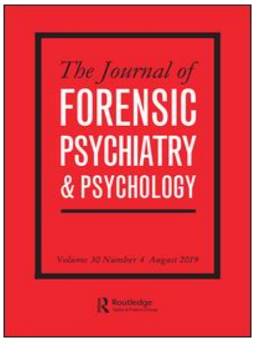


The Journal of Forensic Psychiatry & Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjfp20

# Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in motor vehicle accident patients

**Esteban Puente-López , David Pina , José Antonio Ruiz-Hernández & Bartolomé Llor-Esteban**

THE JOURNAL OF FORENSIC PSYCHIATRY & PSYCHOLOGY
https://doi.org/10.1080/14789949.2020.1833073

Routledge
Taylor & Francis Group

Check for updates

# Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in motor vehicle accident patients

Esteban Puente-López [a], David Pina [a], José Antonio Ruiz-Hernández [a,b] and Bartolomé Llor-Esteban [a,b]

[a]External Service of Forensic Sciencies and Techniques - Servicio Externo de Ciencias y Técnicas Forenses (SECyTEF), Murcia, Spain; [b]Murcia´s Institute of Biosanitary Research - Instituto Murciano de Investigación Biosanitaria (IMIB-Arrixaca), Murcia, Spain; [c]Facultad de Medicina, Universidad de Murcia, Murcia, Spain; [d] Departamento de Psicología Social y Psiquiatría, Universidad de Murcia, Murcia, Spain; [e]Departamento de Enfermería, Universidad de Murcia, Murcia, Spain

**ABSTRACT**

The SIMS is used in the medico-legal context to assess people who have suffered a traffic accident without proven scientific evidence to support this procedure. The objective of this research is to evaluate the SIMS' ability to discriminate instructed malingerers, general population, patients who have suffered a traffic accident, and overreporting patients. A simulation design was used to evaluate a total of 650 subjects divided into the four mentioned groups. Our results indicate that the cut-off scores proposed by the authors of the English and Spanish versions (14 and 16) produce moderate sensitivity and specificity values. With a cut-off score of 10, specificity decreases, but sensitivity significantly increases. These results suggest that the SIMS needs to be adapted to the context's particularities, either by using the proposed cut-off score and other instruments that compensate for the low specificity or by designing new subscales that include symptoms that are typical for the usual conditions in the context.

**ARTICLE HISTORY** Received 15 April 2020; Accepted 1 October 2020

**KEYWORDS** Malingering; traffic accidents; medico-legal context; Structured Inventory of Malingered Symptomatology; simulation design

## Introduction

Malingering has been defined as the intentional production of physical and/or psychological symptomatology motivated by obtaining an external gain, such as possible financial compensation, obtaining medication, lengthening a work leave, avoiding military duty, etc. (*Diagnostic and statistical manual of mental disorders 5th ed.*, American Psychiatric Association, 2013). This vision of

**CONTACT** David Pina david.pina@um.es

malingering has been widely criticized for being categorical and ignoring the evidence that malingering is a continuous and multidimensional construct (Rogers, 2018). On the contrary, it has been pointed out that malingering should not be focused from an all-or-nothing dichotomous proposition in which any exaggeration is considered an intentional production or a deliberate attempt at deception, but as one of the multiple response styles that an evaluee can adopt (Merckelbach et al., 2019; Rogers et al., 2014).

These response styles can be studied following one of the designs specifically proposed for malingering research. As Rogers (2018), the two most commonly used designs are simulation design and known group design. The first compares a relevant clinical group with an experimental group of asymptomatic participants, randomly assigned to honest response conditions (asked to respond honestly) and malingering (they are asked to feign the studied condition). The internal validity of this design is high, but external validity can be low due to the limitation that malingerers do not face real-world consequences. For this reason, some studies choose to use the second design, the known group design. In this design, the relevant clinical group is divided into subgroups (usually 'undifferentiated,' 'genuine patient,' 'possible malingerer,' 'probable malingerer,' and 'definite malingerer') following a rigorously validated highly accurate criterion, and are subsequently compared to each other. This design has lower internal validity than the previous one, but the external validity is higher because it has a group of genuine malingerers. However, the rigorous classification criterion required to develop this design means that it is not always possible to use it.

Alternatively, the other two context-specific designs are available, the partial-criteria design and the differential prevalence design. The former is similar to the known group design, except that an unknown precision classification criterion is used, typically, a set of scales or indicators. On the other hand, the latter uses as a method to classify the groups according to a specific characteristic, assuming that the response styles of each group will be different. For example, the use of financial compensation or state of litigation is common, assuming that litigating patients will malinger more than non-litigators.

Malingering is considered to be highly present in the forensic field, reaching a prevalence ranging from 20 to 40% (Greve et al., 2009; Young, 2015) and amounting to 59.7% in cases involving financial compensation (Chafetz & Underhill, 2013). Controlling for the effects of response styles on clinical patient evaluation is critical in the forensic context because of their possible legal and economic implications (Arce et al., 2015; Fariña et al., 2014; Rogers et al., 2014). It has therefore been pointed out that the forensic professional should consider a complete clinical assessment that contemplates all possible variables of interest in the condition or pathology assessed, and rules out the possibility of malingering (Osuna et al., 2015).

Although multi-method strategies are recommended for adequate differential diagnosis of malingering, the reality is that self-report-based evaluation predominates (González Ordi et al., 2010). These instruments are usually classified as symptom validity tests (SVT), which fall into two categories: performance validity tests (PVTs), tools designed to study the presentation of symptoms by evaluating underperformance in a number of neurocognitive tasks (Rogers, 2018); and self-reported symptom validity tests (SRSVTs), which evaluate whether the presentation of symptoms by the evaluee reflects their true experience of them (Larrabee, 2015). Within the latter category is one of the most commonly used standardized instruments for detecting malingering, the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997; Widows & Smith, 2005).

The SIMS is a self-report made up of 75 dichotomous questions designed to detect the feigning of psychiatric symptomatology and cognitive problems. It was based on the idea that malingerers tend to express a greater number of symptoms and/or claim to suffer from rare, atypical, or extreme symptoms (Smith & Burger, 1997; Van Impelen et al., 2014). The results of the instrument are reflected in an total score of malingering risk and five subscales that aim to indicate to the evaluator in which area malingering may be occurring: Low Intelligence (LI), Affective Disorders (AF), Neurologic Impairment (N), Psychosis (P), and Amnesia (AM). The original publication of Smith and Burger (1997) established a cut-off point of 14 for the total score and, subsequently, a cut-off point of 16 was proposed (Rogers et al., 1996). Various works have shown that the psychometric properties of the scale are adequate ($\alpha$ =.96; sensitivity = .93; specificity = .98), and it has proven to be effective in differentiating between honest evaluees and malingerers (Rogers, 2018; Van Impelen et al., 2014).

