



The European Journal of Psychology Applied to Legal Context

<https://journals.comadrid.org/ejpalc>



Self-report Measures for Symptom Validity Assessment in Whiplash-associated Disorders

David Pina^a, Esteban Puente-López^b, José Antonio Ruiz-Hernández^{a,c}, Bartolomé Llor-Esteban^{a,c}, and Luis E. Aguerrevere^d

^aUniversidad de Murcia, Spain; ^bUniversidad Nebrija, Spain; ^cInstituto Murciano de Investigación Biosanitaria (IMIB-Arrixaca), Spain; ^dDepartment of Human Services, Stephen F. Austin State University, USA

ARTICLE INFO

Article history:

Received 8 January 2021

Accepted 5 May 2022

Available online 26 May 2022

Keywords:

Malingering

Feigning

Neck injury

Symptom validity test

Simulation design

ABSTRACT

Background/Objective: Whiplash-Associated Disorders (WAD) are one of the most complex conditions to evaluate because several of its symptoms are not observable with current diagnostic methods and cannot be quantified or evaluated correctly. No method is currently available to assess the risk of malingering in the aforementioned condition efficiently. Our aim is to study the capacity of several biopsychosocial psychometric self-report instruments, such as the Brief Pain Inventory (BPI), the Cervical Disability Index (NDI), the SF-36 Health Questionnaire, the Beck Anxiety and Depression Inventories (BDI-II and BAI), or the Brief Illness Perception Questionnaire (BIPQ), to discriminate between patients diagnosed with WAD following a vehicle accident and non-clinical participants with malingering instructions. **Method:** A simulation design was used with 630 participants: 200 non-clinical controls with honest responding condition, 201 instructed malingerers, and 229 WAD clinical outpatients. **Results:** Our results showed an AUC range of .60 to .90, with the highest value being that of the BPI (.90), followed by the NDI (.88), and the lowest value that of the BIPQ (.60), followed by the BAI (.71). **Conclusions:** Overall, the BPI, the NDI, and SF-36 can correctly discriminate between groups with a good specificity (> 90%), while the BAI, BDI, and BIPQ showed a lower capacity, with a high rate of false positives in the case of the BDI and of false negatives in the other two. Practical and research implications are discussed.

Medidas de autoinforme para la evaluación de la validez de los síntomas del síndrome del latigazo cervical

RESUMEN

Antecedentes/Objetivo: El Síndrome del Latigazo Cervical (WAD) es una de las condiciones más complejas de evaluar debido a que varios de los síntomas que presenta no son objetivables con los métodos diagnósticos actuales y no puede cuantificarse ni evaluarse correctamente. En la actualidad no se dispone de ningún método eficiente para valorar el riesgo de simulación en la citada condición. Nuestro objetivo es estudiar la capacidad de varios instrumentos psicométricos biopsicosociales de autoinforme, como el Inventario Breve de Dolor (BPI), el Índice de Discapacidad Cervical (NDI), el Cuestionario de Salud SF-36, los Inventarios de Ansiedad y Depresión de Beck (BDI-II y BAI) o el Cuestionario Breve de Percepción de la Enfermedad (BIPQ) para discriminar entre pacientes diagnosticados con WAD tras un accidente de circulación y participantes no-clínicos con instrucciones de simulación. **Método:** Se utilizó un diseño de simulación con 630 participantes: 200 controles no clínicos con condición de respuesta honesta, 201 simuladores instruidos y 229 pacientes clínicos con WAD. **Resultados:** Nuestros resultados mostraron un rango de AUC de .60 a .90, siendo el valor más alto el del BPI (.90), seguido del NDI (.88), y el valor más bajo el del BIPQ (.60), seguido del BAI (.71). **Conclusiones:** En general, el BPI, el NDI y el SF-36 pueden discriminar correctamente entre grupos con una buena especificidad (> 90%), mientras que el BAI, el BDI y el BIPQ mostraron una menor capacidad, con una alta tasa de falsos positivos en el caso del BDI y falsos negativos en los otros dos. Se discuten además las implicaciones prácticas y de investigación.

Malingering is defined as the intentional presentation of physical and/or psychological symptomatology motivated by obtaining an external gain, including possible financial compensation, obtaining medication, lengthening a work leave, avoiding military

service, etc. (American Psychiatric Association, 2013; Young, 2015). The examination of the differential diagnosis of malingering is mandatory in the forensic setting (Arce, 2017), where it has a ratio of approximately 15 ± to 15% of the cases (Young, 2015). Pain-related

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

Correspondence: epuentel@nebrija.es (E. Puente-López).

ISSN: 1889-1861/© 2022 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

disorders are considered to be one of the most feigned conditions in compensable settings because pain is a subjective experience that is difficult to assess objectively (Greve et al., 2009; Monaro, Bertomeu et al., 2021a; Tuck et al., 2019). Some of the pain conditions often feigned are whiplash injury, fibromyalgia, or lower back pain, which do not usually have an identifiable organic source, and their diagnosis is based on patients' self-reports (Monaro, De Rosario et al., 2021).

Within pain-related disorders, whiplash-related injuries are one of the most confusing, controversial, and complex conditions to diagnose (Elliott et al., 2009). Whiplash-related injuries are usually caused by motor vehicle accidents, or MVA (Walton & Elliot, 2017), where the impact generates an acceleration-deceleration mechanism (whiplash) that transfers energy to the neck and can result in soft tissue injury (whiplash injury) (Spitzer et al., 1995). The whiplash injury is characterized by a high variability of symptoms, like neck stiffness and pain, migraine or headache, fatigue, dizziness, etc., which form a clinical presentation known as whiplash associated disorders (WAD) (Monaro, Bertomeu et al., 2021; Monaro, De Rosario et al., 2021). Several of these symptoms are difficult to objectify with the diagnostic methods currently available, and cannot be quantified or measured correctly (Cassidy et al., 2018; Represas et al., 2020). Therefore, the diagnosis of WAD is usually made according to patients' manifestations and without medical evidence (Carroll et al., 2008; Monaro, Bertomeu et al., 2021).

WAD is considered to be one of the leading sources of disability in the world (Kamper et al., 2008). Specifically, cervical pain of traumatic and non-traumatic origin is the fourth cause of disability worldwide (Walton & Elliot, 2017). In some countries, WAD is compensated financially (Cassidy et al., 2018) and, in countries with high compensation rates for whiplash injuries, chronic whiplash is highly prevalent (Monaro, Bertomeu et al., 2021). Although updated data are not available, its annual incidence ranges, approximately, from 16 to 300 cases per 100,000 inhabitants, with a high variability depending on the country. For example, the USA reaches approximately 300 cases per 100,000 inhabitants, Canada 70 per 100,000, Australia 106 per 100,000, the Netherlands from 188 to 325 per 100,000, and Spain where it amounts to 60 per 100,000 (Holm et al., 2008; Pataskia & Kumar, 2011; Regal Ramos, 2011). The economic cost is approximately 42 billion dollars in the United States and 10 billion euros per year in Europe, with estimates of approximately 3 billion pounds per year in the United Kingdom and 10,000 million euros in Spain (Crouch et al., 2006; Kamper et al., 2008; Noll-Hussong, 2017; Pink et al., 2016).

As Monaro, De Rosario et al. (2021) and Puente-López et al. (2021) indicate, the diagnostic difficulty of WAD, as well as the possibility of obtaining an external compensation, make it a very attractive condition to feign. However, despite the "variety of techniques that have been developed to identify malingering in forensic contexts, there are very few rigorous methodologies for the assessment of WAD that take into account both the heterogeneous nature of the syndrome and the possibility of malingering" (Monaro, Bertomeu et al., 2021, p. 2017). In this sense, for the assessment of symptom validity and malingering, it is advisable to follow a multimethod strategy (Arce et al., 2015; Fariña et al., 2014; Gancedo, Sanmarco et al., 2021) that uses multiple sources of information, including self-report measures (Sherman et al., 2020; Sweet et al., 2021). However, there are still relatively few validated self-report measures to professionals for assessing overreporting in pain-related injuries, especially for whiplash-related injuries (Monaro, Bertomeu et al., 2021), and the studies conducted offer preliminary results that need to be further investigated before being applied in clinical or forensic practice (e.g., Monaro, De Rosario et al., 2021; Puente-López et al., 2021; Sartori et al., 2003).

