



ASOCIACIÓN NACIONAL  
DE  
MÉDICOS FORENSES

# REVISTA ESPAÑOLA DE MEDICINA LEGAL

[www.elsevier.es/mlegal](http://www.elsevier.es/mlegal)



ORIGINAL

## Rating criteria of the causal nexus in Whiplash Associated Disorders: A study of interrater reliability<sup>☆</sup>



Carlos Represas-Vázquez<sup>c</sup>, Esteban Puente-López<sup>b</sup>, David Pina<sup>b,\*</sup>,  
Aurelio Luna-Maldonado<sup>a,b</sup>, José Antonio Ruiz-Hernández<sup>a,b</sup>, Bartolomé Llor-Esteban<sup>a,b</sup>

<sup>a</sup> Murcia's Institute of Biosanitary Research - Instituto Murciano de Investigación Biosanitaria (IMIB-Arrixaca), Murcia, Spain

<sup>b</sup> External Service of Forensic Sciences and Techniques - Servicio Externo de Ciencias y Técnicas Forenses (SECyTEF), Murcia, Spain

<sup>c</sup> Private Medical Expert in the Area of "Body Injury Assessment", A Coruña, Spain

Received 6 February 2020; accepted 21 June 2020

### KEYWORDS

Causal nexus;  
Whiplash associated  
disorders;  
Interrater reliability;  
Psychosocial model;  
Causality criterion

### Abstract

**Introduction:** The diagnosis of Whiplash Associated Disorders is subject to various criteria, some of a subjective nature, which impedes a high percentage of agreement among raters of the causal nexus. It is necessary to develop an evidence-based methodology to maximize the uniformity of this diagnosis. Our objective is to assess the interrater reliability of four experts in legal medicine who establish the causal nexus of traffic accident patients.

**Material and method:** 300 traffic accident patients evaluated by 4 legal medicine experts. They conducted two ratings, one without and one with a study of the biomechanics of the accident. The degree of concordance of the ratings was studied by the intraclass correlation coefficient (ICC) and the concordance index (CI).

**Results:** The results of both the intraclass correlation coefficient and the concordance index of the rating prior to the biomechanical study of the injury yield moderate-high values (ICC = .70; CI = .60). Likewise, the values observed in the rating with the study of biomechanics also indicate moderate-high agreement (ICC = .78, CI = .67).

**Conclusions:** The study of biomechanics slightly improves concordance. However, the aforementioned study does not help to reduce the proportion of doubtful cases and the levels of improvement are not acceptable. There are significant differences between the classifications, suggesting that the proposed valuation methodology is not uniform enough to reach a consensus. © 2020 Asociación Nacional de Médicos Forenses. Published by Elsevier España, S.L.U. All rights reserved.

<sup>☆</sup> The present investigation has not received specific aid from public sector agencies, commercial sector or non-profit entities.

\* Corresponding author.

E-mail address: [david.pina@um.es](mailto:david.pina@um.es) (D. Pina).

**PALABRAS CLAVE**

Nexo causal;  
 Síndrome de latigazo cervical;  
 Concordancia interjueces;  
 Modelo biopsicosocial;  
 Criterios de causalidad

## Criterios para el establecimiento del nexo causal en síndrome de latigazo cervical: un estudio de concordancia interobservadores

**Resumen**

**Introducción:** El diagnóstico del síndrome de latigazo cervical está sujeto a criterios ambiguos, algunos de ellos de naturaleza subjetiva, que impide un alto porcentaje de acuerdo interjueces, y compromete seriamente la replicabilidad de los diagnósticos. El objetivo de este estudio es evaluar la concordancia interjueces de 4 expertos en medicina legal, quienes establecen el nexo causal de los pacientes que han sufrido un accidente de tránsito.

**Material y método:** Trescientos pacientes que sufrieron un accidente de circulación fueron evaluados por 4 expertos en medicina legal. Realizaron 2 valoraciones del nexo causal, una sin estudio de la biomecánica y otra con este estudio. El grado de concordancia de las valoraciones fue calificado con el coeficiente de correlación intraclase (ICC) y el índice de concordancia (IC).

**Resultados:** Los resultados de la valoración previa al estudio biomecánico de la lesión arrojan valores moderados-altos (ICC = 0,70; CI = 0,60). Asimismo, los valores observados en la valoración con el estudio de la biomecánica también indican un acuerdo moderado-alto (ICC = 0,78, CI = 0,67).

**Conclusiones:** El estudio de la biomecánica mejora la concordancia de forma leve. A pesar de ello, el citado estudio no ayuda a reducir la proporción de casos dudosos y la mejora no alcanza niveles aceptables. Se observan diferencias significativas entre las clasificaciones, lo que sugiere que la metodología de valoración propuesta no es lo suficientemente uniforme como para alcanzar consenso.

