



UNIVERSIDAD DE MURCIA

ESCUELA INTERNACIONAL DE DOCTORADO

**Detección de simulación de síntomas
asociados al esguince cervical**

**D. Esteban Puente López
2021**



UNIVERSIDAD DE MURCIA
ESCUELA INTERNACIONAL DE DOCTORADO

Línea de investigación: Ciencias Forenses

TESIS DOCTORAL

Detección de simulación de síntomas asociados al
esguince cervical

AUTOR:

D. Esteban Puentes López

DIRECTOR:

Dr. D. José Antonio Ruiz Hernández

CODIRECTOR:

Dr. D. Bartolomé Llor Esteban

MURCIA, 2021

“Money, so they say, is the root of all evil today”

— Roger Waters

Agradecimientos

Cuando llega este momento uno puede agradecer, sobretodo, haber llegado al momento en cuestión con vida, que no es poco. No obstante, sería quedarnos cortos ya que si estoy escribiendo este apartado es gracias a las personas que, en este duro viaje, han permanecido a mi lado.

En primer lugar y sin duda alguna, agradecer a mis padres Joaquín y Esther, mis tios Mercedes y Paco y mis abuelos Esteban y Mercedes. El apoyo de los familiares y seres queridos es el más puro e incondicional, no importa que les hables de la importancia de la razón de verosimilitud encadenada en en la evaluación de baterías psicométricas o de las virtudes del paparajote murciano, siempre te atenderán con ilusión, te animarán a seguir con lo que haces y te ayudarán, sin pensárselo, a levantarte de los tropiezos.

En segundo lugar, no puede estar otro que David Pina, que comenzó como un conocido de la facultad y ha acabado como un hermano. Es una de las personas mas talentosas que conozco, y una de las pocas a las que le otorgaría el título de “genio”. David es pura inventiva, capacidad de resolución, resiliencia, talento, sentido del humor nefasto y fidelidad. Apostó por mi sin dudarlo en un momento en el lo necesitaba enormemente y puedo garantizar que esta tesis doctoral ha salido adelante gracias a su esfuerzo e interés. Que los audios científicos de wasap a las 8 de la mañana nos duren muchos años más, compañero.

En tercer lugar, darles las gracias a mis directores, José Antonio y Bartolomé. José y Bartolo son ese tipo de director que procesa un amor serbo-bosnio, del duro, del que curte y endurece. Te ayudarán en todo lo que esté en su mano, pero también querrán que des tu 200%, y si es el 300%, mejor. No cabe duda que son dos personas muy exigentes, y trabajar con ellos no siempre ha sido fácil, pero son el golpe de realidad que un alumno

falto de motivación necesita, y esforzarme día tras día en cumplir sus expectativas me ha llevado a este momento. Gracias a ellos he pasado de ser una persona con nulo interés por la disciplina a una que le apasiona y, si algún día puedo afirmar que soy un buen profesional, se lo deberé a ellos.

En cuarto lugar, agradecer a Aurelio Luna su inestimable ayuda. Aurelio, al igual que David, creyó en mí y me ayudó de forma altruista, apostando por mí sin apenas conocerme, y poniendo a mi disposición todo lo que estuviera en su mano (que no es poco) con una amplia sonrisa. Aurelio es una fuente de inspiración, de buenos momentos y de verdad que sólo deseo haberlo conocido unos cuantos años antes. Reiteraré en este párrafo una frase que ya le he dicho en muchas ocasiones “Aurelio, de mayor quiero ser como tú”.

En quinto lugar, agradecer a mis dos mejores amigos, Juan y a Javi, su apoyo y paciencia. Los amigos no son los que te dicen palabras bonitas, sino esos que te dicen que “jamás ganarás tanto como un odontólogo”, “pues haber estudiado menos” o “vaya psicólogo que no sabes leer mentes”. Pocas personas hay en mi vida que sepan sacarme una sonrisa como ellos, y me apoyen pase lo que pase y en cualquier cosa que me proponga.

En sexto lugar a Débora. Nuestros caminos se han separado, pero no podría haber hecho esto sin tu ayuda, apoyo incondicional y sonrisa. Gracias, por todo.

Finalmente, me gustaría dar las gracias a Carmen Godoy, Abelardo, Simina, María Pagán, Luis Aguerrevere, Thomas Merten, Richard Frederick, y todas aquellas personas que me han ayudado y animado a lo largo de la elaboración de esta tesis doctoral.

Índice

Marco teórico

1. La simulación	Pág.1
1.1. Conceptos básicos	Pág.1
1.2. Estrategias de detección	Pág.11
1.3. Herramientas e instrumentos para la detección de la simulación	Pág.20
1.3.1. Entrevistas para la detección de la simulación/fingimiento	Pág.24
1.3.2. Instrumentos psicométricos para la evaluación de la simulación/fingimiento	Pág.28
2. Métodos de investigación en el campo de la simulación	Pág.41
2.1. Diseños de investigación	Pág.41
2.1.1. Diseño de simulación	Pág.41
2.1.2. Diseño de grupo conocido	Pág.48
2.1.3. Diseño autosuficiente y diseño de criterio parcial	Pág.50
2.1.4. Diseño de prevalencia diferencial	Pág.51
2.2. Análisis de datos en el campo de la simulación	Pág.52
2.2.1. Comparación entre grupos y estimadores del tamaño del efecto	Pág.52
2.2.2. Sensibilidad y especificidad	Pág.53
2.2.3. Poder predictivo positivo y negativo	Pág.55
2.2.4. Odds ratio diagnósticos	Pág.56
2.2.5. Razón de verosimilitud	Pág.57
2.2.6. Razón de verosimilitud encadenada	Pág.57
2.2.7. Curva Característica Operativa del Receptor	Pág.59
3. La simulación en el Síndrome de Latigazo Cervical	Pág.61
3.1. ¿Qué condiciones son las más simuladas?	Pág.61
3.2. El esguince cervical y el Síndrome de Latigazo Cervical	Pág.65
3.3. ¿Por qué se cronifica el Síndrome de Latigazo cervical?	Pág.69
3.4. Diferencias en la prevalencia/incidencia del Síndrome de Latigazo Cervical	Pág.84
3.5. ¿Por qué se simula el Síndrome de Latigazo Cervical?	Pág.86
3.6. ¿Cómo se estudia la simulación en el Síndrome de Latigazo Cervical?	Pág.88

Parte empírica

4. Objetivos generales y específicos	Pág.93
5. Bloque 1 - Estudio 1. Criterios para el establecimiento del nexo causal en SLC: Un estudio de concordancia inter-observadores	Pág.97
5.1. Metodología	Pág.98
5.1.1. Participantes	Pág.98
5.1.2. Instrumentos	Pág.99
5.1.3. Procedimiento	Pág.102
5.1.4. Análisis de datos	Pág.103
5.2. Resultados	Pág.104
5.3. Discusión	Pág.110
6. Bloque 1 - Estudio 2. Valoración de la simulación en SLC: Una revisión sistemática	Pág.115
6.1. Metodología	Pág.116
6.1.1. Criterios de selección	Pág.116
6.1.2. Estrategia de búsqueda	Pág.117
6.1.3. Selección de los estudios	Pág.118
6.1.4. Evaluación del riesgo de sesgo	Pág.119
6.1.5. Tabulación y análisis de datos	Pág.120
6.2. Resultados	Pág.121
6.2.1. Dimensión biológica	Pág.124
6.2.2. Dimensión psicológica	Pág.127
6.2.3. Dimensión social	Pág.129
6.2.4. Elementos de apoyo a la evaluación	Pág.129
6.3. Discusión	Pág.138
7. Bloque 2 – Estudio 3. Perfiles biopsicosociales de percepción del dolor de pacientes con SLC	Pág.146
7.1. Metodología	Pág.147
7.1.1. Participantes	Pág.147
7.1.2. Procedimiento	Pág.147
7.1.3. Instrumentos y variables	Pág.152
7.1.4. Análisis de datos	Pág.155

7.2. Resultados	Pág.156
7.2.1. Clasificación de los grupos y características de la muestra	Pág.156
7.2.2. Análisis cluster	Pág.158
7.2.3. Análisis discriminante (validación cruzada)	Pág.162
7.3. Discusión	Pág.162
8. Bloque 3 - Estudio 4. Capacidad discriminativa de un protocolo biopsicosocial para la detección de simulación en SLC.	Pág.168
8.1. Metodología	Pág.168
8.1.1. Participantes	Pág.168
8.1.2. Instrumentos	Pág.169
8.1.3. Procedimiento	Pág.170
8.1.4. Análisis de datos	Pág.171
8.2. Resultados	Pág.173
8.2.1. Característica de los grupos	Pág.173
8.2.2. Puntuaciones de las escalas y comparación entre grupos	Pág.173
8.2.3. Razón de verosimilitud encadenada	Pág.181
8.3. Discusión	Pág.184
9. Bloque 3 - Estudio 5. Checklist de inconsistencias e indicadores de simulación en SLC	Pág.189
9.1. Metodología	Pág.191
9.1.1. Participantes	Pág.191
9.1.2. Instrumentos	Pág.192
9.1.3. Procedimiento	Pág.193
9.1.4. Análisis de datos	Pág.194
9.2. Resultados	Pág.195
9.3. Discusión	Pág.199
10. Bloque 3 - Estudio 6. Capacidad discriminativa del Inventario Estructurado de Simulación de Síntomas (SIMS)	Pág.206
10.1. Metodología	Pág.207
10.1.1. Participantes	Pág.207
10.1.2. Instrumentos	Pág.209

10.1.3. Procedimiento	Pág.209
10.1.4. Análisis de datos	Pág.211
10.2. Resultados	Pág.212
10.2.1. Clasificación de los grupos y variables médico-legales del grupo de pacientes clínicos y sobredimensionados	Pág.212
10.2.2. Puntuaciones del SIMS y comparación entre grupos	Pág.214
10.2.3. Precisión de la clasificación	Pág.216
10.3. Discusión	Pág.221
11. Discusión general	Pág.225
12. Limitaciones	Pág.241
13. Conclusiones	Pág.243
14. Referencias	Pág.246
Anexos	Pág.301

Índice de tablas y figuras

Tablas

Tabla 1. Resumen de los criterios de Slick <i>et al.</i> (1999) y Bianchini <i>et al.</i> (2005), extraído de González-Ordi <i>et al.</i> (2012)	Pág.7
Tabla 2. Manipulation <i>check</i> para diseños de simulación (adaptado de Rogers y Bender, 2018)	Pág.47
Tabla 3. Media y desviación típica de las características médico-legales de los pacientes	Pág.105
Tabla 4. Frecuencias de las características médico-legales de los pacientes	Pág.106
Tabla 5. Clasificación de pacientes de cada uno de los evaluadores	Pág.108
Tabla 6. Coeficiente de correlación intraclase (ICC), Índice de concordancia y ANOVA	Pág.109
Tabla 7. Características de los estudios incluidos	Pág.131
Tabla 8. Protocolo y fuentes de información – primera versión	Pág.143
Tabla 9. Componentes del protocolo definitivo	Pág.149
Tabla 10. Factores relacionados con el accidente	Pág.157
Tabla 11. Media, SD y ANOVA del análisis cluster	Pág.160
Tabla 12. Diferencias intercluster en las variables de perfilado	Pág.161
Tabla 13. Media, desviación típica, ANOVA y <i>d</i> de Cohen para las puntuaciones de las escalas de los grupos	Pág.175
Tabla 14. Capacidad discriminativa del punto de corte óptimo	Pág.177
Tabla 15. Capacidad discriminativa de los instrumentos	Pág.179
Tabla 16. Poder predictivo positivo y negativo con un 10%, 30% y 50% de prevalencia	Pág.180
Tabla 17. Razón de verosimilitud encadenada con una prueba positiva (RB 30%)	Pág.182
Tabla 18. Razón de verosimilitud encadenada con dos pruebas positivas (RB 30%)	Pág.182
Tabla 19. Razón de verosimilitud encadenada con tres pruebas positivas (RB 30%)	Pág.183
Tabla 20. Criterios para determinar sospecha de simulación en el protocolo	Pág.187
Tabla 21. Factores relacionados con el accidente	Pág.196

Tabla 22. Resultados prueba t de Student entre Pacientes clínicos y Simuladores instruidos	Pág.198
Tabla 23. Resultados de la regresión logística binaria	Pág.199
Tabla 24. Resultados de la prueba ANOVA y Magnitudes del efecto para la comparación entre grupos	Pág.215
Tabla 25. Estimadores de utilidad con un punto de corte de 14	Pág.217
Tabla 26. Poder predictivo positivo y negativo, y <i>Accuracy Index</i> para una prevalencia del 10% al 50% (Punto de corte de 14)	Pág.217
Tabla 27. Estimadores de utilidad con un punto de corte de 16	Pág.219
Tabla 28. Poder predictivo positivo y negativo, y <i>Accuracy Index</i> para una prevalencia del 10% al 50% (Punto de corte de 16)	Pág.219

Figuras

Figura 1. Curva Característica Operativa del Receptor. Reproducido de Kim et al (2014) con permiso de los autores	Pág.60
Figura 2. Modelo de diátesis-estrés. Adaptado de Turk (2002) con permiso del autor	Pág.75
Figura 3. Modelo de miedo-evitación del dolor. Adaptado de Leeuw <i>et al.</i> (2007) con permiso del autor	Pág.77
Figura 4. Modelo integrado de SLC crónico. Adaptado de Walton y Elliot (2017) con permiso del autor	Pág.79
Figura. 5. El proceso de selección (siguiendo la guía PRISMA, Moher <i>et al.</i> , 2009)	Pág.123
Figura 6. Curvas COR para los instrumentos del protocolo	Pág.176
Figura 7. Curva COR para la comparación entre simuladores instruidos y pacientes clínicos	Pág.220

MARCO TEÓRICO

1. La simulación

1.1. Conceptos básicos

Fingir malestar de algún tipo con el objetivo de obtener un beneficio es un comportamiento de carácter multisistémico, intrínseco del ser humano, que incluye aspectos cognitivos, subjetivos, fisiológicos, somáticos y motor-comportamentales (Wessely, 2003). La Real Academia Española de la Lengua (2014) ofrece una definición genérica de la simulación, entendiéndose como la alteración aparente de la causa, la índole o el objeto verdadero de un acto o contrato. A pesar de estar recogida en los principales manuales de diagnóstico clínico, la simulación no es considerada un trastorno mental, aunque puede estar relacionada o ser comórbida de alguno de éstos, entendiéndose más, en términos generales, como una conducta fraudulenta asociada a una enfermedad o discapacidad (Kropp y Rogers, 1993).

En el sistema de clasificación internacional de enfermedades CIE-10, en su codificación Z76.5, afirma que la simulación debe entenderse como la producción intencional de síntomas o incapacidad motivada por incentivos externos, quedando asociado al código Z76 perteneciente a personal que entra en contacto con los servicios de salud en otras circunstancias (Organización Mundial de la Salud, 1996). Por otro lado, en la quinta edición del DSM, la simulación es operativizada como la presentación de sintomatología de corte físico o psicológico y/o la representación de síntomas físicos o psicológicos falsos o muy exagerados, motivada por incentivos externos (American Psychiatric Association, 2013), reconociéndose la idea de que, en determinadas ocasiones, puede constituir un comportamiento adaptativo.

Específicamente, según el manual de la APA (APA, 2000, 2013), se debe sospechar de simulación cuando se observe la combinación de: 1) Contexto médico legal

de la presentación, 2) discrepancia entre las quejas o discapacidad mencionada por el individuo y los hallazgos y observaciones objetivas del profesional, 3) Falta de cooperación durante el proceso de valoración y para cumplir las pautas de tratamiento prescritas, 4) presencia de un trastorno de la personalidad antisocial. Además, pone especial énfasis por un lado en la importancia de un incentivo externo para el diagnóstico diferencial con el trastorno facticio y, por otro lado, en la generación intencionada de síntomas para diferenciarlo del trastorno de conversión y de síntomas somáticos. Sintetizando las definiciones propuestas, nos encontramos con que la simulación es la producción intencional de síntomas motivada por la consecución de un incentivo externo. Se plantea que simular es un acto plenamente consciente, y la ganancia obtenida será principalmente material, diferenciándose de una ganancia puramente psicológica, en el caso del trastorno facticio, y de una producción no intencional, en el caso de los trastornos somatomorfos.

Esta conceptualización del término ha facilitado que no sea considerado como un trastorno mental ya que, en sí misma, la simulación es una estrategia comportamental que personas sanas o enfermas pueden usar en determinadas situaciones (Merten y Merckelbach, 2020; Widows y Smith, 2005). A pesar de ello, la operativización como un constructo categorial (presencia *VS* ausencia de criterios) realizada por los anteriores manuales, tiene poca utilidad clínica (Bass y Halligan, 2014) y va en contraposición de lo observado en la evidencia, donde la simulación es conceptualizada como un constructo multidimensional cuyo diagnóstico debería ser realizado en función a grados de certeza (Bianchini *et al.*, 2005; González-Ordi *et al.*, 2012; Larrabee, 2007; Slick *et al.*, 1999). De la misma forma, se ha observado que existen diferentes niveles de fabricación o exageración de sintomatología: de simulación muy exagerada y poco consistente a simulación refinada y relativamente bien ajustada; de simulación específica y contingente

a un contexto a simulación generalizada a diversos ámbitos, así como simulación en aquellas patologías que tienen diferentes grados de dificultad para fingir los síntomas (Rogers y Bender, 2018). La simulación no debería ser enfocada desde una proposición dicotómica de “todo o nada” en la que toda exageración sea entendida como una producción intencional o un intento deliberado de engaño, sino como uno de los múltiples estilos de respuesta que puede adoptar un evaluado (Merckelbach *et al.*, 2019; Rogers *et al.*, 2014).

Fue Rogers (1988) en la primera edición de su libro “La valoración clínica de la simulación y la mentira”, del inglés “*Clinical assessment of malingering and deception*” el que definió por primera vez el concepto adaptado al contexto clínico-forense e introdujo el término “estilo de respuesta”. Para entender la propuesta de Rogers, imaginemos una supuesta valoración clínico-forense y dividámosla en varias fases. En la primera, surge la sospecha de exageración o producción de síntomas. Las anteriores definiciones, DSM o CIE, contemplan tres posibilidades en esta situación: simulación (intencional y con incentivo externo no psicológico, principalmente material), trastorno facticio (intencional y con incentivo externo de carácter psicológico) y trastorno somatomorfo (no intencional). Una vez descartadas estas dos últimas, pasamos a la segunda fase, donde solo queda la posibilidad de simulación. Rogers explica que, en esta segunda fase, no es la simulación lo que debe ser considerado, sino el “fingimiento”. A lo largo de sus múltiples publicaciones (ver por ejemplo Rogers, 1997, 2008; Rogers y Bender, 2018; Correa y Rogers, 2012), el autor ha hecho hincapié en la diferencia entre los términos fingimiento y simulación, en inglés “*Feign/Feigning*” y “*Malingering/Malingering*”. La definición literal de estos dos conceptos, proporcionada por Rogers y Bender (2018), es la siguiente:

Feigning is the deliberate fabrication or gross exaggeration of psychological or physical symptoms, without any assumptions about its goals (Rogers & Bender, 2003, 2013). This term was introduced because standardized measures of response styles (e.g., psychological tests) have not been validated to assess an individual's specific motivations. Therefore, determinations can often be made for feigned presentations but not their underlying motivations. To underscore this point, psychological tests can be used to establish feigning but not malingering (p.6).

Malingering has been consistently defined by DSM nosology as “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives” (American Psychiatric Association, 2013, p. 726). An important consideration is magnitude of the dissimulation; it must be the fabrication or gross exaggeration of multiple symptoms. The presence of minor exaggerations or isolated symptoms does not qualify as malingering. Its requirement of external incentives does not rule out the co-occurrence of internal motivations (p.6).

Correa y Rogers (2012) realizaron una traducción prácticamente literal de estos dos conceptos para el manual de González-Ordi *et al.* (2012):

Fingir se define como la producción deliberada o la exageración evidente de síntomas. El término se emplea generalmente dentro del ámbito de las pruebas psicológicas, las cuales no permiten establecer motivación última de un paciente. Fingir se diferencia por tanto de simular en que, pese a que en los dos casos es necesaria una producción o exageración de síntomas por parte del evaluado, en el caso de fingir no se establece cuál es la motivación (p.154).

La **simulación** es definida por el DSM-IV-TR (American Psychiatric Association, 2000) como la producción intencionada de síntomas físicos o psicológicos desproporcionados o falsos, motivados por incentivos externos. Es realmente importante remarcar el hecho de que la mera presencia de incentivos externos no es suficiente; los profesionales deberán siempre poder demostrar que existe intencionalidad por parte del evaluado (p.155).

En la bibliografía española, el término “fingimiento” ha sido traducido en ocasiones como “exageración”. Por ejemplo, en el manual de González-Ordi *et al.* (2012), Santamaría y González-Ordi (2012) escriben un capítulo en el que hablan de “exageración de síntomas” y definen dicho término como “aquel estilo de respuesta en el que el sujeto busca deliberadamente presentarse como alterado o deteriorado, exagerando o fabricando síntomas para ello y maximizando las características negativas de sí mismo” (p. 192). Aunque los autores no exponen explícitamente que la traducción del término sería la anterior, es coherente con el concepto de “fingimiento” y, además, lo utilizan para realizar la traducción de la tabla de estrategias de detección de Rogers (2008), en la que habla de “feigning” en la versión inglesa. Ramón Arce, Francisca Fariña, Manuel Vilariño y otros investigadores españoles de la temática, si bien no diferencian explícitamente entre simulación y fingimiento, hacen hincapié en la importancia de la intencionalidad para el diagnóstico diferencial de simulación y en múltiples publicaciones los autores indican que el diagnóstico de simulación es coherente con otras hipótesis alternativas (por ejemplo, Arce, 2017; Arce *et al.*, 2015; Seijo *et al.*, 2014; Vilariño *et al.*, 2013).

