



How to Distinguish Feigned from Genuine Depressive Symptoms: Response Patterns and Content Analysis of the SIMS Affective Disorder Scale

Eleonora Ricci¹ · Marco Colasanti² · Merylin Monaro³ · Cristina Mazza¹ · Alessandra Cardinale⁴ · Francesca Bosco⁵ · Giuliana Mazzoni⁶ · Clelia Rossi Arnaud⁷ · Stefano Ferracuti⁵ · Alberto Di Domenico² · Paolo Roma⁵

Received: 19 May 2023 / Accepted: 27 July 2023 / Published online: 1 September 2023
© The Author(s) 2023

Abstract

In civil and forensic evaluations of psychological damage, depression is one of the most commonly identified disorders, and also one of the most frequently feigned. Thus, practitioners are often confronted with situations in which they must assess whether the symptomatology presented by a patient is genuine or being feigned for secondary gains. While effective, traditional feigning detection instruments generate a high number of false positives—especially among patients presenting with severe symptomatology. The current study aimed at equipping forensic specialists with an empirical decision-making strategy for evaluating patient credibility on the basis of test results. In total, 315 participants were administered the Beck Depression Inventory-II (BDI-II) and SIMS Affective Disorders (SIMS AF) scales. Response patterns across the experimental groups (i.e., Honest, Simulators, Honest with Depressive Symptoms) were analyzed. A machine learning decision tree model (i.e., J48), considering performance on both measures, was built to effectively distinguish Honest with Depressive Symptoms subjects from Simulators. A forward logistic regression model was run to determine which SIMS AF items best identified Simulators, in comparison with Honest with Depressive Symptoms subjects. The results showed that the combination of feigning detection instruments and clinical tests generated incremental specificity, thereby reducing the risk of misclassifying Honest with Depressive Symptoms subjects as feigners. Furthermore, the performance analysis of SIMS AF items showed that Simulators were more likely to endorse three specific items. Thus, computational models may provide effective support to forensic practitioners, who must make complex decisions on the basis of multiple elements. Future research should revise the content of SIMS AF items to achieve better accuracy in the discrimination between feigners and honest subjects with depressive symptoms.

Keywords Psychological damage evaluation · Forensic · Affective disorders · Machine learning · BDI-II

Eleonora Ricci and Marco Colasanti contributed equally.

✉ Eleonora Ricci
eleonora.ricci@unich.it

¹ Department of Neuroscience, Imaging and Clinical Sciences, G. d'Annunzio University of Chieti-Pescara, 66100 Chieti, Italy

² Department of Psychological, Health and Territorial Sciences, G. d'Annunzio University of Chieti-Pescara, 66100 Chieti, Italy

³ Department of General Psychology, University of Padova, 35121 Padua, Italy

⁴ National Nuclear Physics Institute, 00186 Rome, Italy

⁵ Department of Human Neuroscience, Sapienza University of Rome, 00185 Rome, Italy

⁶ Department of Dynamic and Clinical Psychology, and Health Studies, Sapienza University of Rome, 00185 Rome, Italy

⁷ Department of Psychology, Sapienza University of Rome, 00185 Rome, Italy

Introduction

Depression is the most prevalent mental disorder, affecting 4.4% of the global population (equivalent to 300 million people worldwide; World of Health Organization, 2017). It is also the leading cause of disability and one of the most common disorders encountered in civil and forensic evaluations of psychological damage (Cuijpers & Smit, 2008; Druss et al., 2000). Consequently, the disorder has a tremendous socio-economic cost. The WHO (2018) estimated the annual direct cost of depression in the European Union as approximately €617 billion in 2013, taking into account treatment, absenteeism, unemployment, disability benefits, and compensatory measures by insurance companies. In medico-legal settings, individuals frequently fabricate or exaggerate psychological symptoms of a wide array of mental disorders, in an attempt to obtain compensation; such behavior is known as “feigning” (Rogers & Bender, 2018). Depression is one of the most commonly feigned psychiatric disorders in forensic evaluations (16.08%; Mittenberg et al., 2002), because most individuals have experienced a low mood at least once (Bush et al., 2014), and can fabricate the symptoms on the basis of this experience (Sullivan & King, 2010). Therefore, it is of paramount importance that practitioners can effectively evaluate the reliability of the presented symptomatology.

Various strategies have been proposed to identify feigners of a broad range of psychopathological conditions, including depression. Among these strategies, the administration of symptom validity tests (SVTs) is most common. SVTs are self-report questionnaires that analyze the credibility/validity of a subject’s presented symptoms and psychological problems. Most SVTs rely on a rare or quasi-rare symptoms approach, presenting test-takers with items that are rare or improbable, in combination (Rogers & Bender, 2018; for a review, see Giromini et al., 2022; Sherman et al., 2020).

One of the most widely applied SVTs is the Structured Inventory of Malingered Symptomatology (SIMS; Dandachi-FitzGerald et al., 2013; Martin et al., 2015; Smith & Burger, 1997). The SIMS is a standalone instrument that is used to screen for feigned psychiatric symptoms and/or cognitive deficits. It describes implausible, rare, atypical, or extreme symptoms, which respondents must either endorse or reject. Although the instrument is undoubtedly serviceable, its validity in medico-legal settings is controversial. Specifically, the SIMS is prone to generating false positives, especially within the clinical population (Harris & Merz, 2022). In fact, a meta-analysis including more than 4,000 protocols highlighted the high sensitivity but low specificity of the SIMS, when applied to severely disturbed patients (van Impelen et al., 2014). In other words, the instrument identifies feigners at a high rate, but misclassifies many legitimate patients as feigners. This

occurs because severe patients consistently manifest many symptoms of high intensity, and, likewise, feigners tend to endorse a wide range of symptoms at heightened degrees of symptom intensity (Chafetz & Underhill, 2013). Thus, severe patients show similar response patterns to feigners, making it difficult to differentiate between the two profiles.

With respect to feigned depression, specifically, the SIMS Affective Disorder (AF) scale assesses the degree to which test-takers report atypical symptoms of affective disorders. Recently, scholars have questioned the efficacy of this scale in distinguishing between feigners and genuine patients with severe symptomatology, arguing that there is considerable overlap between item content and true symptomatology (e.g., trouble sleeping; Shura et al., 2022). Indeed, some authors, disputing the content validity of the SIMS AF scale, have argued that its items do not reflect atypical, improbable, inconsistent, or illogical symptoms, as reported by the SIMS professional manual (Widows & Smith, 2005). Further empirical research is needed to settle this debate.

The present study aimed at making an empirical contribution to the challenge of differentiating the response profiles of feigners, honest subjects, and honest subjects with depressive symptoms. To that end, we used machine learning techniques to identify decision-making rules that effectively distinguished honest respondents from feigners. Over the past decade, research has shown that the application of artificial intelligence algorithms to psychometric tests results in a superior detection of feigning, relative to traditional methods (Mazza et al., 2019a; Orrù et al., 2021). Specifically, machine learning algorithms are more effective at delineating the subtle response patterns typical of feigners, and psychometric instruments (e.g., the SIMS) that incorporate these algorithms tend to show better performance, relative to the traditional cut-offs (Orrù et al., 2021). A further goal of the research was to provide a strategy for medico-legal practitioners to analyze the responses to questionnaires that are routinely used in clinical-forensic practice, in order to support their evidence-based decision-making about patient credibility. Finally, the research aimed at evaluating the role of single items of the SIMS AF scale in distinguishing between feigners and honest subjects with depressive symptoms, in order to generate insight into the content validity of the scale.