The use of multiple sources of information for the assessment of symptom validity, as well as self-report measures, can be especially effective in WAD, as it is compatible with the biopsychosocial model of the disease (Campbell et al., 2018; Pina et al., 2021; Sterling, 2011). Pain and other psychosocial variables, such as anxiety,

depression, perception of disability, alteration in the quality of life, or attitude toward the illness situation play a major role in the development of WAD (Björnsenius et al., 2020; Campbell et al., 2018; Sterling, 2011). Thus, the evaluator should consider a wide variety of variables relevant to the condition, with multiple psychometric sources of information including symptom validity determination to determine the presence of WAD. Given the aforementioned lack of methodologies for the assessment of symptom validity in WAD, our main objective is to use a simulation design to study the ability of a biopsychosocial battery of psychometric instruments to discriminate between patients diagnosed with WAD after an MVA, non-clinical controls, and non-clinical instructed malingerers.

Method

Participants

An initial sample of 651 participants divided into three groups was used for this study: non-clinical controls with honest responding condition (hereinafter general population), non-clinical uncoached instructed malingerers (with no symptom information, warning or symptom validity test (SVT)-specific coach, hereinafter instructed malingerers), and MVA WAD clinical outpatients (hereinafter clinical controls). The following inclusion criteria were used for each group: General population and malingerers had to: (1) sign informed consent, (2) pass the pre-manipulation check (all questions answered correctly), (3) pass the post manipulation check (no item with a score higher than 3), (4) provide complete answers on all the instruments administered, (5) not to suffer, or have suffered, from any psychological or medical alteration that could alter the responses to the instruments, and (6) not to have any person in their close socio-familial context who is suffering, or has suffered, from WAD. Clinical controls had to: (a) sign the informed consent, (b) have received a diagnosis of WAD, (c) be of legal age (≥ 18 years), and (d) be classified as a clinical patient (see Procedure section). Failure to meet the criteria in each group resulted in exclusion from the study; 18 clinical controls were excluded due to criterion d (classified as overreporting patient) and 3 instructed simulators were excluded due to criterion 2.

The final sample was composed of a total of 630 participants, with an average age of 31.17 ($SD = 10.17$) with a range of 18 to 65 years. The group division was: 200 participants from the general population, 116 women (58%), with a mean age of 24.55 ($SD = 7.81$); 201 instructed malingerers, 106 women (52.74%), with a mean age of 23.92 ($SD = 7.51$); and 229 clinical controls, 119 women (52%), with a mean age of 37.21 ($SD = 10.65$).

Finally, 16 WAD patients refused to participate after the initial explanation of the objectives of the study, and it was not possible to collect data other than sex (10 women and 6 men) and the reason for non-participation (they did not want to undergo any kind of evaluation that was not imposed by the insurance companies).

Variables and Instruments

Brief Pain Inventory (BPI; Cleeland, 1989, 1990, 1991; Cleeland & Ryan, 1994)

The BPI is an 11-item self-report developed to assess the severity of clinical pain suffered, as well as the degree of social disturbance caused by it. Each of the items is scored from 0 to 10 on a visual scale (Cleeland, 1991). It offers two indices, pain severity and pain interference. It has an excellent internal consistency with a Cronbach alpha of $\alpha = .91$. To represent the results of the instrument, the final score was calculated by adding all the items (Poquet & Lin, 2016). For our sample, the observed internal consistency was good ($\alpha = .89$).

Neck Disability Index (NDI; [Vernon & Mior, 1991](#))

The NDI is a 10-item self-report that was developed to assess a patient's perceived disability, on a score from 0 (no disability) to 34 (full disability). Unlike the interference dimension of the BPI, which seeks to assess pain disturbance in general, the NDI is designed to specifically measure interference in different social areas caused by pain in the cervical region. It has an excellent internal consistency with a Cronbach alpha of $\alpha = .92$. For our sample, the observed internal consistency was also excellent ($\alpha = .93$).

36-Item Short Form Survey (SF-36; [Ware, 2000](#))

The SF-36 is a 36-item self-reported instrument that assesses health-related quality of life (HRQoL). The instrument is divided into the following subscales: physical function (ability to perform physical tasks), physical role (ability to fulfill the physical role), body pain, general health, vitality (energy/fatigue), social function (ability to perform social activities and tasks), emotional role (role limitations due to emotional problems), and mental health. Each subscale produces a score from 0 to 100, with higher scores indicating better quality of life. Cronbach Alpha is $\alpha = .85$ for all dimensions except for social functioning ($\alpha = .75$). For our sample, the internal consistency of the subscales was good (α ranging from .79 to .88).

Beck Depression Inventory (second version, BDI-II; [Beck et al., 1996](#))

The BDI-II is a self-report inventory that measures the presence and severity of depression with a range of scores from 0 (minimal depression / no depression) to 63 (severe depression). It consists of 21 items rated from 0 to 3 based on separate anchors for each item. Respondents choose the statement that best describes their situation during the previous two weeks. It has a good internal consistency with a Cronbach alpha, of $\alpha = .86$. For our sample, the observed internal consistency was also good ($\alpha = .83$).

Beck Anxiety Inventory (BAI; [Beck et al., 1988](#))

The BAI is a self-report inventory that measures the presence and severity of anxiety with a range of scores from 0 (minimal anxiety/no anxiety) to 63 (severe anxiety). It consists of 21 multiple-response items in which the respondent chooses the statement that best describes their situation during the previous two weeks. It has an excellent internal consistency with a Cronbach's alpha of $\alpha = .93$. For our sample, the observed internal consistency was good ($\alpha = .87$).

Brief Illness Perception Questionnaire

The BIPQ is a short version of the Illness Perception Questionnaire (IPQ; [Weinman et al., 1996](#)), designed to evaluate cognitive and emotional representations of the disease. It consists of nine items, with a score ranging from 1 to 10, which measure consequences, duration, personal control, and control of treatment, emotional representations, coherence, and causes. Internal consistency was not evaluated in the study used to design the scale, but it showed a high convergent validity with the IPQ ([Broadbent et al., 2006](#)). For our sample, the observed internal consistency was good ($\alpha = .82$).

Participants' Sociodemographic Variables (sex and age) **and Variables of Medico-Legal Interest**

Time of day when the accident occurred, seat occupied, location of impact, type of road, buckled seat belt, head position, car condition, time since the accident, symptomatology described, and

biomechanical report of the accident were also collected, using an ad-hoc checklist and requesting the information from the assessing physician.

Procedure

A simulation design was used with participants recruited during 2017, 2018, and 2019. For the clinical control group, a sample of outpatients diagnosed with WAD after suffering an MVA was recruited at a multidisciplinary medical center in Spain. All patients had been evaluated first in primary care (emergency) within 72 hours of the accident. They came to the clinic for a personal injury assessment requested by a patient's insurance company or the opposing party, as part of a financial compensation process. The medico-legal assessment performed in the context of the present investigation was carried out within the framework of a compensation assessment, and was requested by the patient's insurance company. Physicians (medical experts in personal injury assessment with over 25 years of experience) at the clinic who agreed to participate in the study verified that their patient met the inclusion criteria and invited them to participate. Patients who agreed to participate signed the informed consent and were evaluated by one of the authors, who applied the prepared battery of scales and conducted a brief assessment interview. Special emphasis was placed on the anonymous nature of the study, and it was indicated that under no circumstances study-related information would be provided to third parties, especially their physician. All necessary medico-legal information (causal link, impact details, biomechanical study of the MVA, etc.) was subsequently requested from the physician.

To avoid the inclusion of possible over-reporters in the clinical patient group, the existence of contradictory evidence in the medico-legal evaluation was used as classification method (clinical controls inclusion criterion d). For this purpose, the three physicians who participated in the study analyzed each case and assessed whether the patient's symptom presentation was consistent with the anatomical-structural indicators expected in the condition (pain severity, pain location, cervical movement range, active/passive joint balance, etc.). Since all patients had a biomechanical report of the accident, their results were also included in this assessment. The biomechanical reports were designed by experts from an external consultancy with extensive experience in vehicle damage analysis. The results of the report were reflected in two values: the magnitude indicator of a collision in the absence of cabin intrusion (Δv) and the mean acceleration (a). These values were interpreted by the evaluators using the thresholds proposed by [Represas-Vázquez et al. \(2016\)](#). Once assessed, the patients were listed as: 1) clinical patient (coherent clinical presentation and biomechanical production of the injury), 2) overreporting patient (incoherent clinical presentation or biomechanical production of the injury), 3) probable malingerer (incoherent clinical presentation and biomechanical production of the injury), or 4) doubtful (the assessing physician doubted which category to assign the patient to). Disagreements were resolved by consensus. Detailed classification method is available upon request to the corresponding author.