Puede observarse que la diferencia entre los dos conceptos radica en el objetivo o meta que se persiga. Volviendo a la situación de ejemplo expuesta con anterioridad, y utilizando los conceptos actualizados, el/la profesional habrá detectado una presentación deliberada o exageración evidente de síntomas, ante la cual deberá plantearse una serie

de hipótesis alternativas, entre las que se encuentran la secuencia de los ítems y los test aplicados, falta de información sobre los síntomas que puede padecerse, un patrón de respuesta a los instrumentos inatento, rasgos de personalidad (Merckelbach *et al.*, 2019), un patrón de actuación defensivo, la deseabilidad social y la gestión de la impresión (Rogers y Bender, 2018). Al descartar las citadas hipótesis, únicamente quedará el “fingimiento” como posible explicación a la exageración de síntomas. Ante esta situación, cabrá preguntarse ¿qué meta busca el/la evaluado? En el caso de que la meta sea externa, por ejemplo, un incentivo económico, se planteará la simulación como posible explicación, mientras que, si la meta es interna, por ejemplo, la asunción del rol de enfermo, se planteará la presencia de un trastorno facticio (Merten y Merckelbach, 2020).

Bajo esta premisa, parece evidente concluir que tanto para entender como evaluar la simulación, es necesario una estrategia completa, rigurosa y específica, así como la aplicación de técnicas y herramientas con demostrada validez que aborden el fenómeno sistemáticamente (Holmquist y Wanlass, 2002). En respuesta a esto, los investigadores en la temática han ido proponiendo alternativas que se adapten mejor a lo hallado en la investigación, haciendo énfasis en la exploración de los datos y la concurrencia de la información, entendiendo la simulación como un fenómeno no solo multidimensional, sino también multidisciplinar. En este sentido, dos de las propuestas más influyentes, basadas en esta visión actualizada de la simulación, son las de Bianchini *et al.*, (2005) y Slick *et al.* (1999), los cuales establecieron una serie de criterios diagnósticos que buscaban ofrecer a los profesionales una alternativa actualizada a los criterios de los manuales estadísticos. Dichas propuestas se desarrollan siguiendo tres premisas: 1) Deben existir incentivos externos significativos; 2) debe realizarse un diagnóstico diferencial, es decir, lo observado no se explica mejor por factores psicopatológicos,

neurológicos o evolutivos; y 3) deben estar centrados en probabilidad de ocurrencia en función a la evidencia disponible (simulación definitiva, posible o probable). En la tabla 1 se resumen los criterios de ambas propuestas y las conclusiones en función al cumplimiento o no de estos.

Tabla 1. *Resumen de los criterios de Slick et al. (1999) y Bianchini et al. (2005), extraído de González-Ordi et al. (2012)*

Slick et al. (1999)	Bianchini et al. (2005)
<p>Centrada en la detección de simulación de síntomas neurocognitivos. Los criterios son:</p> <p>a- Presencia de incentivos externos sustanciales al momento de la evaluación (estatus de litigancia pensión por incapacidad, acusación criminal, etc.)</p> <p>b- Evidencia de exageración o fabricación del deterioro neurocognitivo en pruebas de evaluación neuropsicológica, demostrada por al menos uno de los subcriterios siguientes:</p> <p><i>b1</i> - sesgo de respuesta negativo definitivo: ejecución por debajo del azar ($p < .05$) en una o más medidas cognitivas de elección forzosa.</p> <p><i>b2</i> - sesgo de respuesta probable: las respuestas en uno o más test o indicadores psicométricos empíricamente bien validados y diseñados para evaluar la exageración o fabricación de déficits cognitivos, es consistente con el engaño.</p> <p><i>b3</i> - discrepancia entre los datos de los tests y los patrones de funcionamiento cerebral</p> <p><i>b4</i> - discrepancia entre los datos de uno o más tests neuropsicológicos que evalúan un mismo dominio y la conducta observada</p> <p><i>b5</i> - discrepancia entre los tests neuropsicológicos de un dominio y los informes colaterales fiables</p>	<p>Posible aplicación tanto en simulación de síntomas neurocognitivos como psicológicos y físicos. Los criterios son:</p> <p>a- Evidencia de incentivos externos significativos contingentes con un patrón de exageración o fabricación de síntomas (estatus de litigancia, pensión por incapacidad, indemnizaciones por parte de aseguradoras, prolongación de bajas laborales, etc.).</p> <p>b- Evidencia procedente de la evaluación física. Evidencia de que las habilidades, capacidades o limitaciones demostradas o largo de una evaluación física formal son consistentes con la exageración de las quejas y síntomas y la disminución exagerada de la capacidad física. Este aspecto puede operativizar mediante la detección de: 1) sesgos en el rendimiento físico: realización de pruebas de rendimiento que evidencian la existencia de una disminución de la capacidad física que no resulta concordante con la patología diagnosticada; 2) discrepancia entre la valoración subjetiva de dolor y la reactividad psicofisiológica concomitante; 3) los síntomas o las quejas informadas por el paciente son sustancialmente divergentes con la evidencia fisiopatológica existente; y 4) discrepancia entre la actividad física mostrada por el</p>

Slick <i>et al.</i> (1999)	Bianchini <i>et al.</i> (2005)
<p><i>b6</i> - discrepancia entre los tests neuropsicológicos y la historia clínica documentada</p>	<p>paciente durante la evaluación formal con la actividad física demostrada cuando el individuo no sabe que está siendo evaluado.</p>
<p>c- Evidencia procedente del autoinforme. Los siguientes criterios son indicadores de una posible simulación de déficits cognitivos, aunque no son suficientes para establecer el diagnóstico.</p>	<p>c- Evidencia procedente de la evaluación neuropsicológica. Evidencia de que los déficits neurocognitivos demostrados a lo largo de una evaluación formal son consistentes con la exageración o el falseamiento de la discapacidad cognitiva. Este criterio se operativiza en: 1) rendimiento por debajo del azar en pruebas de elección forzada que evalúan funciones cognitivas y perceptivas; 2) rendimiento consistente con la exageración de la discapacidad en pruebas que evalúan funcionamiento cognitivo o perceptivo; 3) discrepancias entre los resultados de las pruebas neuropsicológicas y el conocimiento existente del funcionamiento cerebral: discrepancias entre los resultados de las pruebas neuropsicológicas y la conducta observable y objetivable del paciente.</p>
<p>Sin embargo, la presencia de uno o más de estos criterios aporta evidencia adicional para apoyar el diagnóstico de simulación. Dichos criterios implican inconsistencias y discrepancias significativas en los síntomas autoinformados por el paciente que sugieren un intento deliberado por exagerar o fabricar déficits cognitivos.</p>	<p>d- Evidencia procedente de la evaluación mediante autoinforme. Evidencia de que los síntomas, quejas y limitaciones autoinformadas son consistentes con un patrón de exageración de síntomas y discapacidad a nivel físico, cognitivo o emocional. Este criterio puede operativizarse a través de: 1) la discrepancia de la conducta del paciente cuando es formalmente evaluado frente a cuando no sabe que está siendo evaluado, 2) las quejas autoinformadas no son consistentes con la evolución de la historia clínica del paciente; 3) los síntomas autoinformados son discrepantes con los patrones conocidos de funcionamiento fisiológico y neurológico del dolor y la discapacidad referida a la patología en cuestión; 4) los síntomas autoinformados</p>
<p><i>c1</i> - La historia referida por el paciente es discrepante con la historia documentada</p>	
<p><i>c2</i> - los síntomas referidos son discrepantes con los patrones conocidos de funcionamiento cerebral.</p>	
<p><i>c3</i> - los síntomas referidos son discrepantes con la observación conductual.</p>	
<p><i>c4</i> - los síntomas referidos son discrepantes con la información obtenida de informadores colaterales.</p>	
<p><i>c5</i> - evidencia de exageración o fabricación de alteraciones psicológicas a través de índices o escalas de autoinforme empíricamente validadas que evalúan ajuste psicológico (p.ej., MMPI-2)</p>	
<p>d- Los criterios mencionados no se explican mejor por factores psicopatológicos, neurológicos o evolutivos.</p>	
<p><u>Conclusiones</u></p>	
<p>Simulación definitiva = A + B1 + D.</p>	

Slick <i>et al.</i> (1999)	Bianchini <i>et al.</i> (2005)
<p>Simulación probable = A + dos o más criterios entre B2 y B6, o bien, un criterio entre B2 y B6 y uno o más criterios entre C1 y C5 + el criterio D.</p>	<p>son discrepantes con la conducta observable y objetivable; y 5) presentación de un patrón de exageración de síntomas en pruebas de personalidad de amplio espectro y pruebas de Screening de simulación.</p>
<p>Simulación posible = A + uno o más criterios entre C1 y C5 + el criterio D.</p>	<p>e- Los criterios mencionados no se explican mejor por factores psicopatológicos, neurológicos o evolutivos.</p>
<p>Puede ocurrir que se confirmen los criterios de simulación definitiva o probable pero no se pueda determinar el criterio D, en cuyo caso, deberán explorarse hipótesis alternativas a la simulación.</p>	<p><u>Conclusiones</u></p>
	<p>Simulación definitiva = confirmación de la mayoría de los criterios y subcriterios A, B, C, D y E.</p>
	<p>Simulación probable = A y E + dos o más subcriterios de B, C o D.</p>
	<p>Simulación posible = criterio A + un subcriterio de B, C o D, no es necesario confirmación de E.</p>

Estos dos grupos de criterios forman una de las primeras propuestas de aproximación estructurada para detectar la simulación, ofreciendo una metodología que incluye los principales avances en la temática y todas las variables que el/la profesional debía considerar para poder mencionarse sobre la presencia o ausencia de simulación. Resulta de especial interés la inclusión del criterio “e” de ambos listados, “los criterios mencionados no se explican mejor por factores psicopatológicos, neurológicos o evolutivos”. Este criterio, junto al “a”, donde se habla de incentivo “significativo” operativiza la conceptualización de Rogers e indica que el profesional debe enfocar el diagnóstico diferencial de simulación como un estudio de los múltiples estilos de respuesta que un/a paciente puede mostrar en el contexto clínico-forense (Merckelbach *et al.*, 2019; Rogers *et al.*, 2014).

Como ya exponíamos, descartar la existencia de hipótesis alternativas es fundamental en la valoración de la simulación, pero también es un procedimiento de gran importancia en el contexto forense en general (Arce *et al.*, 2015). Se puede tomar como ejemplo el procedimiento que se sigue en el análisis de la validez de una declaración con el SVA (*Statement Validity Assessment*; Köhnken y Steller, 1988; Steller y Köhnken, 1989). El citado instrumento se utiliza para evaluar criterios o atributos de memoria y determinar si un relato ha sido vivido o ha sido fabricado (Arce, 2017), pero, aunque haya sido vivido, debe tenerse en consideración que el relato obtenido puede haberse visto condicionado por variables externas, como aspectos relativos a la entrevista o una posible intención secundaria del entrevistado. Así, el SVA contiene un elemento llamado listado de validez mediante el cual se busca descartar posibles hipótesis alternativas que ofrezcan una mejor explicación a los resultados que la hipótesis principal (Horowitz, 1991). De esta manera, el/la evaluador/a puede plantear la simulación como hipótesis principal y valorar las posibles hipótesis alternativas (distintos estilos de respuesta) y, descartándolas, puede llegar a una conclusión sobre la existencia de simulación, la cual se valoraría en función de grados de certeza.

Así, sintetizando la información anteriormente expuesta, para realizar un diagnóstico diferencial de simulación es necesario obtener evidencia de que el paciente elige conscientemente (volición) invertir una cantidad significativa de esfuerzo hacia el objetivo a corto plazo de fingir síntomas y déficits, los cuales tienen el objetivo a largo plazo de proporcionar un incentivo externo en forma de ganancia substancial personal o evitación de deberes u obligaciones (Slick *et al.*, 2010). Además, el/la profesional deberá adoptar un modelo multidimensional en el que se servirá de diferentes fuentes de información relacionadas con la condición y descartará otros posibles estilos de respuesta, o hipótesis alternativas.

Finalizaremos este apartado señalando que, además, resulta de gran importancia que el/la profesional sea conocedor de los principales errores que suelen cometerse en el contexto de la simulación. Siguiendo la síntesis de Rogers y Bender (2018), en primer lugar, la simulación no es un evento raro o poco frecuente, debiendo ser evaluada sistemáticamente, considerando su omisión una negligencia. En segundo lugar, la simulación no es un estilo de respuesta estático, es decir, si un individuo ha simulado en alguna ocasión no necesariamente lo hará en el resto de las ocasiones y viceversa. En tercer lugar, la percepción errónea heredada de la conceptualización de simulación del DMS-5 lleva a pensar que la simulación está relacionada con actos y personalidad antisocial. Esto sería un error que puede llevar a conductas discriminatorias. En cuarto lugar, la simulación no necesariamente es un acto de maldad. En quinto lugar, la simulación no es similar al fenómeno iceberg, es decir, cualquier detección de conducta fingida no necesariamente representa un patrón generalizado y oculto más grande. En sexto lugar, la simulación y un trastorno genuino no son mutuamente excluyentes, pudiendo darse ambos a la vez. En séptimo lugar, las escalas psicométricas de simulación no se asemejan con precisión a lo que reflejan en sus descripciones por lo que deben interpretarse con cautela. Y finalmente, en octavo lugar, la simulación no tiene una tasa base estable. Incluso dentro de un mismo contexto es posible que se produzcan variaciones en función de las preguntas realizadas o las circunstancias individuales.

1.2. Estrategias de detección

Rogers (2008) define una estrategia de detección como un “método estandarizado que tiene una base conceptual y está validado empíricamente para diferenciar un estilo de respuesta específico” (p. 16). El autor expone que, para que una estrategia de detección del fingimiento pueda utilizarse, debe incluir necesariamente los siguientes cinco criterios:

- **1. El uso de métodos estandarizados.** Toda estrategia de detección debe estar operativizada y estandarizada, de manera que el profesional disponga de un conjunto de ítems homogéneos, conozca las normas de administración y sea capaz de interpretar las puntuaciones de manera uniforme y adaptada a la población estudiada.
- **2. Disponer de un marco conceptual.** El profesional debe operar bajo un marco teórico bien definido y limitado para poder elegir y/o validar una estrategia específica y una hipótesis concreta.
- **3. Contar con validación empírica.** El evaluador debe disponer de medios que hayan seguido una metodología de validación correcta. En el campo de la simulación es necesario utilizar diseños de investigación adaptados al mismo, los cuales presentan una serie de particularidades que deben tenerse en consideración. En el capítulo 2 del presente marco teórico se desarrollarán en profundidad los citados diseños de investigación.
- **4. Diferenciación sistemática.** Debe valorarse la capacidad discriminativa de los instrumentos mediante estadísticos diseñados para tal fin, no limitándose el análisis del instrumento al estudio de la significación estadística. En el segundo apartado del capítulo 2 se presenta una síntesis de dichos estadísticos, conocidos como “estimadores de utilidad” (Rogers, 1997).
- **5. Delineación adecuada de los estilos de respuesta.** Rogers (2008) expone el ejemplo de la confusión del término “rendimiento subóptimo” con “simulación”, como suele ocurrir con “fingimiento”, “exageración” o “sobredimensión”. Presentar una vivencia de la condición/enfermedad más severa de lo esperable, o incluso atípica, suele asociarse con simulación.

El evaluador debe ser conocedor de los diferentes estilos de respuesta y los instrumentos utilizados deben ser capaces de discriminar entre ellos.

Este planteamiento no dista de lo esperable en la evaluación de un/a psicólogo/a ya que, aunque a veces se olvide, el código deontológico español de la profesión, en su Artículo 18, refiere:

Sin perjuicio de la legítima diversidad de teorías, escuelas y métodos, el/la Psicólogo/a no utilizará medios o procedimientos que no se hallen suficientemente contrastados, dentro de los límites del conocimiento científico vigente. En el caso de investigaciones para poner a prueba técnicas o instrumentos nuevos, todavía no contrastados, lo hará saber así a sus clientes antes de su utilización (Colegio Oficial de Psicólogos de Madrid, 2004, p.8).

Se establece así que el/la profesional encargado/a de utilizar una estrategia de detección de la simulación debe conocer los posibles estilos de respuesta y los instrumentos, herramientas o metodologías empleadas. De la misma forma, es un requisito fundamental que éstas presenten unas propiedades psicométricas óptimas y una capacidad discriminativa adecuada, y hayan sido validadas empíricamente mediante la metodología apropiada. Ahora bien, en el apartado anterior exponíamos una serie de criterios diagnósticos que el/la profesional podía seguir para realizar un diagnóstico diferencial de simulación, pero, en el caso de utilizarlos el/la profesional aún tendrá que dar respuesta a dos cuestiones: ¿qué estrategias dispone para valorar esos criterios?, y ¿qué instrumentos puede utilizar?

Para responder la primera pregunta, Rogers (1984) realizó una propuesta sistemática y basada en la evidencia de estrategias de detección. Dicha propuesta ha sido revisada y actualizada por Rogers y Bender (2003, 2018) y Rogers (1997, 2008), los

cuales exponen 12 estrategias diferentes divididas en dos grupos: *presentación improbable de síntomas* y *estrategias de detección amplificadas*. Originalmente se disponía de 10 estrategias divididas en los citados grupos, pero en la última revisión de Rogers y Bender (2018) se añadieron dos aportes adicionales (*aproximaciones cercanas a síntomas genuinos* y *síntomas excesivamente específicos*) bajo la advertencia de que era necesario más evidencia para considerar su uso. Exponemos a continuación una síntesis de estas 12 estrategias:

Presentación improbable de síntomas. La estrategia más utilizada para la detección del fingimiento/simulación, basada en la búsqueda de síntomas o signos diagnósticos improbables que rara vez pueden ser observados en condiciones genuinas (Rogers y Bender, 2012). Dentro de esta categoría se encuentra (Rogers y Bender, 2018; Walczyk *et al.*, 2018):

1. **Síntomas inusuales.** Se trata de una estrategia basada en la detección de síntomas con una frecuencia de aparición muy inusual en la población clínica genuina (<5%). Es una de las estrategias más comunes en los instrumentos psicométricos utilizados para detectar fingimiento, y suele generar magnitudes del efecto muy elevadas ($d > 1.25$). Puede generar problemas de clasificación si se utiliza en poblaciones muy homogéneas con un criterio muy laxo (<20%).
2. **Síntomas improbables.** Se trata de una variante extrema de la estrategia de síntomas inusuales y se basa en el uso de sintomatología absurda. Debido a su condición fantástica, los ítems utilizados para describir dichos síntomas difícilmente serán reales y no representarán un perfil clínico genuino. La principal limitación que presenta esta estrategia es que dichos ítems pueden despertar sospecha en simuladores más precavidos. De la misma forma, son más fáciles de detectar, por lo que también son más vulnerables a la preparación o al *coaching*.

3. **Síntomas poco frecuentes.** Se basa en la búsqueda/uso de síntomas poco comunes, pero que pueden aparecer con más frecuencia que los analizados en la anterior estrategia. También presenta la ventaja de que produce magnitudes del efecto elevadas ($d > 1.25$), pero, al contrario que la anterior, la presencia de estos síntomas también puede indicar la presencia de un trastorno real, ya que varios de estos síntomas pueden ser detectados con más frecuencia en pacientes con psicopatología genuina, como la esquizofrenia o el trastorno de estrés postraumático.

4. **Combinación de síntomas.** Esta estrategia utiliza síntomas frecuentes o comunes en la población clínica que difícilmente se presentarán juntos o resultan incompatibles entre sí. Por ejemplo, la aparición conjunta de sensibilidad extrema e insensibilidad en el mismo dermatoma. Se trata de una de las estrategias más robustas, ya que es resistente a la preparación o al *coaching*, y se puede incluir con facilidad en cualquier metodología de evaluación.

5. **Patrón falso de psicopatología.** Se trata de una versión extendida de la anterior estrategia basada en el uso de combinaciones de escalas o subescalas específicas. Al ser más compleja y elaborada, es mucho más resistente a la preparación o al *coaching*, pero requiere de un extenso proceso de validación cruzada con diferentes muestras clínicas.

Estrategias de detección amplificadas. Las anteriores estrategias estaban basadas principalmente en la aparición de síntomas atípicos o raros. En las de detección amplificadas se presta atención a la frecuencia o severidad, o en lo que Rogers y Bender (2018) denominan “grado de exceso” (p.27). Dentro de esta categoría se encuentran las siguientes estrategias (Rogers y Bender, 2018; Walczyk *et al.*, 2018):

- 6. Presentación indiscriminada de síntomas.** Se basa en la hipótesis del *symptom overendorsement*¹ en la cual se considera que los simuladores tienden a presentar un número mucho más elevado de síntomas en comparación con los pacientes clínicos genuinos (Rogers, 1997). Esta estrategia ha sido evaluada principalmente mediante el uso de entrevistas estructuradas y el evaluador debe ser precavido al aplicar instrumentos psicométricos que evalúen múltiples síntomas, ya que una interpretación errónea puede llevar a falsos positivos.
- 7. Severidad de síntomas.** Estrategia basada en el hecho de que incluso pacientes de severidad muy elevada experimentan un número muy limitado de síntomas “extremos” o “insoportables”, pero los simuladores tienden a presentar un perfil compuesto por múltiples síntomas de esta gravedad. Tiene la ventaja de ser una estrategia fácil de implementar, pudiéndose adaptar a una amplia gama de instrumentos y metodologías.
- 8. Síntomas obvios.** Esta estrategia se centra en el hecho de que los simuladores tienden a presentar los síntomas más conocidos o frecuentes en el contexto y que claramente indican que se padece un trastorno severo. Por ejemplo, en psicopatología resulta más común que los simuladores presenten sintomatología ansioso-depresiva, o síntomas positivos (delirios y alucinaciones) de la esquizofrenia, que otros síntomas o signos menos conocidos, como la anhedonia, la apatía o la afectividad plana.
- 9. Síntomas informados vs síntomas observados.** Basada en la búsqueda de inconsistencias o incongruencias entre los síntomas o signos que el paciente

¹ *Overendorsement* no tiene una traducción exacta al español, ya que literalmente significa “sobre-endosar”. Debido a que no estamos seguros de que el término sea correcto, y no hemos encontrado una alternativa en la bibliografía española de la temática, utilizaremos la palabra original.

afirma padecer y los que el/la profesional puede observar o identificar. Esta estrategia puede ser de gran utilidad en trastornos del área cervical, como las cervicalgias o el esguince cervical, ya que las limitaciones en el rango de movimiento pueden ser contrastadas con facilidad por el evaluador.

