GRUPO EDITORIAL DE PSICOFUNDACIÓN

Psicothema (2025) 37(2) 74-85

Psicothema

https://www.psicothema.com/es • ISSN 0214–9915 • eISSN 1886-144X

Colegio Oficial de Psicología del Principado de Asturias

Psicothema

Article

Prevalence of Symptom Overreporting in the Structured Inventory of Malingered Symptomatology (SIMS) in Clinical Patients: A Meta-Analysis

Esteban Puente-López[1], David Pina[2], Brechje Dandachi-FitzGerald[3], Luciano Giromini[4], Rubén López-Nicolás[5], María Dolores Nieto-Cañaveras[6] and Thomas Merten[7]

1 Universidad de Valladolid (Spain)
2 Universidad de La Rioja (Spain)
3 Maastricht University (The Netherlands)
4 University of Turin (Italy)
5 Universidad de Murcia (Spain)
6 Universidad Antonio de Nebrija (Spain)
7 Vivantes Klinikum im Friedrichshain (Germany)

ARTICLE INFO

Received: June 19, 2024

Accepted: October 3, 2024

Keywords:

Structured Inventory of Malingered Symptomatology

Malingering

Meta-analysis

Clinical assessment

Symptom validity

ABSTRACT

Background: Failure on symptom validity tests may occur in a variety of contexts and situations, including routine clinical settings. To date, no meta-analysis has targeted the failure rate of the Structured Inventory of Malingered Symptomatology (SIMS) in clinical assessments, nor the factors that may moderate this rate. **Method:** We used a binomial-normal random-effects meta-analysis to estimate the pooled failure rate of SIMS among patients with a clinical diagnosis who were evaluated in a non-forensic setting. **Results:** 34 studies and 40 samples were included. The total sample size was 8844 patients. The mean total SIMS score was 15.9 (*SD* = 5.2). The estimated overall failure rate of SIMS was 36% (95% CI: 30%–43%; $I^2 = 96.6\%$, $p < .001$). **Conclusions:** There is an elevated failure rate on the SIMS in clinical patient populations; however, these positive results are not necessarily *false* positives. The methodological challenge to tell true and false positives apart appears to be of primary importance and should dictate both careful planning of future studies and circumspection when interpreting rates of validity test failure in clinical assessments.

Prevalencia de la Sobreinformación de Síntomas en el Inventario Estructurado de Simulación de Síntomas (SIMS) en Pacientes Clínicos: un Metaanálisis

Palabras clave:

Inventario Estructurado de Simulación de Síntomas

Simulación

Meta-análisis

Evaluación clínica

Validez de síntomas

RESUMEN

Antecedentes: Los fallos en las pruebas de validez de síntomas (puntuar por encima del punto de corte establecido) pueden producirse en diversos contextos y situaciones, incluidos los entornos clínicos rutinarios. Hasta la fecha, ningún metaanálisis se ha centrado en la tasa de fallos del Inventario Estructurado de Simulación de Síntomas (SIMS) en evaluaciones clínicas. **Método:** Se realizó un meta-análisis de efectos aleatorios binomial-normal para estimar la tasa de fallos combinada del SIMS entre pacientes con un diagnóstico clínico que fueron evaluados en un entorno no forense. **Resultados**: Se incluyeron 34 estudios y 40 muestras ($n = 8844$). La puntuación media total del SIMS fue de 15.9 (*DE* = 5.2). La tasa global estimada de fallo de la SIMS fue del 36% (IC del 95%: 30%-43%; $I^2 = 96.6\%$, $p < .001$). **Conclusiones:** Existe una elevada tasa de fallo en el SIMS en poblaciones de pacientes clínicos; sin embargo, estos resultados positivos no son necesariamente falsos positivos. El reto metodológico de diferenciar los verdaderos de los falsos positivos es vital y debería dictar tanto la planificación cuidadosa de futuros estudios como la circunspección en la interpretación de las tasas de fallo de las pruebas de validez en las evaluaciones clínicas.

Cite as: Puente-López, E., Pina, D., Dandachi-FitzGerald, B., Giromini, L., López-Nicolás, R., Nieto-Cañaveras, M. D., & Merten, T. (2025). Prevalence of symptom overreporting in the Structured Inventory of Malingered Symptomatology (SIMS) in clinical patients: A meta-analysis. *Psicothema, 37*(2), 74-85. https://doi.org/10.70478/psicothema.2025.37.18
Corresponding author: David Pina, david.pina@unirioja.es

In clinical settings, psychological assessments are performed to better understand a patient's symptoms, coping style, and personality traits so as to inform diagnostic decision-making and treatment planning (e.g., Meyer, 2001). However, because psychological assessments rely heavily on self-reported symptoms, complaints, and impairments, the obtained results will only be valid in as far as patients are able and willing to present their problems accurately. This prerequisite for accurate assessment is not always met. Symptom overreporting is a common problem whenever primary or secondary gain is a potential issue, but patients may as well underreport or embellish symptoms or functional impairments (Dandachi-FitzGerald et al., 2024; Giromini et al., 2022; Pina et al., 2022).

When an exaggerated symptom presentation goes undetected and obtained data are wrongly considered valid, erroneous conclusions may be drawn, potentially leading to adverse consequences such as misdiagnosis and harmful therapeutic interventions (e.g., Roor et al., 2016; van der Heide et al., 2020). Therefore, when interpreting self-report data about symptoms, a necessary step is to exclude significant response distortions. Based on clinical judgment alone, it can be difficult to discriminate between a test profile indicating genuine problems and one indicating symptom exaggeration (Dandachi-FitzGerald & Martin, 2022; Sweet et al., 2021). Fortunately, to aid in this discrimination, various specialized instruments have been developed. These are symptom validity tests (SVTs) and performance validity tests (PVTs) that measure the credibility of symptom reporting and cognitive test profiles, respectively

Historically, these validity tests were developed and applied in the forensic context. The idea behind them was that failing on a validity test (e.g., scoring above or below the established cutoff scores) indicated malingering (i.e., the intentional gross exaggeration or invention of symptoms motivated by obtaining an external incentive). Due to extensive research as well as conceptual refinement within the scientific and professional communities, the consensus currently is that validity tests measure a certain class of behaviors (i.e., symptom exaggeration and cognitive underperformance), and that an interpretation in terms of malingering requires additional inferences about the intent and motives behind these behaviors (Merten & Dandachi-FitzGerald, 2022). In fact, malingering is but one possible causal antecedent of validity tests failure (Merckelbach et al., 2019). Thus, there might be other reasons as to why people engage in overreporting such as inattentive responding (e.g., Ward & Meade, 2023) or factitious motives (e.g., Chafetz et al., 2020).

Along with this conceptual refinement, interest shifted from the detection of malingering within the forensic context to trying to understand distorted symptom presentations in general, regardless of context (Schroeder & Martin, 2022). As a result, in recent years, the importance of symptom and performance validity assessment in clinical and rehabilitation context has been significantly strengthened (e.g., Carone & Bush, 2018; Schroeder & Martin, 2022), with a growing acceptance of the need to determine the credibility of test profiles and symptom reports outside the forensic arena. In contrast to forensic referrals, in patient care the clinical worker's primary obligation is to serve the best interest of their patients. However, taking clinical test results and subjective symptom claims at face value may lead to wrong diagnostic decisions, to wrong treatment options, and be potentially harmful to patients (e.g., van der Heide et al., 2020). The most obvious constellation in which this may occur is certainly a clinical patient with factitious disorders (e.g., Chafetz et al., 2020; Merten & Merckelbach, 2020). Hidden agendas and external gain expectations, which can hinder accurate diagnosis and negatively impact treatment outcome, may play a significant role in distorted symptom presentations (van Egmond & Kummeling, 2002). In addition, clinicians should generally be careful not to make false statements about their patients and thereby violate both legal standards and ethical professional obligations (e.g., Bush et al., 2006; Iverson, 2006).

The overwhelming majority of empirical studies in this domain have focused on PVTs in the context of neuropsychological assessments and, consequently, there is now a rich literature on this topic, including scoping reviews and meta-analytic papers (e.g., Lippa, 2018; McWhirter et al., 2020; Roor et al., 2024). However, information on the performance of SVTs in the clinical setting is very limited, especially that related to prevalence estimates or base rates. Base rate estimations about validity test failure in different groups and referral contexts is important because it is essential for the clinician in determining the positive and negative predictive value of validity test results for an individual case.

Specifically, the likelihood that a deviant test score on a validity test signals a noncredible symptom presentation depends on both the sensitivity and specificity of the test, and on the base rate of the condition (noncredible report) in the setting where the evaluation took place (Dandachi-FitzGerald & Martin, 2022; Tiemens et al., 2020). The base rate statistics obtained from PVTs do not easily generalize to SVTs, as both type of validity tests assess interconnected tap into overlapping but yet distinct concepts (e.g., Giromini, Barbosa et al., 2020, Giromini, Viglione et al., 2020; Ord et al., 2021; Shura et al., 2021). Empirically derived estimates of the prevalence of SVT failure in clinical assessments are lacking.