The SIMS is also used in the medical-legal context by professionals hired by insurance companies to assess people who have suffered a motor vehicle accident (hereafter MVA) (Cernovsky et al., 2019a, 2019b). This type of accident has as main consequences a trauma of the spine, cervical type, such as neck pain, whiplash injury, or Whiplash Associated Disorders (WAD), or lumbar type, such as lower back pain (Bartsch, et al. 2008). To a lesser extent, emotional symptomatology such as anxiety, depression, phobia, or post-traumatic stress (Elliott et al., 2009) may also be observed. Although the study Capilla Ramírez et al. (2014) applied SIMS as part of a multidisciplinary protocol to evaluate malingering in whiplash injury, its use in patients of MVA is problematic because was not prepared to be applied in that population and, to this day, there is no proven scientific evidence that validates it . As Cernovsky et al. (2019a, 2019b) indicate, in these conditions, there is a risk of misclassifying genuine patients as malingerers (false positives), generating a iatrogenic impact on them. Similarly, as the effectiveness of the instrument in that population is unknown, the opposite effect may occur, and it may not detect possible malingerers (false negatives), with the consequences that this entails.

It is essential for the instruments or techniques used in the legal context, especially in those cases where financial compensation is granted, to have empirical support and for their psychometric properties to be known, in particular the false positive ratio (Arce, 2017; Cernovsky et al., 2019a).

To respond to the above need, in this study, we used a simulation design to assess the discriminative capacity of the SIMS in four groups: general population (asymptomatic participants who were asked to respond honestly), instructed malingerers (asymptomatic participants who responded following specific malingering instructions), clinical patients (patients who had suffered an MVA), and overreporting patients (patients who have suffered an MVA and express symptomatology with an atypical or rare severity in the diagnosed condition).

## Method

### *Participants*

A total of 650 subjects participated in the study, with a mean age of 30.29 (*SD* = 11.27) and a range of 18 to 65 years. They were divided into 4 groups: general population (asymptomatic patients who responded honestly) of 200 participants, 84 men (42%) and 116 women (58%), with a mean age of 24.55 (*SD* = 7.81); instructed malingerers (asymptomatic participants instructed to feign), consisting of 201 participants, 95 men (47.26%) and 106 women (52.74%), with a mean age of 23.92 (*SD* = 7.51); clinical population (patients who have suffered an MVA), including 195 patients, 102 men (52.31%) and 93 women (47.69%), with a mean age of 37.45 (*SD* = 10.97); and 54 overreporting patients (patients who have suffered an MVA and show symptomatology of atypical severity), 25 men (46.3%) and 29 women (53.7%), with a mean age of 35.27 (*SD* = 11.21). Regarding the educational and economical level, in the group of clinical patients, 58.4% had university studies (level 4), 13.1% had high school or vocational training (level 3), 23.5% had secondary education (level 2), and the remaining 5% had no studies of any kind, whereas, in the group of overreporting patients, 55.3% had university studies (level 4), 15.7% had high school or vocational training (level 3), 20.6% had secondary education (level 2), and 8.4% had no studies. All participants of the malingering and population groups were recruited at the University of Murcia, so all of them had at least high school education (level 3). Of this group, 3% had a second college degree (level 4). On the other hand, the two groups of patients showed a similar economic level, predominantly medium level (between 15,000 and 20,000 euros per year) with 48.2% in the clinical patients and 51.7% in the overreporting patients, followed by the low level (<15,000 euros

**Table 1.** Sociodemographical variables of the groups.

| | Clinical patients | Overreporting patients | General population | Malingerers |
|---|---|---|---|---|
| Age[1] | 37.45 (10.97) | 35.27 (11.21) | 24.55 (7.81) | 23.92 (7.51) |
| Sex[2] | | | | |
| Men | 52.31 (102) | 46.3 (25) | 42 (84) | 47.26 (95) |
| Women | 47.69 (93) | 53.7 (29) | 58 (116) | 52.74 (106) |
| Educational level[2] | | | | |
| Level 4 | 58.4 (114) | 55.3 (30) | 2 (4) | 1 (2) |
| Level 3 | 13.1 (25) | 15.7 (9) | 100 | 100 |
| Level 2 | 23.5 (46) | 20.6 (11) | - | - |
| Level 1 | - | - | - | - |
| No studies | 5 (10) | 8.4 (4) | - | - |
| Economical level[2] | | | | |
| High | 23.6 (46) | 18.8 (10) | 8 (16) | 7 (14) |
| Medium | 48.2 (94) | 51.7 (28) | 11 (22) | 9 (18) |
| Low | 28.2 (55) | 29.5 (16) | 81 (162) | 84 (169) |

[1]Mean (SD); [2]% (n).

per year), in 28.2% of the clinical patients and 29.5% of the overreporting patients. The vast majority of participants in the malingering group and general population had a low level (84% and 81%, respectively), followed by the medium level (9% and 11%, respectively). All the sociodemographic variables can be found in Table 1.

The following inclusion criteria were used for each group: (1) *General population and malingerers*: Criterion a: Agreeing to sign an informed consent and Criterion b1: Passing a manipulation check. (2) *Clinical patients*: Criterion a: Agreeing to sign an informed consent and Criterion b2: The person was being evaluated to assess the physical and psychological harm caused after an MVA, and Criterion b3: Classified as a clinical patient. (3) *Overreporting patients*: Criterion a: Agreeing to sign an informed consent and Criterion b2: The person was being evaluated to assess the physical and psychological damage caused after an MVA, and Criterion b4: Classified as an overreporting patient. Failure to comply with the criteria in each group meant exclusion from the study. No patient was ruled out when applying the criteria described, but 24 patients attended to refused to participate after the initial explanation of the objectives of the study, and no data could be collected beyond sex (10 women and 14 men) and the reason for non-participation (16 of them claimed to be in a hurry and 8 did not want to undergo any type of assessment that was not imposed by the insurance companies).

## *Instruments*

### Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997*)*

Is a self-report screening instrument containing 75 true and false items, which evaluates the exaggeration or feigning of neurological and psychiatric symptoms. Its result indicates the overall risk of malingering. It presents five subscales (Psychosis, Neurological Impairment, Amnesic Disorders, Low Intelligence, and Affective Disorders) that provide information on the specific area in which malingering is detected. We used the Spanish version of the scale, adapted by González-Ordi and Santamaría (2009). Studies conducted for the adaptation in a total of 1,005 participants indicated that the inventory had a Cronbach alpha of .94 for the total test. In this study, a Cronbach alpha of .86 was obtained for the total sample. For this study, we used the cut-off points recommended by the authors of the two versions, 14 of the English version (Smith & Burger, 1997) and 16 of the Spanish version (González-Ordi & Santamaría, 2009).

### *Clinical interview*

The clinical evaluation routine of the physicians consisted of a semi-structured clinical interview that includes an exploration of the characteristics of the accident, a medical assessment of bodily harm, and a study of possible psychological damage. This last study was conducted only via the interview, without the application of psychometric instruments and, in case of severe psychological symptomatology, the patient was referred to another center for a complete psychological assessment. This evaluation routine was designed by the physicians of the centers, following the Spanish standars of damage assessment (see Villanueva Cañadas, 2018).