For the experimental group, the participants were recruited from the degrees of nursing, medicine, and psychology of the university of one of the authors. They registered by email after the presentation of the study and were randomly assigned to the proposed conditions: instructed 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. The instructed malingerers were asked to feign physical and psychological harm suffered after an alleged MVA to obtain financial compensation. As motivation, they were offered an internal reward (being able to cheat on the tests) and an external reward (an extra

Table 1. Mean, ANOVA, and Cohen's *d*, PSEs and PIS for the Groups' Scale Scores

	Gen.	Cli.	Mal.	<i>F</i> (2, 647)	<i>d</i> ₁ [95% CI]	<i>d</i> ₂ [95% CI]	<i>d</i> ₃ [95% CI]	PS _{ES1}	PIS ₁	PS _{ES2}	PIS ₂	PS _{ES3}	PIS ₃
	<i>M</i>	<i>M</i>	<i>M</i>										
BPI				1141.01**	3.01 [2.73, 3.27]	3.94 [3.60, 4.27]	1.39 [1.18, 1.60]	0.98	0.00	0.99	0.00	0.82	0.08
NDI	11.15	45.54	67.23	639.06**	1.89 [1.66, 2.11]	2.44 [2.17, 2.69]	1.15 [0.94, 1.34]	0.90	-0.02	0.99	0.00	0.79	0.12
BDI	8.72	18.50	27.50	320.68**	1.34 [1.13, 1.54]	2.23 [1.98, 2.47]	1.18 [0.98, 1.38]	0.82	0.09	0.98	0.01	0.79	0.11
BAI	4.15	11.33	21.70	231.12**	2.05 [1.82, 2.28]	1.83 [1.59, 2.06]	0.45 [0.27, 0.65]	0.92	0.02	0.90	0.03	0.62	0.32
SF-36 PH	6.34	17.74	14.66	659.09**	2.88 [2.61, 3.13]	4.41 [4.06, 4.76]	1.60 [1.38, 1.81]	0.97	0.00	0.99	0.00	0.87	0.05
SF-36 MH	85.28	60.20	38.15	374.34**	1.68 [1.46, 1.89]	3.17 [2.87, 3.46]	1.53 [1.31, 1.73]	0.88	0.04	0.98	0.00	0.85	0.06
BIPQ	87.27	70.29	47.48	189.86**	2.10 [1.87, 2.33]	1.67 [1.24, 1.89]	0.26 [0.07, 0.45]	0.93	0.01	0.88	0.04	0.57	0.39

Note. Gen. = general population; Cli. = clinical controls; Mal. = malingering; *M* = mean; *d*₁ = magnitude of the effect of the comparison between the instructed malingerers and general population; *d*₂ = magnitude of the effect of the comparison between the instructed malingerers and clinical controls; *d*₃ = magnitude of the comparison effect between the instructed malingerers and the clinical controls; PS_{ES1} and PIS₁ = probability of superiority and probability of an inferiority score, respectively, of the *d*₁ effect size; PS_{ES2} and PIS₂ = probability of superiority and probability of an inferiority score, respectively, of the *d*₂ effect size; PS_{ES3} and PIS₃ = probability of superiority and probability of an inferiority score, respectively, of the *d*₃ effect size.

point in the final grade of a subject). Failure was penalized, so that only those who completed the battery according to the assigned role would receive the extra point. The instructions they received were designed following the requirements of clarity, specificity, contextualization, and motivation proposed by Rogers and Cruise (1998).

To ensure understanding of the roles, a questionnaire of several multiple-choice questions was administered (pre-experimental manipulation check). All the questions were related to the scenario that the participants had read beforehand: the general population group answered 3 questions and the instructed malingerers answered 6 questions. Also, at the end of the study, participants completed another questionnaire of several multiple-choice questions (post-experimental manipulation check), where scoring from 1 (indicating high levels) to 5 (indicating low levels), memory, understanding, compliance with instructions, effort, and motivation were assessed. Participants who obtained low scores (4 or 5) on any of the questions were excluded from the experiment. Also, after the post-experimental manipulation check, instructed malingerers were asked, with 3 short open-ended questions, how they had prepared for the assigned condition.

No information or training was provided on how to feign, and they were instructed to prepare the assigned role themselves. The importance of being consistent with what was expected in the selected condition was emphasized. Participants assigned to general population status were asked to respond to the scales

with standard instructions (respond sincerely). The same external reward conditions and manipulation check as in the previous group (instructed malingerers) were applied. Both manipulation checks, and instructions, are available upon request to the corresponding author.

This study was approved by the Research Ethics Committee of the authors' university and followed the ethical considerations proposed by the American Psychological Association (2002, 2010).

Data Analysis

Parametric statistics were used, analyzing the differences between the groups with a one-factor analysis of variance (ANOVA). The effect size was calculated using Cohen's *d* statistic (Cohen, 1988). The effect size was interpreted quantitatively as the probability of superiority of the effect size (PS_{ES}; Arce et al., 2020; Arias et al., 2020). PS_{ES} is the probability with which a specific effect size exceeds all other effect sizes observed in the study, and it has demonstrated its practical utility (Gancedo, Fariña et al., 2021; Gancedo, Sanmarco et al., 2021; Ruiz-Hernández et al., 2020). The statistical model error was estimated in terms of the probability of an inferiority score (PIS) (Fandiño et al., 2021). It is interpreted as the probability in the instructed malingerers group of scoring lower than the mean of the clinical controls or the general population groups. To calculate the discriminative ability of the instruments, the receiver operating characteristic curve (ROC curve) was calculated, also obtaining the area under the curve (AUC)

Table 2. Discriminative Ability of the Instruments in the Comparison between Clinical Controls (*n* = 229) and Instructed Malingerers (*n* = 201)

Scale (CoS)	AUC [95% IC]	SEAU	J	SEN (%)	[95% CI]	SPEC (%)	[95% CI]	L+	[95% CI]	L-	[95% CI]
BPI (64)	.90 [.86, .94]	.02	.69 ¹	74.39	63.6, 83.4	94.12	89.1, 97.3	12.65	6.6, 24.1	0.27	0.2, 0.4
BPI (65)	.90 [.86, .94]	.02	.67	71.95	60.9, 81.3	95.42	90.8, 98.1	15.73	7.5, 32.8	0.29	0.2, 0.4
BPI (67)	.90 [.86, .94]	.02	.68	71.95	60.9, 81.3	96.08	91.7, 98.5	18.35	8.3, 40.7	0.29	0.2, 0.4
NDI (29)	.88 [.83, .92]	.02	.69 ¹	71.41	60.7, 81.1	98.04	94.4, 99.6	36.70	11.9, 113.4	0.29	0.2, 0.4
NDI (31)	.88 [.83, .92]	.02	.67	68.29	57.1, 78.1	98.69	95.4, 99.8	52.24	13.1, 208.6	0.32	0.2, 0.4
NDI (32)	.88 [.83, .92]	.02	.64	64.63	53.3, 74.9	99.35	96.4, 100	98.89	13.9, 702.1	0.36	0.3, 0.5
BDI (11)	.87 [.82, .92]	.02	.61 ¹	85.37	75.8, 92.2	75.16	67.5, 81.8	3.44	2.6, 4.6	0.19	0.1, 0.3
BDI (12)	.87 [.82, .92]	.02	.57	81.71	71.6, 89.4	75.16	67.5, 81.8	3.29	2.5, 4.4	0.24	0.2, 0.4
BDI (13)	.87 [.82, .92]	.02	.54	76.83	66.2, 85.4	77.12	69.6, 83.5	3.36	2.5, 4.6	0.30	0.2, 0.4
BAI (13)	.71 [.63, .77]	.03	.37 ¹	43.90	33.0, 53	93.46	88.3, 96.8	6.72	3.5, 12.8	0.60	0.5, 0.7
BAI (14)	.71 [.63, .77]	.03	.33	43.90	33.0, 55.3	88.89	82.8, 93.4	3.95	2.4, 6.6	0.63	0.5, 0.8
BAI (15)	.71 [.63, .77]	.03	.36	51.22	39.9, 62.4	84.97	78.3, 90.2	3.41	2.2, 5.2	0.57	0.5, 0.7
SF-36 (34/45)	.86 [.81, .91]	.02	.62 ¹	70.73	59.6, 80.3	90.85	85.1, 94.9	7.73	4.6, 13.0	0.32	0.2, 0.4
SF-36 (33/43)	.86 [.81, .91]	.02	.60	68.29	57.1, 78.1	91.50	85.9, 95.4	8.04	4.7, 13.8	0.35	0.3, 0.5
SF-36 (28/37)	.86 [.81, .91]	.02	.55	60.98	49.6, 71.6	94.12	89.1, 97.3	10.37	5.4, 20.0	0.41	0.3, 0.5
BIPQ (6)	.60 [.54, .67]	.04	.28 ¹	46.34	35.3, 57.7	81.70	74.6, 87.5	2.53	1.7, 3.8	0.66	0.5, 0.8
BIPQ (7)	.60 [.54, .67]	.04	.17	23.17	14.6, 33.8	94.12	89.1, 97.3	3.94	1.9, 8.3	0.82	0.7, 0.9
BIPQ (8)	.60 [.54, .67]	.04	.13	18.29	10.6, 28.4	94.77	90.0, 97.7	3.50	1.5, 7.9	0.86	0.8, 1.0