Below, we focus on one specific freestanding SVT that is widely used across a variety of referral settings and countries (Dandachi-FitzGerald et al., 2013; Martin et al., 2015; Nijdam-Jones & Rosenfeld, 2017). The Structured Inventory of Malingered Symptomatology (SIMS) is a 75 true/false items self-reported SVT that was originally developed as a *malingering test* for forensic settings (e.g., Smith & Burger, 1997). The primary score is the SIMS total score, with higher scores indicative of endorsement of unlikely symptomatology. It also contains five subscales of 15 items each: Psychosis (P), Neurological Impairment (NI), Amnestic Disorders (AM), Low Intelligence (LI), and Affective Disorders (AF). Smith and Burger recommended a cutoff of > 14 for the SIMS total score, meaning that a patient's presentation of their symptoms is classified as noncredible whenever more than 14 symptoms listed by the SIMS are endorsed. However, due to the high false positive rate associated with this cutoff point, many foreign-language adaptations of the SIMS recommend using a cut score of > 16 for decision-making at screening level, as is the case with the Dutch and the German versions. A first meta-analysis that included predominantly experimental simulation studies and forensic assessments (van Impelen et al., 2014) found that the SIMS as a screening instrument attained satisfactory sensitivity (.87–1.00) and specificity (.60–.93) in these contexts. However, concerns were raised about the specificity in clinical patients, with unacceptably low specificity rates of .37 to .59.

An important caveat, however, is that the studies included in this meta-analysis did not check but assumed honest responding in clinical patients (i.e., bona-fide patients), and therefore all scores above the cutoff on the SIMS in patients were considered false positives. As mentioned above, this is no longer considered a reasonable or unassailable conclusion, because in real-world clinical

assessments, professionals should never assume that the clinical profile presented is credible without objective evidence. Although several other clinical studies using the SIMS have been published following the meta-analysis by van Impelen et al. (2014; for a recent review, see Shura et al., 2022), no meta-analysis to date has specifically examined the SIMS failure rate in clinical assessments, nor the factors that may moderate this rate (e.g., type of symptoms assessed, evaluation context, etc.). The present study therefore seeks to address this gap in the literature by examining the prevalence of SIMS failure rates in patients with a clinical diagnosis assessed in a clinical (non-forensic) setting, as well as the impact of several potentially moderating variables identified in previous work such as the known presence of external incentives (Aparcero et al., 2021; Detullio et al., 2019; Roor et al., 2024; van Impelen et al., 2014).

Method

We performed a systematic review and meta-analysis following the recommendations of the PRISMA guidelines (Page et al., 2021). Specific recommendations for meta-analyses of proportions of Barker et al. (2021) and Migliavaca et al. (2022) were also followed. The PRISMA criteria checklist can be found at the link included at the end of the Meta-analytic Plan section.

Eligibility Criteria and Study Selection

We searched for studies that included patients with a clinical diagnosis who were not in a forensic setting (i.e., samples of clinical patients, including rehabilitation patients, in treatment or clinical assessment contexts). We used the following inclusion criteria: (1) conducted a psychological assessment in the context of a clinical evaluation (i.e., non-forensic); (2) used an unmodified version of the SIMS; (3) reported the number of positive SIMS results in the group of clinical patients with SIMS cutoff score >16 or provided sufficient information to calculate it; (4) the full text was available. As exclusion criteria, we used the following: (1) literature or theoretical reviews; (2) included forensic referrals; (3) included a small simple size ($N < 20$).

Information Sources and Search

An electronic search was performed in the following databases: the core collection of Web of Science, ERIC, ProQuest, Medline, Tripdatabase, EBSCO host (Academic Search Premier, Psychology and Behavioral Sciences Collection, APA PsychArticles, APA PsychInfo, Education Source, PsicoDOC), and the Cochrane Library. The search strategy included the following terms: (*"Structured Inventory of Malingered Symptomatology"* OR *"SIMS"*) AND (*"clinic*"* OR *"patient"*) OR (*"SIMS"* AND *"malinger*"* OR *"feign*"*). The search time range was 1997 (initial SIMS publication) to 2022, and was not limited by language of publication. The last search prior to the data analysis and writing of the document was carried out on April 12, 2024.

Data Collection Procedure

Double and independent coding was performed by EPL and DP for the relevant variables of the studies: name of the article (id), symptoms, type of symptoms, type of setting, inpatient or outpatient, country of administration, patients with possible external incentive excluded, economic compensation for participating in the study, cutoff score used, risk of bias, *N* of positive results (above the cutoff score of 16) and total *N*. In the variables type of condition, inpatient or outpatient, type of setting, country, and financial compensation, the category "mixed" was included for those studies in which the positive results were mixed (for example, patients from Spain and Italy were included in the same group and the results were not provided divided according to country). Discrepancies were resolved by consensus. The coding table used, as well as the codebook showing the coding system used for each variable, can be found at the link included at the end of the Meta-analytic Plan section. The interrater reliability of the coding process was satisfactory, varying between $\kappa = .81$ and 1 (mean $\kappa = .92$) for categorical variables, and between an intraclass correlation coefficient (ICC, Two-Way Mixed-Effects Model) of = .93 and 1 (mean *ICC* =.98) for continuous variables.

When the studies did not report the cutoff point at >16 ($n = 4$), the authors of the publication were contacted to request the information. If no response was received, a missing values imputation process was carried out by fitting a linear regression model and a beta regression model (Ferrari & Cribari-Neto, 2004), using the mean SIMS as the predictor variable and the proportion of positive results with cutoff score as the criterion. The performance of these two models were compared. The linear regression model showed a better performance in terms of R^2, *AIC, MSE and RMSE* (full result of model comparison can be found at the link included at the end of the Meta-analytic Plan section). Hence, the regression equation generated by the linear regression model was used to predict the proportion of positive results in those studies that did not report the cutoff point of > 16. All imputed values were checked to ensure that they were restricted to the interval (0, 1).

Originally, it was also intended to include positive results with cutoff points higher than 16, such as 19, 21 or 24, but the low number of studies ($k = 4$) providing such information led to this option being discarded.

Because no instrument has yet been developed to assess risk of bias in symptom and performance validity studies, the Puente-Lopez et al. (2023) checklist was used. The checklist was composed of 13 items assessing different aspects of the study design that, if absent, constituted a source of bias. The items were scored with *0 = item present*; *1 = item not present* and *= *not applicable or doubtful*. The checklist total score ranged from 0 to 13, and high values indicated higher risk of bias. The analysis was conducted by two independent raters (EPL and DPL) and the reliability of the coding process was assessed using Cohen's kappa coefficient. The interrater reliability of the coding process was satisfactory, varying between .63 and 1 (mean =.87).

Meta-Analytic Plan

A binomial-normal random-effects meta-analysis (Hamza et al., 2008; Stijnen et al., 2010) using maximum-likelihood estimation was performed. This approach outperforms standard meta-analytic methods by fully accounting for within-study uncertainties, avoiding the bias due to the correlation between estimate and standard error, and being able to deal with cases with zero events. Basically, this approach is a *generalized linear mixed model* (GLMM) with logit link function and a random intercept.

To assess study heterogeneity, the I^2 statistic and 95% prediction intervals (Higgins et al., 2009) were used. I^2 values between 25%- and 50% are considered low, between 50% and 75% moderate, and above 75% high (Higgins et al., 2003). In addition, to explore possible sources of heterogeneity, different variables were also included in our model to assess its potential moderation effect. Variables that were identified beforehand as possible moderators were included (type of symptoms, setting of the evaluation, inpatient or outpatient, country, exclusion of patients with possible external incentives, financial compensation for participation and risk of bias). Also, an influence analysis was conducted to determine the influence of each individual study on the overall result by omitting studies one by one. Sensitivity analyses were carried out using a t-distribution for confidence intervals with k-1 degrees of freedom and an F-distribution with k-p degrees of freedom for the omnibus test. The results were very similar. All analyses were performed by EPL and RLN in R environment (4.2.0, R Core Team, 2022) using the *metafor* package (Vietchbauer, 2010). All data, script codes, and relevant materials are openly available at: https://osf.io/u7yvt/?view_only=7b1a81de02a3483f965ea3511a7e7991. Full results for sensitivity analyses are available at: https://osf.io/dmabw and https://osf.io/49d35

Results

Figure 1 shows the selection process. Initially, 1212 publications of interest were identified, of which 774 were eliminated because they were duplicates. Next, the titles and abstracts of the remaining 438 articles were read and 392 were excluded because they did not meet the inclusion criteria. Subsequently, 46 remaining full-text articles were read and 12 of them were excluded, four articles were eliminated for including patients in litigation status (Kobelt-Pönicke et al., 2020; Marín-Torices et al., 2018; Merten et al., 2016; van Impelen et al., 2017), three for not providing sufficient information (Chen et al., 2011; Freeman et al., 2008; Modiano et al., 2021), two for using an incomplete or modified version of the SIMS (van der Heide et al., 2017; van der Heide & Merckelbach, 2016), two for using the same sample as other study included (Ord et al., 2021; Rowland et al., 2020), and one because the sample size of the patients was below 20 (Merckelbach & Smith, 2003). Thirty-four studies were finally selected. In six of these studies (Barbosa-Torres et al., 2023; Benge et al., 2012; Dandachi-Fitzgerald et al., 2016; Giromini et al., 2018; Göbber et al., 2012; van Beilen et al., 2009) two different samples each were investigated, so number of samples for the analysis rose to 40.