© 2020 Asociación Nacional de Médicos Forenses. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

**Introduction**

The Whiplash syndrome, understood as the damage caused by a sudden acceleration/deceleration movement that generates a hyperextension in the cervical region, has become the most common consequence of motor vehicle crash and is estimated to affect 83% of the patients involved in this type of claims.<sup>1</sup> This injury not only generates physical damage, but also causes a wide range of biopsychosocial symptoms, considered under the term Whiplash Associated Disorders (WAD). The incidence of Cervical Whiplash Syndromes in North America and Western Europe has increased in the last 30 years, with at least 300 cases per 100,000 inhabitants, which implies an annual cost of approximately 9 trillion dollars in the USA and 10 trillion euros in Europe.<sup>2</sup>

WAD stands out due to its tremendous diagnostic complexity, because it lacks the objective physical indicators evaluated through the currently available diagnostic methods, and it depends completely on the expression of patients' symptoms. Likewise, the habitual symptoms are common to various highly prevalent pathologies and it is not possible to attribute them to the accident without detailed knowledge of the patient's previous state, and such information is very rare.<sup>3</sup>

The assessment of WAD in the medico-legal context focuses on the parameters of the injury to be compensated, and the study of the cause-effect relationship has been relegated to the background, although the evidence points to the important role it has in the evaluation of WAD. When the expert goes to ratify his report, he must answer the question: "What is the cause of the injury?" and a record of

injuries does not solve this issue.<sup>4–6</sup> Due to this negligence, the methodology of the appraisal of causality is outdated and tends to follow an approach based on the use of criteria that look for unspecific exploratory signs and that rarely pass the statistical tests of interrater reliability.<sup>3</sup>

It is necessary to update this methodology, and it is recommended to use a model based on scientific evidence. In this sense, one of the emerging disciplines in Legal Medicine, the biomechanics of injuries, is highlighted.<sup>7</sup> This discipline allows us to know the mechanism of how an injury is produced and to distinguish between injuries that are the product of a direct traumatism, those that are product of inertial damage, or those that may occur as a consequence of both mechanisms.<sup>8</sup> It also provides solid, highly specific indicators of the likelihood of injury and the onset of symptoms.<sup>6</sup> In general, the evidence indicates that a multidisciplinary WAD study that includes a biomechanical reconstruction of the dynamics of the event will allow considering the problem in its entirety and will greatly facilitate the study of causality.<sup>9</sup>

Drawing on the above, the need for a methodology based on scientific evidence, both for the study of causality and the diagnosis of WAD, is evident. With the aim of responding to the aforementioned problems, the main objective of this study is to evaluate the interrater reliability of four experts in legal medicine who establish the causal nexus, with and without the study of the biomechanics, in a sample of patients who have suffered WAD from a traffic accident. Considering this objective, we hypothesize that the rating that includes the study of the biomechanics will be higher than the one without this study.

## Methodology

### Participants

For the elaboration of this study, we used a sample of 300 clinical patients (55.1% males) evaluated in a clinic specializing in the assessment of bodily harm in A Coruña (Spain), from January to September 2017. The average age of the patients was 32.93 (SD = 15.24) with a range of 22–89 years. All cases went to the clinic to undergo an assessment of an injury following a traffic accident. It is unknown whether any of the patients were in a state of litigation, but all of them could obtain a secondary gain in the form of an economic incentive.

The patients were sent to the clinic by their respective insurance companies and they presented all the relevant information, including a biomechanical report of the injury.

Inclusion criteria were being diagnosed of WAD, being of age, signing the informed consent, and not being diagnosed of another disease, anomaly, or medical alteration that could distort the responses of the methodology applied. Patients who did not meet any of these criteria were excluded from the experiment.

### Instruments

#### Causality criteria

The four criteria set out in Section 135 of Law 35/2015 of the scale of traffic accidents<sup>10</sup> were used for the study of the patients' causality:

- Criterion a. Exclusion, which means that there is no other cause that fully justifies the pathology.
- Criterion b. Chronological, which consists of the symptomatology appearing within a medically explainable time. In particular, it is especially relevant for this criterion for the symptoms to manifest within seventy-two hours after the accident or for the injured person to be the subject of medical care within this period.
- Criterion c. Topographic, which means that there is a relationship between the body area affected by the accident and the injury suffered unless a pathogenic explanation justifies otherwise.
- Criterion d. Intensity, consisting of the adequacy between the injury suffered and the mechanism of its production, taking into account the intensity of the accident and the other variables affecting the likelihood of the injury's existence.

As the experts considered some of the criteria to be unspecific, they were operationalized by including the following subcriteria:

- Criterion a. Exclusion.
  - a1. Specific objective studies. The presence of objective evidence to establish without a doubt the existence of an alternative cause explaining the observed symptomatology.<sup>3</sup>
- Criterion c. Topographic.