Note. CoS = cut-off score; 95% CI = 95% confidence interval; AUC = area under the curve; SEAU = standard error of area under curve; J = Youden Index; ¹optimal cut-off point; SEN = sensitivity; SPEC = specificity; L+ and L- = positive and negative likelihood.

and the standard error of that area (SE_{AUC}) (Loainaz & de Sousa, 2020). Youden index (J) was used to determine the optimal cut-off point for each instrument, calculating its sensitivity, specificity, positive and negative likelihood index ($L+$ and $L-$), and the positive/negative predictive power (PPP and NPP) for a prevalence of 10%, 30%, and 50%. The analyses described above were also performed at the two cut-off points after the optimal one, to offer a wider range of decisions.

Results

Characteristics of the Groups

With regard to the clinical controls, the medico-legal assessment was carried out after an average of 38.24 days ($SD = 9.25$) since the accident, with a range of 26 to 60 days. All patients claimed to have cervical pain, 96.1% ($n = 237$) had excessive sensitivity in the cervical region, 20.6% ($n = 51$) manifested dizziness, and 4.4% ($n = 11$) indicated another symptom in addition to those mentioned, with all of them manifesting pain in the lumbar area. None of the patients had severe neurological signs, nor were any fractures or severe musculoskeletal lesions observed. Only 12 patients (4.8%) presented information on the pre-accident status (previous state). The mean biomechanical values of the accident were $\Delta v = 7.24$ km/h ($SD = 3.41$) and $a = 5.2$ g ($SD = 2.11$), respectively. Finally, no significant gender differences were found in any of the scales included in the study (BPI score, $p = .14$; NDI score, $p = .21$; BDI score, $p = .48$; BAI score, $p = .53$; SF-36 PH score, $p = .66$; SF-36 MH score, $p = .70$; and BIPQ score, $p = .73$).

Regarding the instructed malingerers and general population, for the preparation of the role, the majority (64%) used non-scientific internet pages (e.g., groups of lawyers or insurers) as a source of information, located through a search engine using terms such as “cervical whiplash, cervical sprain, or cervical whiplash syndromes”, “traffic accident and compensation” and/or “common damage from road accidents.” Of the remaining 36%, 30% used a combined search of 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.

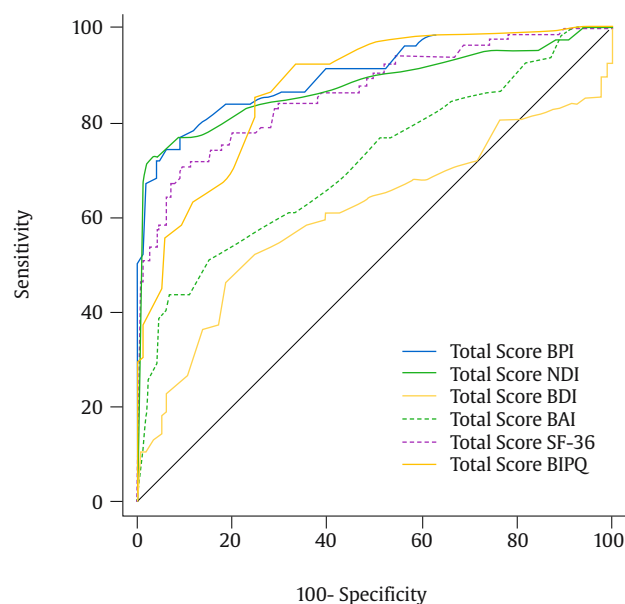


Figure 1. ROC Curves for the Comparison between Clinical Controls ($n = 229$) and Instructed Malingerers ($n = 201$).

Scale Scores and Group Comparison

As can be seen in Table 1, the one-factor ANOVA indicated the existence of significant differences between groups in all the scales used. The highest scores were observed in the group of instructed malingerers (BPI, $M = 67.23$, $SD = 18.12$; NDI, $M = 27.50$, $SD = 9.93$; BDI, $M = 21.70$, $SD = 10.77$; BAI, $M = 14.65$, $SD = 5.98$; SF-36 PH, $M = 38.15$, $SD = 14.65$; SF-36 MH, $M = 47.48$, $SD = 15.58$; BIPQ, $M = 5.26$, $SD = 2.51$), with the average scores of that group exceeding the Youden index optimal cut-off point (Table 2), with the exception of the NDI and the BIPQ. Regarding the effect sizes (Table 1), the comparison between the group of instructed malingerers and the general population (d_1) obtained the highest values (d_1 ranges between 1.34 and 3.01). Specifically, in the comparison between clinical controls and instructed malingerers (d_3), the highest values were observed in the SF-36 PH ($d = 1.60$; an effect size above 87%, $PS_{ES} = 0.87$, $PIS = 0.05$), SF-36 MH

Table 3. Positive and Negative Predictive Power with 10%, 30%, and 50% Prevalence