### *Visual Analogue Scale (VAS)*

For the evaluation of pain severity, medical experts used a Visual Analogue Scale (VAS; Aitken, 1969) with a range of scores from 1 to 10. The version used, an Spanish adaptation of the proposal by Zale et al. (2015), included, in addition to numbers, a range of colors (green – yellow – red) and a series of words (none, annoying, uncomfortable, dreadful, horrible, agonizing), both markers ordered according to severity.

### *Manipulation check*

To ensure that participants had properly understood and remembered their instructions, we used a manipulation check with five open-ended questions on which they were evaluated: *recall* (What were your instructions?), *comprehension* (Tell me in your own words what you were asked to do), *compliance with the instructions* (Describe how motivated you felt to do this research), *reported effort* (From 1 to 10, how hard have you tried?), *internal motivation* (Were you motivated by the challenge of 'beating' the test?), and *external incentives* (Did you take your role more seriously for fear of losing the reward?). These instructions were subsequently assessed by one of the researchers, giving them a score of 1 to 5 (*very bad, bad, regular, good* and *very good*) based on the adequacy of the response to a correction template provided to them (available for use by emailing the author). This manipulation check was applied to both groups, with the only difference that the one given to the malingerers also included two items asking what condition they had decided to feign and how they had prepared it ('What source and/or material have you used for this experiment and how did you use it to prepare your role? Please give details about the procedure'). These two questions were not evaluable, specifying this aspect after the questions ('You can answer with total sincerity, these two questions will not be used to assess your performance in the experiment').

## *Procedure*

A simulation design was used for this study. Initially, the other three specific designs mentioned in the introduction were assessed, discarding the known group and the partial criteria because the medical center where the study was conducted imposed a severe temporal limitation (no more than 15 minutes with each patient), and because of the differential prevalence due to their low external and internal validity and the simplistic and biased approach to their classification system (Rogers, 2018).

Participants assigned to the groups of clinical patients and overreporting patients were recruited randomly at a specialized multidisciplinary medical center in the Region of Murcia (Spain), during 2017, 2018, and 2019. These patients went to the clinic after suffering an MVA to assess the possible physical or psychological damage.

They were evaluated by three physicians, experts in the assessment of bodily harm and forensic medicine, with more than 25 years of experience. These professionals applied the SIMS after their clinical evaluation routine (see clinical interview in 'Instruments'), following a detailed explanation of the purpose and procedure of the study. All participants were asked to sign the informed consent.

To classify patients into the two designated groups (general population and malingerers) we used as a criterion the existence of contradictory evidence in the medico-legal evaluation. To assess this criterion, three legal medicine experts who participated in the study analyzed each case and assessed whether the patient's manifestation of their symptomatology was consistent with the anatomical-structural indicators observed or expected in the condition (pain severity, pain location, cervical movement range, active/passive joint balance, etc.). Once assessed, patients were listed as 1. Clinical patient, 2. Overreporting patient, 3. Probable malingerer, 4. Doubtful (the assessing physician doubted about which category to assign the patient to). Disagreements were resolved by consensus. The probable malingerer group was not included in this study because no cases were recorded.

For the experimental group, participants were recruited at the University of Murcia during 2018 and 2019. They registered by email after the presentation of the study and were randomly assigned to the proposed conditions: malingerers and general population. Subsequently, they were given instructions according to their role and an appointment was made for them to come to the university's outbuildings. These instructions were designed according to the requirements proposed by Rogers and Cruise (1998) of clarity, specificity, contextualization, and motivation.

The malingerer were asked to feign physical and psychological harm suffered after an alleged MVA, to obtain financial compensation. The instructions provided can be consulted in Appendix A. As motivation, they were offered an internal (being able to cheat on the test) and an external reward (an extra point on the final grade of a subject). Following the methodology of Rogers (2018), failure was penalized; only those who completed the scale according to the assigned role would receive the bonus. When the experiment was completed, the participants were subjected to the previously presented manipulation check. Those who did not achieve a score of 4 in all of the variables of that manipulation check were excluded from the experiment (Criterion b1). To avoid the problem of coaching (Van Impelen et al., 2014), no information or training was offered on how to malinger, only the importance of being consistent with what was expected from subjects who had suffered an MVA was emphasized.

Participants assigned to general population status were asked to respond to the SIMS scale with standard instructions (respond sincerely). The same external reward conditions and manipulation check as in the previous group (malingerers) were applied. The instructions that were used in the group can be consulted in Appendix B.

## *Statistical analysis*

The experts' degree of agreement or consistency of patient classification was analyzed with the Intraclass Correlation Coefficient (ICC, Model 2.1). Means, standard deviations and frequencies were computed for each one of the variables studied. We also analyzed differences in SIMS scores based on the sociodemographic and medico-legal variables using a one-factor ANOVA. To decide which statistical test should be used to analyze the difference of means between the groups, the normality of the distribution was confirmed with the Shapiro-Wilk test, and the variance homogeneity of the variables was determined with Levene´s test. As the distribution of the sample was normal, these differences were analyzed with a one-factor ANOVA. The magnitude of the effect was calculated using Cohen's *d* statistic. Because the original values proposed by Cohen (1988) are indicative and must be adapted to the studied context (Ato et al., 2013), we used the categorization that Rogers et al. (2003)

proposed specifically for the assessment of malingering (*moderate* ≥ .75; *large* ≥ 1.25; *very large* ≥1.50). Accuracy classification of the SIMS was estimated for the two cut-off scores recommended by the authors (14 and 16) using sensitivity, specificity, diagnostic odds ratio (DOR), Area Under the Curve (AUC) and Positive/Negative likelihood ratio (L+ and L-). The Positive/Negative Predictive Power (PPP and NPP) and the accuracy index for a prevalence of 10% to 50% were also calculated (Ranganathan & Aggarwal, 2018). Because the most relevant comparison in simulation designs is that made between the clinical patient group and the malingerers' group (Rogers, 2018), the ROC curve with state variable 'malingerers' will be presented to decide a possible optimal cut-off point for this context.

The analysis of the discriminative capacity described above was carried out only with the total SIMS score because, as the authors of the instrument warn, the subscales are not designed to detect malingering and serve only to guide the kind of psychopathology the evaluee may be feigning once the cut-off point has been exceeded (Smith & Burger, 1997). However, the means, standard deviations, and group differences of those subscales were calculated for descriptive purposes.

## Results

### *Classification and medico-legal variables*

The results of the degree of agreement analysis between the three evaluators showed an ICC of .96. Seven disagreements were found, all resolved by consensus. In the 7 cases, two of the experts considered that the evaluee was an overreporting patient, whereas the third considered it a dubious case.