Table 1 shows the main characteristics of the 40 samples included. The total sample size was 8844 patients, with a mean age of 47.6 years (SD = 8.8). The mean total SIMS score was 15.9 (SD = 5.2). The majority reported psychological symptoms (62.5%), such as anxious-depressive, traumatic, psychotic or substance use-related symptoms, and approximately half of the participants (48%) were outpatients. The setting in which the evaluation was performed showed a relatively uniform distribution, with the Veteran Administration (VA, 17.5%) predominating in the first variable, followed by psychiatric hospitals (17.5%), clinic non-psychiatric hospitals (17.5%), and rehabilitation (16%). Regarding the country, the majority of studies were conducted in European countries (70%), such as Germany (27.5%) and the Netherlands (25%). The remaining studies (30%) were conducted in the USA. Most of the reports (80.0%) did not clarify if that participants with a possible external incentive had been excluded or not. Also, the majority of reports did not offer financial compensation for participation (82.5%).

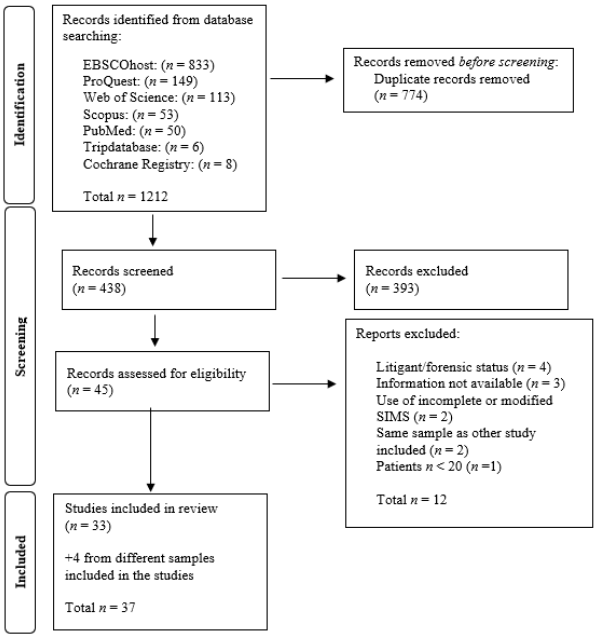
The estimated overall failure rate of SIMS >16 was 36% (95% CI: 30%–43%; Figure 2), with significant heterogeneity between studies (I^2 = 96.6%, p < .001, PI = [0.09-0.76]).

Regarding the possible sources of heterogeneity, the meta-regression analysis showed that SIMS failure rates were independent of the type of the patient (outpatient vs inpatient; p = .954). However, the setting of the evaluation (Q=32.8, p < .001), the country (Q = 24.10, p < .001), the compensation for the participation in the study (Q = 7.30, p < .01), the type of symptoms (Q= 11.3, p < .001), and the exclusion of participants with possible external incentive (Q = 7.29, p < .01) were identified as significant moderator variables.

Several subgroup analyses were performed with the variables identified as significant moderators. In the evaluation setting, prevalence was significantly higher in Veterans' Administration contexts (k = 7; 54.0%; 95% CI: 42.5%–65.1%) and psychiatric hospital settings (k = 7; 50.3%; 95% CI: 38.5%–62.1%). For the country, the prevalence was higher in the U.S. (k = 12; 54.4%; 95% CI: 45.2%–63.4%) and lower in the European countries (k = 27; 27.9%; 95% CI: 23.1%–33.3%). Also, the prevalence was lower in those studies where patients with a possible external incentive had been excluded (k = 8; 21.4%; 95% CI: 13.0%-33.0%) and it was higher in those studies where financial compensation was offered for participation (k = 7; 55.5%; 95% CI: 40.0%–69.9%). As for the type of symptoms, prevalence was higher in studies with psychological conditions/diagnosis, such as posttraumatic stress disorder (PTSD), dissociative identity disorder (DID), or depression (k = 25; 42.4%; 95% CI: 35.0%-50.2%) and lower in studies

Figure 1

Characteristics of the Studies Included



with neuropsychological conditions/diagnosis, such as Korsakoff syndrome, dementia, or Parkinson's disease (k = 6; 16.4%; 95% CI: 0.87%-28.8%). Based on theoretical criteria, we selected a subset of statistically significant moderators to fit a multiple meta-regression model. This model included country (dichotomized as U. S. *vs.* Europe) and symptom type. Only a subset of moderators was included in the model because, as categorical moderators, each one requires the inclusion of as many dummy variables as there are levels – 1. To avoid overfitting, we limited the number of coefficients in the model. The overall model was statistically significant (Q = 69.9, $p < .001$) Figure 3 shows a forest plot displaying the adjusted proportions for each combination of the moderators included in the multiple meta-regression model.

An influence analysis was also performed, excluding each study one-by-one from the analysis. The pooled prevalence of SIMS failure rates did not substantially change. It varied between 35.1% (95% CI: 29.2%–41.6%) excluding Brand et al. (2021) and 37.2% (95% CI: 31.2%–43.7%) excluding Oudman et al. (2020). This indicates that no single study had a disproportional impact on the overall prevalence.