Scale (CoS)	10% PPP	[95% CI]	10% NPP	[95% CI]	30% PPP	[95% CI]	30% NPP	[95% CI]	50% PPP	[95% CI]	50% NPP	[95% CI]
BPI (64)	58.4	42.4, 72.8	97.1	95.8, 98.0	84.4	74.0, 91.2	89.6	85.5, 92.6	92.7	86.9, 96.0	78.6	71.7, 84.2
BPI (65)	63.6	45.6, 78.5	96.8	95.6, 97.7	87.1	76.3, 93.4	88.8	84.9, 91.8	94.0	88.3, 97.0	77.3	70.6, 82.8
BPI (67)	67.1	47.9, 81.9	96.9	95.6, 97.8	88.7	78.0, 94.6	88.9	84.9, 91.9	94.8	89.2, 97.6	77.4	70.7, 82.9
NDI (29)	80.3	56.9, 92.6	96.9	95.7, 97.8	94.0	83.6, 98.0	89.1	85.2, 92.0	97.3	92.2, 99.1	77.8	71.2, 83.2
NDI (31)	85.3	59.2, 95.9	96.6	95.3, 97.5	95.7	84.9, 98.9	87.9	84.1, 90.9	98.1	92.9, 99.5	75.7	69.4, 81.1
NDI (32)	91.7	60.7, 98.7	96.2	95.0, 97.1	97.7	85.7, 99.7	86.8	83.0, 89.8	99.0	93.3, 99.9	73.7	67.7, 79.0
BDI (11)	42.7	28.1, 58.8	93.7	92.5, 94.8	74.2	60.1, 84.6	79.5	76.2, 82.5	87.0	77.9, 92.8	62.5	57.8, 67.0
BDI (12)	30.5	20.9, 42.2	93.4	92.1, 94.6	62.9	50.4, 73.8	78.7	75.2, 81.9	79.8	70.3, 86.8	61.3	56.5, 65.9
BDI (13)	27.5	19.7, 36.8	94.0	92.6, 95.2	59.4	48.7, 69.2	80.3	76.3, 83.7	77.3	68.9, 84.0	63.5	58.0, 68.7
BAI (13)	27.6	22.2, 33.8	97.9	96.5, 98.7	59.6	52.4, 66.3	92.3	87.6, 95.3	77.5	72.0, 82.1	83.7	75.1, 89.7
BAI (14)	26.8	21.4, 32.9	97.4	95.9, 98.3	58.5	51.2, 65.4	90.6	85.7, 93.9	76.7	71.0, 81.5	80.4	72.0, 86.8
BAI (15)	27.2	21.4, 33.8	96.8	95.2, 97.8	59.0	51.2, 66.3	88.6	83.8, 92.1	77.1	71.0, 82.1	76.9	69.0, 83.3
SF-36 (34/45)	46.2	33.8, 59.1	96.5	95.2, 97.5	73.2	63.0, 81.4	87.6	83.4, 90.9	88.5	82.2, 92.8	75.6	68.8, 81.4
SF-36 (33/43)	47.2	34.2, 60.5	96.3	95.0, 97.3	76.2	65.5, 84.3	87.0	82.9, 90.2	88.2	81.6, 92.6	74.1	67.5, 79.8
SF-36 (28/37)	53.5	37.4, 69.0	95.6	94.3, 96.6	81.6	69.7, 89.6	84.9	81.1, 88.1	91.2	84.3, 95.2	70.7	64.7, 76.0
BIPQ (6)	22.0	15.8, 29.7	93.2	91.7, 94.4	52.0	41.9, 62.0	78.0	74.1, 81.5	71.7	62.7, 79.2	60.4	55.1, 65.4
BIPQ (7)	30.4	17.2, 48.0	91.7	90.7, 92.6	62.8	44.5, 78.1	74.1	71.6, 76.4	79.8	65.1, 89.3	55.1	51.9, 58.1
BIPQ (8)	28.0	14.7, 46.8	91.3	90.3, 92.1	60.0	39.9, 77.2	73.0	70.8, 75.1	77.8	60.8, 88.8	53.7	51.0, 56.4

Note. CoS = cut-off score; 95% CI = 95% confidence interval; PPP = positive predictive power; NPP = negative predictive power.

($d = 1.53$; an effect size above 85.9%, $PS_{ES} = 0.85$, $PIS = 0.06$), BPI ($d = 1.39$; an effect size above 82.3%, $PS_{ES} = 0.82$, $PIS = 0.08$), BDI ($d = 1.18$; an effect size above 79.6%, $PS_{ES} = 0.79$, $PIS = 0.11$), and NDI ($d = 1.15$; an effect size above 79.1%, $PS_{ES} = 0.79$, $PIS = 0.12$).

Optimal Cut-off Scores and Discriminative Ability of the Instruments

To obtain a possible optimal cut-off point for comparison between instructed malingerers and clinical controls, a ROC curve was established (Figure 1) and the Youden index was calculated (Table 2). In addition to the Youden index optimal cut-off score, two additional cut-off points were included for a higher decision range.

The AUC had a range of .60 to .90, with the highest value being that of the BPI (.90), followed by the NDI (.88), and the lowest value that of the BIPQ (.60), followed by the BAI (.71). For the BPI, the optimal cut-off point was 64 ($J = .68$), with a sensitivity of 74.39% and a specificity of 94.12% (false positive rate of 5.88%); for the NDI, it was 29 ($J = .69$), with a sensitivity of 71.95% and a specificity of 98.04% (false positive rate of 1.96%); for the BPI, it was 11 ($J = .60$), with a sensitivity of 87.37% and a specificity of 75.16% (false positive rate of 24.84%); for the BAI, it was 13 ($J = .37$), with a sensitivity of 43.90% and a specificity of 93.46% (false positive rate of 6.54%); for the SF-36, PH/MH, it was 34 and 42, respectively ($J = .61$), with a sensitivity of 70.73% and a specificity of 90.85% (false positive rate of 9.15%); and for the BIPQ, it was 6 ($J = .28$), with a sensitivity of 46.34% and a specificity of 81.70% (false positive rate of 18.30%).

With regard to the positive likelihood ratio (L+), the BPI and the NDI also achieved the highest values (12.65, 15.73, and 18.35 for the BPI, and 36.7, 52.24, and 98.89 for the NDI), indicating that scores equal to or higher than the optimal cut-off point in the BPI were between 12 and 18 times more likely in the group of instructed malingerers and, in the NDI, between 36 and 98 times more likely. All other instruments obtained values ranging from 2.53 to 10.37.

The PPP and NPP for a prevalence of 10, 30, and 50% can be seen in Table 3. Generally speaking, assuming a base ratio of 30%, the most likely to be found in forensic practice (Curtis et al., 2019), the BPI, NDI, and SF-36 achieved a PPP greater than 70% (84.4, 87.18, and 88.7 for the BPI; 94, 95.7, and 97.77 for the NDI; and 73.2, 76.2, and 81.6 for the SF-36), and a NPP greater than 85% (89.6, 88.8, and 88.9 for the NPI; 89.1, 87.9, and 86.8 for the NDI; 87.6, 87, and 84.9 for the SF-36). In the case of NPP, the BAI obtained the highest values (92.3, 90.6, and 88.6).

Discussion

The main objective of this study was to evaluate the ability of several self-report tests to discriminate between non-clinical instructed malingerers, non-clinical honest respondents, and clinical patients who were diagnosed with WAD after a MVA. Significant differences in all scales can be seen in the group comparisons. As expected, due to the characteristics of the population, the comparison between instructed malingerers and the general population had the greatest effect size. The instructed malingerers achieved higher scores than the other two groups, which is consistent with the literature, where this group tends to offer a more severe presentation of the feigned condition than expected (Bianchini et al., 2014; Crighton et al., 2014; Curtis et al., 2019; Puente-López et al., 2021; Sánchez et al., 2017). Clinical controls offered an intermediate severity profile, with moderate pain and quality of life impairment, slight perception of disability, mild anxious-depressive symptomatology, and an attitude toward the illness situation at an intermediate point between negative and positive. This profile is consistent with the presentation commonly offered by a mild severity WAD patient (Åhman & Stålnacke, 2008; Beltran-Alacreu et al., 2018; Sullivan et al., 2002).

An exception to the above was found in the instructed malingerers, as the BAI scores were significantly higher in the clinical controls. A possible explanation for these differences is that the instructed malingerers considered that the depressive symptomatology was more consistent with the condition to be feigned, as can be seen in the BDI scores, and they underestimated the role of anxiety in WAD. Due to the biomedical nature of the condition, it is possible that the sources consulted to prepare the role may have followed a primarily anatomical or biological approach, paying little attention to the psychological consequences. The difference in the scores observed in the two SF-36 indices, physical and mental health, is consistent with this explanation, since the instructed malingerers stated that their physical health was more impaired than their psychological health. For future research, it would be advisable to consider the inclusion of several groups of instructed malingerers to study the possible effect of prior training on role preparation.

With regard to the discriminative ability of the instruments, it should be noted that the term "optimal" used in this study has been determined by a statistical criterion (Youden index) and should not be followed without question. The "optimal" cut-off score should be determined by the intended use of the instrument (Giromini et al., 2022). However, the vast majority of research recommends maintaining a specificity above 90% if the instrument is to be used in a high-stakes forensic context (Sweet et al., 2021). Taking this into account, in the comparison between clinical controls and instructed malingerers, the BPI, NDI and SF-36 obtained the highest specificity at the Youden index optimal cutoff scores (94%, 98%, and 91% respectively with a cut-off score of 64, 29, and 34/45, respectively), with a moderate sensitivity (74%, 71%, and 70%, respectively). A more conservative cutoff score (67, 32, and 28/37, respectively) can be taken to increase specificity to more optimal levels (96%, 99%, and 94%, respectively) in exchange for slightly decreasing in sensitivity (72%, 65%, and 61%, respectively). Above these cutoff scores, sensitivity decreases very significantly in exchange for a very slight increase in specificity. Generally speaking, the BPI, NDI, and SF-36 can discriminate effectively between instructed malingerers and clinical controls, and may be promising instruments for a symptom validity protocol for WAD. On the other hand, at the Youden index optimal cutoff scores, the BDI has a high false positive rate in exchange for a high sensitivity (85%), which would limit its application to screening tasks, and the BAI and the BIPQ have a low sensitivity (37% and 28% respectively) that discourage their inclusion in any type of protocol for the assessment of symptom validity. However, anxious-depressive symptomatology and attitudes toward the illness situation are useful for the WAD assessment (Berglund et al., 2006; Björnsenius et al., 2020; Campbell et al., 2018; Falla et al., 2016), so, for future research, it would be advisable to study the performance of other instruments that measure these variables, such as the Anxiety and Depression Scale of Goldberg (EADG; Goldberg et al., 1988), the Anxiety and Stress Depression Scale-21 (DASS-21; Lovibond & Lovibond, 1995), or the Millon Behavioral Medicine Diagnostic (MBMD; Millon et al., 2001) which, while needing more administration time, include integrated validity scales that may be useful as an additional measure.