With regard to medico-legal variables, all participants in the clinical patient and overreporting patient groups underwent a full clinical assessment after suffering an MVA. The assessment was carried out after an average of 38.24 days (*SD* = 9.25) since the accident, with a range of 26 to 59 days. All of the participants claimed to have cervical pain, with an average severity on Visual Analogue Scale (VAS) of 7.56 (*SD = 1.23)* in the overreporting patient group and of 5.15 (*SD* = .92) in the clinical patients. In terms of diagnosis by the professionals, 95.5% of the clinical patients were diagnosed with whiplash injury, the remaining 4.5% were also diagnosed with whiplash injury, but they also reported suffering from lower back pain. For overreporting patients, the diagnosis of whiplash injury was presented in 89% of the cases, and whiplash injury with lower back pain was made in 8% of the cases. The remaining 3% were diagnosed from public emergency services as 'cervicalgia postraumática' (post-traumatic neck pain), but the doctor who assessed them at the clinic confirmed that their clinical picture was compatible with whiplash injury. In none of the cases did professionals assign a grade to the whiplash, but none of the patients suffered neurological signs (Grade 3), except for 6 overreporting patients who claimed to have mild dizziness, 3 of them when moving their heads abruptly. Similarly, no fractures or serious musculoskeletal injuries (Grade 4) were observed.

Regarding the group of instructed malingerers, 78% of them indicated that they had decided to feign whiplash injury, 17% indicated 'neck pain' generically, 3% post-traumatic cervicalgia and 2% spine injury. The vast majority (64%) used non-scientific internet pages (e.g. from groups of lawyers or insurance companies) as a source of information, located through a search engine using terms such as 'traffic accident consequences', 'traffic accident compensation', 'common damage of traffic accidents'. Of the remaining 36%, 30% used a combined way of searching the internet for the main consequences and extending the information with specialized literature, and 6% said they had consulted only specialized literature. In all cases, the aforementioned specialized literature consisted of manuals available in the university library. The form of access was not specified.

## SIMS scores and group comparisons

As can be seen in Table 2, the one-factor ANOVA indicated that there were significant group differences for all subscales (Low intelligence, $F(2,646) = 166.27$, $p < .001$; Affective disorders, $F(2,646) = 132.91$, $p < .001$; Neurological impairment, $F(2,646) = 148.46$, $p < .001$; Psychosis, $F(2,646) = 2994.21$, $p < .001$; Amnesic Disorders, $F(2,646) = 723.89$, $p < .001$). In all four groups, the Affective Disorders subscale obtained the highest score, whereas the Psychosis Subscale obtained the lowest, recording values of 0 in the general population group, overreporting, and clinical patients.

For the total scale score, the mean score was 4.91 ($SD = 2.83$) in the general population group; 9.01 ($SD = 2.95$) in the clinical patient group; 12.87 ($SD = 4.11$) in overreporting patients; and 16.74 ($SD = 6.49$) in the malingerer group. The one-factor ANOVA yielded significant differences in the means of the four groups, $F(2, 646) = 253.09$, $p < .001$. The analysis showed that the malingerers obtained a higher score than the other three groups ($M = 16.74$, $SD = 6.49$), exceeding the two proposed cut-off points. The magnitudes of the effect obtained were higher between the malingerers/overreporting patients and general population ($d_1 = 2.36$ and $d_3 = 2.23$, respectively) than between the malingerers/overreporting patients and clinical patients ($d_2 = 1.53$ and $d_4 = .1.09$, respectively).

Differences in the total SIMS score were also sought based on the sociodemographic variables studied, using an ANOVA. No significant differences were found in any of them [Sex: $F(2,646) = .62$, $p = .42$; educational level: $F$

**Table 2.** One-factor ANOVA for mean contrast of general, overreporting, clinical, and malingering populations.

| | Gen. | | Pat. | | Over. | | Mal. | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* | F (2, 646) | $d_1$ [95% CI] | $d_2$ [95% CI] | $d_3$ [95% CI] | $d_4$ [95% CI] | $d_5$ [95% CI] | $d_6$ [95% CI] |
| SIMS total score | 4.91 | 2.83 | 9.01 | 2.95 | 12.87 | 4.11 | 16.74 | 6.49 | 253.09** | 2.36 [2.10,2.61] | 1.53 [1.31, 1.75] | 2.26 [2.00, 2.50] | 1.09 [0.88, 1.30] | 1.42 [1.20, 1.64] | .64 [.33, .94] |
| Low intelligence* | .94 | .34 | 2.35 | 1.04 | 3.22 | .81 | 3.01 | 1.40 | 166.27** | 2.03 [1.78,2.27] | .53 [.33, .73] | 3.67 [3.35, 3.99] | .93 [.73, 1.14] | 1.83 [1.59, 2.06] | .16 [.14, .46] |
| Affective disorders* | 2.34 | 1.23 | 4.15 | 2.11 | 6.15 | 1.95 | 7.41 | 3.94 | 132.91** | 1.74 [1.50,1.96] | 1.03 [.82, 1.24] | 2.34 [2.08, 2.59] | .98 [.77, 1.19] | 1.05 [0.84, 1.26] | .35 [.05, .65] |
| Neurological impairment* | 1.17 | .72 | 2.51 | 1.21 | 3.50 | .82 | 2.94 | .98 | 148.46** | 2.06 [1.81, 2.30] | .39 [.19, .59] | 3.02 [2.73, 3.30] | .96 [.75, 1.16] | 1.35 [1.13, 1.57] | .59 [.28, .89] |
| Psychosis* | .00 | .00 | .00 | .00 | .00 | .00 | .94 | .41 | 2994.21** | 3.24 [2.93, 3.53] | 3.24 [2.93, 3.53] | n.a. | n.a. | n.a. | 2.58 [2.20, 2.94] |
| Amnesic disorders* | .45 | .17 | .00 | .00 | .00 | .00 | 2.44 | 1.02 | 723.89** | 2.72 [2.44, 2.98] | 3.38 [3.07, 3.68] | 3.74 [3.41, 4.05] | n.a. | 3.72 [3.39, 4.04] | 2.69 [2.30, 3.06] |

Abbreviations: Gen: General population; Pat: Clinical patients; Over: Overreporting patients; Mal: Malingerers; * SIMS Subscales $d_1$: Cohen´s *d* effect size for malingerer vs. general population; $d_2$: Cohen´s *d* effect size for malingerer vs. clinical patient; $d_3$: Cohen´s *d* effect size for overreporting patient vs. general population; $d_4$: Cohen´s *d* effect size for overreporting patient vs. clinical patient.; $d_5$: Cohen´s *d* effect size for general patient vs. clinical patient; $d_6$: Cohen´s *d* effect size for overreporting patient vs. malingerers; 95% CI: 95% confidence interval. ** $p < .001$; n.a.: non applicable.

(2,646) = .77, $p$ = .37; and economical level: $F$(2,646) = .51, $p$ = .60;]. The possibility that the source of information collection (only internet, internet and bibliography, only bibliography) of the malingering group could influence the scores of the instrument used was also assessed, but no significant differences were found [$F$(2,197) = .20, $p$ = .81].