10. Estereotipos erróneos. En la estrategia de los síntomas obvios se indicaba que los simuladores tienden a representar los síntomas más comunes o conocidos del contexto. En esta circunstancia puede ocurrir que dicha representación esté basada en estereotipos o ideas equivocadas asociadas a los trastornos. Si bien es una estrategia altamente resistente a la preparación, la evidencia de su funcionamiento es limitada.

11. Aproximaciones cercanas a síntomas genuinos. Esta estrategia ha sido utilizada únicamente en la escala MAL de la *Malingering Probability Scale* (MPS; Silverton y Gruber, 1998) y se basa en la presentación de ítems que describen síntomas falsos que parecen síntomas genuinos, pero se diferencian de éstos en detalles importantes. Presenta la limitación de que puede confundir a pacientes genuinos, clasificándolos erróneamente.

12. Síntomas excesivamente específicos. Basada en la hipótesis de que los simuladores ofrecerán una presentación de la condición excesivamente precisa y específica. Al igual que la anterior estrategia, requiere validación empírica, ya que presenta el problema de que puede tratarse de una medida de inatención o descuido del paciente, en lugar de una medida de fingimiento o simulación.

Rogers y Bender (2018) incluyen también una adaptación de las anteriores estrategias a las particularidades y necesidades del contexto médico ya que, como exponen los autores, estos pacientes pueden experimentar una actitud ante la situación de

enfermedad negativa, la cual puede producir estrategias de afrontamiento desadaptativas que pueden ser confundidas con simulación. Rogers y Payne (2006) explican que el marco conceptual de Radley y Green (1987) puede ser de utilidad para entender este proceso. Los anteriores autores señalan que un paciente puede afrontar su experiencia de la enfermedad integrando en su identidad la condición o trastorno padecido (*acomodación*) y/o puede perder la creencia de que va a experimentar un cambio positivo en la condición al sentirse abrumado y acepta su condición de “enfermo” (*resignación*). En esta situación, la persona estará menos motivada a seguir su tratamiento, tendrá una participación marginal en las intervenciones y podrá experimentar una cronificación de sus síntomas, lo que puede ser interpretado como un intento deliberado de prolongar la condición médica padecida.

En esta adaptación de Rogers y Bender (2018) se incluyen 8 estrategias, también divididas en las categorías de *presentación improbable de síntomas* y *estrategias de detección amplificadas*. En términos generales, en lugar de centrar su atención en los síntomas del paciente, estas estrategias buscan analizar sus quejas médicas y problemas relacionados con la salud. Así, 6 de ellas coinciden con las expuestas con anterioridad, incluyendo en la denominación y/o definición el cambio mencionado (quejas médicas inusuales, quejas médicas improbables, intensidad de las quejas médicas, combinación de síntomas, presentación indiscriminada de problemas de salud y síntomas informados vs síntomas observados), y se incluyen dos nuevas:

- 13. Dependencia de las quejas médicas.** La definición proporcionada por los autores a esta estrategia puede resultar algo confusa. Exponen que se basa en la “tendencia de los simuladores a reconocer atributos positivos de su condición física o estado de discapacidad” (p. 35) o “a reconocer beneficios inmerecidos de su simulación” (p. 36), lo que se podría entender como que los simuladores reconocerán que su

condición tiene algún aspecto positivo o ventaja. No obstante, ponen como ejemplo la subescala “Dependencia a los síntomas” del Battery for Health Improvement 2 (BHI-2; Bruns y Disorbio, 2003), la cual evalúa la búsqueda del paciente de atención y apoyo cuando experimenta los síntomas de la condición padecida (Bruns y Disorbio, 2003). No se ha localizado, a día de hoy, más evidencia sobre esta estrategia y los propios Rogers y Bender (2018) indican que no parece tener ninguna fortaleza ya que la idea de que un simulador reconozca alguna motivación para simular parece contraintuitiva, por lo que su uso debe realizarse con precaución.

14. Presentación de excesiva virtud. Como exponen Lanyon y Carle (2007), el concepto que da nombre a esta estrategia surgió a partir del trabajo de Gur y Sackeim (1979) en el cual se señalaba la necesidad de dividir el concepto “deseabilidad social” en consciente e inconsciente. El término deseabilidad social consciente fue operativizado por Lanyon (2004) en “autopresentación favorable”, el cual hace referencia al intento deliberado de una persona de ofrecer una visión positiva o falsamente apta de sí misma. Por ejemplo, una pareja que esté realizando una entrevista para una adopción puede fingir tener determinadas características positivas que consideran deseables para la situación. Lanyon (2004) indica que el término tiene múltiples facetas, y una de ellas es la presentación de excesiva virtud (del inglés *claims of excessive/extreme virtue*) la cual consiste en una presentación excesivamente positiva de atributos o características. Rogers y Bender (2018) indican que puede ser utilizada por un/una paciente para tratar de disimular un estilo de respuesta concreto, como la simulación. Si bien su utilidad por sí sola es limitada, puede ser utilizada junto a

otras de las estrategias anteriormente mencionadas para incrementar su capacidad discriminativa.

Los anteriores autores señalan además que, si bien no constituye una estrategia en sí misma, en el contexto médico resulta de gran importancia considerar la posibilidad de que el/la paciente tenga una actitud defensiva hacia la presentación de problemas médicos. Como exponen Sokol *et al.* (2005), la negación o minimización tiene un impacto significativo en la vivencia de la enfermedad del paciente, empeorando el desarrollo de la misma y ralentizando la recuperación, lo que podría ser interpretado como simulación.

1.3. Herramientas e instrumentos para la detección de la simulación

Una vez conocidas las estrategias, el/la profesional se enfrentará a la segunda cuestión que planteábamos: elegir instrumentos, herramientas o metodologías que permitan realizar un diagnóstico diferencial eficiente de simulación. A grandes rasgos, los tipos de herramientas disponibles se pueden agrupar en tres categorías:

- **A) Instrumentos diseñados para detectar simulación.** Son recursos que tratan, específicamente, de detectar simulación/fingimiento. En esta categoría se situaría el *Inventario Estructurado de Simulación de Síntomas* (SIMS; Smith y Burger, 1997; Widows y Smith, 2005), cuyo objetivo es la detección de simulación de síntomas psicopatológicos y/o neuropsicológicos, o la *Entrevista Estructurada de Síntomas Informados* en su segunda versión (SIRS 2; Rogers *et al.*, 2010), la cual persigue el mismo fin. Denominaremos estas herramientas “Tipo A”, para referirnos a ellas de ahora en adelante.
- **B) Instrumentos que incluyen control de la distorsión en los estilos de respuesta.** Se trata de herramientas que no han sido diseñadas específicamente para detectar simulación, pero incluyen lo que se conoce como *escalas de validez*

integradas, diseñadas para analizar una posible distorsión en los estilos de respuesta de la persona, entre los que se incluye la simulación. Por ejemplo, el *Inventario Multifásico de la Personalidad de Minnesota* (MMPI-2-RF; Ben-Porath y Tellegen, 2008) incluye una escala para detectar mentira, otra para detectar incoherencia o validez y otra para detectar un estilo de respuesta defensivo, el *Inventario Clínico Multiaxial de Millon* (MCMI-IV; Millon *et al.*, 2015), incluye un índice de validez, una escala de deseabilidad social, otra de simulación y otra de disimulación, o el *Inventario de Evaluación de la Personalidad* (PAI; Morey, 1991) con las escalas de Inconsistencia, Infrecuencia, Impresión negativa e Impresión positiva. Resulta frecuente, además, que las escalas de validez, como las mencionadas, sean sometidas a un proceso de validación posterior en el que se evalúa su capacidad para discriminar entre pacientes genuinos y simuladores (por ejemplo, Sánchez *et al.*, 2017 o Wygant *et al.*, 2011). Denominaremos estas herramientas “Tipo B”, para referirnos a ellas de ahora en adelante.

- **C) Instrumentos validados para detectar la simulación.** Si bien la evidencia indica que los simuladores obtendrán puntuaciones más elevadas y ofrecerán un perfil de la condición más severo que los pacientes clínicos, no debe darse por supuesto que cualquier instrumento puede diferenciar adecuadamente entre las dos poblaciones, por lo que su capacidad discriminativa tiene que ser estudiada y validada empíricamente (Gonzalez Ordi *et al.*, 2008). Así, en esta categoría entran herramientas cuyo objetivo no es detectar simulación o estilos de respuesta anómalos, ni incluyen escalas de validez integradas, si bien se ha comprobado que pueden ser utilizadas para tal fin. Por ejemplo, Crighton *et al.* (2014) y Bianchini *et al.* (2014), estudiaron la capacidad discriminativa del *Modified Somatic*

Perception Questionnaire (MSPQ; Main, 1983) y del *Pain Disability Index* (PDI; Pollard, 1984), o Curtis *et al.* (2019), los cuales estudiaron la capacidad discriminativa de la *Escala de Catastrofización del Dolor* (PCS; Sullivan *et al.*, 1995). Denominaremos estas herramientas “Tipo C”, para referirnos a ellas de ahora en adelante.

A raíz de estas tres categorías podemos añadir una serie de matices al primer criterio de Rogers (2008), y es que no basta con que los instrumentos que se utilicen en el proceso de valoración estén estandarizados, sino que también deben haber sido validados para detectar simulación, y el contexto de la validación/objetivo del instrumento debe ser coherente o coincidir con la condición valorada. Por ejemplo, Cernovsky *et al.* (2019a, 2019b) realizaron una serie de publicaciones en las que denunciaban el uso del SIMS para llevar a cabo “pseudo-diagnósticos” de simulación en pacientes que habían sufrido accidentes de circulación, población para la cual el instrumento no había sido diseñado inicialmente ni había sido validado. Los autores indican que, en estas circunstancias, los pacientes genuinos pueden ser clasificados erróneamente, ya que una gran cantidad de los ítems del instrumento pueden confundir a los pacientes clínicos al exponer sintomatología común en el contexto. En este sentido, Rogers y Bender (2018) proponen una serie de cuestiones que el profesional debe considerar a la hora de elegir un instrumento de *screening*. Consideramos que estas cuestiones pueden resultar de utilidad para la elección de instrumentos en general, por lo que los exponemos a continuación y recomendamos su uso en la fase inicial de la configuración del protocolo de valoración clínico-forense:

- ¿El instrumento tiene unas propiedades psicométricas apropiadas?
- ¿El instrumento tiene una fiabilidad adecuada?
- ¿Hay evidencia sólida de validez convergente y discriminante?

- ¿Las magnitudes del efecto y los estimadores de utilidad, como la sensibilidad o la especificidad, apoyan su uso?
- ¿Ha sido el instrumento evaluado con una población clínica relevante?
- ¿Ha sido el instrumento validado en el contexto, enfermedad o condición que se pretende evaluar?
- ¿Dispone el instrumento de evidencia empírica sólida que apoye su uso en el contexto, enfermedad o condición que se pretende evaluar?
- ¿Para qué se va a utilizar el instrumento (screening o diagnóstico diferencial)?

La publicación de Cernovsky *et al.* (2019b) también señala otro elemento a considerar y es que, según los autores, el SIMS, un instrumento de *screening*, era el único instrumento utilizado para realizar el diagnóstico de simulación. Al igual que en la evaluación psicológica en general, en el contexto de la simulación existen herramientas de *screening*, cuyos resultados indican sospecha o riesgo de simulación, y de diagnóstico diferencial, los cuales se utilizan habitualmente tras aplicar una batería de *screening*, y sirven para valorar de forma definitiva la presencia de simulación (Walczyk *et al.*, 2018). El/la profesional nunca debe mencionarse sobre la existencia o ausencia de simulación únicamente aplicando instrumentos de *screening* y deberá realizar una evaluación más pormenorizada utilizando instrumentos de diagnóstico (Lewis *et al.*, 2002).

Ahora bien, ¿qué instrumentos tiene un/a profesional a su disposición, a día de hoy, para detectar simulación? Arce *et al.*, (2013) y Sánchez *et al.* (2017) indican que el proceder más apropiado para la evaluación de la simulación es el uso de una estrategia multimétodo que combine técnicas de entrevista e instrumentos psicométricos. El uso exclusivo de las primeras no es suficiente, ya que no se controla totalmente el error tipo II, y tampoco sería correcto el uso exclusivo de las segundas, ya que proporciona impresiones diagnósticas, y no diagnósticos, no controla del todo el error tipo II, no

permite establecer un nexo causal entre hechos y daño, y puede constituir una tarea conducente a la simulación. A continuación, exponemos los instrumentos disponibles divididos en las categorías de técnicas de entrevista e instrumentos psicométricos.

1.3.1. Entrevistas para la detección de la simulación/fingimiento.

El uso de la entrevista en la evaluación es una herramienta de uso común indistintamente del ámbito u objetivo del profesional. Desde los profesionales del campo clínico o la gestión de los recursos humanos, hasta el forense y judicial, la entrevista es un recurso que permite al/la experto/a profundizar en el caso al que se enfrenta. En la valoración de la simulación de síntomas, las entrevistas también juegan un papel fundamental y su aplicación ha supuesto un gran avance en este campo (Rogers, 2003). Tradicionalmente, las entrevistas se han clasificado en tres grandes grupos: estructuradas, semiestructuradas y abiertas (Díaz-Bravo *et al.*, 2013). Como exponen González-Ordi *et al.* (2012), en el contexto de la valoración de la simulación, las entrevistas estructuradas y las semiestructuradas son las más utilizadas. Las primeras permiten estandarizar la formulación de preguntas y cómo se presentan, así otorgan un sistema de clasificación para evaluar la presencia y grado de cada síntoma. Por otro lado, las segundas permiten la formulación de cuestiones que clarifiquen las respuestas ambiguas del evaluado, lo que permite alcanzar indicadores de concordancia adecuados entre evaluadores. Las entrevistas abiertas, por su propia naturaleza, no son recomendadas para esta práctica y no han sido utilizadas, a día de hoy, en la temática.

En la actualidad, las tres herramientas de entrevista más utilizadas en el campo de la simulación son el SIRS-2 (Rogers *et al.*, 2010), el *Miller Forensic Assessment of Symptoms Test* (M-FAST; Miller, 2001) y la entrevista clínico-forense de Arce y Fariña (2001) (Arce *et al.*, 2013; Archer *et al.*, 2006).

Entrevista Estructurada de Síntomas Informados (SIRS; Rogers, 1986 y SIRS 2; Rogers *et al.*, 2010). El SIRS es una herramienta de evaluación mediante entrevista, diseñada para detectar fingimiento en psicopatología. Es considerada una de las herramientas mejor validadas para la detección de simulación en el contexto forense y cuenta con una fiabilidad interna adecuada, con un Alfa de Cronbach de .76 a .93, magnitudes del efecto elevadas ($d > 1.25$), una alta validez discriminante y una excelente capacidad para clasificar simuladores con una baja tasa de falsos positivos (sensibilidad y especificidad de .80 y .975 respectivamente, poder predictivo positivo y negativo .91 y capacidad clasificatoria global de .91) (Rogers *et al.*, 2010). Está compuesta por 172 ítems divididos en tres tipos de preguntas (a) preguntas detalladas, que valoran síntomas concretos y su severidad, (b) preguntas repetidas, que miden la consistencia de las respuestas y (c) preguntas generales, las cuales evalúa sintomatología y problemas psicológicos específicos y patrones de síntomas (Seijo *et al.*, 2014). Sus resultados se agrupan en 8 subescalas primarias que incluyen: Síntomas raros (RS), Síntomas improbables o absurdos (IA), Síntomas combinados (SC), Síntomas obvios (BL), Síntomas sutiles (SU), Selección de síntomas (SEL), Gravedad de síntomas (SEV) y Síntomas informados versus síntomas observados (RO) y cuatro subescalas suplementarias. El desarrollo de la versión SIRS-2 se realizó con una muestra procedente del contexto forense y clínico con la finalidad de reducir la tasa de falsos positivos (Becke *et al.*, 2019). Se mantuvo la estructura original y los criterios de corrección y se incorporó un nuevo modelo de decisión y escalas adicionales, entre las que se incluyen la *Escala Total de Síntomas Raros*, el *Índice modificado total* y el *Índice de escalas suplementarias*. Para la corrección se contempla que el paciente responda con un “Sí definitivo”, puntuándose con un “2”, “Sí cualificado/A veces”, puntuándose con un “1” y “No”, que se puntúa con un “0”. En función de los resultados obtenidos en la suma de las

puntuaciones, los pacientes son categorizados como “genuino”, “probable fingimiento” y “fingimiento definitivo” en cada una de las escalas mencionadas con anterioridad. Los resultados de las escalas principales se complementan con las suplementarias para proveer de una descripción clínica más exhaustiva y, además, se utilizan para ofrecer una clasificación sobre el estilo de respuesta del/a evaluado/a (Becke *et al.*, 2019). Como exponen Walczyk *et al.* (2018), el uso del SIRS puede resultar complejo debido a que se trata de un instrumento difícil de aplicar y requiere un tiempo de administración de entre 30 y 60 minutos.

Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). Se trata de una herramienta de *screening* breve, diseñada para detectar simulación de psicopatología en el contexto forense. Tiene un tiempo de aplicación aproximado de 10 minutos y está compuesto por 25 ítems, entre los que se incluyen 15 en formato de respuesta de verdadero y falso, 5 en formato Likert, 2 de *sí* o *no* y 3 diseñados para detectar discrepancias entre las respuestas de los pacientes y las observaciones realizadas por el/la profesional. Consta de siete escalas que evalúan Síntomas reportados frente a observados (RO), Sintomatología extrema (ES), Combinaciones atípicas (RC) Alucinaciones inusuales (UH), Curso de síntomas inusuales (USC), Imagen negativa (NI) y sugestionabilidad. Para su interpretación, deben considerarse puntuaciones de 6 o superiores como sospecha de simulación. Otros estudios sitúan este punto de corte en 5 con mejores indicadores de especificidad y sensibilidad (Montes y Guyton, 2014). Vitacco *et al.* (2007) indican que su aplicación debe hacerse con precaución ya que las escalas individuales presentaban problemas con la homogeneidad y estimadores de utilidad más bajos que la puntuación total. De la misma forma, presenta una ratio de falsos positivos del 10%, cifra que consideran “inaceptable” (p.258). El reciente meta-análisis de Detullio *et al.* (2019) indicó que, si bien el instrumento presenta una magnitud del

efecto elevada en general, los resultados de los estudios analizados son altamente heterogéneos debido a las diferentes metodologías empleadas.

Entrevista clínico-forense (Arce y Fariña, 2001; Vilariño *et al.*, 2013). Probablemente, la entrevista clínico forense de Ramón Arce y Francisca Fariña sea uno de los instrumentos más conocidos y utilizados en el contexto forense español. Los autores defienden el hecho de que, en el contexto clínico, no se sospecha del fingimiento o simulación, por lo que la principal herramienta utilizada, la entrevista clínica, no incluye ningún mecanismo estandarizado para su detección (Arce *et al.*, 2015; Rogers, 2008). Para ello, Arce y Fariña (2001) propusieron la entrevista clínico-forense, la cual somete a los evaluados a “una tarea de conocimiento; no contamina la prueba con preguntas del entrevistador; posibilita el establecimiento de un nexo causal entre los hechos a enjuiciar y el daño psicológico, y contiene una herramienta estandarizada para el diagnóstico diferencial de simulación” (Arce *et al.*, 2015, p. 74).

Esta entrevista no se realiza en forma de interrogatorio cerrado, ni es directiva, sino que está orientada a la reinstauración de contextos, utilizando un procedimiento abierto y en formato de discurso libre (Seijo *et al.*, 2014). Ha sido validada en múltiples ocasiones para la detección de la simulación (por ejemplo Arce *et al.*, 2009, 2013; Vilariño *et al.*, 2013) y utiliza para tal fin las estrategias propuestas por Rogers (2008): no cooperación con la evaluación, síntomas sutiles, síntomas improbables, síntomas obvios, síntomas (cuasi) raros, combinación de síntomas, severidad de síntomas, inconsistencia entre los síntomas informados y los síntomas observados, estereotipos erróneos y presentación indiscriminada de síntomas (Seijo *et al.*, 2014). Si bien su uso ha sido ampliamente reconocido, los autores advierten que no debe utilizarse como una única herramienta para la valoración de la simulación debido a que produce una tasa de falsos positivos de entre el 3 y el 5% (Arce *et al.*, 2015).