- c1. Plausibility. The relationship between the affected area and the production mechanism is consistent and biologically logical (adapted from the scientific plausibility criterion of Bradford-Hill<sup>11</sup>).
- c2. The specificity of symptoms and signs. The relationship between the observed and/or expressed symptoms and signs and the production mechanism is consistent and biologically logical (adapted from the specificity criterion of Bradford-Hill<sup>11</sup>).
- Criterion d. Intensity.
  - d1. Biomechanical efficacy. The dynamics of the injured person coincide with the injury's production mechanism.<sup>3</sup>
  - d2. Biomechanical compatibility. The physical parameters of the accident are compatible with the population thresholds defined for the injury suffered.<sup>3</sup>

No subcriterion was included for the temporal criterion because it was considered sufficiently specific. Each criterion, as well as its respective subcriteria, was rated as "Yes", "No", or "Doubtful". To classify the topographic and intensity criteria as "Yes", the two subcriteria had to be positively assessed. If one of them was negative, it was classified as "No", and if one was "Doubtful" (and the other one yes), it was classified as such.

The causal nexus was rated as "Yes" when at least three criteria were positively assessed, being the fourth one "doubtful". It was rated as "doubtful" when at least 2 of these criteria were rated as "doubtful", but there was no criterion rated as "No". Finally, it was rated as "not present" when at least one criteria were rated as "No".

#### Abbreviated injury scale (AIS)<sup>12</sup>

An anatomical coding system used to classify and describe the severity of injuries, in order to measure an immediate life threat. A 7-digit code is given where the first number is the body region, the second is the type of anatomical structure, the third and the fourth are the specific anatomical structures, the fifth and sixth are the levels (fracture, rupture, laceration, etc.), and the seventh is the severity. For the severity code, a classification system ranging from 1 (*low severity injury*) to 6 (*fatal injury*) is used, also including the value 9 for injuries about which limited information is available. For this study we used the severity of the injuries.

#### Spanish scale of Injuries in Traffic Accidents of Law 35/2015

This is a single article that reforms the system for the assessment of injuries produced in people in road accidents in Spain. The main objective is to make the compensation more consistent with the victim's situation, and it incorporates personal, work, family, and professional variables in the calculation. The scale covers basic personal damage (non-impaired and impaired days), particular personal harm (impediment or limitation caused by injuries or their treatment in anatomy or personal development), and property damage (emerging damage and loss of profit). This study, in addition to the causality criteria of Article 135 detailed above, establishes the non-impaired and impaired days and particular personal harm, divided into esthetic

injury, ranged in groups from 1–6 (light) to 41–50 (very important), and functional alteration, where a score from 1 to 100 is given depending on the damage. These scores are later used to calculate the financial compensation, adjusted for other variables such as age.

Analogic Visual Scale (AVS). An ad-hoc analogic visual scale was used to assess patients' pain, with scores ranging from 1 (*no pain*) to 10 (*extreme pain*). This scale incorporated a color guide for easy comprehension.

## Procedure and design

An analytical-descriptive cross-sectional design was used for the elaboration of this work. For the reliability analysis, four experts in legal and forensic medicine, specialized in the assessment of bodily injury and with more than 30 years of experience in the field, participated as examiners. The experts were blind to the objectives of the study, which was presented as an investigation of the general characteristics of WAD and the causality criteria of the Article 135.

One researcher was in charge of contacting the patients at the clinic where they went to be evaluated. The procedure that would be followed was explained to these patients and, after obtaining their verbal agreement, they signed an informed consent and resolved any possible doubts. The patient was then referred to the first examiner, who performed a complete clinical assessment, analyzed the clinical history, and assessed the causal nexus with the above-mentioned criteria (Rating 1). All information collected during the evaluation, except for Rating 1, was provided to the other three raters separately for them to apply the causality criteria and issue their rating. After the first rating was made, they were given the study of biomechanics and asked to reassess the causal nexus, using this new information (Rating 2).

The biomechanical report of the injury was 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 ( $\Delta v$ ) and the mean acceleration ( $a$ ).<sup>3</sup> These values were interpreted by the evaluators using the thresholds proposed by Represas, Muñoz, and Luna.<sup>6</sup>

## Data analysis

Descriptive statistics were calculated for the socio-demographic and medico-legal variables collected, including the mean and the standard deviation of the continuous variables and the frequencies of the categorical ones. The differences between the two trials of each of the four experts were studied with the chi-squared test. For the analysis of interrater reliability, the intraclass correlation coefficient (ICC, Model 2.1) and the 95% confidence interval were calculated for the two ratings. The results of the ICC were with 1.0 indicating absolute agreement, .80–.99 excellent agreement, .60–.79 suitable agreement, .20–.59 moderate agreement, and <.19 minimum agreement.