Table 1
Characteristics of the Studies Included

Authors	Type of symptoms	Setting	Type of patient	Country	External incentive excluded	Compensation	*M*	*SD*	nposi	ntotal	F.r.
BarbosaTorres et al. (2023) sample 1	Med.	Hosp.	Out	Spa	No	No	19.3	4.2	11	42	.26
BarbosaTorres et al. (2023) sample 2	Mental	M.h.Clin.	Out	Spa	No	No	18.8	5.8	12	36	.33
Benge et al. (2012) sample 1	Mental	VA	Out	USA	No	No	22.4	10.5	65	91	.71
Benge et al. (2012) sample 2	Neuro.	VA	Out	USA	No	No	14.4	7.1	9	29	.31
Bodenburg et al. (2022)	Mental	Clin.	Out	Ger	No	No	n.r.	n.r.	54	221	.24
Brand et al. (2021)	Mental	Psy.Hosp	Out	Mix	No	Yes	23.5	11.0	48*	63	.76
Czornik et al. (2021)	Neuro.	Clin.	Out	Aus	No	No	9.1	5.5	8	54	.15
Dandachi-FitzGerald et al. (2011)	Mixed	M.h.Clin.	Mix	NL	No	No	n.r.	n.r.	38	183	.21
Dandachi-FitzGerald et al. (2016) sample 1	Mixed	Hosp.	Out	NL	No	No	n.r.	n.r.	54	193	.28
Dandachi-FitzGerald et al. (2016) sample 2	Mixed	Hosp.	Out	NL	Yes	No	n.r.	n.r.	38	264	.14
Dandachi-FitzGerald et al. (2020)	Neuro.	Hosp.	Out	NL	Yes	No	n.r.	n.r.	15	120	.12
Giromini et al. (2018) sample 1	Mental	Psy.Hosp	Mix	Ita	No	No	22.7	10.1	59	89	.66
Giromini et al. (2018) sample 2	Mental	Psy.Hosp	Mix	Ita	No	No	17.3	8	53	127	.42
Göbber et al. (2012) sample 1	Mental	Rehab.	In	Ger	No	No	n.r.	n.r.	151	457	.33
Göbber et al. (2012) sample 2	Mental	Rehab.	In	Ger	No	No	n.r.	n.r.	69	138	.50
Graue et al. (2007)	Mental	M.h.Clin.	Out	USA	No	Yes	28.7	12.5	20	26	.77
Harris et al. (2022)	Mixed	Clin.	Out	USA	No	No	14.3	7.4	42	110	.38
Kaminski et al. (2020)	Mental	Rehab.	In	Ger	No	No	14.1	7.7	50	147	.34
Kirchhoff & Steinert (2019)	Mental	Hosp.	Mix	Ger	No	Yes	n.r.	n.r.	9	30	.30
Kobelt et al. (2012)	Mental	Rehab.	In	Ger	No	No	n.r.	n.r.	81	329	.25
Lace et al. (2021)	Neuro.	Clin.	Out	USA	No	No	17.5	8.3	35	67	.52
Merten et al. (2020)	Mental	Rehab.	In	Ger	No	No	14.4	8.1	185	537	.34
Miskey et al. (2020)	Mental	VA	Out	USA	No	No	n.r.	n.r.	78	110	.71
Niesten et al. (2017)	Mental	M.h.Clin.	Out	NL	No	No	14.7	12	9	27	.33
Nübling et al. (2020)	Mental	Rehab.	In	Ger	No	No	15.8	9	1539	3800	.40
Oudman et al. (2020)	Neuro.	Clin.	In	NL	Yes	No	7	4.0	0	20	.02
Peters et al. (2013)	Mental	Psy.Hosp.	In	NL	Yes	Yes	12.9	6.9	12	41	.29
Praus et al. (2021)	Mental	Clin.	Mix	Ger	Yes	No	12.7	7.5	27	77	.35
Puente-López et al. (2022)	Mental	Clin.	Out	Spa	Yes	No	8.7	4.0	6	58	.10
Rogers et al. (2014)	Mental	M.h.Clin.	In	USA	Yes	Yes	22.5	11.3	39*	54	.63
Rogers et al. (2021)	Mental	Psy.Hosp	In	USA	No	Yes	18.6	9.7	26*	41	.49
Shura et al. (2021)	Mixed	VA	Out	USA	No	No	14.85	9.2	122	338	.36
Spencer et al. (2021)	Mixed	VA	Out	USA	No	No	17.7	8.8	121	249	.49
Tierney et al. (2021)	Mixed	VA	In	USA	No	No	n.r.	n.r.	68	122	.56
van Beilen et al. (2009) sample 1	Mental	Hosp.	Out	NL	No	No	11.4	7.4	6	26	.23
van Beilen et al. (2009) sample 2	Neuro.	Hosp.	Out	NL	No	No	7.8	3.9	1	26	.04
van Minnen et al. (2020)	Mental	Psy.Hosp	Mix	NL	No	No	n.r.	n.r.	84	205	.41
Wertz et al. (2021)	Mixed	Psy.Hosp	In	Ger	No	No	18.4	10.8	n.r.*	30	.47
Wolf et al. (2020)	Mental	VA	Out	USA	No	Yes	n.r.	n.r.	102	171	.60
Zimmermann et al. (2013)	Mental	Hosp.	Mix	Ger	Yes	No	11.7	7.3	27	96	.28

Note. Mental = Psychological symptoms; Neuro.= Neuropsychological symptoms; Med.= Medical symptoms; M.h.Clin.= Mental health clinic; Hosp.= Hospital; VA= Veteran Administration; Rehab= Rehabilitation; Psy.Hosp.= Psychiatric hospital; Clin.= Health clinic; Out= Outpatient; In = Inpatient; USA = United States of America; Ger = Germany; NL= Netherlands; Aus= Austria; Ita= Italy; Spa= Spain; Mix= Mixed sample; Ex. Incentive excluded= External incentive excluded; n.r.= non reported; nposi= Number of SIMS positive results with cutoff score >16; ntotal= Total sample.; F.r. = Failure rate with cutoff score >16; *= Studies in which the number of positives for the cutoff score >16 was calculated using the regression model mentioned in the Method section.

Figure 2

Forest Plot of the Estimated Overall Failure Rate of the Structured Inventory of Malingered Symptomatology

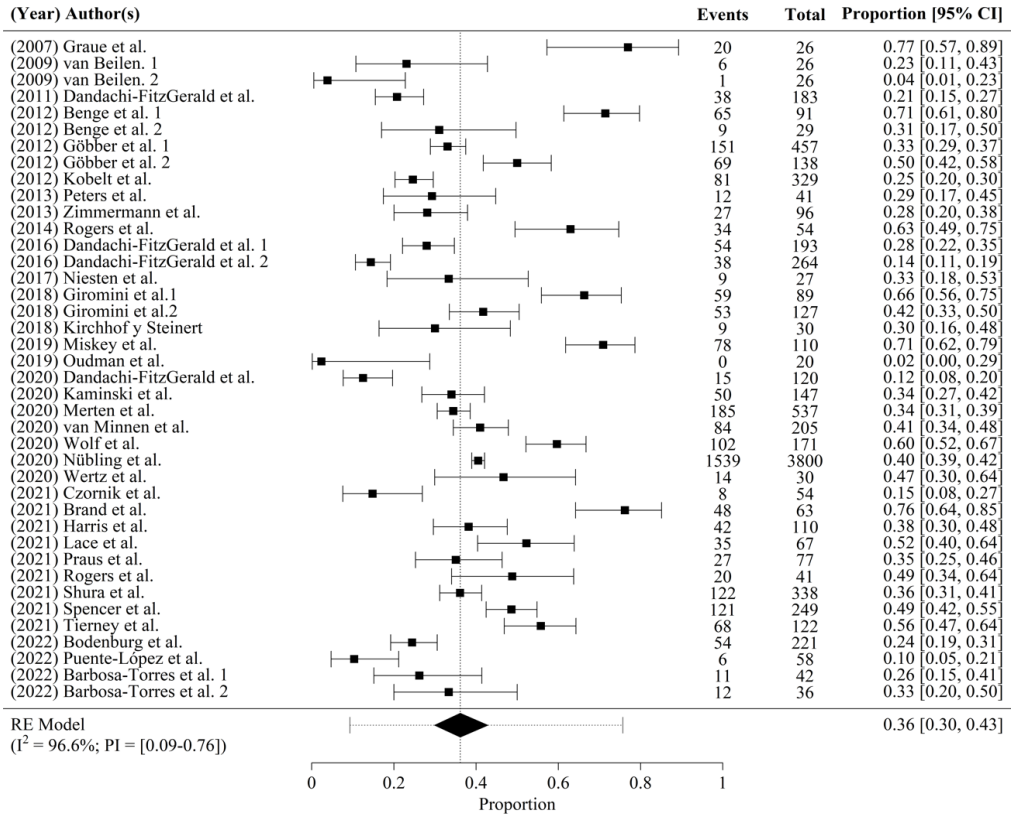
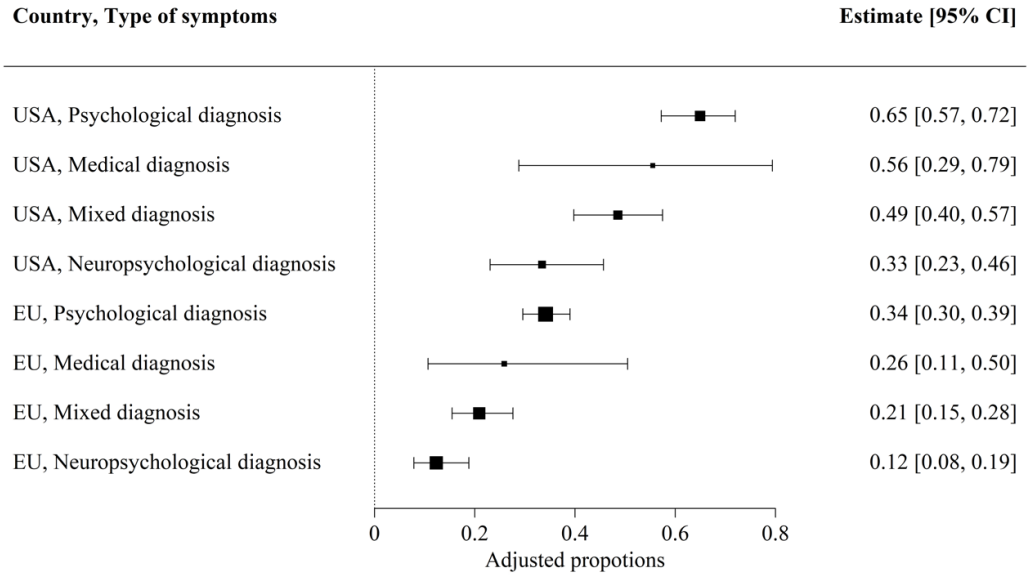


Figure 3

Adjusted Proportions of Each Combination of Moderators Included in the Multiple Meta-Regression Model



Discussion

Our meta-analysis of 40 samples from 34 studies yielded a pooled SIMS failure rate in patients evaluated in a clinical setting of 36%. Also, the mean SIMS score across the total of 8844patients was elevated (M = 15.9; SD = 5.2). Taken together, these data indicate that failing the SIMS at the >16 cutoff is far from rare, challenging the notion that a SIMS score greater than 16 is anomalous and thereby noncredible. Importantly, this is the first SIMS meta-analysis focused exclusively on patients assessed in clinical, rather than forensic, contexts, making this finding particularly noteworthy.