1.3.2. Instrumentos psicométricos para la evaluación de la simulación/fingimiento.

Al igual que la entrevista, los instrumentos psicométricos constituyen otro de los grandes recursos de los que se dispone en la evaluación de múltiples constructos en las Ciencias de la Salud en general. Cuando estos constructos no son fácilmente observables y pueden considerarse originalmente subjetivos, este tipo de herramientas nos permiten medirlos y objetivarlos. El uso de instrumentos psicométricos, en especial los autoinformes, es uno de los métodos más utilizados para detectar el fingimiento o la simulación (Acka *et al.*, 2020). Además de la clasificación que realizábamos con anterioridad (Tipos A, B y C), otra forma de categorizar estas herramientas es mediante el uso de los términos *Symptom Validity Test (SVT)*, *Self-report Symptom Validity Test (SRSVT)* y *Performance Validity Test (PVT)*, lo que se traduciría como *Test de Validez de Síntomas*, *Test de Validez de Síntomas autoinformado* y *Test de Validez del Rendimiento*. La primera categoría, SVT, es la categoría general y hace referencia a medidas integradas en un instrumento, como las escalas de validez integradas del MMPI, o a un instrumento en sí mismo, como el SIMS (Acka *et al.*, 2020). Dicha categoría se divide en las otras dos mencionadas, las SRSVT, que evalúan si la presentación de los síntomas por parte del evaluado refleja su verdadera experiencia de los mismos y las PVT, que estudian la presencia de bajo rendimiento en una serie de tareas neurocognitivas (Larrabee, 2015). Esta terminología es utilizada principalmente en la literatura inglesa y, a día de hoy, no tenemos constancia de que sea utilizada en la española, donde sencillamente se habla de “autoinformes” (por ejemplo, González-Ordi *et al.*, 2012).

A continuación, expondremos aquellos instrumentos utilizados en el campo del fingimiento/simulación que cuentan con mayor evidencia empírica y que han sido adaptados a población española. En otro apartado adicional, mencionaremos aquellos que disponen de un menor número de evidencia y que, si bien no cuentan con adaptación

española, han resultado ser prometedores y sería de interés trabajar en su validación en futuras publicaciones. La gran mayoría de las herramientas que expondremos a continuación pertenecen a la categoría SVT integrado o SRVT, siendo PVTs únicamente aquellos que miden variables neuropsicológicas como la memoria.

Inventario Multifásico de Personalidad de Minnesota (MMPI; Ben-Porath y Tellegen, 2008; Butcher *et al.*, 1989), validado en España por Santamaría Fernández (2009). Es un instrumento autoinformado de tipo B, diseñado para la evaluación de variables psicopatológicas y relacionadas con la personalidad. En su versión revisada (Ben-Porath y Tellegen, 2008) está compuesto por 338 ítems y 50 escalas, “42 escalas sustantivas que se engloban en cuatro grandes categorías (escalas de segundo orden, escalas clínicas reestructuradas, escalas de problemas específicos, escalas de intereses y las cinco de personalidad psicopatológica) y por 8 escalas de validez” (Blasco Saiz y Pallardó Durá, 2013, p. 179). Entre las citadas escalas de validez han sido utilizadas para evaluar fingimiento:

La **Escala F**, la cual evalúa síntomas casi-frecuentes con ítems sobre síntomas extravagantes o inusuales en psicopatología severa (ideación paranoide, procesos psicóticos, esquizoides, falta de implicación, cinismo, actitudes antisociales, hostilidad familiar y síntomas somáticos). Altas puntuaciones (por encima de 90 en puntuaciones T) en esta escala se asocian con el reconocimiento de un amplio rango de síntomas poco habituales en población general. Además de esto, las puntuaciones altas pueden señalar problemas al comprender o rellenar el cuestionario, o psicopatología grave de corte psicótico. Para concluir sobre cuál es el motivo de estas puntuaciones, el evaluador debe descartar la presencia de problemas al cumplimentar el cuestionario y analizar las escalas VRIN (Inconsistencias en las respuestas variables con valor crítico en 70 o más) y TRIN (Inconsistencias de las respuestas verdaderas, con valor crítico en 79 o superiores). Tras

esto, se valorará la posible psicopatología teniendo en cuenta el historial, con entrevistas o como el profesional considere oportuno (González-Ordi *et al.*, 2012). Tradicionalmente, la escala *F* ha destacado por obtener las magnitudes del efecto más grandes en las investigaciones, no obstante, se considera que varios de los ítems que la componen no pueden discriminar adecuadamente entre fingimiento y respuesta honesta, ya que fue diseñada únicamente para detectar un estilo de respuesta atípico (Walczyk *et al.*, 2018).

La **escala *Fb***, formada por 40 ítems redactados con la misma intencionalidad que los de la escala *F*, pero relacionados principalmente con trastornos afectivos graves (pánico o miedo, depresión, baja autoestima, suicidio o autolesión, abuso de alcohol, separación familias y 3 ítems de procesos psicóticos). Como señalan González-Ordi *et al.*, 2012, para puntuaciones altas en estas escalas se deben tener las mismas consideraciones que en la anterior. Sin embargo, si tanto la escala *F* como la *Fb* muestran puntuaciones superiores a T 90 debe sospecharse de un patrón de exageración de síntomas. Si, por el contrario, esto sucede solamente en la escala *Fb*, puede indicar una exageración de un trastorno afectivo emocional o un individuo genuino grave. Al igual que ocurría con la anterior, esta escala valora alteraciones en los estilos de respuesta y no discrimina entre fingimiento y respuesta honesta (Walczyk *et al.*, 2018).

Las otras cuatro escalas de validez también utilizadas para detectar fingimiento son **Respuestas psicopatológicas infrecuentes (Fp-r)**, **Respuestas somáticas infrecuentes (Fs)**, **Validez de síntomas (FBS-r)** y **Escalas de sesgo de respuesta (RBS)**. La primera mide respuestas infrecuentes, indicando una respuesta elevada un intento por parte de la persona de exagerar psicopatología y mostrar una impresión negativa de sí misma; la segunda mide ítems que son infrecuentes en población médica, considerando que la presentación de un amplio número de estos síntomas puede indicar fingimiento; la tercera sigue la misma premisa que la anterior y fue diseñada para analizar

síntomas somáticos y cognitivos que pocas veces aparecen en reclamaciones de lesiones personales; y, por último, la cuarta estudia posibles quejas no creíbles sobre problemas de memoria (Sánchez *et al.*, 2017). El meta-análisis de Ingram y Ternes (2016) encontró evidencia de que estas subescalas discriminaban adecuadamente entre simuladores y personas con respuesta honesta con magnitudes del efecto elevadas, con un rango de 1.08, para **FBS-r**, a 1.43 para **FP-r**. Los autores señalan que las citadas variables pueden verse afectadas por el efecto de variables moderadoras comunes y específicas del contexto, como la persona evaluada o las agrupaciones de las comparaciones, siendo FBS-r y RBS las que menos afectadas se ven.

La reciente publicación de Sánchez *et al.* (2017) aporta resultados coherentes con el meta-análisis de los autores anteriores, observándose que las escalas diferenciaban apropiadamente entre grupos, siendo las escalas FP-r y F-r las que mayor capacidad discriminativa presentaban. A pesar de ello, los autores advierten que la ocurrencia de falsos positivos y negativos es demasiado elevada, y consideran que el uso de estas escalas es insuficiente para la práctica forense.

Inventario de Evaluación de la Personalidad (PAI; Morey; 1991, 2007) El PAI es un instrumento autoinformado de tipo B diseñado para la valoración multidimensional de la personalidad. Se compone de 344 ítems divididos en 18 escalas, de las cuales 11 son clínicas, 5 de tratamiento y 2 interpersonales, subdividiéndose cada una de ellas en aspectos relacionados con el constructo evaluado siguiendo la estructura del DSM. Fue adaptado al español por Ortiz-Tallo *et al.* (2011). Según González-Ordi *et al.* (2012) entre sus principales fortalezas se encuentra un reducido tiempo de aplicación, no se necesita una comprensión lectora, ausencia de solapamiento entre ítems y escalas, uso de 4 categorías de respuesta, concordancia con la terminología diagnóstica y fácil

interpretación. Además, posee escalas e indicadores específicos de la evaluación de respuesta inconsistente, minimización y exageración de síntomas. Estas son:

- Inconsistencia - INC. Evalúa la respuesta inconsistente a lo largo de toda la prueba mediante ítems relacionados.
- Infrecuencia – INF. Explora la respuesta azarosa o peculiar mediante ítems neutrales con respecto a la psicopatología.
- Impresión negativa – IMN. Señala si el sujeto se ha mostrado extremadamente desfavorable exagerando síntomas.
- Índice de simulación – SIM. Indica si se han mostrado características más comunes de los simuladores que de población genuina.
- Función discriminante de Rogers – FDR. Utilizada para maximizar las diferencias entre simuladores instruidos y pacientes genuinos.
- Impresión positiva – IMP. Evalúa si se ha presentado una imagen excesivamente favorable.
- Índice de defensividad – DEF. Señala la presencia de las ocho características más comunes que muestran las muestras de simuladores.
- Función discriminante de Cashel – FDC. Utilizada para calcular la tendencia de los simuladores a presentarse idealmente.

Las puntuaciones T inferiores a 73 en el IMN indican respuesta honesta. Las puntuaciones que se encuentran entre el rango 73 y 84 sugerirían precaución en la interpretación de las escalas clínicas por posible simulación. En el rango de 81 a 101 se encontrarían las puntuaciones de aquellos evaluados con alta distorsión en la respuesta y tras estos, ya por encima de 101, se encontrarían los sujetos que han presentado

indicadores claros de distorsión de la imagen o exageración. Sin embargo, estudios recientes recomiendan que el punto de corte de alto riesgo debe situarse en 110 siendo las puntuaciones localizadas entre 73 y 110 catalogadas como sospechosas de simulación (Hawes y Boccaccini, 2009). Por otro lado, el Índice SIM (Morey, 1996) es un indicador que funciona con cierta independencia del nivel de psicopatología mostrada y estudio características comunes en simuladores. El punto de corte de este índice en 82 como indicador de exageración de síntomas y puntuaciones superiores a 107 señalarían alta probabilidad (Hawes y Boccaccini, 2009). González-Ordi *et al.* (2012) indican que, para el estudio de la simulación, se ponderan las puntuaciones de 20 escalas del PAI. Valores superiores a 0 advierten de posible exageración. Aquellos superiores a 1.8 como alta probabilidad. Para la interpretación conjunta de esta junto a IMN y SIM se debe de tener en cuenta que, existirá bajo riesgo cuando $IMN < 73$, $Sim < 82$ y baja puntuación en FDR. Deberemos de sospechar de simulación cuando $IMN > 92$, $SIM > 94$ y tengamos alta puntuación en FDR. La probabilidad de exageración de síntomas será alta cuando $IMN > 110$, $SIM > 107$ y $FDR > 1.8$ (Rogers, 2008).

Inventario Estructurado de Simulación de Síntomas (SIMS; Widows y Smith, 2005). El SIMS es un autoinforme de tipo A, compuesto por 75 preguntas dicotómicas de verdadero o falso, y fue diseñado para detectar simulación/fingimiento de sintomatología psiquiátrica y problemas cognitivos. Fue adaptado a población española por González-Ordi y Santamaría (2009). Se construyó bajo la idea de que los simuladores tienden a expresar una mayor cantidad de síntomas y/o afirman padecer síntomas raros, atípicos o extremos (Smith y Burger, 1997; van Impelen *et al.*, 2014). Diversos trabajos han mostrado que las propiedades psicométricas y los estimadores de utilidad de la escala son adecuados ($\alpha = .96$; sensibilidad = 93%; especificidad = 98%) y ha resultado ser eficaz

diferenciando entre honestos y simuladores (Rogers y Benders, 2018; van Impelen *et al*, 2014).

Los resultados del instrumento se plasman en una escala general de riesgo de fingimiento y/o simulación y cinco subescalas que tienen como finalidad indicar al evaluador en que área puede estar ocurriendo la simulación/fingimiento. La primera de estas escalas (Psicosis – PS) explora presencia de síntomas psicóticos inusuales o atípicos. La segunda (Deterioro Neurológico – Dn) se centra en síntomas atípicos o ilógicos de corte neurológico. En tercer lugar, evalúa inconsistencias en problemas de memoria producidos por disfunción o daño cerebral (Trastornos amnésicos – Am). En cuarto lugar, estaría la escala centrada en exageración del déficit intelectual (Baja inteligencia – Bi). Finalmente, la escala que evalúa síntomas atípicos relacionados con depresión y ansiedad (Trastornos afectivos – Af). Aunque cuenta con un amplio respaldo empírico tanto es su versión original como la adaptación que se realizó en población española (González-Ordi y Santamaria, 2009), siempre es recomendable acompañar el cuestionario con una entrevista de evaluación. La publicación original de Smith y Burger (1997) establecía un punto de corte de 14 para la escala general y, posteriormente, se propuso un punto de corte de 16 (Rogers *et al.*, 1996), punto que también se recomendó en la adaptación español. Los indicadores obtenidos en dicha adaptación fueron excelentes ($\alpha = .94$; sensibilidad entre el 94% y el 100%, y especificidad entre el 54% y 98%. Para su interpretación, González-Ordi y Santamaría (2009), exponen que, cuando la puntuación total es menor o igual a 16 se podrá descartar la exageración de síntomas. Cuando es superior a 16 se puede considerar sospecha de exageración y deberá realizarse un diagnóstico diferencial con pruebas complementarias. Finalmente, cuando la puntuación es superior a 25 se podrá tener alta sospecha de simulación, aunque al tratarse de un instrumento de *screening* debe realizarse la subsiguiente valoración pormenorizada.

Van Impelen *et al.* (2014) consideran en su meta-análisis del instrumento que presenta fortalezas como una alta sensibilidad, consistencia interna sólida, un bajo tiempo de administración (15 minutos aproximadamente) y una interpretación sencilla, pero señalan que la alta variabilidad en la especificidad observada en los estudios (desde el 23% al 100%) es una cuestión problemática que debe tenerse en consideración.

Inventario Clínico Multiaxial de Millon III (MCMI-III; Millon, 1994, Millon *et al.*, 1997, 2009). Se trata de un instrumento autoinformado de tipo B, compuesto 175 ítems diseñados para evaluar síndromes clínicos y personalidad patológica mediante 28 subescalas. Ha sido adaptado al español por Cardenal y Sánchez (2007). El instrumento incluye 3 escalas de validez diseñadas para medir sinceridad, simulación y disimulación. Walczyk *et al.* (2018) lo incluyen en su listado de instrumentos comentando que las escalas de validez pueden indicar la presencia de simulación, no obstante, la evidencia que respalda su utilidad para tal fin es limitada. Los estudios de Daubert y Metzler (2000), Schoenber *et al.* (2003) y Aguerrevere *et al.* (2011) indican que su utilidad clínica en este aspecto es reducida, ya que en ambos trabajos se obtuvieron valores generales de sensibilidad por debajo del 60% y de especificidad entre el 66.3% y el 96%. De la misma forma, estudios como el de Melton *et al.* (2007), Morgan *et al.* (2002), o Sellbom y Bagby (2008) desaconsejan su uso, hasta tal punto que los autores expresan explícitamente que bajo ninguna circunstancia se debería utilizar en el contexto forense con la finalidad de evaluar estilos de respuesta.

Test de Simulación de Problemas de Memoria (TOMM; Tombaugh, 1996), adaptado a población española por Vilar-López *et al.* (2011). Se trata de un PVT de tipo A diseñado para discriminar entre pacientes que presentan problemas de memoria y personas que simulan dichos problemas. Como expone Frederick (2018), el instrumento está compuesto por 50 ítems, en forma de líneas de dibujos simples, que tienen que

memorizarse en 3 segundos, con un intervalo de un segundo entre exposiciones. Una vez expuestos, estos ítems se presentan a continuación en pares y se pregunta al evaluado/a cuáles ha visto con anterioridad (Fase 1). Inmediatamente después, el proceso se repite, pero se modifica el orden de presentación (Fase 2) y, una vez han transcurrido 15 minutos, el/la evaluador/a dispone de una tercera fase opcional que repite el mismo proceso (Fase de retención). La publicación original de Tombaugh (1996) utilizó como punto de corte el 90% de aciertos (<45), lo que produjo una sensibilidad y especificidad apropiada (82% y 93% respectivamente para la Fase 2 y 89% - 97% respectivamente para la Fase de retención). A pesar de ello, Frederick (2018) indica que no se realizó una validación cruzada, (entendiendo como validación cruzada el uso de un diseño de grupo conocido, el cual se explicará en el siguiente apartado) y, en los subsiguientes estudios realizados aplicando este método, como el de Greve *et al.* (2006) o Schroeder *et al.* (2013), se observó que, con el punto de corte original, los resultados obtenidos eran mediocres, con una sensibilidad especialmente baja (por debajo del 50%), lo que indicaba que el punto de corte requería ser ajustado para maximizar los estimadores de utilidad. Así, las publicaciones de los anteriores autores estimaron que el punto de corte óptimo era ≤ 48 , ya que se mantenía la especificidad en el 90%, se incrementaba la sensibilidad al 61% en la Fase 2, y al 68% en la Fase de retención. Para mejorar la capacidad discriminativa del instrumento, además de estudiar el punto de corte óptimo, Gunner *et al.* (2012) desarrollaron el *Índice de Consistencia Albany* (Albany Consistency Index o ACI), el cual se calcula mediante la suma de los ítems que se responden de forma inconsistente a lo largo de las tres fases del instrumento. En la publicación original de los autores, con un punto de corte ≥ 10 se alcanzaba una sensibilidad del 71% y una especificidad del 100%. Schroeder *et al.* (2012) estudiaron dicho índice y encontraron valores similares, aunque algo inferiores (sen. 65 y espec. 97). De la misma forma, el punto de corte de 48 arrojó

resultados similares a los obtenidos en anteriores publicaciones, con una sensibilidad y especificidad del 65 % y del 97% para la Fase 2, y del 77 % y del 92% respectivamente para la Fase de retención.

Debido a que la estimación del punto de corte óptimo ha sido una línea de investigación que ha producido una cantidad significativa de evidencia, recientemente Martin *et al.* (2019) realizaron una revisión sistemática y meta-análisis de la cuestión. Los autores indican que con un punto de corte <49, el instrumento mantiene una especificidad adecuada en las Fases 1 y 2 (91% y 97% respectivamente) y una sensibilidad moderada (59% y 70% respectivamente), considerando que es el punto de corte más apropiado, especialmente para aquellos contextos con una alta tasa de invalidez.

Otros instrumentos para la detección de la simulación. Como se puede observar, todos los instrumentos incluidos en el anterior listado pertenecen a los tipos A y B. De la misma forma, prácticamente todos los instrumentos han sido diseñados para medir variables psicopatológicas, cognitivas o neuropsicológicas. Los instrumentos de tipo C han sido estudiados, en su gran mayoría, en otras áreas de interés, como el dolor o las lesiones personales, donde no se dispone, a día de hoy, de herramientas con las características de las de tipo A y B. Entre los principales instrumentos señalaremos el *Modified Somatic Perception Questionnaire* (MSPQ; Main, 1983) y el *Pain Disability Index* (PDI; Pollard, 1984), validados ambos por Bianchini *et al.* (2014), Crighton *et al.* (2014) y Larrabee (2003); la *Escala de Catastrofización del Dolor* (PCS; Sullivan *et al.*, 1995), validada por Curtis *et al.*, (2019); y el *Cuestionario de Dolor Cervical* (NPQ; Leak *et al.*, 1994) y el *Cuestionario de Salud SF-36* (Ware, 2000), validados por Capilla Ramírez *et al.* (2014).

El MSPQ y el PDI son dos autoinformes breves, habiéndose diseñado el primero para medir percepción autonómica y somática en pacientes con dolor crónico y el segundo

para medir las áreas particulares de la vida del paciente que han sido afectadas por un trastorno relacionado con el dolor (Crighton *et al.*, 2014). Ambos instrumentos fueron validados como herramientas de *screening* bajo la premisa de que instrumentos de medida breves, pueden ser de especial utilidad en el contexto médico-legal, ya que herramientas como el MMPI-2 tienen un tiempo de administración demasiado elevado (Bianchini *et al.*, 2014). En términos generales, se observan diferencias entre los resultados de las tres publicaciones. El estudio de Larrabee (2003) estableció para el MSPQ, con los puntos de corte 10, 11 y 14, una sensibilidad moderada-elevada (90%, 86% y 69% respectivamente), y una especificidad elevada (90%, 95% y 99%), y para el PDI, con los puntos de corte 54, 57 y 63, una sensibilidad moderada-baja (59%, 38% y 10%), y una especificidad elevada (90%, 95% y 99%). Con esos mismos puntos de corte, Crighton *et al.* (2014) obtuvieron valores inferiores en ambas escalas, presentando el MSPQ una sensibilidad del 84%, 81% y 71% y una especificidad del 55%, 61% y 73%, y el NDI una sensibilidad del 59%, 41% y 31% y una especificidad del 67%, 74% y 91%. Bianchini *et al.* (2014), por su parte, indican que su tasa de falsos positivos fue significativamente más elevada que la de Larrabee (2003), ya que con un punto de corte superior a 10 la citada tasa era del 32%, frente al 10% del anterior autor. Establecen que puntuaciones por encima de 13 en el MSPQ, con una sensibilidad en el citado punto de corte del 61% y una especificidad del 88%, y por encima de 57 en el PDI, con una sensibilidad del 54% y una especificidad del 89%, pueden considerarse indicadores aceptables de simulación.

Por otro lado, el PCS, es un instrumento diseñado para estudiar el constructo de “catastrofización del dolor”, entendido como el desarrollo de un estado mental catastrofista acontecido durante el transcurso de una experiencia dolorosa (Sullivan *et al.*, 2001). Curtis *et al.* (2019) obtuvieron una sensibilidad del 45% y una especificidad del 93% con un punto de corte ≥ 81 . Los autores indican que era 7 veces más probable que

aquellas personas que obtuvieron puntuaciones superiores al citado punto de corte pertenecieran al grupo de los simuladores y establecieron que el porcentaje global de clasificación era del 74%. Dicho porcentaje es similar al alcanzado por Capilla Ramírez *et al.* (2014) en su estudio del NPQ y el SF-36, donde ambos instrumentos alcanzaron un porcentaje de clasificación global del 76%. El primer instrumento es un autoinforme diseñado para medir cómo afecta el dolor a las áreas de la vida cotidiana y, con un punto de corte ≥ 21 , obtuvo una sensibilidad del 73% y una especificidad del 80%. El segundo, también un autoinforme, fue diseñado para valorar estados positivos y negativos de la salud y, con un punto de corte de ≤ 35 , obtuvo los mismos valores de sensibilidad que el anterior instrumento.