Alternatively, the degree of Interrater agreement was also valued using the concordance index [ $CI = \text{agreements} / (\text{agreements} + \text{disagreements})$ ], which

**Table 1** Mean and standard deviation of the Medico-legal characteristics.

	M	SD
Days since the accident <sup>a</sup>	47.8	9.52
Pain severity	5.87	1.21
Abbreviated injury scale (AIS)	1.26	.60
Hospitalization days	4.35	16.03
Impaired days	43.85	67.48
Non-impaired days	87.86	77.09
Functional alteration points	4.22	7.47
Esthetic injury points	1.09	3.03
Delta v ( $\Delta v$ ) <sup>b</sup>	8.75	5.41
Mean acceleration ( $a$ ) <sup>c</sup>	7.11	4.21

<sup>a</sup> In the moment of the evaluation.

<sup>b</sup> Measured in km/h.

<sup>c</sup> Measured in g.

is more restrictive than the ICC.<sup>13,14</sup> The decision to use an additional assessment criterion is based on the high responsibility demanded in the forensic context, in which minimizing the margin of error of the tools used is an obligation.<sup>13</sup>

The statistical analysis was carried out using the SPSS program, version 24.

## Results

Concerning the medico-legal characteristics (Tables 1 and 2), the patients were evaluated in a mean of 47.8 days (SD=9.52) since the accident. 78.2% of them had Grade I WAD, whereas 21.8% had Grade II. 100% of them reported suffering pain in the cervical area, with an average severity of 5.87 (SD=1.21) in the AVS. 41% had mild limitations in the range of cervical movements, 15% moderate, 4% severe limitations, and the remaining 40% had no limitations. Similarly, mild muscular contractions were observed in 47.4% of the patients, moderate in 13.3%, and severe in 3.3%. The remaining 36% had no muscular contractions at the time of the evaluation.

Of the 300 patients included, 75.3% used a public assistance model, and 24.7% used a private or mixed model. 28% of the patients had to be transferred by ambulance, 4% suffered loss of consciousness and 16.3% were involved in a complex collision. On the other hand, 16.7% of these patients had to be hospitalized for more than 24 h, with an average hospitalization time of 4.35 days (SD = 16.03), while the rest was discharged in less than the aforementioned time ( $M = 12.24$  h,  $SD = 4.35$ ). Of this 16.7% of patients, 12.5% were hospitalized for minor muscle injuries in the arms, and the remaining 4.2% because they were concerned about severe dizziness and vertigo. 10.3% was classified as an occupational accident.

Concerning the variables of the Spanish scale of injuries in traffic accidents, the patients obtained a mean of 43.85 (SD = 67.48) of impaired days, 87.86 (SD = 7.09) non-impaired days, 4.22 (SD = 7.47) points in functional alteration and 1.09 (SD = 3.03) points in esthetic injury. All of the functional alteration points were given because of the limitations in the range of cervical movements described.

**Table 2** Frequencies of the medico-legal characteristics.

	Frequency	%
<i>Grade of WAD</i>		
Grade I	235	78.2
Grade II	65	21.8
<i>Range of movement limitation</i>		
None	120	40
Mild	123	41
Moderate	45	15
Severe	12	4
<i>Muscular contraction</i>		
None	108	36
Mild	142	47.4
Moderate	40	13.3
Severe	10	3.3
<i>Loss of consciousness</i>		
Yes	12	4
No	288	96
<i>Ambulance transfer</i>		
Yes	84	28
No	216	72
<i>Admission to UCI</i>		
Yes	2	.7
No	298	99.3
<i>Hospitalization &gt;24 h</i>		
Yes	50	16.7
No	250	83.3
<i>Complex collision</i>		
Yes	49	16.3
No	251	83.7
<i>Safety belt</i>		
Yes	268	89.3
No	22	7.3
Not applicable	10	3.3
<i>Assistential model</i>		
Public	226	75.3
Mixed/private	74	24.7
<i>Rehabilitation</i>		
Yes	263	88.7
Not needed	34	11.3
<i>Occupational accident</i>		
Yes	31	10.3
No	144	48
Not applicable	125	41.7
<i>Sick leave</i>		
Yes	124	41.3
No	51	17
Not applicable	125	41.7
<i>Patients with known previous state</i>		
Yes	38	12.7
No	262	87.3

**Table 2 (Continued)**

	Frequency%	
<i>Previous psychiatric diagnosis (of the known previous state patients)</i>		
Yes	8	21
No	30	79

88.7% of the patients had received rehabilitation at the time of evaluation, while the remaining 11.3% had not needed rehabilitation. On another hand, information about the physical and psychological state prior to the accident was only available in 12.7% of the patients. 21% of these patients had a pre-accident psychiatric diagnosis, in all cases, depressive-type disorders.