Although the BPI, NDI, and SF-36 are promising instruments, this study should be replicated following a criterion group paradigm before applying the findings in professional clinical and forensic settings (see, for example Curtis et al., 2019). Also, their individual use is not recommended. As we explained in the introduction, WAD is a condition consistent with the biopsychosocial model of the disease (Sterling, 2011). The individual use of one of the instruments does not provide a complete overview of the condition and may lead to an erroneous classification. In this regard, for the assessment of symptom validity, multiple psychometric and information sources should be used, and a decision should never be made solely on the basis of the results of a psychometric test (Sweet et al., 2021). A measure such as the proposals "may help to define an endpoint

(i.e., symptom exaggeration or overreporting), but it is silent as to the external and intrinsic factors contributing to this endpoint” (Merckelbach et al., 2019, p. 322).

In addition to being promising for potential application in professional practice, the use of these measures may also be useful for research, specifically for the criterion group paradigm, where clinical patients are classified into different groups (e.g., genuine patients, possible simulators, probable simulators, or definite simulators), using different classification systems, to increase the external validity of the design with the inclusion of probable genuine malingerers. Bianchini et al. (2014), in a similar study with brief measures, indicated that the Modified Somatic Perception Questionnaire (MSPQ; Main, 1983) and the Pain Disability Index (PDI; Pollard, 1984) could be integrated into a system for malingering detection/classification, like the criteria for malingered symptom presentations of Bianchini et al. (2005). Together with the second edition of the Structured Interview of Reported Symptoms (SIRS-2; Rogers et al., 2010), this system is one of the most commonly used to perform group classification in the criterion group paradigm (e.g., Bianchini et al., 2018; Curtis et al., 2019; Greve et al., 2013) and usually uses a large psychometric battery to assess its criteria (see, for example, Table 2 of the Appendix of Curtis et al., 2019). Both the SIRS-2 and the psychometric battery of the previous authors imply an important economic and temporal cost that is not accessible to all researchers, and limits the possibility of applying a criterion group design in the medico-legal context. In particular, the time cost is especially problematic, as, in certain geographical regions, such as that of this research, it is virtually unfeasible to apply instruments that require such extensive administration time, also adding the time required by the application of the battery or instrument to be validated. Having a short and low-cost battery, composed of instruments such as those studied in this publication, where scores above the recommended cut-off point can be considered as “psychometric findings” (Bianchini et al., 2005), would be of interest to provide researchers with alternative classification options. In this regard, Monaro, De Rosario et al. (2021) recently developed a proposal of classification system, using a linear discriminant analysis classification model that incorporates “the mechanical approach and the qualitative analysis of the symptomatology - to obtain a malingering detection model based on a wider range of indices, both biomechanical and self-reported” (p. 1639). Although the model’s ability to discriminate malingerers from clinical patients is moderate (AUC = .84, sensitivity = 77.8%, and specificity = 84.7%), the proposal is highly promising, as it incorporates resources available in the medico-legal context, such as the biomechanical assessment of the injury. Despite the fact that the most complex conditions to diagnose, and considered as more “problematic” and “easy to feign”, such as WAD, have a high prevalence in the medico-legal system, the research of symptom validity assessment in these conditions and context is still far behind other areas such as neuropsychology. Having severe limitations for the application of criterion paradigm classification criteria commonly used in the literature significantly limits the applicability of the results obtained. Therefore, for future research, special emphasis should be placed on the development of a classification system that allows the application of a criterion group paradigm that follows a system adapted to the characteristics of the medico-legal context, such as that of Monaro, De Rosario et al. (2021).

Limitations

Despite the fact that the present research followed recommendations for simulation studies, such as the use of pre and post-manipulation check, use of positive and negative incentives, and comparison with a sample of genuine clinical patients, the results obtained should be interpreted based on a series of limitations. First, as Czornik et al. (2021) states, “the

dilemma arising when special patient populations are studied is that it is difficult to distinguish between true or false positives” (p. 8). Although we tried to objectively discard possible feigners for the clinical control group, the classification criteria used cannot guarantee that all clinical controls are bona fide. As detailed above, in Spain, the medico-legal assessment of bodily injury is carried out in a short period of time, which significantly hinders the application of malingering classification systems that require extensive application time, making the use of a criterion group design highly complex. With the means currently available, this question remains a significant challenge for researchers. Second, our battery of instruments was designed for WAD assessment. While application to other pain-related disorders may be considered, our results cannot be generalized to other conditions without adequate validation. Third, the age of participants differs significantly from the age of the clinical control group, and that can influence the results, as the presentation of symptoms may differ as a function of age. With a view to future research, we consider that it would be advisable to have a sample of instructed malingerers with a wider age range. Fourth and last, the possible effect that different forms of coaching (symptom information, warnings, or SVT-specific coaching) may have on the performance of the applied instruments was not assessed in the present study. Given that coaching can significantly influence the discriminative ability of the applied tests (Giromini et al., 2022), it would be of great interest for future research to replicate the design used by including additional experimental conditions, related to the different forms of coaching.

Conclusions

In this study, a simulation design was used to evaluate the capacity of a battery of short scales, composed of the BPI, the NDI, the SF-36, the BAI, the BDI, and the BIPQ, to differentiate between genuine patients diagnosed with WAD and non-clinical instructed malingerers. Our findings indicate that the BPI, the NDI, and the SF-36 had adequate discriminatory ability and can be usefully integrated into a system, methodology, or battery intended for malingering screening. The BAI, BDI, and BIPQ showed a lower capacity, with a high rate of false positives in the case of the BDI and of false negatives in the other two, so we believe that they are not appropriate for this purpose. However, the evidence indicates that the variables measured by these tests play an important role in the development of the condition studied, so it would be advisable for future research to validate other instruments that assess these variables. In general, our results indicate that self-report measures such as those used herein can be useful both in the forensic context and research, where they can be used integrated in classification systems such as that of Bianchini et al. (2005).

Conflict of Interest

The authors of this article declare no conflict of interest.

References

- Ahman, S., & Stålnacke, B. M. (2008). Post-traumatic stress, depression, and anxiety in patients with injury-related chronic pain: A pilot study. *Neuropsychiatric Disease and Treatment*, 4(6), 1245-1249. <https://doi.org/10.2147/NSD.S4104>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
- American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. American Psychological Association.
- American Psychological Association. (2010). *Ethical principles for psychologists and code of conduct*. American Psychological Association.
- Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: evaluación de la validez científica y judicial de la hipótesis y la prueba [Content analysis of witness statements: Assessment of the scientific