Para finalizar este apartado, consideramos oportuno hacer una breve reseña de los instrumentos y enfoques desarrollados recientemente en la bibliografía inglesa y que, bajo nuestro punto de vista, pueden resultar de gran interés para futuras investigaciones, especialmente para su validación en población española.

Probablemente, uno de los SRSVT de Tipo A más prometedores es el ***Self-Report Symptom Inventory*** (SRSI; Merten *et al.*, 2016). Se trata de un instrumento diseñado para detectar fingimiento de síntomas, compuesto por 107 ítems dicotómicos en formato verdadero o falso divididos en dos escalas, pseudosíntomas y síntomas genuinos. Cada una de las citadas escalas se divide en cinco subescalas que estudian síntomas plausibles o implausibles en los dominios clínicos, psicológicos y cognitivos (Acka *et al.*, 2020). Ha demostrado ser especialmente útil en el contexto civil y, en términos generales, tiene una capacidad de clasificación del 80%. Los autores advierten que aún resulta necesario disponer de más evidencia empírica para poder recomendar su uso en la práctica clínica, pero se trata del primer instrumento desarrollado específicamente para evaluar la posibilidad de fingimiento que incluye escalas relacionadas con el dolor.

Por otro lado, van Impelen *et al.* (2017) desarrollaron la ***Symptom and Disposition Interview (SDI)***, entrevista diseñada como *screening* de fingimiento, pero que incluye un análisis de los incentivos externos potenciales. Está compuesta por 42 ítems divididos en una escala de validez de síntomas, una escala clínica regular, una escala que analiza la necesidad de asumir el rol de enfermo y una escala que mide ansiedad relacionada con la enfermedad y sintomatología somatomorfa (van Impelen *et al.*, 2017). Como indican los autores, su principal fortaleza es que se trata del primer instrumento que incluye una aproximación psicométrica para el estudio de los citados incentivos externos. Los resultados obtenidos indican que la escala es capaz de discriminar significativamente entre grupos, con un Área Bajo la Curva que varía entre .65 y .80, y señalan que la evaluación de los incentivos puede ser incorporada en un SVT sin sacrificar su capacidad para detectar fingimiento.

2. Métodos de investigación en el campo de la simulación

2.1. Diseños de investigación

Investigar en el campo del fingimiento y la simulación es una tarea compleja debido a que disponer de un grupo de simuladores reales sobre los que se tenga total certeza es muy poco frecuente. En condiciones normales, deben plantearse otras opciones para la formación de los grupos que sean capaces de reflejar fielmente las características de la población de interés (Rogers y Cavanaugh, 1983). Los diseños de investigación utilizados tradicionalmente no cubren la totalidad de las particularidades del contexto, por lo que prácticamente desde el inicio de la disciplina se han propuesto y estudiado diseños específicos para el mismo. Rogers y Bender (2012) exponen que, a día de hoy, se dispone de un total de cinco diseños de investigación: diseño de simulación, diseño de grupo conocido, diseños autosuficientes, diseño de criterio parcial y diseño de prevalencia diferencial, detallándose los mismos a continuación.

2.1.1. Diseño de simulación

El diseño de simulación es el más utilizado en la investigación en la temática, y fue uno de los primeros planteados para dar respuesta a la necesidad de disponer de un grupo de simuladores (Rogers y Bender, 2018). Se trata de un diseño experimental en el que se asigna a los participantes, habitualmente no clínicos o asintomáticos, a dos condiciones, una de control, donde se debe responder honestamente, y otra de simulación, donde se le pide al participante que finja padecer una condición específica (Rogers y Gillard, 2011). Entre las ventajas que presenta este diseño es que tiene una elevada validez interna, por lo que resulta más configurable y controlable, y es más accesible para los investigadores por el tipo de muestra que utiliza (González-Ordí *et al.*, 2012).

El principal problema que presenta este diseño es que cuando las comparaciones se realizan con participantes no clínicos, la validez externa disminuye y la generabilidad de los resultados es desconocida (Rogers *et al.* 1993). Como exponen Rogers y Cavanaugh (1983), en este tipo de comparación se produce una paradoja, ya que se pide a sujetos que obedezcan instrucciones de falsear, con la finalidad de estudiar a personas que falsean cuando tienen que obedecer. De la misma forma, se presenta el problema de que los sujetos a los que se les pide simular difícilmente alcanzan el mismo nivel de motivación por fingir que los simuladores “genuinos”, ya que no disponen del posible incentivo psicolegal o socioeconómico de estos últimos (Rogers y Cavanaugh, 1983). Además, una comparación como la planteada no aporta un valor real a la investigación, ya que no permite analizar las diferencias entre simuladores genuinos y pacientes clínicos (González-Ordi *et al.*, 2012). Para solventar parte de esta problemática, se incluyó en el diseño un tercer grupo compuesto por una muestra clínica de referencia, formada por pacientes clínicos genuinos que hayan sido diagnosticados con la condición a estudiar (McCaffrey y Weber, 2000). Siguiendo a Rogers y Bender (2018), la inclusión de esta muestra clínica es fundamental, ya que la comparación cuasiexperimental entre simuladores instruidos y pacientes clínicos es la que aportará información relevante y, de no incluirse, el estudio debe considerarse “fatalmente defectuoso” (p. 594).

Además de disponer de una muestra clínica de referencia, la adecuada preparación de los participantes asignados a la condición de simulación es un factor de gran importancia en este diseño (Rogers y Bender, 2012). Para ello, los elementos que deben tenerse en cuenta son:

Uso de incentivos y motivación. Los diseños de simulación pueden ser fortalecidos significativamente mediante la inclusión de diversos métodos para motivar a los participantes, así como la valoración de la efectividad de los incentivos externos

(Rogers y Bender, 2018). Al igual que ocurre en condiciones reales, no todos los participantes estarán igual de motivados por la posible compensación o recompensa que vayan a recibir (Rogers y Granacher, 2011). Así, Rogers y Gillard (2011) exponen que, para lograr que los participantes se impliquen puede utilizarse la motivación interna y los incentivos externos. Dentro de los primeros, habitualmente se utiliza la motivación positiva, mediante los cuales se plantea un desafío y se reta al participante a superarlo con preguntas como “¿Eres capaz de superar el test?” o “¿Eres lo suficientemente bueno para pasar la prueba?, creando así la motivación interna de ser un “ganador” (Rogers y Bender, 2018). Este último autor señala que también existe la posibilidad de utilizar motivación negativa, (por ejemplo “Al finalizar este estudio te diremos si has fallado o has conseguido superarlo. Por favor, haz todo lo posible por evitar fallar” p. 598) pero a día de hoy no se dispone de evidencia que valore su efecto.

En lo relativo a los segundos, los incentivos externos positivos, suelen aplicarse ofreciendo a los/as participantes una recompensa en forma de pequeña cantidad de dinero o de bono de puntuación en una asignatura. Debido a que los investigadores tienen límites éticos y económicos para ofrecer sumas de dinero que podrían resultar más motivadoras, se recomienda que los incentivos externos se utilicen junto a los motivadores internos detallados con anterioridad, ya que puede incrementar significativamente su eficacia (Gonzalez-Ordi *et al.*, 2012, Rogers y Bender, 2018). Por otro lado, el investigador también puede considerar el uso de incentivos externos negativos. Rogers y Cruise (1998) realizaron un estudio en el que se incluían dos grupos experimentales, uno con incentivos positivos, como los detallados con anterioridad, y otro con incentivos negativos, donde se les avisaba de que, si fallaban, no recibirían el bono en la asignatura que se ofrecía de recompensa y, además, sus nombres se publicarían en el tablón de la facultad -propuesta realizada originalmente por Patrick e Iacono (1989)-. Los resultados obtenidos indicaron

que los participantes que tenían incentivos negativos estuvieron más centrados en la tarea asignada. Peace y Richards (2014), en un experimento similar, encontraron resultados parecidos a los de los anteriores autores, ya que aquellos participantes que tenían incentivos negativos (pérdida de recompensa y publicación de nombres) presentaron puntuaciones más bajas en los instrumentos aplicados, obteniendo un perfil cercano al esperable en pacientes clínicos. Los autores consideraron que estos resultados indicaban que, ante la posibilidad de “pérdida”, los participantes se esforzaban en ofrecer una presentación más plausible y realista de la condición asignada. El uso de incentivos negativos puede estar basado únicamente en la pérdida de la recompensa, realizando las indicaciones pertinentes en el consentimiento informado desarrollado y, en caso de querer implementar la “amenaza” de publicar los nombres sin los problemas éticos que supone dicha acción, sencillamente se puede utilizar una lista de nombres falsos (Rogers y Bender, 2018).

Instrucciones. Como señalan González-Ordi *et al.* (2012), el diseño de simulación depende en gran medida de la calidad de las instrucciones que se les proporcione a los participantes. Dichas instrucciones deben ser capaces de replicar de manera precisa las condiciones reales y deben ser claras, motivadoras y concisas. Rogers y Cruise (1998) identificaron seis elementos básicos a considerar para la construcción de instrucciones en el contexto de la simulación:

- **Comprensibles.** Los/as participantes deben ser capaces de comprender correctamente las instrucciones proporcionadas y entender qué se espera de ellos. Los/as investigadores/as tienen a su disposición múltiples herramientas para evaluar la comprensión lectora, destacando el índice Fernández-Huerta (1959) y el índice Flesch-Szigriszt (Szigriszt Pazos, 1993) para textos en español y la

prueba de legibilidad de Flesch-Kincaid (Kincaid *et al.*, 1975) para textos en inglés.

- **Específicas.** Las instrucciones deben ser lo suficientemente específicas para que el/la investigador/a sea capaz de llevar a cabo una interpretación significativa de los datos recogidos (González-Ordi *et al.*, 2012).
- **Adaptadas al escenario o contexto.** Rogers y Gillard (2011) exponen que los/as participantes deberían tener un mínimo de familiaridad con el contexto en el que se realiza el estudio, ya que es posible que no conozcan las posibles particularidades del mismo y no sean capaces de recrear el rol asignado adecuadamente. Rogers y Bender (2018) ponen el ejemplo de una persona a la que se le pide que actúe como miembro de otra cultura diferente a la suya.
- **Motivadoras e incentivadoras.** Como ya exponíamos con anterioridad, la presencia de elementos motivadores e incentivos es fundamental para lograr que el/la participante se implique en la ejecución del estudio.
- **Relevantes.** Otra manera de hacer que los/as participantes se impliquen en la investigación es hacerles ver que los resultados del estudio pueden ser relevantes para ellos, o para sus familiares en un futuro próximo, indicándoles, por ejemplo, que la mejora de los tests para discriminar con precisión entre simuladores y pacientes genuinos puede ser de utilidad para evitar que personas en su misma situación no puedan tener acceso a un servicio debido a una clasificación errónea (Wooley y Rogers, 2015).
- **Creíbles.** Es recomendable avisar a los/as participantes de que la presentación que realicen del rol asignado debe ser lo más creíble posible (por ejemplo, Vilar López *et al.*, 2013). Gorny y Merten (2005) realizaron un estudio en el que incluyeron

cuatro condiciones experimentales, 2 sin la advertencia de resultar creíble y 2 con la misma (grupo C con la advertencia de no exagerar la presentación de síntomas y Grupo D con formación específica para simular correctamente). Los resultados obtenidos indicaron que tanto el Grupo C como el Grupo D fueron capaces de realizar una presentación de la condición mucho más realista que los otros dos grupos. Además de advertir sobre la necesidad de ser creíbles, formar a los participantes (*coaching* en inglés) es otra opción que puede ser considerada a la hora de configurar las condiciones experimentales. Existen dos estrategias para desarrollar esta formación, la basada en la detección, la cual está enfocada a formar a los/as participantes en el funcionamiento de los estilos de respuesta y los tests de validación de síntomas, y la basada en el trastorno, la cual busca preparar a los/as participantes para que conozcan las características y síntomas de la condición que tienen que simular (Jelicic *et al.*, 2011; Rogers *et al.*, 1993).

Manipulation check. Sellbom y Bagby (2008) y Rogers (2008) recomendaron la inclusión de un cuestionario post-test, llamado *manipulation check*², con la finalidad de evaluar si los participantes habían entendido y seguido correctamente las instrucciones que les habían sido otorgadas. De manera específica, este último autor consideró de interés que se valorara la comprensión de las instrucciones, la capacidad de recuerdo y el esfuerzo invertido en desarrollar el rol asignado. Crighton *et al.* (2016), en un estudio donde buscaban analizar si los participantes cumplían las instrucciones, además de examinar el impacto de la disimulación, confirmaron que la propuesta era de utilidad, ya que permitió detectar a los participantes que no cumplieron adecuadamente las instrucciones proporcionadas. Rogers y Bender (2018) desarrollaron un modelo de

² No existe una traducción al español del término, por lo que hemos decidido conservar el original. No obstante, creemos que podría utilizarse “valoración o evaluación de la manipulación”.

manipulation check, el cual exponemos a continuación en la Tabla 2, para diseños de simulación que incluye diversos apartados cuyo uso puede considerarse según las necesidades y características de la investigación a realizar.

Tabla 2. *Manipulation check para diseños de simulación* (adaptado de Rogers y Bender, 2018)

Elementos de la investigación	Preguntas de ejemplo
<u>Instrucciones</u>	
1. Recuerdo de las instrucciones	¿Qué se te pidió que hicieras? ¿Cuáles eran tus instrucciones?
2. Comprensión	¿Qué significan para ti estas instrucciones?
3. Aceptación de las instrucciones	Algunas personas hacen lo que se les dice y otras pueden estar menos motivadas ¿Cuál crees que te describe mejor en esta investigación?
4. Relevancia de las instrucciones	¿Alguna vez te has enfrentado a una situación como la que se describe en las instrucciones? ¿Te imaginas a ti mismo/a haciendo en la vida real lo que se te ha pedido en el presente estudio?
<u>Preparación</u>	
1. Esfuerzo en la preparación	¿Qué has hecho durante tu tiempo de preparación?
<u>Implicación</u>	
1. Motivación interna	(Positiva) ¿Te motivaba el desafío de “ganar” al test? (Negativa) ¿Te motivaba el miedo a no poder “ganar” al test?
2. Incentivos externos	(Positiva) ¿Te esforzaste más por la recompensa ofrecida? (Negativa) ¿Te esforzaste más por el miedo a fallar?
<u>Esfuerzos por disimular</u>	
1. Representación de la condición	¿Cuánto esfuerzo has invertido en ser convincente a la hora de interpretar tu rol?
2. Verosimilitud	¿Te has esforzado en emular “lo suficiente” de la condición para alcanzar el objetivo asignado?

Elementos de la investigación	Preguntas de ejemplo
3. Evitación de los ítems de detección	¿Localizaste algún ítem diseñado para cazar a los simuladores? Si lo localizaste, ¿qué lo delató?
<u>Limitaciones en la participación</u>	
1. Autorrevelación	A pesar de tus esfuerzos por aparentar ser diferente a lo que realmente eres ¿Cuánto de tu “auténtico” tú has mostrado?
2. Interferencia emocional	A veces, las personas pasan por momentos difíciles y creen que es duro participar. ¿Te ha pasado?
3. Propósito del estudio	¿Te preocupaba que el propósito del estudio fuera diferente al que se te dijo?
<u>Autovaloración de los estilos de respuesta</u>	
1. Éxito percibido	¿Cómo de exitoso crees que has sido “engañando” al test?
2. Habilidad para disimular percibida	¿Cómo de bueno eres engañando a los demás?

2.1.2. Diseño de grupo conocido

El diseño de grupo conocido es considerado una versión “más sofisticada” del diseño de simulación (Merten *et al.*, 2013). Supone además el “segundo paso” a seguir en el proceso de investigación, ya que resulta recomendable que los resultados obtenidos en el citado diseño de simulación sean replicados con grupos conocidos antes de ser aplicados en la práctica clínica (González-Ordí *et al.*, 2012). En esta aproximación, se dispone de un grupo de pacientes clínicos que se clasifica en diferentes categorías aplicando una serie de criterios altamente fiables, como pueden ser instrumentos psicométricos validados o expertos ajenos a la investigación (Rogers *et al.*, 1993). Habitualmente, dichas categorías incluyen un grupo de pacientes clínicos, otro de simuladores, el cual se divide en diferentes grados de riesgo de simulación (posible,

probable y definitivo) y otro de “indeterminados”, que suele eliminarse de la investigación (Bianchini *et al.*, 2005; Rogers y Bender, 2012).

Si bien este diseño destaca por su validez externa y por sus resultados altamente generalizables, ya que se dispone de un grupo de “simuladores reales”, la validez interna es limitada, al no poder realizar una asignación aleatoria de las condiciones experimentales (Rogers *et al.* 1993). De la misma forma, presenta la dificultad de que depende de un método fiable y riguroso de clasificación de los grupos (McCaffrey y Weber, 2000), debido a que únicamente pueden utilizarse aquellos mejor validados y que dispongan de ratios de clasificación comprobados empíricamente (Rogers y Bender, 2012). En este sentido, si bien se trata de una cuestión de gran importancia en el diseño, no parece existir consenso sobre qué metodología utilizar para la citada clasificación. Rogers y Bender (2018) señalan que el SIRS-2 es uno de los instrumentos más apropiados para realizar esta tarea y ha sido utilizado en estudios como el de Sellbom *et al.* (2010) o Vittaco *et al.* (2007), pero resulta habitual que cada autor utilice el sistema de clasificación que considere conveniente. Por ejemplo, De Marchi y Balboni (2018) utilizaron el DSM-IV-TR (American Psychiatric Association, 2000) y el SCL-90-R (Derogatis *et al.*, 1973); González-Ordi y Santamaría (2008) utilizaron los criterios de sospecha clínica de presencia o no de posible simulación de Samuel y Mittenberg (2006); Wisdom *et al.* (2010) utilizaron los criterios de Slick *et al.* (1999) expuestos con anterioridad, usando para su valoración una amplia batería de instrumentos compuesta por instrumentos como el *Test de Simulación de Problemas de Memoria* (TOMM; Tombaugh, 1996) o el *Cuestionario de Dolor* de McGill (Melzack, 1975), y el equipo de Bianchini, Greve y Aguerrevere suelen utilizar en sus múltiples publicaciones los criterios de Bianchini *et al.* (2005) con una batería similar a la utilizada por Wisdom *et al.* (2010)

(ver por ejemplo, Bianchini *et al.*, (2014); Curtis *et al.*, 2019; Etherton *et al.*, 2006; Greve *et al.*, 2009; Greve *et al.*, 2010).

2.1.3. Diseño autosuficiente y diseño de criterio parcial

El diseño autosuficiente sigue un procedimiento similar al del diseño de grupo conocido, utilizando indicadores validados, habitualmente instrumentos de medida con especificidad cercana al 100%, para detectar la probabilidad de presentar un estilo de respuesta específico (Rogers y Bender, 2012). Su uso se recomienda principalmente para evaluar la eficacia general de una medida, ya que carece del rigor necesario para ser utilizado en diseños que evalúen estimadores de utilidad, como la sensibilidad o la especificidad (Rogers, 2008). Incluimos también en este mismo apartado el diseño de criterio parcial debido a que está relacionado con el autosuficiente, aunque esta relación ha variado en las diferentes publicaciones en las que se menciona. En el trabajo de Rogers y Bender (2012), los autores indican que el diseño de criterio parcial es uno de los dos diseños “deficientes” utilizados en algunos estudios, ya que utiliza un planteamiento similar al del diseño de grupo conocido o diseño autosuficiente, pero la metodología de clasificación es limitada o incompleta. Así, puede entenderse que este término se utiliza para identificar y definir aquellos diseños que tienen carencias metodológicas. No obstante, Rogers y Bender (2012) indican que se trata de un nombre más descriptivo para los diseños antiguamente conocidos como diseños autosuficientes, y en la cuarta edición del libro de Rogers y Bender (2018) el citado diseño autosuficiente no se incluye en la clasificación, y se repite la indicación de que es el nombre antiguo que se le daba al diseño de criterio parcial.

Así, en las últimas publicaciones de Rogers, la categorización de “deficiente” que se le otorga al diseño de criterio parcial desaparece, al igual que el término “diseño autosuficiente”, y se propone una metodología de investigación que ocupa un punto

intermedio entre el diseño de simulación y el de grupo conocido, en la cual se clasifica a los simuladores en grados de certeza con criterios más flexibles que los utilizados en el citado grupo conocido. Esto hace que el diseño sea más accesible y sencillo de ejecutar, pero disminuye la precisión de clasificación de los grupos, lo que podría producir falsos positivos.

2.1.4. Diseño de prevalencia diferencial

El diseño de prevalencia diferencial es el segundo diseño “deficiente” que señalaban Rogers y Bender (2012). Este diseño asume que determinados grupos, que presentan una característica concreta, van a tener un mayor riesgo de presentar un estilo de respuesta específico (Rogers y Bender, 2012). Por ejemplo, se considera que aquellos pacientes de esguince cervical que opten a una compensación económica van a simular más que aquellos que no, por lo que el estudio se enfoca en comparar dichos grupos. Como exponen Rogers y Gillard (2011), la utilidad de dichas comparaciones es muy poco significativa debido a que se desconoce el porcentaje de simulación que pueden tener los grupos, además de en qué medida las puntuaciones obtenidas representan realmente un intento de simulación. Al igual que el diseño de simulación sin grupo clínico de referencia, este diseño se considera “fatalmente defectuoso” y no debería utilizarse en investigación (Rogers y Bender, 2018).