Finally, the results of the biomechanical report indicated mean Delta v ( $\Delta v$ ) values of 8.75 km/h (SD = 5.41) and mean acceleration (a) of 7.11 g (SD = 4.21).

With regard to the classification carried out by each of the evaluators (Table 3), in general terms, it was considered that there was causal link in approximately half of the patients, both in the analysis without a study of the biomechanics (Evaluator 1: 56.7%; Evaluator 2: 84%; Evaluator 3: 69.7%; Evaluator 4: 49%) and with that study (Evaluator 1: 52.3%; Evaluator 2: 32%; Evaluator 3: 60.3%; Evaluator 4: 61.3%). Similarly, the analysis indicates significant differences between these analyses in the four evaluators (Evaluator 1:  $X^2(4) = 525.23$ ,  $p < .001$ ; Evaluator 2:  $X^2(4) = 12.68$ ,  $p = .013$ ; Evaluator 3:  $X^2(4) = 304.55$ ,  $p < .001$ ; Evaluator 4:  $X^2(4) = 180.68$ ,  $p < .001$ ).

The concordance indexes (CI) and the reliability coefficients of the ratings (ICC) with their corresponding 95% confidence interval and the analysis of variance (ANOVA) can be found in Table 4. The results of both the intraclass correlation coefficient and the concordance index of the rating prior to the biomechanical study of the injury yielded adequate values (ICC = .70, CI = .60). Likewise, the values observed in the rating with the study of the biomechanics also indicated adequate agreement (ICC = .78, CI = .67). With regard to criterion-by-criterion classification, in the analysis without the study of biomechanical, criteria a (exclusion) and b (temporal) obtained high values (ICC = .96, CI = .81; ICC = 1, CI = 1, respectively), criterion c (topographic) obtained moderate values (ICC = .71, CI = .57), and criterion d (intensity) obtained low values (ICC = .14, CI = .11). After the biomechanical study, the first 3 criteria remained the same, but criterion d reached moderate-low values (ICC = .46, CI = .31).

## Discussion

The objective of this research is to evaluate the degree of agreement of four experts in Legal Medicine in the establishment of the causal nexus with and without the study of biomechanics. Our results indicate that the degree of concordance between the ratings of the four experts was moderate-high both for Rating 1, conducted without the study of the biomechanics of the injury, and for Rating 2, performed after that study. The Rating 2 reached higher interrater reliability, so our research hypothesis is

**Table 3** Patient classification of each of the evaluators.

	Yes		No		Doubtful		$\chi^2$	df	p
	n	%	n	%	n	%			
<b>WOB</b>									
Evaluator 1	170	56.7	62	20.7	68	22.7			
Evaluator 2	252	84	5	1.7	43	14.3			
Evaluator 3	209	69.7	45	15	46	15.3			
Evaluator 4	147	49	102	34	51	17			
<b>WB</b>									
Evaluator 1	157	52.3	67	22.3	76	25.3			
Evaluator 2	96	32	162	54	42	14			
Evaluator 3	181	60.3	70	23.3	49	16.3			
Evaluator 4	184	61.3	65	21.7	51	17.0			
<b>Comparison WOB and WB</b>									
Evaluator 1							525.23	4	<.001
Evaluator 2							12.68	4	.013
Evaluator 3							304.55	4	<.001
Evaluator 4							180.68	4	<.001

Abbreviations: WOB, without biomechanical study; WB, with biomechanical study;  $\chi^2$ , Chi-squared test for the differences between each expert's ratings; df, degrees of freedom; p, significance level.

**Table 4** Intraclass correlation coefficient (ICC), concordance index (CI) and ANOVA.

	n	ICC (2,1)	CI	95% CI Upper lower	F	p
Rating 1 WOB	300	.70	.60	.67 .74	22.96	<.001
Criterion a (a1)	300	.96	.81	.90 .98	8.19	<.001
Criterion b	300	1	1	1 1	*	*
Criterion c	300	.71	.57	.67 .79	4.52	<.001
Criterion c1	300	.85	.64	.82 .88	6.93	<.001
Criterion c2	300	.69	.45	.61 .73	4.20	<.001
Criterion d	300	.14	.11	.11 .19	1.17	<.001
Criterion d1	300	.21	.14	.17 .24	1.41	<.001
Criterion d2	300	.039	.010	-.01 .02	1.04	.33
Rating 2 WB	300	.78	.67	.74 .82	13.20	<.001
Criterion a (a1)	300	.96	.81	.90 .98	8.19	<.001
Criterion b	300	1	1	1 1	*	*
Criterion c	300	.71	.57	.67 .79	4.52	<.001
Criterion c1	300	.85	.64	.82 .88	6.93	<.001
Criterion c2	300	.69	.45	.61 .73	4.20	<.001
Criterion d	300	.46	.31	.40 .51	2.37	<.001
Criterion d1	300	.40	.28	.37 .46	2.12	<.001
Criterion d2	300	.67	.55	.63 .74	4.19	<.001

Abbreviations: WOB, without biomechanical study; WB, with biomechanical study; ICC, intraclass correlation coefficient; CI\*, concordance index; 95% CI, 95% confidence interval of the ICC; F, Fisher's F; p, significance level.

confirmed, and the study of biomechanics generated an increase, slight in our case, of general interjudge concordance.