- and judicial validity of the hypothesis and the evidence]. *Acción Psicológica*, 14(2), 171-190. <https://doi.org/10.5944/ap.14.1.21347>
- Arce, R., Arias, E., Novo, M., & Fariña, F. (2020). Are interventions with batterers effective? A meta-analytical review. *Psychosocial Intervention*, 29(3), 153-164. <https://doi.org/10.5093/pi2020a11>
- Arce, R., Fariña, F., & Vilariño, M. (2015). Daño psicológico en casos de víctimas de violencia de género: un estudio comparativo de las evaluaciones forenses [Psychological injury in intimate partner violence cases: A contrastive analysis of forensic measures]. *Revista Iberoamericana de Psicología y Salud*, 6(2), 72-80. <https://doi.org/10.1016/j.rips.2015.04.002>
- Arias, E., Arce, R., Vázquez, M. J., & Marcos, V. (2020). Treatment efficacy on the cognitive competence of convicted intimate partner violence offenders. *Anales de Psicología/Annals of Psychology*, 36(3), 427-435. <https://doi.org/10.6018/analesps.428771>
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.
- Beck, A. T., Steer, R. A., & Garbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory. Twenty-five years of evaluation. *Clinical Psychological Review*, 8(1), 77-100. [https://www.doi.org/10.1016/0272-7358\(88\)90050-5](https://www.doi.org/10.1016/0272-7358(88)90050-5)
- Beltran-Alacreu, H., López-de-Uralde-Villanueva, I., Calvo-Lobo, C., La Touche, R., Cano-de-la-Cuerda, R., Gil-Martínez, A., Fernández-Ayuso, D., & Fernández-Carnero, J. (2018). Prediction models of health-related quality of life in different neck pain conditions: A cross-sectional study. *Patient Preference and Adherence*, 12, 657-666. <https://doi.org/10.2147/PPA.S162702>
- Berglund, A., Bodin, L., Jensen, I., Wiklund, A., & Alfredsson, L. (2006). The influence of prognostic factors on neck pain intensity, disability, anxiety and depression over a 2-year period in subjects with acute whiplash injury. *Pain*, 125(3), 244-256. <https://doi.org/10.1016/j.pain.2006.05.026>
- Bianchini, K. J., Aguerrevere, L. E., Curtis, K. L., Roebuck-Spencer, T. M., Frey, F. C., Greve, K. W., & Calamia, M. (2018). Classification accuracy of the MMPI-2-Restructured Form Validity Scales in detecting malingering pain-related disability. *Psychological Assessment*, 30(7), 857-869. <https://doi.org/10.1037/pas0000532>
- Bianchini, K. J., Aguerrevere, L. E., Guise, B. J., Ord, J. S., Etherton, J. L., Meyers, J. E., Soignier, R. D., Greve, K. W., Curstis, K. L., & Buy, J. (2014). Accuracy of the Modified Somatic Perception Questionnaire and Pain Disability Index in the detection of malingering pain-related disability in chronic pain. *Clinical Neuropsychologist*, 28(8), 1376-1394. <https://doi.org/10.1080/13854046.2014.986199>
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingering pain-related disability: Lessons from cognitive malingering research. *Spine Journal*, 5(4), 404-417. <https://doi.org/10.1016/j.spinee.2004.11.016>
- Björnsenius, V., Löfgren, M., & Stålnacke, B. M. (2020). One-year follow-up after multimodal rehabilitation for patients with whiplash-associated disorders. *International Journal of Environmental Research and Public Health*, 17(13), Article 4784. <https://doi.org/10.3390/ijerph17134784>
- Broadbent, E., Petrie, K. J., Main, J., & Weinman, J. (2006). The brief illness perception questionnaire. *Journal of Psychosomatic Research*, 60(6), 631-637. <https://www.doi.org/10.1016/j.jpsychores.2005.10.020>
- Campbell, L., Smith, A., McGregor, L., & Sterling, M. (2018). Psychological factors and the development of chronic whiplash-associated disorder(s): A systematic review. *Clinical Journal of Pain*, 34(8), 755-768. <https://doi.org/10.1097/AJP.0000000000000597>
- Carroll, L. J., Holm, L. W., Hogg-Johnson, S., Côté, P., Cassidy, J. D., Haldeman, S., Nordin, M., Hurwitz, E. L., Carragee, E. J., van der Velde, G., Peloso, P. M., & Guzman, J. (2008). Course and prognostic factors for neck pain in whiplash-associated disorders (WAD): Results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *European Spine Journal*, 17(Suppl 1), 83-92. <https://doi.org/10.1007/s00586-008-0628-7>
- Cassidy, J. D., Leth-Petersen, S., & Rotger, G. P. (2018). What happens when compensation for whiplash claims is made more generous? *Journal of Risk and Insurance*, 85(3), 635-662. <https://doi.org/10.1111/jori.12169>
- Cleeland, C. S. (1989). Measurement of pain by subjective report. In C. R. Chapman & J. D. Loeser (Eds.), *Issues in pain measurement* (pp. 391-403). Raven Press.
- Cleeland, C. S. (1990). Assessment of pain in cancer: Measurement issues. In K. M. Foley, J. J. Bonica, & V. Ventafridda (Eds.), *Proceedings of the Second International Congress on Cancer Pain* (pp. 47-55). Raven Press.
- Cleeland, C. S. (1991). Pain assessment in cancer. In D. Osoba (Ed.), *Effect of cancer on quality of life* (pp. 293-305). CRC Press.
- Cleeland, C. S., & Ryan, K. M. (1994). Pain assessment: Global use of the Brief Pain Inventory. *Annals of the Academy of Medicine, Singapore*, 23(2), 129-138.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Crighton, A. H., Wygant, D. B., Applegate, K. C., Umlauf, R. L., & Granacher, R. P. (2014). Can brief measures effectively screen for pain and somatic malingering? Examination of the Modified Somatic Perception Questionnaire and Pain Disability Index. *Spine Journal*, 14(9), 2042-2050. <https://doi.org/10.1016/j.spinee.2014.04.012>
- Crouch, R., Whitewick, R., Clancy, M., Wright, P., & Thomas, P. (2006). Whiplash-associated disorder: Incidence and natural history over the first month for patients presenting to a UK emergency department. *Emergency Medicine Journal*, 23(2), 114-118. <https://doi.org/10.1136/emj.2004.022145>
- Curtis, K. L., Aguerrevere, L. E., Bianchini, K. J., Greve, K. W., & Nicks, R. C. (2019). Detecting malingering pain-related disability with the Pain Catastrophizing Scale: A criterion groups validation study. *Clinical Neuropsychologist*, 33(8), 1485-1500. <https://doi.org/10.1080/13854046.2019.1575470>
- Czornik, M., Merten, T., & Lehrner, J. (2021). Symptom and performance validation in patients with subjective cognitive decline and mild cognitive impairment. *Applied Neuropsychology, Adult*, 28(3), 269-281. <https://doi.org/10.1080/23279095.2019.1628761>
- 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://doi.org/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>
- Fandiño, R., Basanta, J., Sanmarco, J., Arce, R., & Fariña, F. (2021). Evaluation of the executive functioning and psychological adjustment of child to parent offenders: Epidemiology and quantification of harm. *Frontiers in Psychology*, 12, Article 616855. <https://doi.org/10.3389/fpsyg.2021.616855>
- 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>
- Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., & Arce, R. (2021). Reality monitoring: A meta-analytical review for forensic practice. *The European Journal of Psychology Applied to Legal Context*, 13(2), 99-110. <https://doi.org/10.5093/ejpalc2021a10>
- Gancedo, Y., Sanmarco, J., Marcos, V., & Seijo, D. (2021). Estimating the efficacy of Fptsd Scale to report malingering of PTSD: A meta-analytic review. *Revista Iberoamericana de Psicología y Salud*, 12(1), 44-57. <https://doi.org/10.23923/j.rips.2021.01.044>
- Giromini, L., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: New articles, new issues. *Psychological Injury and Law*, 15(1), 1-21. <https://doi.org/10.1007/s12207-022-09444-2>
- Goldberg, D., Bridges, K., Duncan-Jones, P., & Grayson, D. (1988). Detecting anxiety and depression in general medical settings. *BMJ*, 297(6653), 897-899. <https://doi.org/10.1136/bmj.297.6653.897>
- Greve, K. J., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *Clinical Neuropsychologist*, 27(1), 108-137. <https://doi.org/10.1080/13854046.2012.739646>
- Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychological 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., Carroll, L. J., Cassidy, J. D., Hogg-Johnson, S., Côté, P., Guzman, J., Peloso, P., Nordin, M., Hurwitz, E., van der Velde, G., Carragee, E., Haldeman, S., & Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders (2008). The burden and determinants of neck pain in whiplash-associated disorders after traffic collisions: Results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine*, 33(4 Suppl), S52-S59. <https://doi.org/10.1097/BRS.0b013e3181643ece>
- Kamper, S. J., Rebbeck, T. J., Maher, C. G., McAuley, J. H., & Sterling, M. (2008). Course and prognostic factors of whiplash: A systematic review and meta-analysis. *Pain*, 138(3), 617-629. <https://www.doi.org/10.1016/j.pain.2008.02.019>
- Loinaz, I., & de Sousa, A. M. (2020). Assessing risk and protective factors in clinical and judicial child-to-parent violence cases. *The European Journal of Psychology Applied to Legal Context*, 12(1), 43-51. <https://doi.org/10.5093/ejpalc2020a5>
- Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the Depression Anxiety Stress Scales* (2nd ed.). Psychology Foundation.
- Main, C. J. (1983). The Modified Somatic Perception Questionnaire (MSPQ). *Journal of Psychosomatic Research*, 27(6), 503-514. [https://doi.org/10.1016/0022-3999\(83\)90040-5](https://doi.org/10.1016/0022-3999(83)90040-5)
- 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>
- Millon, T., Antoni, M., Millon, C., Meagher, S., & Millon, G. S. (2001). *Millon Behavioral Medicine Diagnostic*. NCS Assessments.
- Monaro, M., Bertomeu, C. B., Zecchinato, F., Fietta, V., Sartori, G., & De Rosario Martínez, H. (2021). The detection of malingering in whiplash-related injuries: A targeted literature review of the available strategies. *International Journal of Legal Medicine*, 135(5), 2017-2032. <https://doi.org/10.1007/s00414-021-02589-w>