Una posible solución a las deficiencias metodológicas del diseño de prevalencia diferencial, y una posible alternativa a la clasificación por criterios, fue sugerida por Frederick (2000) en lo que se conoce como “validación de grupo mixto” (del inglés *mixed-group validation*). En este diseño se determina la eficiencia diagnóstica utilizando información específica sobre el grupo, concretamente el conocimiento preciso sobre la proporción de simuladores, en lugar del estado respecto a la condición (Frederick, 2000). Como exponen Rogers y Bender (2018) si bien se trata de un diseño apropiado a nivel

metodológico y matemático, depende de conocer la ratio base de simulación en la población y asume que es consistente, supuestos difícilmente alcanzables debido a que dicha ratio base es altamente variable.

2.2. Análisis de datos en el campo de la simulación

Exponemos a continuación las estrategias de análisis de datos utilizadas habitualmente en los estudios de la temática. Si bien no existe un consenso sobre cuales utilizar, Rogers y Bender (2018) indican que los estudios en el campo de la simulación deberían incluir, como mínimo, un análisis de la magnitud del efecto de las comparaciones y 4 estimadores de utilidad.

2.2.1. Comparación entre grupos y estimadores del tamaño del efecto

El nivel de análisis más básico es el estudio de las comparaciones entre grupos para valorar si existen diferencias estadísticamente significativas. No entraremos en detalle en este tipo de análisis, debido a que es de sobra conocido y no necesita ninguna consideración en especial para el contexto, pero sí consideramos de interés hablar de los estimadores de la magnitud del efecto.

El tamaño efecto es el “reflejo cuantitativo de la magnitud de un fenómeno que se utiliza con el propósito de abordar una cuestión de interés” (Kelley y Preacher, 2012, p.140). Se trata de uno de los resultados más importantes de los estudios científicos y permiten a los/as investigadores/as comunicar en una medida estandarizada la significación práctica de sus resultados (Lakens, 2013). Existen diversos estadísticos para calcular la magnitud del efecto (para una revisión detallada ver Kelley y Preacher, 2012), pero en el contexto de la simulación resulta de especial interés la d de Cohen (1988). Se trata de una de las medidas de magnitud del efecto más empleadas en las investigaciones empíricas y se utiliza para describir la diferencia de medias estandarizada de un efecto

(Ledesma *et al.* 2008). Tiene un rango de 0 a infinito y puede ser utilizada para comparar efectos entre estudios incluso cuando las variables dependientes son medidas de forma diferente (Lakens, 2013). Se calcula con la fórmula $[d=(M_1-M_2)/SD \text{ agrupada}]$, donde M es la media de cada grupo y la DT agrupada es la desviación estándar agrupada, calculada mediante la fórmula $\sqrt{[(DT1^2+ DT2^2) / 2]}$ (Cohen, 1988). Para interpretar el estadístico, se suele utilizar la propuesta del anterior autor, en la que se considera que la magnitud del efecto es pequeña con $d= .2-.49$, moderada con $d= .5-.79$ y alta $d= \geq .80$. No obstante, se considera que estos valores son arbitrarios y resulta recomendable que, en la medida de lo posible, se interpreten con ajuste al contexto y a la literatura previa existente (Thompson, 2007). Así, Rogers *et al.* (2003) propusieron que, debido a que en la literatura de la simulación se producen magnitudes del efecto sustanciales, la interpretación debería ajustarse a los siguientes valores: $d \geq .75$ para moderada, ≥ 1.25 para alta y ≥ 1.75 para muy alta.

2.2.2. Sensibilidad y especificidad

La sensibilidad y la especificidad son dos de los estimadores de utilidad más frecuentes para valorar la capacidad diagnóstica de un instrumento (por ejemplo, Bianchini *et al.*, 2014; Curtis *et al.*, 2019; Etherton *et al.*, 2006; Greve *et al.*, 2010; Greve *et al.*, 2009). Siguiendo las definiciones de Lange y Lippa (2017), la sensibilidad es la ratio de auténticos positivos de un test o medida, y responde a la pregunta de qué porcentaje del grupo de personas que sufren una condición concreta, o en nuestro caso, están clasificados como simuladores, han sido correctamente diagnosticados por el instrumento. Por otro lado, la especificidad es la ratio de auténticos negativos de un test o medida, y responde a la pregunta de qué porcentaje del grupo control (sin la condición estudiada) han sido clasificados correctamente por el instrumento utilizado. Ambos se pueden expresar en porcentaje, con un rango del 0 a 100, o como fracción decimal con

un rango del 0 al 1 y, al restar el valor obtenido al valor máximo, se puede obtener la tasa de falsos negativos, en el caso de la sensibilidad, y de falsos positivos, en el caso de la especificidad. Para calcularlas, se emplean las siguientes fórmulas (Zhou *et al.*, 2002):

- Sensibilidad: $[a / (a + c)]$
- Especificidad: $[d / (b + d)]$

Donde “a” corresponde al número de auténticos positivos (la condición está presente en el sujeto y ha sido clasificado como positivo), “b” al número de falsos negativos (la condición está presente en el sujeto y ha sido clasificado como negativo), “c” al número de falsos positivos (sujeto del grupo control clasificado como positivo) y “d” corresponde al número de auténticos negativos (sujeto de grupo control clasificado como negativo).

Resulta complejo responder a la pregunta de qué valores de sensibilidad y especificidad deben presentar una escala, ya que no existe un criterio consensuado para determinar si un valor es óptimo o no y, habitualmente, se señala que deben ser lo más altos posibles (Florkowski, 2008). Ciertamente, como discute Ramón Arce en varias de sus publicaciones (por ejemplo, Arce, 2017) no debe admitirse en el contexto forense una tasa de falsos positivos elevada debido a las severas consecuencias que una conclusión errónea pueden ocasionar. González-Ordi *et al.* (2012) proponen un gradiente de certidumbre para el contexto de simulación que puede ser utilizado como método de valoración donde se seguiría el siguiente criterio:

- < 10% de falsos positivos – clara respuesta simulada
- < 25% de falsos positivos – probable respuesta simulada
- >25% de falsos positivos – indeterminado

- < 10% de falsos negativos – clara respuesta honesta
- < 25% de falsos negativos – probable respuesta honesta

De esta manera puede entenderse que el valor óptimo de sensibilidad y especificidad para un instrumento desarrollado para detectar simulación será mayor del 90% y el valor aceptable puede ser el 75%. Este tema será analizado con mayor profundidad en la discusión general de la presente tesis doctoral

2.2.3. Poder predictivo positivo y negativo

Si bien la sensibilidad y la especificidad son estadísticos de gran utilidad, no ofrecen información sobre cómo de confiable es el test, sistema o instrumento a la hora de detectar o descartar una condición para una persona valorada en un contexto específico (Lange y Lippa, 2017). Por este motivo, se recomienda que no sean interpretados de manera aislada y se tengan en consideración otras medidas adicionales de utilidad clínica como el *poder predictivo positivo* (PPP) y el *poder predictivo negativo* (PPN) (Trevethan, 2017). El primero está relacionado con la sensibilidad del instrumento y se define como la probabilidad de que una persona realmente tenga la condición dado un resultado positivo (Iverson, 2011). Este estadístico nos informa de cómo de seguro se puede estar de que el instrumento o test realmente sea capaz de predecir lo que pretende predecir, y responde la pregunta de, si una persona ha sido clasificada como simuladora por un instrumento, ¿qué probabilidad hay de que esa persona de verdad sea simuladora? (Lange y Lippa, 2017). Por otro lado, el segundo está relacionado con la especificidad y se define como la probabilidad de que una persona realmente no tenga la condición dado un resultado negativo (Iverson, 2011). El PPP y el PPN son dos de los estadísticos que más valor informativo presentan para la toma de decisiones en el contexto clínico-forense, no obstante, debe tenerse en consideración que dependen de la prevalencia de la condición

(Slick, 2006). Debido a que la prevalencia de la simulación es altamente variable y los datos disponibles para algunos contextos son antiguos, o no se pueden transpolar de una población a otra, el proceder habitual en los estudios en la temática es calcular los dos estadísticos para una prevalencia del 10 al 50 %, o para el 10, 30 y 50 % (por ejemplo, ver van Impelen *et al.*, 2014). Para calcular estos estadísticos se utilizan las siguientes fórmulas (Altman *et al.*, 2000):

- **PPP.** $\text{Sensibilidad} \cdot \text{prevalencia} / \text{sensibilidad} \cdot \text{prevalencia} + (1 - \text{especificidad}) \cdot (1 - \text{prevalencia})$
- **PPN.** $\text{Especificidad} \cdot (1 - \text{prevalencia}) / (1 - \text{sensibilidad}) \cdot \text{prevalencia} + \text{especificidad} \cdot (1 - \text{prevalencia})$

2.2.4. Odds ratio diagnósticos

Si bien los dos anteriores estadísticos ofrecen información de gran utilidad, el hecho de que dependan de la prevalencia supone que, a no ser que se conozca con exactitud, los resultados serán estimados. Como alternativa, Sanchez *et al.* (2017) utilizan los Odds ratio diagnósticos (del inglés *diagnostic odd ratio* o DOR) y consideran que es un mejor estimador ya que no depende de la ratio base. Los DORs se definen como la ratio entre la probabilidad de una clasificación correcta de la condición, simulación en nuestro caso, y la probabilidad de una clasificación incorrecta de la condición (Kraemer, 1992). Se calculan con la siguiente fórmula (Glas *et al.*, 2003):

- **DOR.** $[\text{Sensibilidad} / (1 - \text{sensibilidad})] / [(1 - \text{especificidad}) / \text{especificidad}]$

El valor de los DOR tiene un rango de 0 a infinito, indicando valores más elevados una mayor capacidad discriminativa del instrumento.

2.2.5. Razón de verosimilitud

La razón de verosimilitud es un estadístico utilizado para evaluar la utilidad potencial de un test de clasificación y se define como “la razón de la probabilidad de tener determinado resultado del test en la población con la condición versus tener el mismo resultado en la población sin la condición” (Salech *et al.*, 2008, p. 1206). La razón de verosimilitud positiva es la ratio entre la probabilidad de un resultado positivo dada la presencia de la condición, y la probabilidad de un resultado positivo dada la ausencia de la condición, mientras que la razón de verosimilitud negativa es la ratio entre la probabilidad de un resultado negativo dada la presencia de la condición, y la probabilidad de un resultado negativo dada la ausencia de la condición (Gardner y Greiner, 2006). Como exponen Rogers y Bender (2018), en el contexto que nos ocupa la primera se calcula cuando el resultado del test es positivo e indica una mayor probabilidad de simulación, y la segunda se calcula cuando el resultado del test es negativo e indica una menor probabilidad de simulación. Para calcularlos se utilizan las siguientes fórmulas:

- **Razón de verosimilitud positiva (L+).** Sensibilidad / (1-especificidad)
- **Razón de verosimilitud negativa (L-).** (1-sensibilidad) / especificidad

2.2.6. Razón de verosimilitud encadenada

Larrabee (2008) propuso como método para valorar los resultados del uso conjunto de múltiples indicadores de simulación el encadenado de las razones de verosimilitud de los mismos, de manera que se podría determinar la probabilidad de simulación con uno o múltiples resultados positivos. Para realizar el citado encadenado, el anterior autor calculó el odd pretest de simulación, mediante la conversión de la ratio base de simulación del contexto con la fórmula $RB / (1-RB)$ y lo multiplicó por la razón de verosimilitud positiva del primer instrumento, obteniendo el odd posttest de

simulación. Para comprobar la probabilidad con dos instrumentos se sustituía la ratio base de simulación por el odd posttest del primer instrumento y se multiplicaba por la razón de verosimilitud del segundo instrumento. Con tres instrumentos se multiplicaría el odd posttest de la operación anterior por la razón de verosimilitud del tercer instrumento y así se repetiría el proceso cuantas veces fuera necesario. Para obtener la probabilidad de simulación en cada proceso sencillamente habría que convertir el odd posttest en probabilidad posttest mediante la fórmula $\text{odds posttest}/(1+\text{odd posttest})$. Meyers *et al.* (2014) replicaron la metodología de Larrabee (2008) con 11 instrumentos y llegaron a la conclusión de que el procedimiento era de utilidad y podía ser aplicado tanto en el contexto clínico como en el contexto forense.

Bender y Frederick (2018) analizaron el método utilizado por Larrabee (2008) y llegaron a la conclusión de que, si bien el planteamiento era correcto, la fórmula empleada no lo era, ya que obviaba la razón de verosimilitud negativa de los instrumentos. En términos generales, la lógica puede indicar que el uso de múltiples instrumentos será un enfoque más robusto para detectar simulación, pero lo cierto es que este proceder incrementa significativamente el riesgo de falsos positivos debido a que la ratio de error se acumula con la administración de cada instrumento adicional, por lo que el/la evaluador/a debe utilizar la información de todos los instrumentos para calcular las probabilidades (Bender y Frederick, 2018). Así, estos autores proponen una nueva fórmula en la que se incluye la razón de verosimilitud negativa de cada instrumento que haya obtenido un resultado negativo. De esta manera, el cálculo con un protocolo de, por ejemplo, cinco instrumentos sería:

- $\text{Odds pretest (RB/1-RB)} * \text{LR+} * \text{LR-} * \text{LR-} * \text{LR-} * \text{LR-} = \text{Odds posttest} = \text{probabilidad de simulación (odds posttest}/1+\text{oddpotest})$.

El L+ es la razón de verosimilitud positiva de la escala que se prueba como resultado positivo y el L- es la razón de verosimilitud negativa del resto de instrumentos. Dicha fórmula corresponde a la prueba para un resultado positivo, para calcular con dos resultados positivos se incluirían 2 L+ y 3 L-, para calcular tres resultados positivos se incluirían 3 L+ y 2 L-, etc.

2.2.7. Curva Característica Operativa del Receptor

El análisis de la *Característica Operativa del Receptor* (COR) se utiliza en el contexto clínico para cuantificar con qué precisión un test o sistema diagnóstico puede diferenciar entre dos posibles estados, como simulador y paciente genuino en nuestro caso (Gardner y Greiner, 2006). El análisis COR solventa la principal limitación que presenta el uso de índices asociados al rendimiento del test, como la sensibilidad y la especificidad, que es el hecho de que no ofrecen información sobre el impacto de la alteración en los puntos de corte, lo que puede conducir a una visión sesgada de la capacidad predictiva de los instrumentos analizados (Nicholson *et al.*, 1997). Como expone Hajian-Tilaki (2013), la representación de la sensibilidad contra 1-especificidad toma la forma de una curva, conocida como curva COR, basada en el concepto de valor de decisión de una variable, de manera que la frecuencia de los resultados positivos y negativos de un instrumento o sistema diagnóstico cambiarán si se cambia el citado valor de decisión, o el punto de corte. Dicha curva representa el cambio progresivo en la capacidad discriminativa del test y, a mayor capacidad, más cerca se sitúa la curva de la zona superior izquierda del espacio.

Una forma de sintetizar los resultados del análisis COR es el *Área Bajo la Curva* (ABC o AUC en inglés), la cual refleja la capacidad discriminativa del instrumento a lo largo de todo el rango de posibles puntos de corte (Cerdeja y Cifuentes, 2012). Puede tomar un valor de 0 a 1, indicando 0 que se trata de un test perfectamente impreciso y 1 perfectamente preciso (Mandrekar, 2010). Para su interpretación se considera que un

valor de .50 indica no discriminación, de .70 a .80 aceptable y por encima de .80 excelente (Hosmer y Lemeshow, 2000). Un ejemplo de curva COR de un test con una capacidad discriminativa adecuada sería la que exponemos a continuación:

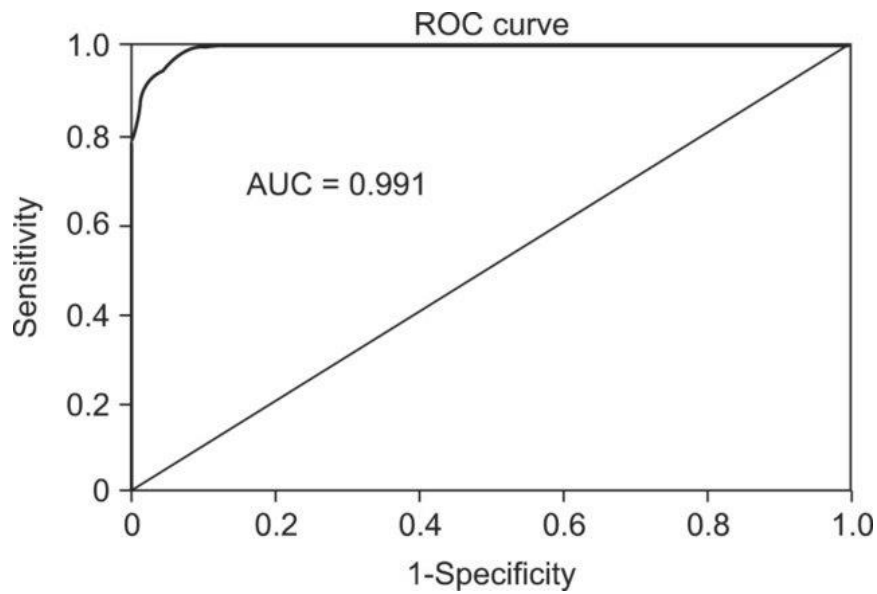


Figura 1. Curva Característica Operativa del Receptor. Reproducido de Kim et al (2014) con permiso de los autores. ROC Curve: Curva COR; Sensitivity: Sensibilidad; Specificity: Especificidad; AUC: Area Under the Curve, Area Bajo la curva.

3. La simulación en el Síndrome de Latigazo Cervical

3.1. ¿Qué condiciones son las más simuladas?

La prevalencia y/o incidencia de la simulación es altamente variable y resulta difícil de determinar (Alozai y McPherson, 2020). Como exponen Resnick y Knoll (2008), la precisión de los datos obtenidos hasta el momento es cuestionable, pues los simuladores exitosos, por pura definición, no se detectan y por lo tanto no están incluidos en los datos obtenidos. De la misma forma, al igual que ocurre en el campo de la detección de mentiras, donde resulta realmente complejo identificar a aquellas personas deshonestas que integran una mentira en un relato verídico (Vrij, 2008), detectar aquellos individuos que simulan leve o parcialmente su condición es una tarea compleja (Bush *et al.*, 2005).

Las posibles deficiencias metodológicas y conceptuales también son una potencial fuente de variabilidad. González-Ordi *et al.* (2012) realizaron una síntesis de las prevalencias registradas en las principales investigaciones en la temática identificadas hasta la fecha, entre las que destacan las de Mittenberg *et al.* (2002), Greve *et al.* (2009) o Larrabee *et al.* (2009), y hallaron los siguientes valores:

- Trastornos psicopatológicos, entre el 13.5% y 32%, teniendo los trastornos somatomorfos alrededor del 31.4%, los trastornos disociativos alrededor del 15.5% y los trastornos por ansiedad o depresión entorno al 13.5% y el 14.9% respectivamente.
- Trastornos de índole neuropsicológica, entre el 11% y el 40%.
- Lesiones corporales, incluido el traumatismo cráneoencefálico leve, entre el 20% y el 38.5%.

- Trastornos relacionados con el dolor, entre el 20% y el 50%, siendo la fibromialgia (25-30%), el dolor crónico en general (25-30%) y esguince cervical (alrededor del 50%) las condiciones con valores más elevados.

Con estos valores se consideró que la ratio base de la simulación en el contexto forense se podía estimar con la conocida cifra de $40 \pm 10 \%$ (Greve *et al.*, 2009), siendo los trastornos relacionados con el dolor los que más prevalencia presentan. No obstante, Young (2014, 2015) realizó un análisis tanto de los estudios incluidos en el libro de González-Ordi *et al.* (2012) como los que se publicaron entre los años 2013 y 2015 (por ejemplo, Axelrod *et al.*, 2014; Guise *et al.*, 2014; Lindley *et al.*, 2014 o Whiteside *et al.*, 2015) e identificó que en la gran mayoría de los estudios se catalogaban a personas como “probable simulador”, habitualmente mediante el uso de criterios de clasificación como los explicados en el apartado anterior, y dicha categoría era incluida en la estimación final junto con la de “simulador definitivo”, produciendo así ratios base excesivamente elevados. Teniendo en consideración únicamente aquellas personas clasificadas como esta última categoría, los valores se reducen significativamente. Por ejemplo, en el estudio de Greve *et al.* (2009) sobre el dolor, un 10.4% de la muestra era clasificada como simulador definitivo, frente al 36% obtenido de las dos categorías combinadas (o 32.4%, según los criterios utilizados). Así, en base a este análisis, Young (2015) estimó que una cifra más realista sobre la prevalencia de la simulación en el contexto forense sería $15 \pm 15 \%$.

Dentro de los aspectos metodológicos también puede considerarse el proceder seguido para la recogida de información como posible fuente de sesgo. Según González-Ordi *et al.* (2012), tradicionalmente se han seguido dos estrategias. En primer lugar, están los estudios que realizan encuestas a profesionales cualificados que se enfrentan a esta problemática, como por ejemplo el de Santamaría Fernández *et al.* (2013), uno de los

últimos estudios realizados en población española, donde se preguntó con un cuestionario estructurado a 161 médicos para valorar qué patologías se percibían como más simuladas, en qué contexto se simulaban, qué mecanismo se utilizaba y qué indicios hacían sospechar de la presencia de simulación. En segundo lugar, están los estudios de revisión de historias archivadas de individuos a los que se les ha sometido a evaluación de la simulación, como el artículo de Greve *et al.* (2009), uno de los artículos de referencia en la investigación de la simulación en dolor, donde se analizaron 508 casos consecutivos relacionados con dolor y con un incentivo económico identificable, que habían sido sometidos a una evaluación psicológica. En función de la estrategia seguida las cifras obtenidas pueden variar significativamente, aún dentro de la misma población.