Specifically, our findings indicate that the inclusion of the biomechanical study did not affect the classification given by the experts in the first three criteria, but it did have an impact on the intensity criterion, significantly improving the values of both the general criterion and of the two subcriteria, in particular, Criterion d2 (Biomechanical compatibility). Nonetheless, it is important to note that the

biomechanical study did not help to reduce the proportion of doubtful cases, as there are hardly any changes between the two ratings, but it generally helped the evaluators to adjust their rate of uncaused cases.

In this regard, it would be expected that, by operationalizing the criterion and providing experts with thresholds for assessing it, higher levels of concordance than those obtained would have been achieved, and the number of doubtful cases would have decreased. We think that an explanation for these results may be that the proposed

operationalization was not objective enough, and thresholds of Represas, Muñoz, and Luna<sup>6</sup> may not reach the level of specificity necessary to achieve a uniform interpretation. Several variables may influence the possibility of injury, such as risk factors, security, etc. This can lead to some cases—which are very close to the threshold points and which are affected by those variables—generating different opinions because the expert must make a subjective assessment of the possible influence of such variables.

Despite this, we think that it is necessary to emphasize that the intensity criterion is subjective, and, for future publications, it would be advisable to explore new ways of operationalizing it. In this regard, we point out that the thresholds used were a theoretical proposal that synthesized the available evidence for this issue and, although they have not allowed us to achieve high levels of concordance, they have produced some improvement. Therefore, we consider that it may be useful for the assessment of this criterion and future research, it would be of interest to develop a broader, more concrete classification with more specific thresholds.

Concerning the proportion of excluded cases, we consider the observed values (between 20 and 30%) are remarkable, as the evaluation context may influence them. For future research, we consider that it may be of interest to study possible differences between the classifications made by expert appraisers and by experts of the institutions of legal medicine.

The results obtained in the other three criteria (exclusion, temporal, and topographic) indicate a high convergence, which does not change when introducing the biomechanical study, suggesting that they are independent of other possible criteria and useful in themselves. In our study, the temporal criterion did not generate any disagreement, and the value obtained in the exclusion criterion indicated hardly any disagreement, but the topographic criterion obtained moderate-high values that are of interest. The topographic criterion was operationalized as two subcriteria because the evaluators considered that it would be of interest to assess both the body area affected by the injury suffered (criterion c1) and the coherence of the observed and/or expressed symptoms and/or signs (criterion c2). If a classic topographic assessment was made through the application of criterion c1, high concordance values were obtained because, in the vast majority of cases, the location of the injury was clear, and there was only a lack of consensus in those cases in which the pain radiated unusually. But, by including criterion c2, the level of agreement decreases. We believe that an explanation for these results is the broad symptomatic heterogeneity of the condition studied and the difficulty to objectify part of the symptoms. It is possible that, by not assessing psychosocial variables following a standardized procedure, the experts interpreted the consistency of some symptoms or signs based solely on their own experience. For example, one of the professionals might have considered that a possible anxious-depressive syndrome—which may have originated in the patient's interpersonal characteristics and not in the magnitude of the accident—was inconsistent with the intensity of the collision and did not need a more extensive analysis, whereas another expert might consider it to be a valid symptom after a more detailed examination. For future research, we think that it may be of interest to analyze the consistency

of the symptomatology expressed by patients (criterion c2), including it in the topographic criterion or as an independent criterion, but it would be advisable to provide experts with a standardized assessment methodology that includes a list of all the variables that WAD may present, including psychosocial ones.

On the other hand, we consider of interest the differences in the classifications made by the experts. It can be observed that, despite achieving moderate-high concordance values, these classifications differ from each other, especially those of evaluator 2, who, after the biomechanical study, significantly changed his/her classification and dramatically increased the classification of uncaused cases. We think that the observed differences could indicate, as has already been pointed out above, that Criterion C2 and Criterion d—those that presented less concordance—are still subjective, and the proposed operationalization was not entirely effective for the evaluators to reach a consensus. It should be recalled that, in general terms, the methodology applied for expert assessment is eminently personal, and the results obtained are rarely subject to a critical review. In this situation, if the professionals do not have a standardized procedure, the assessment will be based on their own professional experience, which may differ significantly among experts.