Materials and Methods

Participants

A total of 340 respondents participated in the study. The inclusion criteria were: (a) aged 18 years and older, (b) living in Italy, and (c) able to read questions on a computer monitor and understand the meaning of those questions. Data were collected over 15 days (i.e., November 16–30, 2020).

The questionnaires were administered cross-sectionally on an online survey platform, which participants accessed via a designated link that was disseminated over email using convenience sampling. Nine participants were excluded because they did not understand the instructions (see the “Research Design” section) and 16 were excluded because they did not complete the questionnaires. The final sample consisted of 315 participants. All participants voluntarily responded anonymously, indicating their informed consent within. The procedures were clearly explained, and participants could interrupt or quit the study at any point without declaring their reasons for doing so. They did not receive any compensation for their participation. The experimental procedure was approved by the local ethics committee (Board of the Department of Human Neuroscience, Faculty of Medicine and Dentistry, Sapienza University of Rome), in accordance with the Declaration of Helsinki.

Participants were grouped into three subsamples: Honest [H], Simulators [S], and Honest with Depressive Symptoms [HDS]. The Honest group included 110 participants, aged 23–68 years ($M=37.30$, $SD=12.52$). Most were male ($n=66$, 60.0%), Italian citizens ($n=105$, 95.5%), and residents of central Italy ($n=70$, 63.6%); 50 held a high school diploma (44.5%) and the majority were employees ($n=56$, 50.9%) and unmarried ($n=71$, 64.5%). The Simulators group was composed of 161 participants, aged 24–75 years ($M=36.30$, $SD=11.87$). Most were female ($n=90$, 55.9%), Italian citizens ($n=160$, 99.4%), and residents of central Italy ($n=105$, 65.2%); 79 held a high school diploma (49.1%), 75 were employees (46.6%), and most were unmarried ($n=104$, 64.6%). The Honest with Depressive Symptoms group comprised 44 participants, aged 25–64 years ($M=35.59$, $SD=12.60$). Most were female ($n=28$, 63.6%), Italian citizens ($n=44$, 100%), and residents of central Italy ($n=26$, 59.1%); 20 held a high school diploma (45.5%), 16 were employees (36.4%), and most were unmarried ($n=32$, 72.7%).

The chi-squared test (χ^2) revealed statistically significant differences between the Honest, Simulators, and Honest with Depressive Symptoms groups, with respect to biological sex [$\chi^2(2)=9.666$, $p=0.008$], with more male participants in the Honest group compared to the other two groups; and employment status [$\chi^2(8)=17.547$, $p=0.025$], with more unemployed participants in the Honest with Depressive Symptoms group compared to the Honest group. No statistically significant differences emerged with respect to the other socio-demographic variables (see Table 1).

Measures

The Beck Depression Inventory – Second Edition (BDI-II)

The BDI-II (Beck et al., 1996; Ghisi et al., 2006) is one of the most widely used instruments for screening depressive symptomatology (von Glischinski et al., 2019). The

self-administered test consists of 21 items that assess the cognitive, affective, motivational, and somatic symptoms of depression: sadness, pessimism, past failure, loss of pleasure, guilty feelings, punishment feelings, self-dislike, self-criticalness, suicidal ideation or wishes, crying, agitation, loss of interest, indecisiveness, feelings of worthlessness, loss of energy, change in sleeping patterns, irritability, change in appetite, concentration difficulty, tiredness or fatigue, and loss of interest in sex (Beck et al., 1996). Each item consists of a list of four statements arranged in order of increasing severity, referring to a particular symptom of depression that respondents may have felt during the prior 2 weeks. Answers are provided on a four-point scale, ranging from 0 to 3. The total score is the summation of respondents' scores for the 21 items, with a maximum of 63. BDI-II items may be grouped into two subscales: the Somatic-Affective subscale, comprised of 12 items that describe the affective, somatic, and vegetative symptoms of depression; and the Cognitive subscale, comprised of 9 items that represent the cognitive symptoms of depression (Beck et al., 1996; Steer et al., 1999). The present study administered the official Italian adaptation of the test (Ghisi et al., 2006), and the internal consistency was excellent ($\alpha=0.972$).

The Structured Inventory of Malingered Symptomatology Affective Disorders Scale (SIMS AF)

The present study administered the Affective Disorders (AF) scale of the SIMS Italian adaptation (La Marca et al., 2011). The SIMS (Smith & Burger, 1997; Widows & Smith, 2005) is a multi-axial self-report questionnaire that aims at identifying respondents' feigning of psychiatric symptoms and/or cognitive deficits. It is comprised of 75 items, describing implausible, rare, atypical, or extreme symptoms that respondents must endorse or reject. The measure has been validated in the clinical-forensic, psychiatric, and non-clinical fields (Harris & Merz, 2022; Monaro et al., 2018; Orrù et al., 2021, 2022). The SIMS AF scale consists of 15 items—each associated with a specific symptom of depression or anxiety; respondents report each symptom via a dichotomous response option (i.e., *true* vs. *false*). The SIMS AF scale has a cut-off of >5 (La Marca et al., 2011). In the present study, it showed good internal consistency ($\alpha=0.768$).

Of note, although the total SIMS score suggests the presence of feigning, the subscale scores suggest the type of psychopathology that is being feigned (e.g., Shura et al., 2022; van Impelen et al., 2014). In the present study, we used only items from the AF scale, in order to mask the study aim. In more detail, the simulation design asked participants to simulate depressive symptoms. Therefore, the inclusion of all SIMS items may have made it immediately obvious which items should be endorsed, in order to comply with the feigning instructions, thereby hindering the content

Table 1 Descriptive Statistics for the Total Sample (N=315) and Each Group (i.e., Honest, Simulators, Honest with Depressive Symptoms)

