24.8.1094.8379
- Rosenhan, D. L. (1973). On being sane in insane places. *Science, 179*, 250–258. doi:10.1126/science.179.4070.250
- Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*, 178–200. doi:10.1037/h0023624
- Schmand, B., & Lindeboom, J. (2005). *The amsterdam short-term memory test. Manuel*. Leiden: PITS.
- Schmidt, J. P. (1989). Why recent researchers have not assessed the capacity of neuropsychologists to detect malingering. *Professional Psychology: Research and Practice, 20*, 140–141. doi:10.1037/h0092792
- Sweet, J. J., & Guidotti Breting, L. M. (2013). Symptom validity test research: Status and clinical implications. *Journal of Experimental Psychopathology, 4*, 6–19.
- Trueblood, W., & Binder, L. M. (1997). Psychologists' accuracy in identifying neuropsychological test protocols of clinical malingerers. *Archives of Clinical Neuropsychology, 12*, 13–27. doi:10.1016/S0887-6177(96)00027-3
- Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *Clinical Neuropsychology, 27*, 1234–1246. doi:10.1080/13854046.2013.835447
- Van Egmond, J., Kummeling, I., & Balkom, T. (2005). Secondary gain as hidden motive for getting psychiatric treatment. *European Psychiatry, 20*, 416–421. doi:10.1016/j.eurpsy.2004.11.012
- Van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta analysis. *The Clinical Neuropsychologist, 28*, 1336–1365. doi:10.1080/13854046.2014.984763
- Wedding, D. (1983). Clinical and statistical prediction in neuropsychology. *Clinical Neuropsychology, 5*, 49–55.
- Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4*, 233–265. doi:10.1016/0887-6177(89)90016-4