Finally, to assess the results obtained, in addition to a statistical approach, the usefulness and sufficiency of the criteria used should be considered. The concordance values obtained would indicate that the a-c criteria (considering Criterion c1), are properly operationalized and are specific, so they may be useful in establishing the causal link. Despite this, we think their sufficiency should be valued with caution. We consider that the causality criteria studied are deterministic. In most cases, traffic accidents may be influenced by external variables that cause variations in the outcome. While it is possible to predict the evolution and risk of complications, these variables will significantly expand the possibilities, moving away from this deterministic nature.<sup>15</sup> Taking as an example the temporal criterion, it is not possible to ensure that no case is erroneously excluded due to the influence of such variables (pain resistance, availability of care facilities, personal and family circumstances, etc.). Therefore, the criteria used may not be sufficient, and it would be advisable to expand the list with scientifically relevant factors. In this regard, it would be of interest to consider for this extension proposals such as that of Represas,<sup>3</sup> who includes three criteria with their respective subcriteria, a biomechanical one, an epidemiological one, and one of the patient's prior state.

Considering the particularities of the WAD could be another approach to develop new criteria. As mentioned in the introduction, WAD is a condition with high diagnostic complexity due to the absence of reliable indicators.<sup>3</sup> The procedure usually applied for evaluating the injury focuses on seeking structural or anatomical divergences through diagnostic practices that fail to achieve adequate efficacy.<sup>16,17</sup> The evidence suggests that WAD should be evaluated following a biopsychosocial model.<sup>18,19</sup> This approach would allow the evaluator to observe the interaction between medical, biomechanical, social, and psychological factors and would offer a much broader view of the injury that would be useful in the study of causality.<sup>20</sup> Future

research should complement the assessment of bodily harm and the study of the biomechanics of the injury with an analysis of the psychosocial variables that affect WAD, such as the perception and expectations of recovery from the disease, the presence of anxious-depressive symptomatology, the perception of disability, or the catastrophizing of pain.<sup>21–23</sup>

We think that the updating of the causality study methodology is an issue that should be addressed. As Luna<sup>15</sup> states, universal or general criteria should not be used for the study of medico-legal causality. With a view to future research, the particularities of the condition should be considered, and specific criteria should be included, extending the list to as many relevant factors as necessary.

The results presented in the preceding paragraph must be interpreted according to the following limitations: due to the conditions and the experimental design, it was not possible to perform probabilistic sampling. Also, the sample collected at one of the clinics was incidental. This could be biasing the results due to provenance. Finally, while it is a difficult issue to address, it would be interesting to propose designs that allow evaluators to personally interview all the patients, thus avoiding some possible biases.

As a conclusion, the results obtained indicate that the inclusion of the biomechanical study slightly improves overall concordance. Specifically, it does not affect the assessment of the a-c criteria, but it significantly improves the concordance in Criterion d. Despite this, the above-mentioned study does not help experts to reduce the proportion of doubtful cases, and the improvement observed in Criterion d does not reach acceptable levels. Similarly, large differences are observed in the classifications made by the evaluators, suggesting that the proposed rating methodology is not uniform enough to reach a consensus. For future research, it would be of interest to try to achieve a higher level of specificity in the operationalization of the criteria, especially c2 and d, as well as to assess the study of biomechanics through more specific thresholds and classification categories. Similarly, it may be of interest to include additional objective criteria and to consider the possible influence of external variables.

## Funding

None declared.

## Conflicts of interest

The authors declare no conflict of interest.

## Acknowledgment

The manuscript submitted does not contain information about medical device(s)/drug(s).