	Total N = 315	Honest	Simulators	Honest with Depressive Symptoms	F/ χ^2	p
Age M (SD)	36.55 (12.12)	37.30 (12.52)	36.30 (11.87)	35.59 (12.16)	.380	.684
Biological sex					9.666	.008
Female	162 (51.4)	44 (14.0) ^a	90 (28.6) ^b	28 (8.9) ^b		
Male	153 (48.6)	66 (21.0) ^a	71 (22.5) ^b	16 (5.1) ^b		
Citizenship					10.397	.581
Italian	309 (98.1)	105 (33.3)	160 (50.8)	44 (14.0)		
Foreign	6 (1.9)	5 (1.5)	1 (0.3)	0 (0.0)		
Educational level					7.095	.526
Primary school	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.3)		
Middle school	21 (6.7)	9 (2.9)	9 (2.9)	3 (1.0)		
High school	149 (47.3)	50 (15.9)	79 (25.1)	20 (6.3)		
Degree	138 (43.8)	49 (15.6)	70 (22.2)	19 (6.0)		
Postgraduate degree	6 (1.9)	2 (0.6)	3 (1.0)	1 (0.3)		
Occupation					17.547	.025
Unemployed	33 (10.5)	5 (1.6) ^a	18 (5.7) ^{a,b}	10 (3.2) ^b		
Student	49 (15.6)	12 (3.8) ^a	28 (8.9) ^a	9 (2.9) ^a		
Employee	147 (46.7)	56 (17.8) ^a	75 (23.8) ^a	16 (5.1) ^a		
Freelancer	79 (25.1)	34 (10.8) ^a	36 (11.4) ^a	9 (2.9) ^a		
Retired	7 (2.2)	3 (1.0) ^a	4 (1.3) ^a	0 (0.0) ^a		
Residence					3.337	.766
Northern Italy	25 (7.9)	10 (3.2)	11 (3.5)	4 (1.3)		
Central Italy	201 (63.8)	70 (22.2)	105 (33.3)	26 (8.3)		
Southern Italy and Islands	87 (27.6)	30 (9.5)	44 (14.0)	13 (4.1)		
Abroad	2 (0.6)	0 (0.0)	1 (0.3)	1 (0.3)		
Marital status					3.886	.692
Single	207 (65.7)	71 (22.5)	104 (33.0)	32 (10.2)		
Married	99 (31.4)	34 (10.8)	54 (17.1)	11 (3.5)		
Separated/Divorced	8 (2.5)	4 (1.3)	3 (1.0)	1 (0.3)		
Widowed	1 (0.3)	1 (0.3)	0 (0.0)	0 (0.0)		

Different letters indicate a significant difference between columns

validity analysis. Furthermore, administering the SIMS in its entirety would have added a disproportionate number of non-believable and unrelated items, considering that the BDI-II is comprised of 21 items, whereas the SIMS and its AF scale include 75 and 15 items, respectively.

Research Design

A between-subjects design was implemented and the informatic system randomly assigned participants to one of two experimental groups, defined by the manipulated factor of instruction: Honest [H] vs. Simulators [S]. In the first group [H], participants completed the tests (i.e., SIMS AF, BDI-II) with the instruction to respond honestly. In the Simulators group [S], participants completed the tests with the instruction to feign depression, according to the DSM-5 criteria for major depressive disorder (American Psychiatric

Association, 2013¹). Of note, the experimental instructions provided to Simulators contained coaching elements—namely symptom preparation and warnings (Puente-López et al., 2022). In fact, participants were clearly instructed to attend to not only the symptoms of major depressive disorder, but also the questionnaire features designed to detect feigning, as their aim was to respond in such a way that their deception would not be detected. We report the experimental instructions in the [Appendix](#) at the end of the manuscript.

At the end of the survey, a final question implemented as a manipulation check asked participants to describe how

¹ The Major Depressive Disorder criteria of the DSM-5-TR (American Psychiatric Association, 2021) are almost the same as those of the previous edition of the manual, with the exception of the fifth criterion, which now states: “Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).”.

they responded to the items: “honestly,” “dishonestly,” or “I don’t remember.” Nine participants in the Simulators group were excluded from the analysis because they answered “honestly” to this question, suggesting that they may not have understood the instructions; and 16 were excluded because they did not complete the questionnaires.

Following the data collection, the descriptive statistics revealed that 28.57% of Honest participants ($n = 44$) scored higher than the cut-off of 12 for the BDI-II total score. According to the Italian technical manual (Ghisi et al., 2006, p. 67), this cut-off should be used as the initial interpretive criterion, for research purposes. Following this recommendation, we grouped participants with a BDI-II total score higher than 12 into a third group, called Honest with Depressive Symptoms [HDS].² This result was not surprising, considering that the data collection occurred during the second wave of the COVID-19 pandemic in Italy, when an increase in depressive symptoms among the general population was observed (Mazza et al., 2022).

Data Analysis

Response Pattern Analysis

A preliminary analysis was run to investigate the response patterns of the three experimental groups (i.e., Honest, Simulators, Honest with Depressive Symptoms) on the SIMS AF scale and the BDI-II (i.e., total score, Cognitive and Somatic-Affective scale scores). We calculated the correlation coefficient (r) between the four scales (i.e., SIMS AF, BDI-II total, BDI-II Cognitive, BDI-II Somatic-Affective) and computed a correlation matrix for each experimental group. Finally, the z -test for comparing sample correlation coefficients was applied to determine differences between groups (with significance set to $\alpha = 0.05$ and the critical value for the z -statistic set to $z = 1.96$).

Machine Learning Model

To determine an interpretable decision model to differentiate the three groups (i.e., Honest, Simulators, Honest with Depressive Symptoms), we built a three-class decision tree using machine learning (ML) methodology. More specifically, we trained a J48 algorithm using tenfold cross-validation. J48 is an algorithm used to generate decision trees (Quinlan, 1993). While it is one of the simplest statistical classifiers, its output

has a high degree of interpretability and explainability (i.e., transparency). In domains such as healthcare, the transparency of ML models is particularly important, especially when artificial intelligence is applied to support clinical decision-making (Adadi & Berrada, 2020). Thus, various ML models imply a different trade-off between accuracy and transparency.

In the present research, we chose a simple but highly interpretable classifier (i.e., J48), as our aim was to produce a model that could concretely support clinicians and forensic experts in their efforts to detect simulated depression. Moreover, we followed a tenfold cross-validation procedure in order to guarantee model generalization and increase the replicability of the results. The k -fold cross-validation consisted of randomly and repeatedly splitting the sample into training and validation sets. This resampling procedure reduced the variance in the model performance estimation with respect to the use of a single training set and a single validation set, thus reducing overfitting (Kohavi, 1995). The sample ($N = 315$) was partitioned into $k = 10$ subsamples of equal size: 9 subsamples were used to train the model and the remaining subsample was used for its validation. This process was repeated 10 times, so that each of the 10 folds was used only once as a validation set. Finally, an estimated validation accuracy was generated by taking an average of the results obtained from the 10 folds. The ML model was run using WEKA 3.9 (Frank et al., 2016).

Content Validity Analysis

To assess the content validity of SIMS AF items, the following analyses were performed, considering only the Simulators and Honest with Depressive Symptoms groups. First, given the imbalance between groups (Simulators: $n = 161$; Honest with Depressive Symptoms: $n = 44$), the Simulators group was subsampled, extracting 44 observations randomly. Subsequently, chi-squared tests were performed to evaluate the associations between groups (Simulators vs. Honest with Depressive Symptoms) and responses to SIMS AF items (i.e., *true* vs. *false*). Concerning the magnitude of the associations, Cramer’s $V \leq 0.20$ was considered indicative of a weak effect, $0.20 < \text{Cramer’s } V \leq 0.60$ of a moderate effect, and Cramer’s $V > 0.60$ of a strong effect.³ Finally, a forward logistic regression model was implemented with only the items that emerged as significant in the previous analysis, in order to determine which items discriminated between groups. Data analyses were computed using the Pandas software library (McKinney, 2010) and SPSS version 28 (IBM Corp., 2021).