## Classification accuracy

The accuracy of the SIMS' classification can be seen in Table 3. Using the cut-off point of 14, the SIMS discriminated between malingerers and the general population with a sensitivity of 59.70% and a specificity of 97.50% and between malingerers and clinical patients with a sensitivity of 59.70% and a specificity of 78.97%. On the other hand, it discriminated between overreporting patients and general population with a sensitivity of 55% and a specificity of 97.50% and between overreporting patients and clinical patients with a sensitivity of 55% and a specificity of 78.97%. Concerning diagnostic accuracy, rated with the DORs, higher accuracy was observed between malingerers/overreporting patients and general population (DOR = 57.77 and 48.75, respectively) than between malingerers/overreporting patients and clinical population (DOR = 5.56 and 4.69, respectively). Similarly, the probability of obtaining a higher SIMS score was greater in malingerers/overreporting patients than in clinical patients (AUC = .94 and .91, respectively) and general population (AUC = .76 and .73, respectively). Regarding the positive likelihood ratio (L+), it was observed that scores of ≥14 in the SIMS were approximately 22 times more likely in the first two groups than in general population and approximately 3 times more likely in the first two groups than in clinical patients.

The PPP, NPP, and accuracy index for a prevalence of 10 to 50% can be seen in Table 4. Overall, assuming a base ratio of 30%, the most likely to be found in clinical practice (Curtis et al., 2019), the results indicated that with a cut-off point of 14, there is a 91% probability of classifying malingerers and overreporting patients when compared to the general population group, and approximately 54% when compared to the group of clinical patients.

Using 16 as a cut-off point (Table 5), sensitivity decreased, both in the comparison between malingerers/overreporting patients and general population, as well as between malingerers/overreporting patients and clinical patients (55.22% and 51.85, respectively for the two cases), but specificity increased (98.50% and 83.08%, respectively). Diagnostic accuracy and L+ also increased, remaining higher between malingerers/overreporting patients and general population (DOR = 80.98 and 70.71; L+ = 39.82 and 34.57, respectively) than between malingerers/overreporting patients

**Table 3.** Classification accuracy with a cut-off score of 14.

| Population | AUC [95%CI] | SEAUC | SEN [95%CI] | SPEC [95%CI] | DOR [95%CI] | L+ [95%CI] | L- [95%CI] |
|---|---|---|---|---|---|---|---|
| Gen/Mal | .94 [.91, .97] | .00 | 59.70 [52.57, 66.54] | 97.50 [94.26, 99.18] | 57.77 [22.76, 146.63] | 23.88 [9.98, 57.17] | .41 [.35, .49] |
| Pat/Mal | .76 [.72, .80] | .02 | 59.70 [52.57, 66.54] | 78.97 [72.58, 84.47] | 5.56 [3.56, 8.68] | 2.84 [2.11, 3.81] | .51 [.42, 61] |
| Gen/Over | .91 [.93, .98] | .01 | 55.00 [47.82, 62.02] | 97.50 [94.26, 99.18] | 48.75 [17.27, 137.56] | 21.78 [8.88, 53.43] | .46 [.34, .62] |
| Pat/Over | .73 [.70, .79] | .03 | 55.00 [47.82, 62.02] | 78.97 [72.58, 84.47] | 4.69 [2.48, 8.88] | 2.64 [1.84, 3.79] | .56 [.41, .76] |

Abbreviations: Gen: General population; Pat: Clinical patients; Over: Overreporting patients; Mal: Malingerers; AUC: Area Under the Curve; 95% CI: 95% confidence intervals; SEAUC: Standard Error of Area Under the Curve; Se: Sensitivity; Sp: Specificity; DOR: Diagnostic Odds Ratio; L+: Positive Likelihood Ratio; L-: Negative Likelihood Ratio.

**Table 4.** Positive/negative predictive power and accuracy index for a prevalence of 10%-50% (Cutoff score 14).

| Population | 10% PPP-NPP (%) | 10% ACC (%) | 20% PPP-NPP (%) | 20% ACC (%) | 30% [95%CI] PPP-NPP (%) | 30% ACC (%) | 40% PPP-NPP (%) | 40% ACC (%) | 50% PPP-NPP (%) | 50% ACC (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Gen/Mal | 72.63–95.61 | 93.72 | 85.65–90.63 | 89.94 | 91.10–84.95 | 86.16 | 94.09–78.40 | 82.38 | 95.98–70.76 | 78.60 |
| Pat/Mal | 23.98–94.63 | 77.05 | 41.52–88.69 | 75.12 | 54.89–82.06 | 73.19 | 64.43–74.62 | 71.27 | 73.95–66.21 | 69.34 |
| Gen/Over | 70.76–95.18 | 93.26 | 84.48–89.77 | 89.07 | 90.32–83.65 | 84.88 | 93.56–76.68 | 80.69 | 95.61–68.68 | 76.50 |
| Pat/Over | 22.70–94.11 | 76.63 | 39.78–87.67 | 74.29 | 53.10–80.57 | 71.95 | 63.79–72.72 | 69.61 | 72.54–63.99 | 67.26 |

Abbreviations: Gen: General population; Pat: Clinical patients; Over: Overreporting patients; Mal: Malingerers; PPP: Positive Predictive Power; NPP: Negative Predictive Power; ACC: Accuracy Index.

**Table 5.** Classification accuracy with a cut-off score of 16.

| Population | AUC [95%CI] | SEAUC | SEN [95%CI] | SPEC [95%CI] | DOR [95%CI] | L+ [95%CI] | L- [95%CI] |
|---|---|---|---|---|---|---|---|
| Gen/Mal | .94 [.91, .97] | .00 | 55.22 [48.07, 62.22] | 98.50 [95.68, 99.69] | 80.98 [25.04, 261.88] | 38.82 [11.89, 113.97] | .45 [.39, .53] |
| Pat/Mal | .76[.72, .80] | .02 | 55.22 [48.07, 62.22] | 83.08 [77.06, 88.06] | 6.05 [3.79, 9.65] | 3.26 [2.33, 4.56] | .54 [.46, .64] |
| Gen/Over | .91[.93, .98] | .01 | 51.85 [37.84, 65.66] | 98.50 [95.68, 99.69] | 70.71 [20.08, 249.05] | 34.57 [10.92, 109.40] | .49 [.37, .65] |
| Pat/Over | .73[.70, .79] | .03 | 51.85 [37.84, 65.66] | 83.08 [77.06, 88.06] | 5.28 [2.75, 10.14] | 3.06 [2.05, 4.59] | .58 [.44, .77] |

Abbreviations: Gen: General population; Pat: Clinical patients; Over: Overreporting patients; Mal: Malingerers; AUC: Area under the Curve; 95% CI: 95% confidence intervals; SEAUC: Standard Error of Area under the Curve; Se: Sensitivity; Sp: Specificity; DOR: Diagnostic Odds Ratio; L+: Positive Likelihood Ratio; L-: Negative Likelihood Ratio.

and clinical population (DOR = 6.05 and 5.28; L + 3.26 and 3.06, respectively).

Lastly, the PPP, NPP, and accuracy index presented in Table 6 indicated that using a cut-off point of 16 increased the probability of classifying malingerers and overreporting patients to approximately 94% when compared to the general population group, and to approximately 58% when compared to the group of clinical patients.