## References

1. Yadla S, Ratliff JK, Harrop JS. Whiplash: diagnosis, treatment, and associated injuries. *Curr Rev Musculoskelet Med*. 2008;1:65–8, <http://dx.doi.org/10.1007/s12178-007-9008-x>.
2. Pink J, Petrou S, Williamson E, Williams M, Lamb SE. Economic and health-related quality of life outcomes of whiplash associated disorders. *Spine*. 2016;41:1378–86, <http://dx.doi.org/10.1097/BRS.0000000000001512>.
3. Represas C. Nexo de causalidad en accidentes de tráfico [Causal nexus in traffic accidents]. *Cienc Forense*. 2017;14:61–76.
4. Ferrari R, Klar L. On matters of causation in personal injury cases: considerations in forensic examination. *Eur J Rheumatol*. 2014;1:150–5, <http://dx.doi.org/10.5152/eurjrheumatol.2014.140061>.
5. Freeman MD, Centeno CJ, Kohles SS. A systematic approach to clinical determinations of causation in symptomatic spinal disk injury following motor vehicle crash trauma. *PM R*. 2009;1:951–6, <http://dx.doi.org/10.1016/j.pmrj.2009.07.009>.
6. Represas C, Muñoz JI, Luna A. 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]. *Rev Esp Med Legal*. 2015;42:45–88, <http://dx.doi.org/10.1016/j.reml.2015.10.001>.
7. Cheng HB, Yang KH, Wang ZG. Biomechanics of whiplash injury. *Chin J Traumatol*. 2009;12:305–14, <http://dx.doi.org/10.3760/cma.j.issn.1008-1275.2009.05.011>.
8. Arregui-Dalmases C, Teijeira R, Rebollo-Soira MC, Kerrigan JR, Crandall JR. La biomecánica del impacto: una herramienta para la medicina legal y forense en la investigación del accidente de tráfico [The biomechanics of impact: a tool for legal and forensic medicine in the investigation of traffic accidents]. *Rev Esp Med Legal*. 2011;37:87–130, [http://dx.doi.org/10.1016/S0377-4732\(11\)70071-2](http://dx.doi.org/10.1016/S0377-4732(11)70071-2).
9. Ferrara SD, Ananian V, Baccino E, Banczerowski P, Bordignon D, Boscolo-Berto R, et al. Whiplash-Associated Disorders: clinical and medico-legal guidelines on the methods of ascertainment. *Int J Legal Med*. 2016;130:13–22, <http://dx.doi.org/10.1007/s00414-015-1243-x>.
10. Ley 35/2015, de 22 de septiembre, de reforma del sistema para la valoración de los daños y perjuicios causados a las personas en accidentes de circulación. *Boletín Oficial del Estado*, 23 de septiembre de 2015, núm. 228. p. 84473–979.
11. Hill AB. The environment and disease: association or causation? *Proc Royal Soc Med*. 1965;58:295–300.
12. Gennarelli T, Wodzin E. The abbreviated injury scale 2005, update 2008. *Des Plaines: American Association for Automotive Medicine (AAAM)*; 2008.
13. Arce R, Fariña F. Evaluación psicológico-forense de la credibilidad y daño psíquico mediante el Sistema de Evaluación Global. In: Rivas P, Barrios GL, editors. *Violencia de género: Perspectiva multidisciplinaria y práctica forense [Gender violence: multidisciplinary perspective and forensic practice]*. Navarra: Thomson Aranzadi; 2015. p. 411–41.
14. Tversky A. Features of similarity. *Psychol Rev*. 1977;84:327–52, <http://dx.doi.org/10.1037/0033-4.327-295X84>.
15. Luna A. Nexo de causalidad en patología forense [Causality nexus in forensic pathology]. *Cienc Forense*. 2017;14:49–60.
16. Valera-Garrido F, Martínez-Rodríguez A, Medina-Mirapeix D, Meseguer-Henarejos AB, Milán-Robles F, Campillo-Pomata JL. El modelo biopsicosocial en los síndromes de dolor vertebral: Implicaciones para la protocolización [The biopsychosocial model in vertebral pain syndromes: Implications for protocolization]. *Rev Iberoam Fisioter Kinesiol*. 2001;4:81–7.
17. Waeyaert P, Jansen D, Bastiaansen M, Scafoglieri A, Buyl R, Schmitt M, et al. Three-dimensional cervical movement characteristics in healthy subjects and subgroups of chronic neck pain patients based on their pain location. *Spine*. 2016;41:908–14, <http://dx.doi.org/10.1097/BRS.0000000000001500>.
18. Phillips L, Carroll L, Cassidy D, Côté P. Whiplash-associated disorders: who gets depressed? Who stays depressed?

- Eur Spine J. 2010;19:945–56, <http://dx.doi.org/10.1007/s00586-010-1276-2>.
19. Walton DM, Elliott JM. An integrated model of chronic whiplash-associated disorder. *J Orthop Sports Phys Ther.* 2017;47:462–71, <http://dx.doi.org/10.2519/jospt.2017.7455>.
  20. Turk DC, Robinson JP, Duckworth MP. The impact of psychosocial and contextual factors on individuals who sustain whiplash-associated disorders in motor vehicle collisions. *Psychol Inj Law.* 2018;11:218–32, <http://dx.doi.org/10.1007/s12207-018-9317-y>.
  21. Carroll L, Holm L, Ferrari R, Ozegovic D, Cassidy D. Recovery in whiplash-associated disorders: do you get what you expect? *J Rheumatol.* 2009;36:1063–70, <http://dx.doi.org/10.3899/jrheum.080680>.
  22. Falla D, Peolsson A, Peterson G, Ludvigsson ML, Soldini E, Schneebeli A, et al. Perceived pain extent is associated with disability, depression and self-efficacy in individuals with whiplash-associated disorders. *Eur J Pain.* 2016;20:1490–501, <http://dx.doi.org/10.1002/ejp.873>.
  23. Holm LW, Carroll LJ, Cassidy D, Skillgate E, Ahlbom A. Expectations for recovery important in the prognosis of whiplash injuries. *PLoS Med.* 2008;5:e105, <http://dx.doi.org/10.1371/journal.pmed.0050105>.