² This group was comprised of subjects with depressive symptoms, according to the cut-off (> 12) reported in the Italian BDI-II manual (Ghisi et al., 2006), who were not seeking treatment.

³ Cramer’s V —IBM documentation <https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-cramrs-v>

Table 2 Means and SDs for the Honest, Simulators, and Honest with Depressive Symptoms groups on the AF SIMS and BDI-II scales

		SIMS AF	BDI-II total	BDI-II Cognitive	BDI-II Somatic-Affective
Honest	-	3.78 (1.69)	6.71 (5.12)	2.28 (2.78)	4.43 (3.16)
	<i>min-max scores</i>	1–10	0–12	0–20	0–20
Simulators	-	10.04 (2.36)	41.22 (14.67)	17.28 (6.79)	23.94 (8.54)
	<i>min-max scores</i>	1–14	0–63	0–27	0–36
Honest with Depressive Symptoms	-	6.30 (2.12)	19.57 (6.94)	7.32 (4.27)	12.25 (4.16)
	<i>min-max scores</i>	3–12	13–44	1–21	6–27

SIMS AF Structured Inventory of Malingered Symptomatology – Affective Disorders scale, BDI-II Beck Depression Inventory – Second Edition

Results

Response Patterns on the Clinical and Symptom Validity Scales

Table 2 reports the average scores obtained by the three experimental groups for the administered scales (i.e., SIMS AF, BDI-II total, BDI-II Cognitive, BDI-II Somatic-Affective).

For each experimental group, the correlation coefficients (*r*) between scales (i.e., SIMS AF, BDI-II total, BDI-II Cognitive, BDI-II Somatic-Affective) were computed. Figure 1 reports the results.

The *z*-test comparison of correlation coefficients revealed no statistical differences between the Honest with Depressive Symptoms and Honest groups for any of the correlations presented in the matrix (see Fig. 2 for the *z* values). In contrast, the Honest and Simulators groups statistically differed with respect to all correlations presented in the matrix, with the exception of that between the SIMS AF and BDI-II total, and that between the SIMS AF and

BDI-II Somatic-Affective. However, the most interesting results emerged in the comparison between the Simulators and Honest with Depressive Symptoms groups, representing the critical situation faced by clinicians and forensic experts. Specifically, these groups statistically differed with respect to all correlation coefficients. While the Honest with Depressive Symptoms group showed a weak correlation ($r \leq 30$) between the SIMS AF and BDI-II (i.e., total, Cognitive, Somatic-Affective), the Simulators group presented a moderate correlation ($50 < r < 70$). Moreover, the Honest with Depressive Symptoms group showed a medium correlation ($30 < r < 40$) between the two BDI-II subscales (i.e., Cognitive, Somatic-Affective), whereas the Simulators group registered a strong correlation ($r > 80$).

Machine Learning: J48 Decision Tree Model

Participants’ scores on the administered scales (i.e., SIMS AF, BDI-II total, BDI-II Cognitive, BDI-II Somatic-Affective) were entered as predictors in the decision tree model (i.e., J48). All

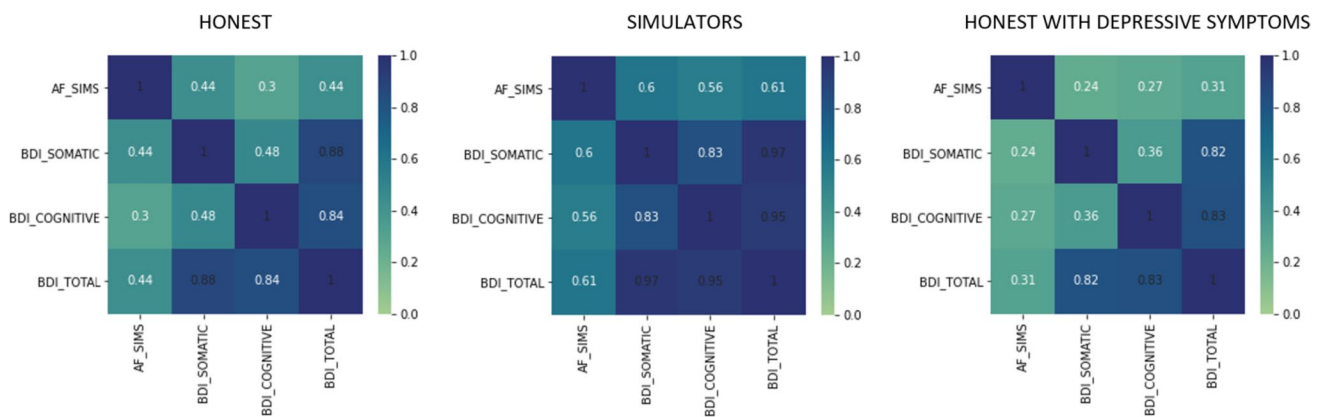
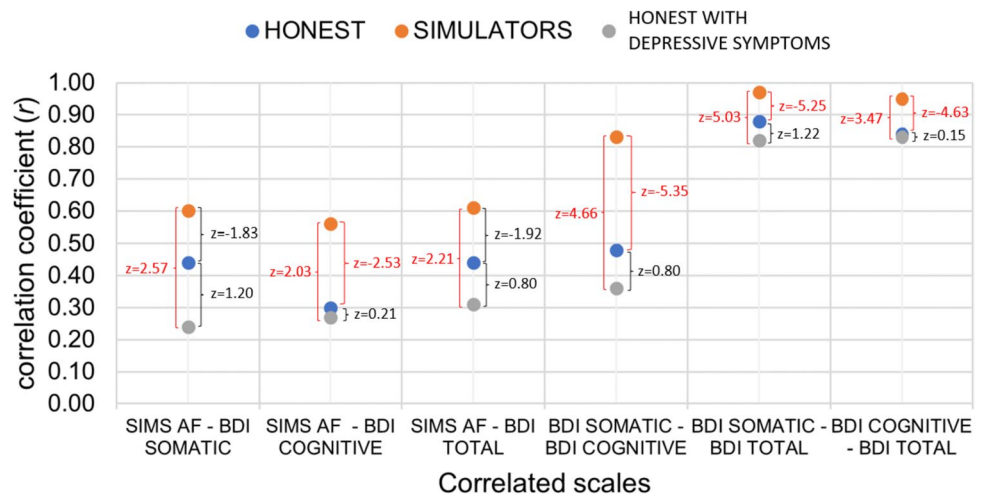


Fig. 1 Correlation Matrices between the Scores Obtained by Each Group on the SIMS AF scale and the BDI-II (Total, Cognitive, Somatic-Affective). Correlation Coefficients (*r*) Reported in the Matrix Squares

Fig. 2 Between-Groups Comparison of Correlation Coefficients between the Administered Scales. Statistically Significant Comparisons ($z = \pm 1.96$) Highlighted in Red



predictors showed a high correlation with the outcome variable (i.e., Honest vs. Simulators vs. Honest with Depressive Symptoms): SIMS AF $r=0.67$, BDI-II total $r=0.68$, BDI-II Cognitive $r=0.66$, BDI-II Somatic-Affective $r=0.67$.