The published failure rates for SIMS varied greatly, ranging from 0% in patients with alcohol-induced Korsakoff amnesia (Oudman et al., 2020) to an overwhelming 77% in patients with intellectual disability from day-treatment settings (Graue et al., 2007). Given the limited cognitive resources of the participating patients and the nature of the SIMS, the high failure rate in the intellectual disability study probably reflects *false positive* classifications with respect to possible feigning or overreporting. Not only does the subscale *Low Intelligence* consist of items that could be part of an intelligence-test (with seemingly low item difficulty, but potentially challenging for participants with MR), but symptom-related items of the other subscales partly use complex linguistic structures, double negations, and conditional formulations. Yet, with respect to non-valid responding, the elevated SIMS scores would, at the same time, *correctly* point at invalid response patterns (*true positives*) and alert the clinician to be highly suspect of the validity of questionnaire results in patients with intellectual disability.

The obvious conclusion from the Graue et al. (2007) study is that the SIMS should not be given to patients with limited intellectual capacity. This conclusion is further supported by results from their honest control group. This group comprised ten community volunteers with a mean full-scale WAIS IQ of only 80.7 (SD = 9.1) and similarly low IQ scores predicted by their performance on the Wechsler Test of Adult Reading (M = 80.2, SD = 8.9). Their mean SIMS score was as high as 18.3 (SD = 11.0; failure rate not reported). This result is in sharp contrast to a variety of studies with honest responders in the normal range of intelligence who usually score low on the SIMS, with zero or low failure rates (e.g., Giger & Merten, 2013; Jelicic et al., 2011; Peters et al., 2013).

A high level of heterogeneity was identified, with a wide prediction interval (PI) ranging from 9% to 76%. This is consistent with what has been observed in meta-analytic studies of prevalence (Migliavaca et al., 2022), and specifically in meta-analyses of symptom validity tests (e.g., Aparcero et al., 2021; Detullio et al., 2019). The main sources of heterogeneity identified were the setting of the evaluation, the country in which the study was performed, the type of symptoms, a compensation given for participating in the study, and the question if participants with a possible external incentive were excluded or not. Regarding the assessment setting, the rate of SIMS failure was significantly higher in evaluations performed in Veterans Administration contexts and in psychiatric hospitals (54% and 50.3%, respectively). High failure rates were also obtained in German rehabilitation centers from where a number of studies included in the analyses stem (Göbber et al., 2012; Kaminski et al., 2020; Kobelt et al., 2012; Merten et al., 2020; Nübling et al., 2020), with potential external gain expectations to be identified in the large majority of the patients (cf. Merten et al., 2020, for details). The German system of psychosomatic rehabilitation is characterized by an interweaving of treatment and medicolegal determinations directly derived from treatment outcome. There are apparent similarities with the Veterans Administration Healthcare System in the U.S. inasmuch as both combine healthcare provision and medicolegal determinations which depend upon treatment outcome, with corresponding embedded external gain expectations in a non-negligible percentage of patients. Accordingly, the country-wise analysis showed the highest SIMS failure rate for the U.S. (54.4%), with a number of studies based on Veterans Administration patients (Benge et al., 2012; Miskey et al., 2020; Shura et al., 2021; Spencer et al., 2021; Thierney et al., 2021; Wolf et al., 2020).

The interpretation of highly elevated SIMS scores obtained in other studies with clinical patients appears to be more complex. Benge et al. (2012) reported 71% positives in patients with psychogenic non-epileptic seizures (in contrast to a much lower failure rate in patients with epileptic seizures). Brand et al. (2021) reported a rate of 86% above the SIMS cutoff in 63 patients diagnosed with dissociative identity disorder when using a cut score of > 14, corrected to 76% at > 16. Wolf et al. (2020) examined trauma-exposed U.S. veterans and found a 74% failure rate on the SIMS for those patients with a confirmed diagnosis of PTSD at the time of the assessment and participants who did not qualify for current PTSD. In contrast, those who did not qualify for current PTSD failed in 37% of cases. Even though we cannot determine whether these positive results are true or false positives, it is hard to believe that such a high number of patients did not attempt to present their mental health condition in a credible, authentic, or genuine manner. Accordingly, this reinforces the notion that the specificity of the SIMS is likely to be problematically low, potentially falling below the 90% level recommended by experts in the field (Sherman et al., 2020; Sweet et al., 2021).

The two prevalent approaches in the current validity assessment research appear to be:

1. either to interpret all positive results on a validity test (or a set of tests) as true positives (interpreting them as prevalence rates of malingering, feigning, symptom overreporting, or cognitive underperformance);
2. or to interpret them all as false positives (interpreting results in the sense of limited test specificity; e.g., Brand et al., 2021; Marín-Torices et al., 2018). In that sense, Palmer et al. (2013) remarked: "Since the participants' responses were obtained in a non-forensic context and under standard instructions – that is, they are presumably honest answers – the classification of the protocol validity indicators as typical of malingering is a false positive". (p. 126)

However, both approaches are likely to be flawed. The first approach is in neglecting of alternative reasons for invalid item endorsement (such as irrelevant responding or test measurement errors) and possible false positives due to factors like low intelligence, insufficient language proficiency in culturally or ethnically diverse patients and recent immigrants, or presence of a relevant reliable diagnosis interfering with response behavior, in particular genuine and severe neuropsychiatric disorders. The second approach can lead to severe misinterpretations for problem populations for which it remains doubtful if patient status is that of bona fide patients who are fully cooperative and honest about their symptom presentations, without hidden agendas or substantial external gain expectations.

The most likely explanation, in our opinion, is that some, but not all, of the positive results observed in these studies could be attributed to invalid, feigned, or perhaps even malingered presentations.

Unfortunately, there are no universally agreed upon and widely accepted guidelines on how to distinguish false positives from true positives in SVTs administered in a clinical context. However, one might hypothesize that a positive SVT result, in the absence of known external incentives and with other SVT results falling within the credible range, is more likely to be a false positive than a true positive. And if we adopt this perspective, our results suggest that a significant percentage (likely larger than 10%) of SIMS scores greater than 16 likely originate from credible, rather than noncredible, presentations. This in turn raises questions about the psychometric soundness of the SIMS, specifically its specificity. For instance, one of the patient samples in the study by Giromini et al. (2018) included 89 patients undergoing treatment for a psychotic disorder who had no known incentives to feign symptoms. Of these, as many as 59 (or 66%) scored > 16 on SIMS. In contrast, the percentages of positive results on the Inventory of Problems–29 (IOP-29; Viglione et al., 2017), another relatively widely used SVT, were only 18% and 7%, when considering the standard and clinical cutoffs, respectively. Therefore, it is highly unlikely that a large proportion of these positive SIMS results in this context are indeed true positives.

Overall, the conclusion that can be drawn with confidence from the analyses is that there is an elevated failure rate on the SIMS in clinical patient populations, but not that the numbers reliably reflect false positives. If the *true* false-positive rate with the cutoff >16 would amount to numbers in the range of 36% in clinical patients (and to much higher numbers in patients with some mental health diagnoses), the validity of the SIMS as an SVT would become questionable because it would yield likely inaccurate classification results with potentially devastating effects. The methodological challenge of distinguishing between true and false positives appears to be of paramount importance and should require both careful planning of future studies and prudent interpretation of validity test failure rates in potentially problematic clinical populations. The first requisite certainly is to control for external gain expectations and hidden agendas.

Whatever the true robustness of the SIMS against the presence of a genuine mental health condition is, in clinical and forensic practice determinations about the credibility of symptom presentations should never be based on one test result, but on wider lines of evidence such as proposed by Sherman et al. (2020). A good SVT should be sensitive to noncredible symptom report and, at the same time, be robust against other factors that may influence test scores, in particular genuine physical, mental, or cognitive pathology, gender, age, education, and racial background. The current data, taken at face value, appear to suggest that the SIMS be severely flawed in this respect with the traditional >16 cutoff score, but as discussed above, it remains highly questionable that all reported positive rates in clinical and rehabilitation patients reflect false positives.