- Monaro, M., De Rosario, M., Baydal-Bertomeu, J. M. Bernal-Lafuente, M., Masiero, S., Macia-Calvo, M., Cantele, F., & Sartori, G. (2021). A model to differentiate WAD patients and people with abnormal pain behaviour based on biomechanical and self-reported tests. *International Journal of Legal Medicine*, 135(4), 1637-1646. <https://doi.org/10.1007/s00414-021-02572-5>
- Noll-Hussong, M. (2017). Whiplash syndrome reloaded: Digital echoes of whiplash syndrome in the European Internet search engine context. *JMIR Public Health and Surveillance*, 3(1), e15. <https://doi.org/10.2196/publichealth.7054>
- Pastakia, K., & Kumar, S. (2011). Acute whiplash associated disorders (WAD). *Open Access Emergency Medicine*, 3, 29-32. <https://doi.org/10.2147/OAEM.S17853>
- Pina, D., Puente-López, E., Ruiz-Hernández, J. A., Ruiz-Cabello, A. L., Aguerrevere, L., & Magalhães, T. (2021). Whiplash-associated disorders. Biopsychosocial profiles of pain perception in forensic cases of victims of motor vehicle accidents. *Frontiers in Psychology*, 12, Article 716513. <https://doi.org/10.3389/fpsyg.2021.716513>
- Pink, J., Petrou, S., Williamson, E., Williams, M., & Lamb, S. E. (2016). Economic and health-related quality of life outcomes of whiplash associated disorders. *Spine*, 41(17), 1378-1386. <https://doi.org/10.1097/BRS.0000000000001512>
- Pollard, C. (1984). Preliminary validity study of the Pain Disability Index. *Perceptual and Motor Skills*, 59(3), 974. <https://doi.org/10.2466/pms.1984.59.3.974>
- Poquet, N., & Lin, C. (2016). The Brief Pain Inventory (BPI). *Journal of Physiotherapy*, 62(1), 52. <https://doi.org/10.1016/j.jphys.2015.07.001>
- Puente-López, E., Pina, D., Ruiz-Hernández, J. A., & Llor-Esteban, B. (2021). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in motor vehicle accident patients. *The Journal of Forensic Psychiatry & Psychology*, 32(1), 131-154. <https://doi.org/10.1080/14789949.2020.1833073>
- Regal Ramos, R. (2011). Síndrome de latigazo cervical. Características epidemiológicas de los pacientes evaluados en la Unidad Médica de Valoración de Incapacidades de Madrid [Whiplash. Epidemiological characteristics of patients evaluated in the Assessment of Disability Unit of Madrid]. *Medicina y Seguridad del Trabajo*, 57(225), 348-360.
- Represas-Vázquez, C., Muñoz-Barús, J. I., & Luna-Maldonado, A. (2016). Importancia de la biomecánica del impacto en la valoración pericial del síndrome del latigazo cervical [Importance of the biomechanics of the impact in the expert appraisal of cervical whiplash syndrome]. *Revista Española de Medicina Legal*, 42(2), 72-80. <https://doi.org/10.1016/j.reml.2015.10.001>
- Represas, C., Puente-López, E., Pina, D., Luna-Maldonado, A., Ruiz-Hernández, J. A., & Llor-Esteban, B. (2020). Rating criteria of the causal nexus in whiplash associated disorders: A study of interrater reliability. *Revista Española de Medicina Legal*, 47(2), 48-56. <https://doi.org/10.1016/j.reml.2020.06.002>
- Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation design: Threats to external validity. *Law and Human Behavior*, 22(3), 273-285. <https://doi.org/10.1023/A:1025702405865>
- Rogers, R., Sewell, K. W., & Gillard, N. D. (2010). *Structured interview of reported symptoms: Professional manual* (2nd ed.). Psychological Assessment Resources.
- Ruiz-Hernández, J., Pina, D., Puente-López, E., Luna-Maldonado, A., & Llor-Esteban, B. (2020). Attitudes towards school violence questionnaire, revised version: CAHV-28. *The European Journal of Psychology Applied to Legal Context*, 12(2), 61-68. <https://doi.org/10.5093/ejpalc2020a8>
- Sánchez, G., Ampudia, A., Jiménez, F., & Amado, B. G. (2017). Contrasting the efficacy of the MMPI-2-RF overreporting scales in the detection of malingering. *European Journal of Psychology Applied to Legal Context*, 9(2), 51-56. <https://doi.org/10.1016/j.ejpal.2017.03.002>
- Sartori, G., Forti, S., Birbaumer, N., & Flor, H. (2003). A brief and unobtrusive instrument to detect simulation and exaggeration in patients with whiplash syndrome. *Neuroscience Letters*, 342(1-2), 53-56. [https://doi.org/10.1016/s0304-3940\(03\)00233-7](https://doi.org/10.1016/s0304-3940(03)00233-7)
- 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>
- Spitzer, W. O., Skovron, M. L., Salmi, L. R., Cassidy, J. D., Duranceau, J., Suissa, S., & Zeiss, E. (1995). Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: Redefining "whiplash" and its management. *Spine*, 20(8), 15-735.
- Sterling, M. (2011). Whiplash-associated disorder: Musculoskeletal pain and related clinical findings. *Journal of Manual & Manipulative Therapy*, 19(4), 194-200. <https://doi.org/10.1179/106698111X13129729551949>
- Sullivan, M. J., Stanish, W., Sullivan, M. W., & Tripp, D. (2002). Differential predictors of pain and disability in patients with whiplash injuries. *Pain Research and Management*, 7(2), 68-74. <https://doi.org/10.1155/2002/176378>
- Sweet, J. J., Heilbronner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Conference Participants (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 35(6), 1053-1106. <https://doi.org/10.1080/13854046.2021.1896036>
- Tuck, N. L., Johnson, M. H., & Bean, D. J. (2019). You'd better believe it: The conceptual and practical challenges of assessing malingering in patients with chronic pain. *The Journal of Pain*, 20(2), 133-145. <https://doi.org/10.1016/j.jpain.2018.07.002>
- Vernon, H., & Mior, S. (1991). The Neck Disability Index: A study of reliability and validity. *Journal of Manipulative and Physiological Therapeutics*, 14, 409-415.
- Walton, D. M., & Elliott, J. M. (2017). An integrated model of chronic whiplash-associated disorder. *Journal of Orthopaedic and Sports Physical Therapy*, 47(7), 462-471. <https://doi.org/10.2519/jospt.2017.7455>
- Ware J. (2000). SF-36 Health Survey update. *Spine*, 25(24), 3130-3139. <https://doi.org/10.1097/00007632-200012150-00008>
- Weinman, J., Petrie, K.J., Moss-Morris, R., & Horne, R. (1996). The Illness Perception Questionnaire: A new method for assessing the cognitive representations of illness. *Psychology and Health*, 11(3), 431-445. <https://doi.org/10.1080/08870449608400270>
- 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>