The model was trained and validated according to the procedure described in the “Machine Learning” section. Table 3 reports the results for the following classification metrics: ROC area, precision, recall, and *F*-measure (*F1* score). The overall classification accuracy of the model was 87.62%. With respect to the distribution of errors between the three classes, the confusion matrix reported more errors for the Honest with Depressive Symptoms group (13/44 = 29.54%; all misclassified as Simulators), followed by the Simulators group (23/161 = 14.29%; 8 misclassified as Honest, 15 misclassified as Honest with Depressive Symptoms), and relatively few errors for the Honest group (3/110 = 2.73%; all misclassified as Simulators). Figure 3 illustrates the decision rules that the model used to classify subjects.

First, the BDI-II total score was considered: low scores (≤ 12) discriminated Honest subjects from Honest with Depressive Symptoms subjects and Simulators. In contrast, subjects with a very high (> 29) BDI-II total score were classified as Simulators. For intermediate BDI-II total scores ($12 < x \leq 29$), the SIMS AF score was analyzed: subjects with a SIMS AF score > 9 were mostly Simulators, and

subjects with a score ≤ 9 were mostly Honest with Depressive Symptoms subjects. The BDI-II Cognitive score and the SIMS AF cut-off of 11 were helpful to fine-tune the discrimination between Honest with Depressive Symptoms subjects and Simulators.

Of note, the decision tree model identified most Simulators using the BDI-II total score, rather than the SIMS AF score. Moreover, the main SIMS AF cut-off proposed by the algorithm to distinguish Simulators from Honest with Depressive Symptoms subjects (i.e., cut-off = 9) differed from the reported value in the SIMS manual (i.e., cut-off = 5). When the cut-off reported in the SIMS manual was applied, 56.82% of Honest with Depressive Symptoms subjects were erroneously classified as Simulators (see Table 4). In contrast, the decision tree classification rules resulted in only 29.54% of Honest with Depressive Symptoms subjects misclassified as Simulators. In other words, considering the two target classes (i.e., Simulators, Honest with Depressive Symptoms), the decision tree specificity was 0.70, whereas the specificity of the SIMS manual cut-off was 0.43. Sensitivity was 0.86 for the decision tree model and 0.94 for the SIMS manual cut-off.

Content Validity of SIMS AF Items

The ability of each SIMS AF scale item to classify Honest with Depressive Symptoms subjects and Simulators was investigated. Chi-squared tests were performed to evaluate the associations between groups and the responses to each SIMS AF item.⁴ Significant ($p < 0.05$) and moderate ($0.237 \leq \text{Cramer’s } V \leq 0.586$) associations

Table 3 Metrics for the Decision Tree Model (J48) that Was Trained and Validated Using 10-Fold Cross-Validation

-	Precision	Recall	F-measure	ROC area
Honest	0.930	0.973	0.951	0.949
Simulators	0.896	0.857	0.876	0.874
Honest with Depressive Symptoms	0.674	0.705	0.689	0.870
Weighted avg	0.877	0.876	0.876	0.900

⁴ All of the listed items belong to the SIMS AF scale. For copyright and test security reasons, item numbers have been blinded. Readers who are interested in the specific item numbers may request this information from the corresponding author.

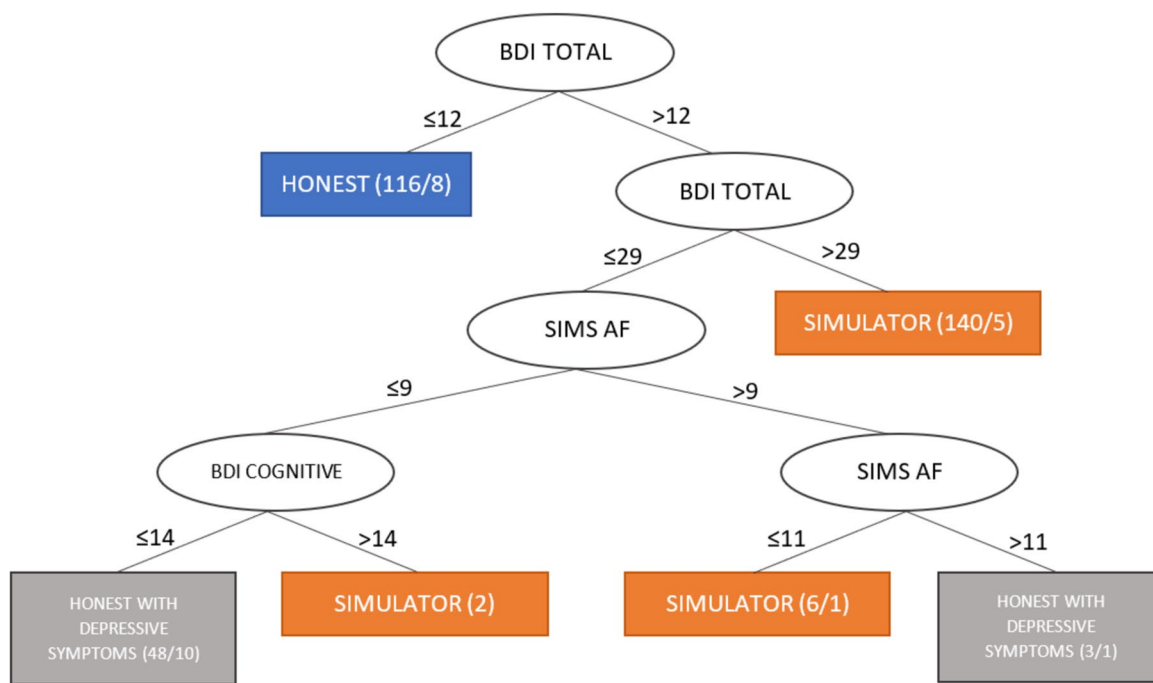


Fig. 3 Decision Tree Model Classification Rules

emerged for six items, with chi-squared values ranging from 4.950–30.171. Furthermore, significant and *strong* ($0.663 \leq \text{Cramer's } V \leq 0.773$) associations emerged for three items, with chi-square values ranging from 38.727–52.545. In detail, endorsement of these nine items was significantly associated with the Simulators group. Conversely, a significant and *moderate* association was recorded for one item ($\chi^2(1) = 7.543$; $V = 0.293$), indicating that endorsement of this item was significantly associated with the Honest with Depressive Symptoms group. The remaining five items were not significantly associated ($p > 0.05$) with either group.

A forward logistic regression model was implemented with the 10 items that were found to be significantly

associated in the previous analysis. Three items were found to be significant, and Table 5 presents the resulting model.