Van Impelen et al. (2014) and Shura et al. (2022) recommended an increase in the cutoff point for samples with a higher risk of false positives (i.e., >19 or >23). While raising cutoff scores has its limitations, it could help avoid misuse and provide a warning against using lower cutoff scores, such as >16 or even >14 in clinical samples. However, increasing the cutoff score of a validity test is automatically accompanied by a decrease in sensitivity and in the ability of the test to detect true noncredible response patterns (i.e., it leads to a higher percentage of false-negative results). Furthermore, the evidence that increasing the SIMS cutoff score reduce the positive rate in clinical patients to 10% or less is extremely scarce. In addition, cutoff scores well above 23 are very likely to be accompanied by a substantial loss of SIMS sensitivity. Consequently, increasing the cutoff scores would not really solve the dilemma. For future research, it would be advisable to report SIMS failure rates at multiple alternative cut points so that it can be empirically analyzed whether modification of the cutoff scores (>19, >23) can effectively manage false positives and, if so, at what cost to sensitivity.

Future research also will have to ascertain where the true limits of applicability of this test are to be posited. Some difficulties are posed by the SIMS itself. The item selection was based on expert judgment and not on empirical item analysis, and several items are linguistically complex and logically flawed. A major weakness is that two different detection strategies (bizarre and uncommon symptoms on four subscales and overreporting of genuine depressive complaints on the fifth subscale) are combined and summarized in one total score. With all this in mind and with the accumulated evidence from a wealth of studies from a variety of geographical regions, referral backgrounds, patient populations, etc., it might be an option to go ahead and develop a revised SIMS version which should correct identified problems and be robust against genuine pathology, within a clearly defined scope of test usability.

Author Contributions

Esteban Puente-López: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Writing – Original Draft, Writing – Review & Editing. **David Pina**: Data Curation, Investigation, Methodology, Supervision. **Brechje Dandachi-FitzGerald:** Conceptualization, Investigation, Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision. **Luciano Giromini**: Conceptualization, Investigation, Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision. **Rubén López-Nicolás:** Methodology, Data Curation, Formal Analysis, Resources, Software, Supervision. **Maria Dolores Nieto-Cañaveras**: Conceptualization, Investigation, Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision. **Thomas Merten:** Conceptualization, Investigation, Methodology, Writing – Original Draft, Writing – Review & Editing, Supervision.

Funding

The present research has not received specific support from public sector agencies, commercial sector or non-profit entities.

Declaration of Interests

The authors declare that there is no conflict of interest.

Data Availability Statement

The data and script codes that support the findings of this study are open available at: https://osf.io/u7yvt/?view_only=7b1a81de02a3483f965ea3511a7e7991

References

References marked with an asterisk indicate studies included in the meta-analysis.

Aparcero, M., Picard, E. H., Nijdam-Jones, A., & Rosenfeld, B. (2021). The impact of coaching on feigned psychiatric and medical symptoms: A meta-analysis using the MMPI-2. *Psychological Assessment, 33*(8), 729–745. https://doi.org/10.1037/pas0001016

*Barbosa-Torres, C., Bueso-Izquierdo, N., & Cubo-Delgado, S. (2023). Analysis of malingered psychological symptoms in a clinical sample for early detection in initial interviews. *European Archives of Psychiatry and Clinical Neuroscience, 273*(2), 427–438. https://doi.org/10.1007/s00406-022-01422-8

Barker, T. H., Migliavaca, C. B., Stein, C., Colpani, V., Falavigna, M., Aromataris, E., & Munn, Z. (2021). Conducting proportional meta-analysis in different types of systematic reviews: a guide for synthesisers of evidence. *BMC Medical Research Methodology, 21*(1), Article 189. https://doi.org/10.1186/s12874-021-01381-z

*Benge, J. F., Wisdom, N. M., Collins, R. L., Franks, R., Lemaire, A., & Chen, D. K. (2012). Diagnostic utility of the Structured Inventory of Malingered Symptomatology for identifying psychogenic non-epileptic events. *Epilepsy & Behavior, 24*(4), 439–444. https://doi.org/10.1016/j.yebeh.2012.05.007

Bodenburg, S., Wendiggensen, J., & Kasten, E. (2022). Scores in self-report questionnaires assessing adult ADHD can be influenced by negative response bias but are unrelated to performance on executive function and attention tests. *Psychological Injury and Law, 15*(2), 189–199. https://doi.org/10.1007/s12207-022-09448-y

*Brand, B. L., Barth, M., Schlumpf, Y. R., Schielke, H., Chalavi, S., Vissia, E. M., Nijenhuis, E. R. S., Jäncke, L., & Reinders, A. A. T. S. (2021). The utility of the Structured Inventory of Malingered Symptomatology for distinguishing individuals with dissociative identity disorder (DID) from DID simulators and healthy controls. *European Journal of Psychotraumatology, 12*(1), Article 1984048. https://doi.org/10.1080/20008198.2021.1984048

Bush, S. S., Connell, M. A., & Denney, R. L. (2006). *Ethical practice in forensic psychology: A systematic model for decision making.* American Psychological Association.

Carone, D. A., & Bush, S. S. (2018). *Validity assessment in rehabilitation psychology and settings.* Oxford University Press.

Chafetz, M. D., Bauer, R. M., & Haley, P. S. (2020). The other face of illness deception: Diagnostic criteria for factitious disorder with proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *34*(3), 454–476. https://doi.org/10.1080/13854046.2019.1663265

Chen, D. K., Izadyar, S., Collins, R. L., Benge, J. F., Lemaire, A. W., & Hrachovy, R. A. (2011). Induction of psychogenic nonepileptic events: success rate influenced by prior induction exposure, ictal semiology, and psychological profiles. *Epilepsia, 52*(6), 1063–1070. https://doi.org/10.1111/j.1528-1167.2011.02985.x

Czornik, M., Merten, T., & Lehrner, J. (2021). Symptom and performance validation in patients with subjective cognitive decline and mild cognitive impairment. *Applied Neuropsychology: Adult, 28*(3), 269–281. https://doi.org/10.1080/23279095.2019.1628761

*Dandachi-FitzGerald, B., Duits, A. A., Leentjens, A. F. G., Verhey, F. R. J., & Ponds, R. W. H. M. (2020). Performance and symptom validity assessment in patients with apathy and cognitive impairment. *Journal of the International Neuropsychological Society, 26*(3), 314–321. https://doi.org/10.1017/S1355617719001139

Dandachi-FitzGerald, B., & Martin, P. K. (2022). Clinical judgement and clinically applied statistics: Description, benefits, and potential dangers when relying on either one individually in clinical practice. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 107–125). The Guilford Press.

Dandachi-FitzGerald, B., Merckelbach, H., & Merten, T. (2024). Cry for help as a root cause of poor symptom validity: A critical note. *Applied Neuropsychology. Adult*, *31*(4), 527–532. https://doi.org/10.1080/23279095.2022.2040025

Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, *28*(8), 771–783. https://doi.org/10.1093/arclin/act073

*Dandachi-FitzGerald, B., Ponds, R. W., Peters, M. J., & Merckelbach, H. (2011). Cognitive underperformance and symptom overreporting in a mixed psychiatric sample. *The Clinical Neuropsychologist, 25*(5), 812–828. https://doi.org/10.1080/13854046.2011.583280

*Dandachi-FitzGerald, B., van Twillert, B., van de Sande, P., van Os, Y., & Ponds, R. W. (2016). Poor symptom and performance validity in regularly referred Hospital outpatients: Link with standard clinical measures, and role of incentives. *Psychiatry Research, 239*, 47–53. https://doi.org/10.1016/j.psychres.2016.02.061

Detullio, D., Messer, S. C., Kennedy, T. D., & Millen, D. H. (2019). A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST). *Psychological Assessment, 31*(11), 1319–1328. https://doi.org/10.1037/pas0000753

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research, 158*(3), 374–380. https://doi.org/10.1016/j.psychres.2007.04.002

Giger, P., & Merten, T. (2013). Swiss population-based reference data for six symptom validity tests. *Clínica y Salud, 24*(3), 153–159. https://doi.org/10.1016/S1130-5274(13)70016-1

Giromini, L., Barbosa, F., Coga, G., Azeredo, A., Viglione, D. J., & Zennaro, A. (2020). Using the Inventory of Problems–29 (IOP-29) with the test of memory malingering (TOMM) in symptom validity assessment: A study with a Portuguese sample of experimental feigners. *Applied Neuropsychology: Adult, 27*(6), 504–516. https://doi.org/10.1080/23279095.2019.1570929

Giromini, L., Viglione, D. J., Zennaro, A., Maffei, A., & Erdodi, L. A. (2020). SVT meets PVT: Development and initial validation of the Inventory of Problems–Memory (IOP-M). *Psychological Injury and Law, 13*(3), 261–274. https://doi.org/10.1007/s12207-020-09385-8

*Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity of SIMS and IOP-29 with an Italian sample. *Psychological Injury and Law, 11*(4), 340–350. https://doi.org/10.1007/s12207-018-9314-1

Giromini, L., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: new articles, new issues. *Psychological Injury Law, 15*(1), 1–21. https://doi.org/10.1007/s12207-022-09444-2