### *Optimal cut-off point to discriminate between clinical patients and malingerers*

In order to obtain a possible optimal cut-off point for comparison of malingerers and clinical patients, Figure 1 presents the ROC curve using 'malingerers' as a state variable. Sensitivity decreased significantly from cut-off point 8 (Sen: 90%; Spec: 30%), achieving an intermediate specificity at cut-off point 10 (Sen: 82%; Spec: 48%). With values >16, sensitivity is below 50%, although specificity increased to 80% (Cutoff point 17, Sen: 42%; Spec: 80%).

## Discussion

The results indicate that the instructed malingerers obtain higher scores on the SIMS than the other three groups. Generally speaking, the SIMS can better discriminate between malingerers/overreporting patients and the general population than between malingerers/overreporting patients and clinical patients. Similarly, the instrument discriminates malingerers better than it does overreporting patients.

Concerning the SIMS results, the Affective Disorders subscale recorded the highest score in the groups of overreporting and clinical patients. These results are consistent with the criticisms made by other authors of this subscale because it includes genuine depressive symptomatology, which is common after an MVA (Sarrami et al., 2016; Van Impelen et al., 2014). These findings contradict the hypothesis of Cernovsky et al. (2019b) that at least 30 items of the instrument would be confusing for clinical patients because they are consistent with the symptomatology commonly observed in patients who have suffered an MVA. However, we agree with these authors' assertion that several of these items are not adequately formulated to be applied in the context of the study.

Concerning the study's objective (assessing the SIMS' ability to discriminate between instructed malingerers, general population, clinical patients, and overreporting patients), applying the proposed cut-off points and taking into account the need to reduce as much as possible the false positive rate, better results are obtained with the cut-off point of 16 than with 14. In

**Table 6.** Positive/negative predictive power and accuracy index for a prevalence of 10%-50% (Cutoff score 16).

| Population | 10% PPP-NPP (%) | 10% ACC (%) | 20% PPP-NPP (%) | 20% ACC (%) | 30% [95%CI] PPP-NPP (%) | 30% ACC (%) | 40% PPP-NPP (%) | 40% ACC (%) | 50% PPP-NPP (%) | 50% ACC (%) |
|---|---|---|---|---|---|---|---|---|---|---|
| Gen/Mal | 80.36–85.19 | 94.17 | 90.20–89.80 | 89.84 | 94.04–83.69 | 85.52 | 96.09–76.74 | 81.19 | 97.36–68.75 | 76.86 |
| Pat/Mal | 26.61–94.35 | 80.29 | 44.93–88.13 | 77.51 | 58.31–81.24 | 74.72 | 68.51–73.57 | 71.94 | 76.54–64.98 | 69.15 |
| Gen/Over | 79.34–94.85 | 93.84 | 89.63–89.11 | 89.17 | 93.68–82.68 | 84.51 | 95.84–75.42 | 79.84 | 97.19–61.17 | 75.18 |
| Pat/Over | 25.40–93.95 | 79.95 | 43.37–87.34 | 76.83 | 57.77–80.10 | 73.71 | 67.13–72.13 | 70.59 | 75.39–63.31 | 67.46 |

Abbreviations: Gen: General population; Pat: Clinical patients; Over: Overreporting patients; Mal: Malingerers; PPP: Positive Predictive Power; NPP: Negative Predictive Power; ACC: Accuracy Index.

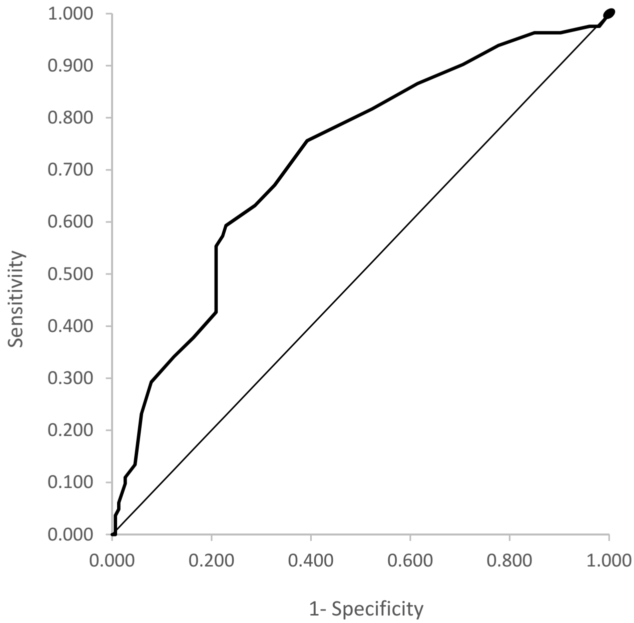


**Figure 1.** ROC curve for malingerers and clinical patients.

general, in the analysis between malingerers/overreporting patients and clinical patients, the scale shows a relatively high specificity and an intermediate sensitivity. These results indicate that the instrument will misclassify approximately half of the malingerers (false negatives) and will have a false positive rate of between 17% and 21%, or one in five cases. On the other hand, in the comparison of overreporting patients with clinical patients, a slight decrease in sensitivity compared to the previous analysis can be observed, while maintaining the false positive rate between 17% and 21%.

These suggest that SIMS cannot accurately discriminate between groups with the cut-off points proposed by Smith and Burger (1997) and González-Ordi and Santamaría (2009) (hereafter, we will refer to them as the original cut-off points), so there may be another cut-off point that matches the characteristics of a population for which the instrument was not originally designed. In this sense, in the manual of the Spanish version of the SIMS, the authors present the study of Capilla Ramírez et al. (2008) where they apply SIMS to patients with chronic pain from the medico-legal context. The aforementioned study suggests using a stricter cut-off point for this type of patient (scores > 6), which allows discriminating with 88% sensitivity and 47% specificity. Taking into account this decision criterion, for our sample, the optimal cut-off point would be 10, as values similar to those of the previous authors were reached, with a sensitivity of 82% and a specificity of 48%.

Despite this, we believe that, to perform a malingering screening with SIMS in populations such as those appraised in our study and in that of Capilla Ramírez et al. (2008), it would be necessary to have additional instruments that compensate for this limitation.

We consider of interest the fact that our findings, applying the original cut-off points, are contrary to those of the meta-analysis of Van Impelen et al. (2014), where it is observed that the sensitivity of the instrument is quite high and the specificity is low, but they are similar to those of Capilla Ramírez et al. (2014), where high specificity and moderate-low sensitivity are presented. Similarly, our total SIMS scores in the malingering group are similar to those of the study of the last authors, while differing significantly from those reported by Van Impelen et al. (2014).

We believe that a possible explanation for these differences is that the aforementioned meta-analysis included studies that evaluate malingering in psychopathology – with some caveats such as the publications of González Ordi and Santamaría Fernández (2008) and González Ordi et al. (2010) who evaluated patients with fibromyalgia and chronic pain without specified cause – whereas, in our study and in that of Capilla Ramírez et al. (2014), malingering in cervical pathology is evaluated (whiplash injury). These differences in the conditions may affect the performance both of the real malingerers and the instructed malingerers, as they would respond to the instrument adapting to the condition to be feigned.