The prediction model indicated that the final model was superior to that which used only the intercept to the observed data ($\chi^2(3) = 71.340$, $p < 0.001$). Nagelkerke's pseudo R^2 of 0.752 illustrated that the significant SIMS AF items explained approximately 75.2% of the variability. In particular, when a respondent answered *true* (vs. *false*) for these three SIMS AF items, their probability of belonging to the Simulators group increased.

The prediction had an accuracy of 87.2%, a specificity of 0.86 (Honest with Depressive Symptoms), and a sensitivity of 0.88 (Simulators).

Table 4 Confusion Matrix of Classification Errors Made by the Decision Tree Model and the AF Cut-Off for the Simulators and Honest with Depressive Symptoms Groups

	% error applying decision tree model		% error applying SIMS AF cut-off (manual)	
	Simulators	Honest with Depressive Symptoms	Simulators	Honest with Depressive Symptoms
Simulators	138/161 = 85.71%	15/161 = 9.32%	151/161 = 93.79%	10/161 = 6.21%
Honest with Depressive Symptoms	13/44 = 29.54%	31/44 = 70.45%	25/44 = 56.82%	19/44 = 43.18%

Considering only these classes, the decision tree model achieved 82.44% accuracy and the SIMS AF cut-off achieved 82.93% accuracy

SIMS AF Structured Inventory of Malingered Symptomatology – Affective Disorders scale

Table 5 Association between Significant SIMS AF Items and the Two Groups (Honest with Depressive Symptoms and Simulators)

Predictor	Estimate	EXP(B)	95% C.I		<i>p</i>
SIMS AF item X	-2.156	0.116	0.020	0.683	0.017
SIMS AF item Y	-2.289	0.101	0.019	0.551	0.008
SIMS AF item Z	-2.353	0.095	0.013	0.721	0.023

The Simulators group served as the baseline

SIMS AF Structured Inventory of Malingered Symptomatology – Affective Disorders scale

Discussion

The present research aimed at equipping forensic specialists with an empirical decision-making strategy for evaluating a patient's credibility on the basis of test results. Although the SIMS is an effective and highly sensitive tool for identifying feigning behavior, it has the unfortunate side effect of generating a large number of false positives, and thereby misclassifying patients—especially those with severe symptomatology—as feigners (van Impelen et al., 2014). Additionally, the literature highlights that a sole reliance on SVTs may yield unreliable results. For this reason, many authors have suggested that multiple instruments (Bush et al., 2014; Chafetz et al., 2015; Rogers & Bender, 2018) and/or deception detection techniques (e.g., Mazza et al., 2020; Monaro et al., 2021) should be applied to detect feigning behavior. For instance, Giromini et al. (2019) found that, when assessing depression-related symptoms, the combined application of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) and the Inventory of Problems-29 (IOP-29) provided incremental validity over the application of either instrument, alone. Accordingly, we jointly analyzed the response profiles of Simulators versus Honest with Depressive Symptoms subjects on clinical and symptom validity scales, finding very different patterns. On average, Honest with Depressive Symptoms subjects scored *lower* than Simulators on all clinical and symptom validity scales. Moreover, they showed a *weak* correlation between the BDI-II scores (i.e., total, Cognitive, Somatic-Affective) and the SIMS AF score, and a *moderate* correlation between scores on the BDI-II subscales (i.e., Cognitive, Somatic-Affective). In contrast, Simulators obtained *higher* scores on all of the administered tests. Furthermore, they showed a *moderate* correlation between the BDI-II scores and the SIMS AF score, and a *strong* correlation between the BDI-II subscale scores. These findings may reflect the tendency of feigners to over-endorse more symptoms and to describe their symptoms as more severe, relative to genuine patients (Walczyk et al., 2018).

Applying artificial intelligence techniques, we implemented a decision tree classification model, trained using a

tenfold cross-validation procedure, to distinguish Honest and Honest with Depressive Symptoms subjects from Simulators. This simple classifier had the advantage of being transparent (Quinlan, 1993)—a quality that is fundamental for supporting clinical and forensic decision-making (Adadi & Berrada, 2020). The decision tree, considering scores on both the BDI-II and the SIMS AF subscale, achieved an overall classification accuracy of 87.62%. For Simulators and Honest subjects, the classification accuracy of the decision tree model (82.44%) was equivalent to the accuracy obtained by the SIMS AF cut-off, as reported in the manual (82.93%). Furthermore, the SIMS AF cut-off outperformed the decision tree model in terms of sensitivity (decision tree sensitivity = 0.86; SIMS AF cut-off sensitivity = 0.94). However, consistent with previous research (van Impelen et al., 2014), the specificity of the SIMS AF cut-off was very low (0.43), and thereby a large number of Honest with Depressive Symptoms subjects were erroneously classified as Simulators.

In contrast, the decision tree model, which considered response patterns on both the SIMS AF and the BDI-II, achieved significantly higher specificity (0.70), thereby generating significantly fewer false positives. This suggests that the combination of traditional feigning detection instruments with clinical tests may result in incremental specificity, thereby limiting the risk of misclassifying respondents as feigners. Moreover, most feigners were identified by the decision tree model on the basis of their BDI-II total score, rather than the SIMS AF score. This suggests that the SIMS AF scale items may not meet their intended aim. Furthermore, a recent study (Fuermaier et al., 2023) used the BDI-II as an SVT with a cut-off > 38, achieving a sensitivity of 0.58 and a specificity of 0.90 in a sample of early retirement claimants undergoing forensic neuropsychological assessment. Of interest, in the present study, the ML model identified a lower feigning-related cut-off (> 29). This suggests possible differences related to the type of pathology simulated, meriting further investigation.

The analysis of the effectiveness of SIMS AF single items in discriminating between Simulators and Honest with Depressive Symptoms subjects provided further confirming results. No significant differences emerged between Honest with Depressive Symptoms subjects and Simulators for 5 (out of 15) items of the SIMS AF scale. Additionally, as one item was frequently endorsed by Honest with Depressive Symptoms subjects, SIMS AF items may portray symptoms of genuine depressive disorder, and not only symptoms that are improbable or implausible, as described in the SIMS manual (Widows & Smith, 2005). Furthermore, most (i.e., 7 out of 10) of the significant associations that emerged from the chi-squared tests had small to moderate effect sizes, indicating that the endorsement of these items was not strongly associated with Simulators.

Finally, only three items of the SIMS AF scale (out of the 10 inserted in the logistic regression) were significantly more likely to predict feigned depression. In detail, compared with Honest with Depressive Symptoms subjects, Simulators tended to describe the presence of a stable and pervasive depressive symptomatology that did not fluctuate, as well as maintained emotional reactivity and symptoms related to alexithymia. These endorsements likely derived from a naive knowledge of depression, which led to an over-endorsement of unrelated symptomatology. Moreover, when only these three SIMS AF items were input into the model, the accuracy in distinguishing Honest with Depressive Symptoms subjects from Simulators increased to 87.2%. In comparison, the accuracy obtained by applying the SIMS AF cut-off (according to the manual) to all items was 82.93%. Importantly, the specificity also improved (i.e., 0.86 for the three SIMS AF items vs. 0.43 for the traditional cut-off with all SIMS AF items). Again, this suggests that the current SIMS form must be optimized to reduce the number of false positives.