La alta variabilidad observada en las anteriores cifras impide responder con exactitud la pregunta que da título al presente apartado, “¿qué condiciones son las más simuladas?”. La prevalencia propuesta por Young (2015) hace referencia al contexto forense en general y, a día de hoy, no se ha publicado ningún artículo posterior al del anterior autor, siendo la actualización de estos datos una necesidad que debería ser tratada (Rumschick y Appel, 2018). No obstante, de cara al diseño de una posible línea de investigación, el/la autor/a puede reformular la pregunta y plantear “¿cuál de las condiciones simuladas habitualmente es la más problemática?”, y sin duda la respuesta será los trastornos relacionados con el dolor, debido a que tienen un alto coste socioeconómico, son difíciles de evaluar y aún no existe una metodología específica que permita realizar un diagnóstico diferencial de simulación fiable (Tuck *et al.*, 2018; Boskovic *et al.*, 2017). Por trastornos relacionados con el dolor se entienden todos aquellos que presentan el dolor como uno de los principales síntomas, destacando en el contexto de la simulación la cervicalgia, la lumbalgia, el síndrome de latigazo cervical (o esguince cervical), la fibromialgia y el dolor crónico (Santamaría Fernández *et al.*, 2013).

Como exponen Tuck *et al.* (2018), el dolor se trata de una experiencia subjetiva que no puede ser valorada de forma objetiva, ya que, en algunas condiciones como el esguince cervical, la fibromialgia o el dolor lumbar, no hay una fuente identificable que produzca dicho dolor. Por este motivo, el diagnóstico de este tipo de condiciones suele estar basado en el relato de la víctima, lo que las hace especialmente susceptibles a la simulación (Greve *et al.*, 2009).

En este sentido, también debe considerarse el hecho de que, en países como el Reino Unido, España o Estados Unidos, varias de estas condiciones están asociadas a un proceso de compensación externa, habitualmente en forma de dotación económica, lo que favorece la demanda de indemnizaciones e incrementa tanto el riesgo de simulación real como el riesgo de simulación percibido (Represas *et al.*, 2008). Estas dos circunstancias generan consecuencias severas tanto a nivel general como individual ya que, en el primer caso, provoca un incremento significativo del coste económico y, en el segundo, puede hacer que los profesionales de la medicina adopten una actitud excesivamente precavida al tratar con estos trastornos sin disponer de herramientas eficaces para la detección de la simulación (Bianchini *et al.*, 2005).

Entre los trastornos asociados al dolor, el Síndrome de Latigazo Cervical o SLC (del inglés *Whiplash Associated Disorders*, o WAD, y a veces traducido como trastornos del latigazo cervical o síndromes asociados al latigazo cervical) es considerado como uno de los más simulados, así como el que mayor coste socioeconómico genera (González-Ordi *et al.*, 2012; Represas *et al.*, 2008; Represas *et al.*, 2020). Otros autores añaden que se trata de una de las condiciones más confusas, controvertidas y complejas de diagnosticar (Elliot *et al.*, 2009).

3.2. El esguince cervical y el Síndrome de Latigazo Cervical

SLC es el término que se utiliza para definir al conjunto de síntomas que se producen a raíz de sufrir un latigazo cervical. Una de las aportaciones históricamente más relevantes es la operativización del latigazo cervical realizada por la Quebec Task Force en 1995 (QTF; Spitzer *et al.*, 1995), entendiéndose como un mecanismo de aceleración-desaceleración que transmite energía al cuello y que puede ocasionar lesiones óseas o en los tejidos blandos (esguince cervical). Si bien no se dispone de datos actualizados, se estima que su incidencia anual varía entre 16 y 200 casos cada 100.000 habitantes, con una alta variabilidad en función del país. Por ejemplo, en España asciende a 60.2 casos cada 100.000 habitantes, Canadá asciende a 70 cada 100.000, Australia 106 cada 100.000 y Países Bajos de 188 a 325 cada 100.000 (Pataskia y Kumar, 2011; Regal Ramos, 2011). Por otro lado, el coste a nivel económico se sitúa en aproximadamente 42 billones de dólares en los Estados Unidos y 10 billones de euros anuales en Europa, con estimaciones aproximadas de 3 billones de libras anuales en el Reino Unido y 10.000 millones de euros en España (Crouch et al, 2006; Kamper et al, 2008; Pink *et al.*, 2016). Además, también se considera que el SLC es una de las principales fuentes de discapacidad a nivel mundial (Kamper *et al.*, 2008). Concretamente, el dolor cervical de origen traumático y no traumático es la cuarta causa global de discapacidad, siendo éste el mismo rango que ostentaba en el año 1990, lo que surge que la investigación en prevención y rehabilitación de este tipo de daño realizada en los últimos años ha tenido un efecto pequeño en su impacto global (Walton y Elliot, 2017).

En función de la severidad de la lesión y los síntomas, se le otorga uno de los siguientes grados (Pastakia y Kumar, 2011):

- **Grado 0.** Sin signos físicos ni quejas sobre el área cervical.

- **Grado 1.** Quejas sobre dolor, exceso de sensibilidad y/o rigidez en el área cervical, pero sin signos físicos.
- **Grado 2.** Quejas sobre el área cervical y signos musculoesqueléticos, entre los que se incluyen puntos de rigidez y disminución del rango de movimientos.
- **Grado 3.** Quejas sobre el área cervical y signos neuropsicológicos, entre los que se incluye alteración de los reflejos tendinosos profundos, debilidad y déficits sensoriales.
- **Grado 4.** Quejas sobre el área cervical y fractura o dislocación.

La principal causa del SLC son los accidentes de circulación con vehículos de motor (Walton y Elliot, 2017), estimándose en España que aproximadamente entre un 15% y un 60% de este tipo de colisiones desembocan en casos de esguince cervical (Regal Ramos, 2011). Para la explicación de los aspectos patomecánicos del SLC resulta frecuente encontrar en la bibliografía dos enfoques: uno basado en la hiperextensión/flexión de la cabeza, y otro basado en la transmisión de energía. Como ejemplo del primero, Regal Ramos (2011), expone que la secuencia clásica de la lesión se iniciaría cuando un vehículo recibe un impacto trasero con una velocidad ≥ 8 km/h y sufre una aceleración hacia delante, experimentando la cabeza, los hombros y el tronco del ocupante una aceleración similar. La cabeza, estática en el espacio, sufre un movimiento de hiperextensión forzada que puede afectar a las facetas articulares cervicales, los discos, los nervios del simpático cervical, los ligamentos y las fibras musculares, que pueden sufrir hemorragias microscópicas y edemas que determinan procesos inflamatorios localregionales. Tras sufrir la hiperextensión, la inercia empuja la cabeza adelante, provocando el movimiento de hiperflexión que puede afectar a la musculatura cervical posterior.

Por otro lado, como ejemplo del segundo, Elliot *et al.* (2009) detallan que la secuencia se iniciaría con un alcance trasero que haría que el torso del ocupante fuera lanzado hacia adelante (entre 0 y 50 milisegundos desde el impacto), provocando una curva cervical en forma de “S” (entre 50 y 75 milisegundos desde el impacto) que fuerza la columna cervical a realizar un movimiento anormal y no fisiológico de extensión del segmento inferior y de flexión del segmento superior. Dicho movimiento produce energía que se almacena en los componentes elásticos de la columna cervical y, a continuación, la libera (>100 milisegundos) produciendo un brusco movimiento de la cabeza y el cuello, provocando un impacto que puede producir potenciales lesiones en un amplio número de tejidos anatómicos de la zona cervical. Como indican los anteriores autores, esta secuencia se desarrolla sin prácticamente ninguna resistencia de la musculatura paraespinal de apoyo, ya que la latencia requerida para producir la actividad reflexiva muscular es mucho más extensa.

En términos generales, los síntomas más característicos del SLC son el dolor y las disfunciones motores y sensorio-motores (Sterling, 2011), pero el cuadro clínico puede cursar con (Cailliet, 2006):

- **Dolor cervical.** Con un rango de dolor de leve a severo. Localizado en puntos específicos o en general, con posibilidad irradiación hacia la zona de los hombros y la zona superior de la espalda o lumbar.
- **Rigidez cervical.** Que produce una reducción significativa en el rango de movimientos que puede realizarse con la cabeza, debido al dolor, tensión de un músculo o problema mecánico.
- **Dolores de cabeza.** Producido por una posible tensión muscular o lesión en algún nervio de la zona cervical.

- **Insensibilidad, debilidad u hormigueos.** Los procesos inflamatorios pueden producir pinzamientos en los nervios espinales que se manifiestan mediante insensibilidad, debilidad y/o hormigueos en la zona afectada.

Además, también puede presentar síntomas menos comunes como mareos, alteraciones visuales y/o auditivas, disfunciones temporomandibulares, fotofobia, disfonía, disfagia, fatiga y sintomatología ansioso-depresiva (Elliot *et al.*, 2009).

La duración de los síntomas varía en función del grado y la interacción entre diversas variables de corte biopsicosocial (Sterling, 2011). La evidencia indica que, en términos generales, la recuperación de los síntomas suele ocurrir en los primeros 2-3 meses desde el accidente, produciéndose el pico máximo de severidad entorno al primer mes, aunque algunos síntomas, como el dolor, pueden tardar hasta 6 meses en desaparecer completamente (Carroll *et al.*, 2009; Meeus *et al.*, 2012). Los síntomas suelen disminuir significativamente en el citado periodo, pero Sterling (2014) señala que hay 3 posibles “vías” de recuperación:

- **Niveles iniciales de discapacidad asociada al dolor de leve a moderados.** Supone aproximadamente un 45% de los pacientes de SLC y tienen una recuperación óptima, dentro de los plazos anteriormente citados.
- **Niveles iniciales de discapacidad asociada al dolor de moderados a severos.** Supone aproximadamente un 39% de los pacientes de SLC y tienen una recuperación relativamente adecuada, pero los niveles de discapacidad permanecen en un nivel moderado hasta 12 meses transcurrido el accidente.
- **Niveles iniciales de discapacidad asociada al dolor severo.** Supone aproximadamente un 16% de los pacientes de SLC y al grado de recuperación es bajo, con una discapacidad de severa a moderada.

Consideramos que, en vista de la temática que perseguimos con la presente tesis, hay dos cuestiones de interés relativas a la condición estudiada: ¿por qué el SLC es una de las condiciones más simuladas? y ¿cómo se evalúa el riesgo de simulación?, pero la respuesta a estas dos cuestiones viene prácticamente en su totalidad de otras dos preguntas que queremos desarrollar a continuación: ¿Por qué el SLC se cronifica?, y ¿Por qué tiene tanta incidencia/prevalencia en unos países y en otros no?

3.3. ¿Por qué se cronifica el Síndrome de Latigazo cervical?

Como se ha podido observar en las 3 posibles vías de desarrollo de la condición expuestas con anterioridad, existe un alto porcentaje de pacientes que no llegan a recuperarse del todo y experimentan síntomas de severidad moderada-alta durante un periodo de tiempo mucho más elevado de lo esperado. Si bien la evidencia indica que el dolor y la discapacidad juegan un papel muy importante, la cronificación del SLC es considerada, a día de hoy, una incógnita, y los mecanismos subyacentes del proceso no están del todo claros o definidos (Regal Ramos, 2011; Sterling, 2014). Una de las principales líneas de investigación en la temática ha estudiado la capacidad de diferentes variables de corte biopsicosocial para predecir el riesgo de cronificación, lo que ha permitido configurar varios modelos explicativos que desarrollamos a continuación.

Modelos de lesiones pato-anatómicas. Como exponen Walton y Elliot (2017), estos modelos tuvieron un importante desarrollo en el siglo XX y estaban basados en un enfoque puramente anatómico, estudiando posibles daños en áreas y tejidos específicos. Los principales resultados indicaban que el mecanismo del latigazo cervical podía causar daños en tejidos periféricos, entre los que se incluían los ligamentos cervicales superiores, el núcleo trigémino y las articulaciones facetarias. Con esto, los autores sugerían que estas lesiones generaban malestar psicológico que cronificaba el dolor, pero no se ofrecía una

explicación concreta que ayudara a entender el mecanismo subyacente del proceso. Por otro lado, otra parte de la literatura se centró en el estudio de los déficits motores y en la reducción del rango de movimientos cervicales (por ejemplo, ver Heikkila et al., 2000; Sterling, 2004 o Sterling *et al.*, 2006). Siguiendo a Sterling (2011) y Sterling *et al.* (2006), esta línea de investigación resulta de interés ya que se han observado alteraciones motoras y déficits en el control de los músculos tanto en pacientes crónicos como en agudos, persistiendo hasta 2 años después, que podría dejar a los pacientes en un estado de mayor vulnerabilidad frente a futuros episodios de dolor cervical. A pesar de ello, la evidencia parece indicar que la presencia de déficits no tiene valor predictivo y tiene poco impacto en el proceso de cronificación del SLC, ya que el tratamiento directo para rehabilitar la disfunción motora y la capacidad de movimiento tiene un efecto leve en los niveles de dolor y discapacidad de los pacientes (Sterling, 2011).

Modelos biopsicosociales. Los modelos con enfoque eminentemente biomédico no han sido capaces de ofrecer un modelo explicativo completo ya que, como señala Sterling (2014) la aproximación “una talla para todos” ignora la heterogeneidad sintomática del SLC. La evidencia sugiere que los pacientes que sufren este tipo de lesión experimentan síntomas de corte psicosocial incluso desde la fase aguda, principalmente sintomatología ansioso-depresiva y alteraciones comportamentales, como una actitud negativa ante la situación de enfermedad, miedo al movimiento o pensamientos catastrofistas relacionados con el dolor (Sterling, 2014; Walton y Elliot, 2017). En términos generales, se ha observado que padecer un dolor cervical más intenso y persistente se asocia con mayor percepción de discapacidad, depresión y peor autoeficacia (Aguilera *et al.*, 2019; Carroll *et al.*, 2009; Falla *et al.*, 2016; Holm *et al.*, 2008). La relación entre la sintomatología física y psicológica no es unidireccional, y en el proceso de vivencia de la enfermedad o condición se produce una interacción cíclica donde dichos

síntomas pueden influir entre sí. Por ejemplo, la presencia de dolor de alta intensidad puede producir una vivencia traumática que genera sintomatología ansioso-depresiva y lleva a los pacientes a adoptar estilos de afrontamiento desadaptativos. Ante esta situación se experimenta un estado emocional negativo en el que se presta más atención al dolor, lo que se traduce en una percepción de severidad mayor (Aparicio *et al.*, 2013; Lumley *et al.* 2011).

Uno de los primeros modelos biopsicosociales explicativos fue propuesto por Ferrari *et al.* (1999) en una publicación en la que exponía la importancia de la actitud ante la situación de enfermedad de los pacientes. Los autores expusieron la hipótesis de que un estilo terapéutico pasivo en el que se trata al paciente como una “persona lesionada” únicamente agrava el problema y puede llevar a la cronificación del mismo. Si se presenta el SLC como un problema crónico y no se da una explicación detallada sobre lo que debería esperar en su transcurso, cualquier problemática nueva que aparezca, aunque no esté relacionada con la condición (mala postura, esfuerzos inadecuados, etc.) pasará a formar parte de ella. El paciente creerá que no está mejorando, que lo que está experimentando es la cronificación de su enfermedad y, en palabras de los autores, vivirá “una profecía autocumplida de dolor crónico y discapacidad” (p. 323).

En su lugar, los autores recomiendan que el profesional siga un rol proactivo y eduque al paciente de lo que debería esperar en el transcurso de su enfermedad y/o condición, explicando, entre otras cosas, el porqué y el pronóstico de sus síntomas. Por ejemplo, pueden explicar que el SLC es un “problema menor, benigno y auto-limitado (...) y que el daño de la lesión aguda no causa el dolor crónico (...) otras cosas lo causan y pueden ser prevenidas” (p.323). Si bien el modelo de los autores es hipotético, los resultados de la revisión sistemática de Meeus *et al.* (2012) lo apoyan, ya que sus resultados indicaron que la gran mayoría de modalidades verbales de educación sobre la

enfermedad/condición del paciente reduce significativamente la severidad del dolor, mejora la movilidad cervical y disminuye el tiempo de recuperación.

Pocos años después de la propuesta de Ferrari *et al.* (1999), Turk (2002) desarrolló el *Modelo de Diatesis-Estrés* para el dolor que Gatchel (1996) comenzó a esbozar al hablar de alteraciones psicológicas y dolor crónico. En este modelo se sugiere que pueden existir una serie de factores predisponentes que interactúan con la exposición al trauma y activan un proceso comportamental y cognitivo. Así, los pacientes se enfrentan a la vivencia de la enfermedad desde unos antecedentes contextuales que actuarían como factores de protección o como factores de vulnerabilidad, los cuales producirían la cronificación. No hace referencia a una “personalidad del dolor”, sino a un estilo desadaptativo de afrontamiento de la enfermedad (Peng *et al.*, 2006).

En este sentido, Turk (2002) incluye los siguientes factores predisponentes:

- Sensibilidad a la ansiedad. Entendida como el miedo a los síntomas ansiosos por la creencia de que tendrán consecuencias negativas (Reiss y McNally, 1985). Aquellas personas que presentan alta sensibilidad a la ansiedad tenderían a desarrollar un estilo de atención selectivo en el que predominaría la hipervigilancia al dolor y a otros estímulos nocivos. Así, estos pacientes prestarán especial atención a las sensaciones físicas y las exacerbarán, magnificando la percepción de severidad de las mismas. Ante esta situación magnificada, los pacientes buscarán medios para paliar y disminuir dichas sensaciones, desarrollando así conductas desadaptativas de evitación e incrementando el consumo de analgésicos.

- Procesos cognitivos. La evidencia indica que una serie de creencias interrelacionadas son de especial importancia para entender el proceso del dolor, entre ellas:
 - a) La anticipación del dolor. La experiencia del dolor puede activar una serie de pensamientos negativos relacionados con el miedo a padecer más dolor, llevando a un estado de vigilancia continua en la que predomina la búsqueda de sensaciones nocivas que puedan indicar que la enfermedad/condición no se está curando. Ante esta situación los pacientes prestan más atención a posibles señales de amenaza venideras y son menos capaces de desviar la atención de la información relacionada con el dolor (Nijs e Ickmans, 2014).
 - b) La catastrofización del dolor. La catastrofización es una distorsión cognitiva caracterizada por un discurso excesivamente negativo y una visión pesimista del futuro (Ruscheweyh *et al.*, 2011). En el contexto del dolor, la catastrofización “hace referencia a una percepción mental negativa y exagerada respecto a la experiencia de dolor, tanto real como anticipada” (Olmedilla Zafra *et al.*, 2013, p. 84). Una de las principales consecuencias de esta distorsión es el desarrollo de miedo al movimiento, donde los pacientes evitarán realizar movimientos que crean que pueden ser dolorosos, lo que se traduce en mayor percepción de discapacidad (Vlaeyen *et al.*, 1995).
 - c) Atribución causal. Como exponen Turk y Holzman (1986), las dos variables anteriores se convierten en un problema especialmente significativo cuando los pacientes atribuyen la causa original de su dolor a una lesión repentina y traumática. Los pacientes desarrollan la

percepción de estar “heridos”, creerán que su lesión es más grave por haber sido producida por un accidente traumático y enfocarán la experiencia de la condición de una forma más negativa, considerando que el proceso de recuperación va a ser más lento y complejo.

- d) La autoeficacia. La autoeficacia hace referencia a “la creencia que tiene una persona de poseer las capacidades para desempeñar las acciones necesarias que le permitan obtener los resultados deseados” (Olivari Medina y Urra Medina, 2007, p. 10). Turk (2002) explica que la percepción sobre la propia capacidad de afrontamiento puede influir en la capacidad de vivir la experiencia del dolor. Así, una percepción positiva de autoeficacia hará que el paciente aplique aquellas habilidades positivas de afrontamiento que haya aprendido y perseverará en sus esfuerzos, mientras que una percepción negativa hará que el paciente dude de su propia capacidad de afrontar la situación y será menos probable que aplique mecanismos adaptativos. De la misma forma, la autoeficacia positiva disminuye la ansiedad por anticipación, lo que evita que se desarrolle el estado de hipervigilancia y activación que exacerba la percepción del dolor y el malestar psicológico.

Así, integrando las variables anteriormente expuestas, el modelo indica que aquellas personas que tienen un estilo de afrontamiento desadaptativo tienen más probabilidad de experimentar más ansiedad ante la experiencia de enfermedad (real y anticipada) y desarrollar respuestas de evitación, lo que lleva a un estado de hipervigilancia y activación emocional donde se presta más atención y se desarrolla temor a las reacciones fisiológicas. Ante este estado, los pacientes se ven envueltos en una visión negativa y catastrofista de la condición sufrida, lo que hace que respondan de forma

negativa tanto al dolor como a cualquier estímulo fisiológico nocivo que puedan experimentar.

Las conductas de evitación que realice el paciente, como no hacer determinados movimientos, incrementan la percepción de discapacidad, lo que disminuye la autoeficacia y aumenta la percepción negativa de la experiencia de la condición. En la Figura 2 se muestra un esquema del proceso.