*Graue, L. O., Berry, D. T., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., & Werline, D. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings. *The Clinical Neuropsychologist, 21*(6), 929–942. https://doi.org/10.1080/13854040600932137

*Göbber, J., Petermann, F., Piegza, M., & Kobelt, A. (2012). Beschwerdenvalidierung bei Rehabilitanden mit Migrationshintergrund in der Psychosomatik [Symptom validation in patients with migration background in psychosomatic medicine]. *Rehabilitation, 51*(5), 356–364. https://doi.org/10.1055/s-0032-1323669

Hamza, T. H., van Houwelingen, H. C., & Stijnen, T. (2008). The binomial distribution of meta-analysis was preferred to model within-study variability. *Journal of Clinical Epidemiology, 61*(1), 41–51. https://doi.org/10.1016/j.jclinepi.2007.03.016

*Harris, M., & Merz, Z. C. (2022). High elevation rates of the Structured Inventory of Malingered Symptomatology (SIMS) in neuropsychological patients. *Applied Neuropsychology: Adult, 29*(6), 1344–1351. https://doi.org/10.1080/23279095.2021.1875227

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Higgins, J. P., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology, 13*(2), 77–90. https://doi.org/10.1207/s15324826an1302_3

Jelicic, M., Ceunen, E., Peters, M. J. V., & Merckelbach, H. (2011). Detecting coached feigning using the Test of Memory Malingering (TOMM) and the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology, 67*(9), 850–855. https://doi.org/10.1002/jclp.20805

*Kaminski, A., Merten, T., & Kobelt-Pönicke, A. (2020). Der Vergleich von drei Beschwerdenvalidierungstests in der stationären psychosomatischen Rehabilitation [Comparison of three symptom validity tests in a sample of psychosomatic inpatients]. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 68*(2), 96–105. https://doi.org/10.1024/1661-4747/a000408

*Kirchhoff, L., & Steinert, T. (2019). Klassifikationsgüte der Beschwerdevalidierungstests Word Memory Test und Strukturierter Fragebogen Simulierter Symptome [Classification accuracy of the symptom validity tests Word Memory Test and the German version of the Structured Inventory of Malingered Symptomatology]. *Nervenarzt 90*(5), 516–522. https://doi.org/10.1007/s00115-018-0637-7

*Kobelt, A., Göbber, J., Bassler, M., & Petermann, F. (2012). Beschwerdenvalidität im Rahmen stationärer psychosomatischer [Symptom Validity in Psychosomatic Rehabilitation]. *Rehabilitation, 51*(5), 349–355. https://doi.org/10.1055/s-0032-1323686

Kobelt-Pönicke, A., Walter, F., Lid, N., & Ramien, N. (2020). Zum Einfluss einer Ehrlichkeitserklärung auf die Authentizität von Beschwerdedarstellungen [The influence of a statement of truth on the authenticity of complaints]. *Rehabilitation, 59*(5), 291–297. https://doi.org/10.1055/a-1195-7986

*Lace, J. W., Merz, Z. C., & Galioto, R. (2021). Exploring the Structured Inventory of Malingered Symptomatology in patients with multiple sclerosis. *Psychological Injury and Law, 14*(4), 291–303. https://doi.org/10.1007/s12207-021-09424-y

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Marín Torices, M. I., Hidalgo-Ruzzante, N., Daugherty, J. C., Jiménez-González, P., & Perez Garcia, M. (2018). Validation of neuropsychological consequences in victims of intimate partner violence in a Spanish population using specific effort tests. *Journal of Forensic Psychiatry & Psychology, 29*(1), 86–98. https://doi.org/10.1080/14789949.2017.1339106

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. https://doi.org/10.1080/13854046.2015.1087597

McWhirter, L., Ritchie, C. W., Stone, J., & Carson, A. (2020). Performance validity test failure in clinical populations: A systematic review. *Journal of Neurology, Neurosurgery, and Psychiatry, 91*(9), 945–952. https://doi-org.mu.idm.oclc.org/10.1136/jnnp-2020-323776

Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science*, *28*(3), 321–326. https://doi.org/10.1177/0963721419837681

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*(2), 145–152. https://doi.org/10.1016/S0887-6177(01)00191-3

Merten, T., & Dandachi-FitzGerald, B. (2022). Symptom and performance validity assessment: European trends in research and practice. *Psychological Injury and Law, 15*(2), 113–115. https://doi.org/10.1007/s12207-022-09454-0

*Merten, T., Kaminski, A., & Pfeiffer, W. (2020). Prevalence of overreporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients. *The Clinical Neuropsychologist, 34*(5), 1004–1024. https://doi.org/10.1080/13854046.2019.1694073

Merten, T., & Merckelbach, H. (2020). Factious disorders and malingering. In J. R. Geddes, N. C. Andreasen, & G. M. Goodwin (eds.), *New Oxford Textbook of Psychiatry* (3rd ed., pp. 1342–1349). Oxford University Press.

Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law, 9*(2), 102–111. https://doi.org/10.1007/s12207-016-9257-3

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., Eisman, E. J., Kubiszyn, T. W., & Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, *56*(2), 128–165. https://doi.org/10.1037/0003-066X.56.2.128

Migliavaca, C. B., Stein, C., Colpani, V., Barker, T. H., Ziegelmann, P. K., Munn, Z., Falavigna, M., & Prevalence Estimates Reviews-Systematic Review Methodology Group (PERSyst) (2022). Meta-analysis of prevalence: I2 statistic and how to deal with heterogeneity. *Research*

Synthesis Methods, 13(3), 363–367. https://doi.org/10.1002/jrsm.1547

*Miskey, H. M., Martindale, S. L., Shura, R. D., & Taber, K. H. (2020). Distress tolerance and symptom severity as mediators of symptom validity failure in veterans with PTSD. *Journal of Neuropsychiatry and Clinical Neurosciences, 32*(2), 161–167. https://doi.org/10.1176/appi.neuropsych.17110340

Modiano, Y. A., Taiwo, Z., Pastorek, N. J., & Webber, T. A. (2021). The Structured Inventory of Malingered Symptomatology Amnestic Disorders scale (SIMS-AM) is insensitive to cognitive impairment while accurately identifying invalid cognitive symptom reporting. *Psychological Injury and Law, 14*(4), 232–247. https://doi.org/10.1007/s12207-021-09420-2

Niesten, I. J. M., Müller, W., Merckelbach, H., Dandachi-FitzGerald, B., & Jelicic, M. (2017). Moral reminders do not reduce symptom overreporting tendencies. *Psychological Injury and Law, 10*(4), 368–384. https://doi.org/10.1007/s12207-017-9303-9

Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, *29*(11), 1321–1336. https://doi.org/10.1037/pas0000438

*Nübling, R., Henn, J., Kaiser, U., Schmidt, J., & Bassler, M. (2020). Zur Validität subjektiver Patientenangaben: Anmerkungen zur Beschwerdenvalidierung [The validity of subjective patient data: Remarks on the validation of complaints]. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie, 68*(2), 81–95. https://doi.org/10.1024/1661-4747/a000407

Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance validity and symptom validity tests: Are they measuring different constructs? *Neuropsychology*, *35*(3), 241–251. https://doi.org/10.1037/neu0000722

*Oudman, E., Krooshof, E., van Oort, R., Lloyd, B., Wijnia, J. W., & Postma, A. (2020). Effects of Korsakoff amnesia on performance and symptom validity testing. *Applied Neuropsychology: Adult, 27*(6), 549–557. https://doi.org/10.1080/23279095.2019.1576180

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research ed.), 372*, Article n71. https://doi.org/10.1136/bmj.n71

Palmer, A., Borrás, C., Pérez-Pareja, J., Sesé, A., & Vilariño, M. (2013). Are patients with chronic pain and fibromyalgia correctly classified by MMPI-2 validity scales and indexes? *European Journal of Psychology Applied to Legal Contexts, 5*(2), 123–129. https://doi.org/10.5093/ejpalc2013a1

*Peters, M. J. V., Jelicic, M., Moritz, S., Hausschildt, M., & Jelinek, L. (2013). Assessing the boundaries of symptom overreporting using the Structured Inventory of Malingered Symptomatology in a clinical schizophrenia sample: Its relation to symptomatology and neurocognitive dysfunctions. *Journal of Experimental Psychopathology, 4*(1), 64–77. https://doi.org/10.5127/jep. 023811

Pina, D., Puente-López, E., Ruiz-Hernández, J. A., Llor-Esteban, B., & Aguerrevere, L. E. (2022). Self-report measures for symptom validity assessment in whiplash-associated disorders. *The European Journal of Psychology Applied to Legal Context, 14*(2), 73–81. https://doi.org/10.5093/ejpalc2022a7