Overall, the results support the criticisms raised by some authors (e.g., Shura et al., 2022) in relation to the content validity of SIMS AF items. Although more research is needed to corroborate the findings (preferably within ecological settings), the results suggest that the content of SIMS AF items may require revision, or the worst performing items should be removed.

The results of the present study may have practical applications in clinical and forensic settings, as they provide medico-legal practitioners with advice and tools for diagnosing depression and identifying feigning behavior. First, the results suggest that the use of multiple instruments (instead of a single questionnaire) to determine feigning behavior may generate more reliable conclusions. Second, forensic practitioners using both the SIMS and the BDI-II may rely on the above-reported decision tree as a valuable tool to support their decision-making about the genuineness of the depressive symptomatology presented by patients. Finally, when the SIMS is used alone, practitioners should give more weight to the three items on the AF scale that emerged as more predictive of feigned depression—particularly in the context of doubtful cases.

Some limitations of the present study should be noted. First, the study was based on a simulation design, with participants instructed to feign depressive symptomatology (as reported in the experimental instructions) while avoiding detection. Thus, the experimental instruction included coaching elements, such as symptoms preparation and warnings. Previous research (Puente-López et al., 2022) has shown that such elements might affect participants' performance, resulting in a more cautious response pattern. Furthermore, the propensity to feign, triggered by the study instruction (rather than a real compensation setting), may have influenced participants' response accuracy, thereby limiting the ecological validity of the findings. Second, the three groups differed in size: the Honest with Depressive Symptoms group had only

44 participants, while the Honest group had 110 and the Simulators group included 161. When the groups in a dataset are unbalanced (i.e., when one is significantly larger or smaller), the groups do not follow a uniform distribution. Thus, there is a risk that models will fail to accurately distinguish between groups, by always predicting that units belong to the larger group. Third, we did not consider multiple forms of feigning (Lipman, 1962; Mazza et al., 2019b; Resnick et al., 2018), but referred only to the main construct. Finally, the results—based on norms for standardized Italian instruments—may not be transferable to other cultures. Therefore, the results of the study should be generalized with caution.

Building on these limitations, future research should investigate the detection and differentiation of various types of feigning (e.g., accentuation, simulation) through an analysis of participant response patterns, combined with multi-method assessment, in ecological settings. Future studies should also seek to verify the results of this study by administering the SIMS in its entirety, thus incorporating both AF scale items and the remaining 60 test items.

To conclude, computational models may provide effective support to forensic practitioners considering multiple elements within a complex decision-making process, such as that aimed at determining whether the symptomatology presented by an individual is genuine or feigned.

Appendix

Experimental Instructions

The Simulators group completed the tests with the instruction to feign depression, according to the DSM-5 criteria for major depressive disorder (American Psychiatric Association, 2013⁵). Specifically, the instructions were as follows:

Imagine being examined by an insurance policy board. The result of their assessment will determine whether you will receive compensation for psychological injury. You must make them believe that the damage caused you severe depression. Symptoms of depression are listed below. Please read carefully.

1. *Depressed mood most of the day, nearly every day.*
2. *Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day.*
3. *Significant weight loss when not dieting or weight gain or decrease or increase in appetite nearly every day.*

⁵ For accuracy, the Major Depressive Disorder criteria of the DSM-5-TR (American Psychiatric Association, 2021) are almost the same as those of the previous edition of the manual, with the exception of the fifth criterion, which now states: “*Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).*”

4. *Sleep disturbance (insomnia or hypersomnia) nearly every day.*

5. *A slowing down of thought and a reduction of physical movement (observable by others, not merely subjective feelings of restlessness or being slowed down).*

6. *Fatigue or loss of energy nearly every day.*

7. *Feelings of worthlessness or excessive or inappropriate guilt nearly every day.*

8. *Diminished ability to think or concentrate, or indecisiveness, nearly every day.*

9. *Recurrent thoughts of death, without a specific plan, or a suicide attempt or a specific plan for committing suicide.*

We want you to take these tests and deliberately feign depressive symptoms. Pay attention, because the questionnaire contains features designed to detect feigning, and you should aim to make your deception undetectable.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Eleonora Ricci, Marco Colasanti, Merylin Monaro, Cristina Mazza, and Alessandra Cardinale. The first draft of the manuscript was written by Eleonora Ricci, Marco Colasanti, Merylin Monaro, and Cristina Mazza. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by Università degli Studi G. D'Annunzio Chieti Pescara within the CRUI-CARE Agreement. The authors did not receive support from any organization for the submitted work.

Data Availability The datasets generated and/or analyzed for the present study are available from the corresponding author upon reasonable request.

Declarations

Ethics Approval Approval was obtained from the ethics committee of the Board of the Department of Human Neuroscience, Faculty of Medicine and Dentistry, Sapienza University of Rome. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to Participate Informed consent was obtained from all individual participants included in the study.

Consent to Publish Consent to publish was obtained from all individual participants included in the study.

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated

otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adadi, A., & Berrada, M. (2020). Explainable AI for healthcare: From black box to interpretable models. In V. Bhateja, S. C. Satapathy, & H. Satori (Eds.). *Embedded systems and artificial intelligence*, 1076, 327–337. Springer Singapore. https://doi.org/10.1007/978-981-15-0947-6_31
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, fifth edition* (5th ed.). American Psychiatric Association Publishing.
- American Psychiatric Association. (2021). *Diagnostic and statistical manual of mental disorders, fifth edition, text revision (DSM-5-TR™)* (5th ed., text revision). American Psychiatric Association Publishing.
- Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck Depression Inventory–II. *Psychological Assessment*. <https://psycnet.apa.org/doi/10.1037/t00742-000>
- Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, 7(3), 197–205. <https://doi.org/10.1007/s12207-014-9198-7>
- Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology*, 28(7), 633–639. <https://doi.org/10.1093/arclin/act038>
- Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration Policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, 29(6), 723–740. <https://doi.org/10.1080/13854046.2015.1099738>
- Cuijpers, P., & Smit, F. (2008). Subclinical depression: A clinically relevant condition? *Tijdschrift Voor Psychiatrie*, 50(8), 519–528.
- Dandachi-FitzGerald, B., Ponds, R. W. H. M., & 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>
- Druss, B. G., Rosenheck, R. A., & Sledge, W. H. (2000). Health and disability costs of depressive illness in a major U.S. corporation. *American Journal of Psychiatry*, 157(8), 1274–1278. <https://doi.org/10.1176/appi.ajp.157.8.1274>
- Frank, E., Hall, M., & Witten, I. (2016). *The WEKA workbench. Online Appendix for 'Data mining: Practical machine learning tools and techniques'*. Morgan Kaufmann Publishers.
- Fuermaier, A. B., Dandachi-Fitzgerald, B., & Lehrner, J. (2023). Validity assessment of early retirement claimants: Symptom overreporting on the Beck Depression Inventory–II. *Applied Neuropsychology: Adult*, 1–7. <https://doi.org/10.1080/23279095.2023.2206031>
- Ghisi, M., Flebus, G., Montano, A., Sanavio, E., & Sica, C. (2006). *Beck Depression Inventory-II. Italian manual* (4th ed.). OS Organizzazioni Speciali Giunti Editore.
- Giromini, L., Lettieri, S. C., Zizolfi, S., Zizolfi, D., Viglione, D. J., Brusadelli, E., Perfetti, B., di Carlo, D. A., & Zennaro, A. (2019). Beyond rare-symptoms endorsement: A clinical comparison simulation study using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) with the Inventory of Problems-29 (IOP-29).