SIMS was built following the overendorsement theory, according to which it is proposed that alleged malingerers will indiscriminately display symptoms of any kind, regardless of the context (Monaro et al., 2018; Rogers, 2018), but, as Merckelbach et al. (2009) state, the influence of such a context should be taken into account, as it modulates the form and intensity with which symptoms are feigned. In this case, real malingerers may only feign expected symptoms and discard those they consider uncommon. A subject with minimal harm after a traffic accident who wants to malinger a whiplash injury will focus on symptoms that are consistent with the condition and will consider that feigning striking and atypical symptomatology, such as psychotic symptoms, may detract from their credibility.

Likewise, instructed malingerers are usually asked to recreate the assigned pathology as realistically and believably as possible. If an analogue participant is asked to pretend to suffer from severe psychopathology, they will find a greater number of items in the SIMS that they will consider of interest to recreate their role, because the instrument was specifically designed to evaluate such symptomatology. On the other hand, a participant assigned to feign a cervical pathology will find atypical symptoms that they will not consider due to the possibility of diminishing their credibility. The results observed in the subscales of our study and in that of Capilla Ramírez et al. (2014) are consistent with this hypothesis, as only the subscale of affective disorders, which contains items consistent with some of the symptoms that could manifest in a condition such as WAD, obtained relatively high values.

We believe that, while the psychometric properties of the tool to detect feigned mental disorders are excellent (Rogers, 2018; Van Impelen et al., 2014), for future research, it would be advisable to try to adapt the instrument to the particularities of contexts for which it was not originally designed. In this sense, additional subscales could be included that evaluate common symptomatology in this context, such as pain. Not only could items relating to pain severity be incorporated, but psychosocial factors related to the experience of pain could be included. The evidence indicates that pain follows a multidimensional model in which many variables interact, such as behavior in the face of the illness, pain catastrophism, or the perception of disability, which may be of interest in detecting malingering (Aguilera et al., 2019; Casey et al., 2015; Falla et al., 2016; Holm et al., 2008).

Despite the above, and regardless of whether the results indicate that the instrument has higher or lower discriminatory capacity, it is essential to note that malingering is a possible explanation for the exaggeration of symptoms, but it is not the only one (Merckelbach et al., 2019). The professional should focus the evaluation considering all possible hypotheses for such exaggeration and apply an evaluation protocol that is capable of analyzing them. In this sense, the SIMS should not be used as the sole evaluation tool because it 'may help to define an end point (i.e. symptom exaggeration), but it is silent as to the external and intrinsic factors contributing to this end point' (Merckelbach et al., 2019, p. 322).

The results we present have some limitations. First, a simulation design was used with healthy subjects who followed malingering instructions. As Rogers & Bender (2018) explains, these subjects do not offer the same experience as real malingerers. While the internal validity of the design is relatively high, the external validity, and the generalization of the results, may be limited. Secondly, the results obtained are limited to the assessment of traffic accidents, so the results could not be generalized to another context. Third, it was not possible to apply a known-group design because the clinic where the study was conducted imposed a strict temporal limitation for each case. Fourth and last, the group of malingerers and general population are composed only of students of the University of Murcia, which can affect their representativeness.

## Data availability statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

## Disclosure statement

The Author(s) declare(s) that there is no conflict of interest.

## Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

## ORCID

Esteban Puente-López http://orcid.org/0000-0001-6367-457X
David Pina http://orcid.org/0000-0001-5944-4683

## References

Aguilera, M., Paz, C., Compañ, V., Medina, J. C., & Feixas, G. (2019). Cognitive rigidity in patients with depression and fibromyalgia. *International Journal of Clinical and Health Psychology*, *19*(2), 160–164. https://doi.org/10.1016/j.ijchp.2019.02.002

Aitken, R. C. B. (1969). A growing edge of measurement of feelings [Abridged]. *Proceedings of the Royal Society of Medicine*, *62*(10), 989–993. https://doi.org/10.1177/003591576906201005

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (*5th* ed.). Author.

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba forense [Content analysis of the witness statements: Evaluation of the scientific and judicial validity of the hypothesis and the forensic proof]. *Acción Psicológica*, *14*(2), 171–190. https://doi.org/10.5944/ap.14.1.21347

Arce, R., Fariña, F., & Vilariño, M. (2015). Daño psicológico en casos de víctimas de violencia de género: Estudio comparativo de las evaluaciones forenses [Psychological harm in cases of victims of gender-based violence: Comparative study of forensic assessments]. *Revista Iberoamericana de Psicología y Salud*, *6*(2), 72–80. https://doi.org/10.1016/j.rips.2015.04.002

Ato, M., López, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología [A classification system of research designs in psychology]. *Anales de Psicología*, *29*(3), 1038–1059. https://doi.org/10.6018/analesps.29.3.178511

Bartsch, A. J., Gilbertson, L. G., Prakash, V., Morr, D. R., & Wiechel, J. F. (2008). Minor crashes and 'whiplash' in the United States. Annals of advances in automotive medicine. Association for the Advancement of Automotive Medicine. Annual Scientific Conference, 52, 117–128

Capilla Ramírez, P., González Ordi, H., Santamaría Fernández, P., & Casado Morales, M. I. (2014). Detección de exageración de síntomas en esguince cervical: Pacientes clínicos versus sujetos análogos [Detection of symptom exaggeration in whiplash: Patients versus analogue participants]. *Trauma Fundación MAPFRE*, *24*(1), 4–10.

Capilla Ramírez, P., González Ordi, H., & Santamaría Fernández, Y. (2008). Detección de simulación en dolor crónico. *Comunicación presentada en el VIII Congreso Nacional* *de la Sociedad Española de Traumatología Laboral – SETLA*, 27-29 de Noviembre de 2008.

Casey, P. P., Feyer, A. M., & Cameron, I. D. (2015). Course of recovery for whiplash associated disorders in a compensation setting. *Injury*, *46*(11), 2118–2129. https://doi.org/10.1016/j.injury.2015.08.038

Cernovsky, Z., Ferrari, J., & Mendonca, J. (2019a). Deceptive clinical diagnosing of malingering via structured inventory of malingered symptomatology. *Archives of Psychiatry and Behavioral Sciences*, *2*(1), 44–49.

Cernovsky, Z., Ferrari, J., & Mendonca, J. (2019b). Pseudodiagnoses of malingering of neuropsychological symptoms in survivors of car accidents by the Structured Inventory of Malingered Symptomatology. *Archives of Psychiatry and Behavioral Sciences*, *2*(1), 55–65.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (*2nd* ed.). Erlbaum.