*Praus, P., Schward, S., Reinhard, I., Zink, M., & Dressing, H. (2021). Validität des "Strukturierten Fragebogens Simulierter Symptome" (SIMS) bei gesunden Probanden und Patienten mit psychiatrischen Erkrankungen außerhalb eines forensischen Kontexts [Validity of the "Structured Inventory of Malingered Symptomatology" (SIMS) in the assessment of healthy individuals and patients with psychiatric disorders outside the forensic context]. *Fortschritte der Neurologie Psychiatrie, 90*(3), 100–107. https://doi.org/10.1055/a-1493-2782

*Puente-López, E., Pina, D., Boskovic, I., Shura, R. D., Martinez-Jarreta, B., & Merten, T. (2022). The impact of different forms of coaching on the Structured Inventory of Malingered Symptomatology (SIMS): A simulation study. *Psicothema, 34*(4), 528–536. https://doi.org/10.7334/psicothema2022.129

Puente-López, E., Pina, D., López-Nicolás, R., Iguacel, I., & Arce, R. (2023). The Inventory of Problems-29 (IOP-29): A systematic review and bivariate diagnostic test accuracy meta-analysis. *Psychological Assessment, 35*(4), 339–352. https://doi.org/10.1037/pas0001209

R Core Team (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Retrieved from https://www.R-project.org

*Rogers, R., Otal, T. K., Velsor, S. F., & Pan, M. (2021). How good are inpatients at feigning Miranda abilities? An investigation of the Miranda Quiz, Inventory of Legal Knowledge, and Structured Inventory of Malingered Symptomatology. *Behavioral Sciences & the Law, 39*(3), 245–261. https://doi.org/10.1002/bsl.2506

Rogers, R., Robinson, E. V., & Gillard, N. D. (2014). The SIMS Screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences & the Law, 32*(4), 455–466. https://doi.org/10.1002/bsl.2131

Roor, J. J., Dandachi-FitzGerald, B., & Ponds, R. W. (2016). A case of misdiagnosis of mild cognitive impairment: The utility of symptom validity testing in an outpatient memory clinic. *Applied Neuropsychology: Adult*, *23*(3), 172–178. https://doi.org/10.1080/23279095.2016.1030018

Roor, J. J., Peters, M. J. V., Dandachi-FitzGerald, B., & Ponds, R. W. H. M. (2024). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychology Review, 34*(1), 299–319. https://doi.org/10.1007/s11065-023-09582-7

Rowland, J. A., Martindale, S. L., Spengler, K. M., Shura, R. D., & Taber, K. H. (2020). Sequelae of blast events in Iraq and Afghanistan war veterans using the Salisbury Blast Interview: A CENC study. *Brain Injury, 34*(5), 642–652. https://doi.org/10.1080/02699052.2020.1729418

Schroeder, R. W., & Martin, P. K. (2022) Validity assessment in clinical settings: How it differs from forensic settings and why it is important. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 3–10). The Guilford Press.

Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology, 35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

*Shura, R. D., Armistead-Jehle, P., Rowland, J. A., Taber, K. H., & Cooper, D. B. (2021). Self-reported neurobehavioral symptoms in combat veterans: An examination of NSI with mBIAS symptom validity scales and potential effects of psychological distress. *Psychological Assessment, 33*(12), 1192–1199. https://doi.org/10.1037/pas0001047

Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured Inventory of Malingered Symptomatology: A psychometric review. *Psychological Injury and Law, 15*(1), 64–78. https://doi.org/10.1007/s12207-021-09432-y

Shura, R. D., Yoash-Gantz, R. E., Pickett, T. C., McDonald, S. D., & Tupler, L. A. (2021). Relations among performance and symptom validity, mild traumatic brain injury, and posttraumatic stress disorder symptom burden in postdeployment veterans. *Psychological Injury and Law, 14(4)*, 257–268. https://doi.org/10.1007/s12207-021-09415-z

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law*, *25*(2), 183–189.

*Spencer, R. J., Gradwohl, B. D., & Kordovski, V. M. (2021). Initial validation of short forms of the SIMS for neuropsychological evaluations. *Psychological Injury and Law, 14*(1), 37–45. https://doi.org/10.1007/s12207-020-09394-7

Stijnen, T., Hamza, T. H., & Ozdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine, 29*(29), 3046–3067. https://doi.org/10.1002/sim.4040

Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

Tiemens, B., Wagenvoorde, R., & Witteman, C. (2020). Why every clinician should know Bayes' rule. *Health Professions Education, 6*(3), 320–324. https://doi.org/10.1016/j.hpe.2020.05.002

*Tierney, S. M., Webber, T. A., Collins, R. L., Pacheco, V. P., & Grabyan, J. M. (2021). Validity and utility of the Miller Forensic Assessment of Symptoms Test (M-FAST) on an inpatient epilepsy monitoring unit. *Psychological Injury and Law, 14*(4), 248–256. https://doi.org/10.1007/s12207-021-09418-w

*van Beilen, M., Griffioen, B. T., Gross, A., & Leenders, K. L. (2009). Psychological assessment of malingering in psychogenic neurological disorders and non-psychogenic neurological disorders: Relationship to psychopathology levels. *European Journal of Neurology, 16*(10), 1118–1123. https://doi.org/10.1111/j.1468-1331.2009.02655.x

van der Heide, D., Boskovic, I., & Merckelbach, H. (2017). Standard symptom inventories for asylum seekers in a psychiatric hospital: Limited utility due to poor symptom validity. *Psychological Injury and Law, 10*(4), 358–367. https://doi.org/10.1007/s12207-017-9302-x

van der Heide, D., Boskovic, I., van Harten, P., & Merckelbach, H. (2020). Overlooking feigning behavior may result in potential harmful treatment interventions: Two cases of undetected malingering. *Journal of Forensic Sciences, 65*(4), 1371–1375. https://doi.org/10.1111/1556-4029.14320

van der Heide, D., & Merckelbach, H. (2016). Validity of symptom reports of asylum seekers in a psychiatric hospital: A descriptive study. *International Journal of Law and Psychiatry, 49*(A), 40–46. https://doi.org/10.1016/j.ijlp.2016.05.007

van Egmond, J. J., & Kummeling, I. (2002). A blind spot for secondary gain affecting treatment outcome. *European Psychiatry, 17*(1), 46–54. https://doi.org/10.1016/S0924-9338(02)00622-3

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*(8), 1336–1365. https://doi.org/10.1080/13854046.2014.984763

van Impelen, A., Merckelbach, H., Jelicic, M., Niesten, I. J. M., & à Campo, J. (2017). Differentiating factitious from malingered symptomatology: The development of a psychometric approach. *Psychological Injury and Law, 10*(4), 341–357. https://doi.org/10.1007/s12207-017-9301-y

*van Minnen, A., van Dalen, B., Voorendonk, E. M., Wagenmans, A., & de Jongh, A. (2020). The effects of symptom overreporting on PTSD treatment outcome. *European Journal of Psychotraumatology, 11*(1), Article 1794729. https://doi.org/10.1080/20008198.2020.1794729

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viglione, D. J., Giromini, L., & Landis, P. (2017). The development of the Inventory of Problems-29: A brief self-administered measure for discriminating bona fide from feigned psychiatric and cognitive complaints. *Journal of Personality Assessment, 99*(5), 534–544. https://doi.org/10.1080/00223891.2016.1233882

Ward, M. K., & Meade, A. W. (2023). Dealing with careless responding in survey data: Prevention, identification, and recommended best practices. *Annual Review of Psychology*, *74,* 577–596. https://doi.org/10.1146/annurev-psych-040422-045007

*Wertz, M., Mader, E., Nedopil, N., Schiltz, K., & Yundina, E. (2021). Antwortverzerrung oder Symptombelastung? Beschwerdeschilderung von psychiatrischen Patienten und sozialmedizinischen Begutachtungsprobanden [Response distortion or symptom severity? Symptom description by psychiatric patients and sociomedical assessment subjects]. *Nervenarzt, 92*(11), 1163–1171. https://doi.org/10.1007/s00115-020-01041-5

*Wolf, E. J., Ellickson-Larew, S., Guetta, R. E., Escarfulleri, S., Ryabchenko, K., & Miller, M. W. (2020). Psychometric performance of the Miller Forensic Assessment of Symptoms Test (M-FAST) in veteran PTSD assessment. *Psychological Injury and Law*, *13*(3), 284–302. https://doi.org/10.1007/s12207-020-09373-y

*Zimmermann, P., Kowalski, J. T., Alliger-Horn, C., Danker-Hopfe, H., Engers, A., Meermann, R., & Hellweg, R. (2013). Detection of malingering in the assessment of occupational disability in the military. *German Journal of Psychiatry, 16*(2), 54–60