- Psychological Injury and Law*, 12(3–4), 212–224. <https://doi.org/10.1007/s12207-019-09357-7>
- Giromini, L., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: Introducing the special Issue. *Psychological Injury and Law*, 15, 1–21. <https://doi.org/10.1007/s12207-022-09444-2>
- 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>
- IBM Corp. (2021). *IBM SPSS statistics for Windows, Version 28.0*. (Version 28). IBM Corp.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 2, 1137–1143.
- La Marca, S., Rigoni, D., Sartori, G., & Lo Priore, C. (2011). *Structured Inventory of Malingered Symptomatology (SIMS): Manual. Italian adaptation*. Giunti O.S. Organizzazioni Speciali.
- Lipman, F. (1962). Malingering in personal injury cases. *Temple Law Quarterly*, 35(2), 141–162.
- 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>
- Mazza, C., Monaro, M., Burla, F., Colasanti, M., Orrù, G., Ferracuti, S., & Roma, P. (2020). Use of mouse-tracking software to detect faking-good behavior on personality questionnaires: An explorative study. *Scientific Reports*, 10(1), 1–13. <https://doi.org/10.1038/s41598-020-61636-5>
- Mazza, C., Monaro, M., Orrù, G., Burla, F., Colasanti, M., Ferracuti, S., & Roma, P. (2019a). Introducing machine learning to detect personality faking-good in a male sample: A new model based on Minnesota Multiphasic Personality Inventory-2 restructured form scales and reaction times. *Frontiers in Psychiatry*, 10, 389. <https://doi.org/10.3389/fpsy.2019.00389>
- Mazza, C., Orrù, G., Burla, F., Monaro, M., Ferracuti, S., Colasanti, M., & Roma, P. (2019b). Indicators to distinguish symptom accentuators from symptom producers in individuals with a diagnosed adjustment disorder: A pilot study on inconsistency subtypes using SIMS and MMPI-2-RF. *PLOS One*, 14(12), e0227113. <https://doi.org/10.1371/journal.pone.0227113>
- Mazza, C., Ricci, E., Colasanti, M., Cardinale, A., Bosco, F., Biondi, S., Tambelli, R., Di Domenico, A., Verrocchio, M. C., & Roma, P. (2022). How has COVID-19 Affected mental health and lifestyle behaviors after 2 years? The third step of a longitudinal study of Italian citizens. *International Journal of Environmental Research and Public Health*, 20(1), 759. <https://doi.org/10.3390/ijerph20010759>
- McKinney, W. (2010). *Data structures for statistical computing in Python*. 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
- 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(8), 1094–1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Monaro, M., Mazza, C., Colasanti, M., Ferracuti, S., Orrù, G., di Domenico, A., Sartori, G., & Roma, P. (2021). Detecting faking-good response style in personality questionnaires with four choice alternatives. *Psychological Research Psychologische Forschung*, 85(8), 3094–3107. <https://doi.org/10.1007/s00426-020-01473-3>
- Monaro, M., Toncini, A., Ferracuti, S., Tessari, G., Vaccaro, M. G., De Fazio, P., Pigato, G., Meneghel, T., Scarpazza, C., & Sartori, G. (2018). The detection of malingering: A new tool to identify made-up depression. *Frontiers in Psychiatry*, 9, 249. <https://doi.org/10.3389/fpsy.2018.00249>
- Orrù, G., De Marchi, B., Sartori, G., Gemignani, A., Scarpazza, C., Monaro, M., Mazza, C., & Roma, P. (2022). Machine learning item selection for short scale construction: A proof-of-concept using the SIMS. *The Clinical Neuropsychologist*, 1–18. <https://doi.org/10.1080/13854046.2022.2114548>
- Orrù, G., Mazza, C., Monaro, M., Ferracuti, S., Sartori, G., & Roma, P. (2021). The development of a short version of the SIMS using machine learning to detect feigning in forensic assessment. *Psychological Injury and Law*, 14(1), 46–57. <https://doi.org/10.1007/s12207-020-09389-4>
- Puente López, E., Shura, R., Boskovic, I., Merten, T., Martínez Jarreta, B., & Pina, D. (2022). The impact of different forms of coaching on the Structured Inventory of Malingered Symptomatology (SIMS). *Psychothema*, 34(4), 528–536.
- Quinlan, J. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers.
- Resnick, P., West, S., & Wooley, C. (2018). The malingering of post-traumatic disorders. In *Clinical assessment of malingering and deception* (Rogers & Bender, pp. 188–211). The Guilford Press.
- Rogers, R., & Bender, S. D. (2018). *Clinical assessment of malingering and deception*. 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/aca019>
- 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>
- Smith, G., & Burger, G. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law Online*, 25(2), 183–189.
- Steer, R. A., Clark, D. A., Beck, A. T., & Ranieri, W. F. (1999). Common and specific dimensions of self-reported anxiety and depression: The BDI-II versus the BDI-IA. *Behaviour Research and Therapy*, 37(2), 183–190. [https://doi.org/10.1016/S0005-7967\(98\)00087-4](https://doi.org/10.1016/S0005-7967(98)00087-4)
- Sullivan, K., & King, J. (2010). Detecting faked psychopathology: A comparison of two tests to detect malingered psychopathology using a simulation design. *Psychiatry Research*, 176(1), 75–81. <https://doi.org/10.1016/j.psychres.2008.07.013>
- 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>
- von Glischinski, M., von Brachel, R., & Hirschfeld, G. (2019). How depressed is “depressed”? A systematic review and diagnostic meta-analysis of optimal cut points for the Beck Depression Inventory revised (BDI-II). *Quality of Life Research*, 28(5), 1111–1118. <https://doi.org/10.1007/s11136-018-2050-x>
- Walczyk, J. J., Sewell, N., & DiBenedetto, M. B. (2018). A review of approaches to detecting malingering in forensic contexts and promising cognitive load-inducing lie detection techniques. *Frontiers in Psychiatry*, 9, 700. <https://doi.org/10.3389/fpsy.2018.00700>
- Widows, M., & Smith, G. (2005). *SIMS: Structured Inventory of Malingered Symptomatology—Professional manual*. PAR Inc.
- World Health Organization. (2017). *Depression and other common mental disorders: Global health estimates*. World Health Organization.
- World Health Organization. (2018). *Health workforce: Fact sheet on Sustainable Development Goals (SDGs): Health targets* (No. WHO/EURO: 2018–2366–42121–58038; World Health Organization. Regional Office for Europe).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.