Curtis, K. L., Aguerrevere, L. E., Bianchini, K. J., Greve, K. W., & Nicks, R. C. (2019). Detecting malingered pain-related disability with the pain catastrophizing scale: A criterion groups validation study. *The Clinical Neuropsychologist*, *33*(8), 1485–1500. https://doi.org/10.1080/13854046.2019.1575470

Elliott, J. M., Noteboom, J. T., Flynn, T. W., & Sterling, M. (2009). Characterization of Acute and Chronic Whiplash-Associated Disorders. Journal of Orthopaedic & Sports Physical Therapy, 39(5), 312–323. https://www.doi.org/10.2519/jospt.2009.2826 doi:10.2519/jospt.2009.2826

Falla, D., Peolsson, A., Peterson, G., Ludvigsson, M. L., Soldini, E., Schneebeli, A., & Barbero, M. (2016). Perceived pain extent is associated with disability, depression and self-efficacy in individuals with whiplash-associated disorders. *European Journal of Pain*, *20*(9), 1490–1501. https://doi.org/10.1002/ejp.873

Fariña, F., Arce, R., Vilariño, M., & Novo, M. (2014). Assessment of the standard forensic procedure for the evaluation of psychological injury in intimate-partner violence. *Spanish Journal of Psychology*, *17*(e32), 1–10. https://doi.org/10.1017/sjp.2014.30

González Ordi, H., & Santamaría Fernández, P. (2008). Detection of malingering in clinical, medicolegal, and forensic settings. In H. González Ordi & P. Santamaría Fernández (Eds.), *Inventario Estructorado de Simulación de Síntomas – SIMS Manual* [Structured Inventory of Malingered Symptomatology – SIMS Manual] (pp. 60–66). Tea Ediciones.

González Ordi, H., Santamaría Fernández, P., & Fernández Marín, P. (2010). Precisión predictiva del Inventario Estructurado de Simulación de Síntomas –SIMS- en el contexto medicolegal [Predictive accuracy of the Structured Inventory of Malingered Symptomatology -SIMS- in the medical context]. *eduPsykhé*, *9*(1), 3–22.

González-Ordi, H., & Santamaría, P. (2009). *Adaptación española del Inventario Estructurado de Simulación de Síntomas* [Spanish adaptation of the Structured Inventory of Malingered Symptomatology]. TEA Ediciones.

Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation*, *90*(7), 1117–1126. https://doi.org/10.1016/j.apmr.2009.01.018

Holm, L. W., Caroll, L. J., Cassidy, D., Skillgate, E., & Ahlbom, A. (2008). Expectations for recovery important in the prognosis of whiplash injuries. *PLoS Medicine*, *5*(5), e105. https://doi.org/10.1371/journal.pmed.0050105

Larrabee, G. J. (2015). The multiple validities of neuropsychological assessment. *American Psychologist*, *70*(8), 779–788. https://doi.org/10.1037/a0039835

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., Smeets, T., & Jelicic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *The Journal of Forensic Psychiatry & Psychology*, *20*(3), 378–386. https://doi.org/10.1080/14789940802456686

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/fpsyt.2018.00249

Osuna, E., López-Martínez, M., Arce, R., & Vázquez, M. J. (2015). Analysis of response patterns on the MMPI-2 in psychiatric prison inmates. *International Journal of Clinical and Health Psychology*, *15*(1), 29–36. https://doi.org/10.1016/j.ijchp.2014

Ranganathan, P., & Aggarwal, R. (2018). Study designs: Part 1 – An overview and classification. *Perspectives in Clinical Research*, *9*(4), 184–186. https://doi.org/10.4103/picr.PICR_124_18

Rogers, R. & Bender, S.D. (Eds.). (2018). *Clinical assessment of malingering and deception* (4th ed.). Guilford Press.

Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, *22*(3), 273–285. https://doi.org/10.1023/A:1025702405865

Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS - PB - Routledge. *Journal of Personality Assessment*, *67*(2), 244–257. https://doi.org/10.1207/s15327752jpa6702_2

Rogers, R., Robinson, E. V., & Gilliard, 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

Rogers, R., Sewell, K. W., Martin, M. A., & Vitacco, M. J. (2003). Detection of feigned mental disorders: A meta-analysis of the MMPI-2 and malingering. *Assessment*, *10*(2), 160–177. https://doi.org/10.1177/107319110301

Sarrami, P., Armstrong, E., Naylor, J. M., & Harris, I. A. (2016). Factors predicting outcome in whiplash injury: A systematic meta-review of prognostic factors. *Journal of Orthopaedics and Traumatology*, *18*(1), 9–16. https://doi.org/10.1007/s10195-016-0431-x

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.

Van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *Neuropsychology, Development and Cognition. Section D: The Clinical Neuropsychologist*, *28*(8), 1336–1365. https://doi.org/10.1080/13854046.2014.984763

Villanueva Cañadas, E. (2018). *Gisbert Calabuig. Medicina legal y toxicológica* [Gisbert Calabuig. Legal and toxicological medicine]. Elsevier.

Widows, M. R., & Smith, G. P. (2005). *SIMS: Structured inventory of malingered symptomatology. Professional manual*. Psychological Assessment Resources.

Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15 ± 15 %. *Psychological Injury and Law*, *8*(3), 188–199. https://doi.org/10.1007/s12207-015-9232-4

Zale, K. E., Klatt, M., Volz, K. R., Kanner, C., & Evans, K. D. (2015). A mixed-method approach to evaluating the association between myofascial trigger points and ankle/foot pain using handheld sonography equipment. *Journal of Diagnostic Medical Sonography*, *31*(4), 210–220. https://doi.org/10.1177/8756479315586396

## Appendix A. Instructions for malingerers

"This study is about malingering in cases of road accidents. We ask you to take the role of a **malingerer** and we want you to imagine that you have received an impact on your car while driving. Even if you're feeling well and it doesn't hurt too much (or at all), you've heard that compensation for physical or psychological harm is very generous and you want to do everything you can to receive it. You're going to go to a medical center so an evaluating physician will evaluate you and report your alleged injuries to the insurance company. This doctor will apply an instrument and you will have to fill it in by exaggerating or inventing the symptoms of a condition of your choice.

**It is essential that the condition you select is consistent with what is expected after a traffic accident**. For example, severe psychopathology such as schizophrenia or bipolar disorder are not consequences of this type of accident. You must prepare your condition before the seminar you have been assigned to, we leave it up to you the method you follow, you can look online, in manuals, etc. We ask you to try to be completely believable, you have to try to get into the role that we have assigned you as much as possible and complete the test as if you really suffered from the condition you have prepared.

The instrument that we will apply to you can detect whether or not you interpret the role that we have assigned you well and, if it 'catches' you, you will not receive the bonus of the subject.

Will you be able to beat it? Go for it!"

## Appendix B. Instructions for the general population

"This study is about malingering in cases of road accidents. We need you to respond to this instrument in a completely honest way **and taking into account your current state.** You don't need to invent or fake any symptoms, just consider how you're feeling right now. We ask you to strive to be completely believable, you have to try as much as possible. The instrument that we will apply to you can detect whether or not you interpret the role that we have assigned you well and, if it 'catches' you, you will not receive the bonus of the subject.

Will you be able to do it right? Go for it!"

*Both instructions were translated from spanish, so there may be some slight difference in content.