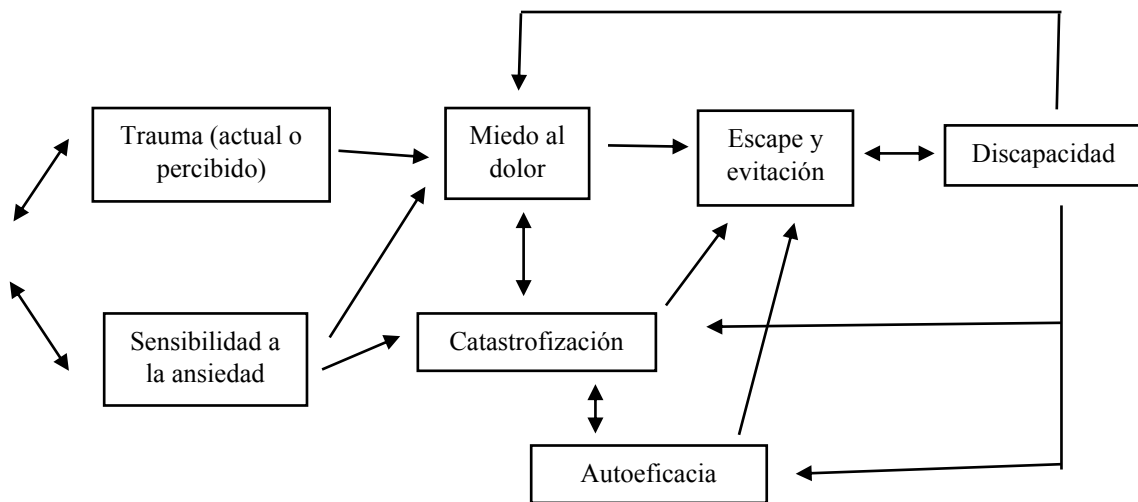


Figura 2. Modelo de diátesis-estrés. Adaptado de Turk (2002) con permiso del autor.

Investigaciones realizadas con posterioridad, como las de Martin *et al.* (2010) o Vachon-Preseau *et al.* (2013), aportaron resultados que proporcionaban apoyo empírico al modelo y, a día de hoy, es considerado como uno de los modelos explicativos más completos. No obstante, se considera que el modelo únicamente tiene en cuenta variables psicológicas y, para explicar el impacto emocional de la condición/enfermedad, sería conveniente incluir además un enfoque fisiológico (Walton y Elliott, 2017).

Otro de los modelos biopsicosociales que se desarrolló a partir de los primeros trabajos sobre el impacto de las variables psicológicas en la cronificación del dolor y la

importancia de las creencias que tienen los pacientes sobre la experiencia del mismo, como los de Fordyce (1976), Lethem *et al.* (1983) o Waddell *et al.* (1993) fue el *Modelo de Miedo-Evitación del dolor* (del inglés *Fear-Avoidance Model* o *FAM*). Vlaeyen y Linton (2000) realizaron una síntesis de los resultados de las investigaciones de los anteriores autores, formulando un modelo teórico que sigue un planteamiento similar al de diátesis-estrés. Según estos autores, la experiencia de la condición tiene dos posibles vías. La primera está caracterizada por la ausencia de miedo, confrontación directa del dolor y resolución, similar al proceso de autoeficacia positiva que describíamos con anterioridad. En esta vía, el paciente aumentará su nivel de actividad, o confrontación, hasta que alcanza la recuperación (Esteve Zarazaga y Ramírez Maestre, 2013). En la segunda vía, el paciente desarrolla catastrofización del dolor y se ve envuelto en el bucle de hipervigilancia, activación emocional, ansiedad y discapacidad que ya exponíamos en el anterior modelo, con la inclusión del descondicionamiento físico o desuso, que hace referencia a la pérdida de facultades físicas causada por la restricción autoimpuesta de los movimientos. En la Figura 3 exponemos una síntesis de la secuencia de las dos vías.

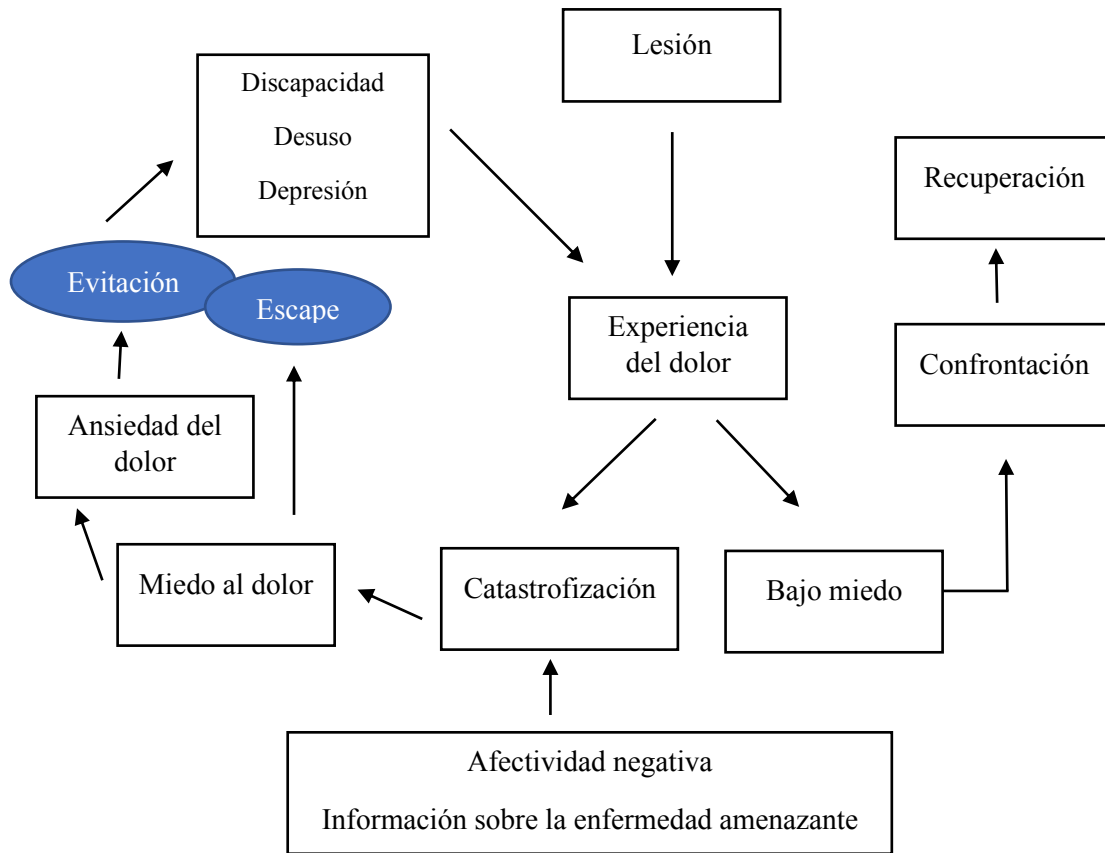


Figura 3. *Modelo de miedo-evitación del dolor*. Adaptado de Leeuw *et al.* (2007) con permiso del autor.

En la actualidad, el FAM cuenta con una amplia popularidad y aceptación en la comunidad científica y, si bien sus inicios fueron hipotéticos, existen a día de hoy numerosas publicaciones empíricas que apoyan su validez (ver por ejemplo las revisiones de Keefe *et al.* 2004 o Pincus *et al.*, 2010). No obstante, como indican Walton y Elliott (2017), aunque la catastrofización del dolor tiene una importancia significativa en el paso a la cronificación, el modelo tiene una naturaleza causal y unidireccional que ha sido cuestionada en investigaciones posteriores, como la de Nieto *et al.* (2009) en la que se identificó un efecto mediador del miedo al movimiento en la relación entre catastrofización y discapacidad y entre catastrofización y depresión. A pesar de ello, el

modelo no fue concebido como una idea incuestionable y está abierto a debate, reflexión, adaptaciones y extensiones (Crombez *et al.*, 2012).

Con el objetivo de solventar los aspectos negativos de los dos modelos anteriores, Walton y Elliott (2017) proponen un modelo, adaptado al SLC y actualizado con la evidencia más reciente, que integra sus propuestas con un enfoque holístico centrado en el bienestar, la salud y la influencia de los factores genéticos, psicológicos, biológicos y/o situacionales. El planteamiento de dos vías del FAM se conserva, pero dichas vías no se entienden como un camino prefijado y se incluye la posibilidad de que en algún momento del proceso el paciente pase de la una a la otra. El modelo se inicia con la diástesis, en la que se consideran factores de riesgo y de protección, como la genética, la dieta, el nivel de masa muscular, hacer ejercicio físico, la propensión a la catastrofización del dolor, la existencia de sintomatología ansioso-depresiva previa, un patrón de confrontación desadaptativo, etc. De la misma forma, también se incluyen factores y creencias culturales relacionadas con la experiencia del dolor, la expresión de sufrimiento, el contexto médico-legal o la calidad del apoyo social. Así, teniendo en consideración las citadas variables, la persona se ve expuesta a un evento traumático (accidente de circulación) que actúa como activador de una cascada de reacciones psicológicas (confrontación con la situación) y físicas (inflamación en los tejidos, daño muscular, etc.), entrando en el proceso de vivencia de la condición ya descrito en el modelo de Turk (2002). Walton y Elliott (2017) incluyen en este proceso factores psicosociales no considerados como la sensación de injusticia, o victimización, el escrutinio social al que se ve sometida la persona y la sensación de vulneración a la autogestión, que actúan en la fase inicial del ciclo incrementando el riesgo de desarrollar un estilo de afrontamiento disruptivo. En la Figura 4 se ofrece un esquema explicativo del modelo.

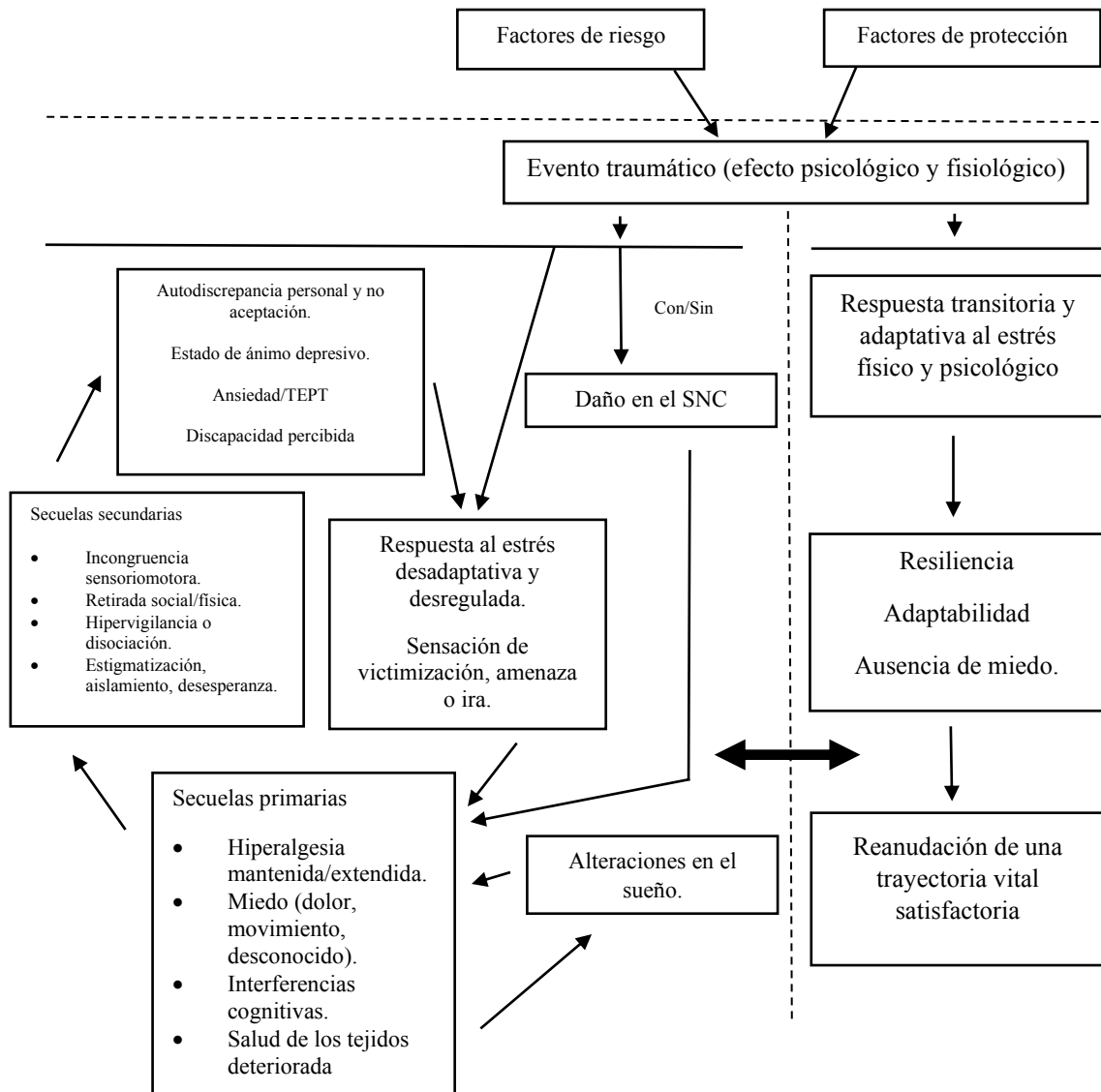


Figura 4. *Modelo integrado de SLC crónico.* Adaptado de Walton y Elliot (2017) con permiso del autor.

La compensación como pseudo-modelo explicativo. Probablemente, una de las variables que más se asocian al SLC es la existencia de un beneficio externo otorgado por padecer la condición, o compensación, el cual suele tomar la forma de remuneración económica o una baja laboral. Tradicionalmente, se considera que la existencia de dicha compensación se traducirá en una peor experiencia de la condición, o peores resultados a largo plazo, y será parcialmente responsable de la cronificación de los síntomas, propuesta que Spearing *et al.* (2012) denominan “hipótesis de compensación”.

Si bien esta hipótesis no ha sido planteada como un modelo explicativo en sí mismo, ha sido citada en múltiples estudios como una variable con una gran influencia en el proceso de cronificación, por lo que hemos decidido incluirla en esta sección. Gran parte de dichos estudios fueron realizados entre los años 90 y la primera década de los años 2000 (por ejemplo, Atherton *et al.*, 2006; Burgess y Dworkin, 1993, Cameron y Gabbe, 2009; o Cassidy *et al.*, 2000) y 4 revisiones sistemáticas analizaron sus resultados concluyendo que existía suficiente evidencia como para afirmar que la hipótesis de compensación era válida (Binder y Rohling, 1996; Harris *et al.*, 2005; Iles *et al.*, 2008; Kamper *et al.*, 2008). Debido a las severas consecuencias sociopolíticas que estas revisiones ocasionaron, Spearing y Connelly (2011), realizaron una meta-revisión de todas las publicaciones hasta la fecha, incluyendo las 4 revisiones sistemáticas citadas, que evaluaban específicamente la relación entre compensación y salud, y otras publicaciones que, si bien no analizaba dicha relación específicamente, las incluía entre las variables estudiadas. Los autores concluyeron que la gran mayoría de los estudios tenían limitaciones metodológicas que afectaban a los resultados obtenidos. Por ejemplo, la mitad de las revisiones sistemáticas no realizaron un análisis de la calidad de los artículos incluidos, observándose que gran parte de los estudios empíricos analizados presentaban sesgo de selección y no realizaban una correcta gestión de las variables perturbadoras. También se observó que, si bien se consideraba como principal medida del resultado la “salud” (se valoraba la mejora del paciente mediante una mejoría en dicha variable), las herramientas utilizadas para evaluarla fueron muy heterogéneas, lo que dificultaba la síntesis de resultados al no existir una visión consensuada del concepto. Por estos motivos, consideraron que no era posible ni confirmar ni refutar la hipótesis de compensación, convirtiéndose en el primer estudio que cuestionaba la validez de la misma. Un año después, Spearing *et al.* (2012) realizaron otra revisión sistemática sobre

la temática, en esta ocasión específicamente en SLC, analizando 11 estudios longitudinales que analizaban la relación entre compensación y salud. Los autores observaron que únicamente la mitad de los estudios ofrecían resultados significativos, pero en ninguno de ellos se había tratado el problema de la causalidad inversa, asumiéndose la dirección de la relación causal. Al no tratarse esta posible fuente de sesgo, no era posible determinar si estos hallazgos indican que la presencia de compensación hace que los pacientes presenten peor salud o, por el contrario, los pacientes que presentan peor salud son los que tienden a buscar compensación económica. Por este motivo, Spearing *et al.* (2012) concluyen que, al igual que en la anterior revisión, no es posible confirmar o refutar la hipótesis de compensación en el SLC.

A pesar de que las dos anteriores revisiones sistemáticas sirvieron para señalar aspectos metodológicos que debían ser cambiados de cara a futuras publicaciones, Murgatroyd *et al.* (2015) indicaron que su presentación de resultados se realizó considerando tanto la compensación como la salud como una única variable, en lugar de categorizarlas en diferentes factores. Así, los autores evaluaron 29 estudios realizando la citada categorización y obtuvieron evidencia de una asociación entre los factores relacionados con la compensación y peor salud después de una lesión musculoesquelética. Específicamente, estar en “estado de compensación” (tener una petición) y tener representación legal se asociaba con una peor función física y psicológica y una severidad de dolor más elevada tras la lesión. A pesar de ello, los autores afirmaron que no era posible descartar la existencia del sesgo de causalidad inversa basándose en sus hallazgos y señalaron que era “imperativo para los investigadores que lo tuvieran en consideración” (Murgatroyd *et al.*, 2015, p. 28). Giummarra *et al.* (2016a) obtuvieron resultados similares en una revisión sistemática que realizaron específicamente para dolor crónico tras un accidente de circulación, indicando que existía una relación entre recibir compensación y

una incidencia/severidad más alta del dolor, peor función física y mayor presencia de sintomatología ansioso-depresiva, pero advirtiendo que no era posible establecer una relación causal entre las citadas variables.

La hipótesis de compensación ha seguido evaluándose hasta la fecha con resultados dispares. Varias de las publicaciones han encontrado una asociación significativa entre las variables, pero parte de las limitaciones ya indicadas no se han solventado, en especial el sesgo de causalidad (Duckworth y Iezzi, 2018; Giummarra *et al.*, 2016b, 2017a, 2017b; Gopinath *et al.*, 2015; 2019; Guest *et al.*, 2018). Debido a ello, habitualmente los autores advierten de este hecho y señalan que sus hallazgos deben tenerse en consideración para otras cuestiones de interés. Por ejemplo, Giummarra *et al.* (2016a, 2016b) señalan que, sea cual sea la dirección causal de la relación, los resultados pueden indicar que el contexto de una lesión con compensación será más complejo y presentará más factores psicosociales que requerirán la atención de los profesionales. Es posible que, en esta situación, el paciente perciba desconfianza y hostilidad por parte del sistema de salud y viva la situación de una forma más estresante. Así, enfrentarse a las barreras y desafíos impuestos por el sistema de compensación crearía un impacto psicológico significativo en las víctimas que afectaría a su capacidad de recuperación (Samoborec *et al.*, 2018).

El reciente estudio de Pozzato *et al.* (2020), uno de los más completos a nivel metodológico realizados hasta la fecha, señaló que estos resultados podían indicar que, en el contexto señalado, se producía una interacción circular compleja entre variables personales previas al accidente, factores contextuales, factores relacionados con la compensación y variables psicológicas y funcionales posteriores al accidente. Para comprobar esta hipótesis, los autores utilizaron un modelo de ecuaciones estructurales para estudiar la relación entre factores relacionados con la compensación y con la salud

en 2019 pacientes, de los cuales 709 tenían un beneficio externo por una lesión. Sus resultados indicaron que el hecho de buscar compensación no lleva inevitablemente a sufrir una peor salud psicológica tras un accidente de circulación. Si bien la búsqueda de compensación se relacionaba con una peor salud psicológica, también se producía una relación inversa significativa. De la misma forma, la compensación no era la única variable que influía en la salud psicológica, ya que la vulnerabilidad individual al estrés y la salud posterior a la lesión también contribuían al modelo y, cuando se incluía la discapacidad y el estado funcional en la ecuación, estas dos variables explicaban completamente los cambios en la salud psicológica, desapareciendo la compensación del modelo.

No cabe duda de que el papel de la compensación en la experiencia de la enfermedad o condición es polémico. Aun siendo una de las variables predictoras más estudiadas, no existe, a día de hoy, un consenso firme sobre su papel real en el proceso de cronificación. Como señala Jull (2016), la evidencia indica que los factores relacionados con la compensación pueden tener cierto efecto en la prognosis de la enfermedad/condición, pero este efecto no es necesariamente importante. A pesar de ello, las investigaciones anteriormente expuestas señalan ciertos aspectos que pueden resultar de interés, tanto para la investigación como para la práctica clínica, y es que, como exponían Giummarra *et al.* (2016a, 2016b), si bien la dirección de la relación causal es incierta, el contexto de la compensación introduce variables adicionales que pueden resultar desafiantes para el paciente e influir en el proceso, por lo que conocerlas y trabajarlas podría resultar beneficioso para todas las partes implicadas.

A pesar de los grandes avances que se han realizado en la disciplina, los mecanismos subyacentes del proceso de cronificación del SLC y sus síntomas aún no son del todo conocidos (Walton y Elliott, 2017). Si bien los modelos explicativos detallados

disponen de evidencia que los respaldan, ninguno de ellos ofrece una explicación que pueda considerarse definitiva. No obstante, el enfoque del SLC como una condición biopsicosocial ha recibido un gran apoyo empírico (Carroll, 2011; Carroll *et al.*, 2008, 2014; Fritz, 2017; Jull *et al.*, 2013; Kongsted *et al.*, 2008; Lamb *et al.*, 2013; Pozzato *et al.*, 2020; Sarrami *et al.*, 2016; Scholten-Peeters *et al.*, 2003; Spearing *et al.*, 2012), y no cabe duda de que las variables psicosociales juegan un papel fundamental en el proceso. A pesar de la evidencia existente sobre la importancia de estas variables, la evaluación del SLC suele enfocarse desde el modelo biomédico y para su tratamiento se suele recurrir a terapia física y/o educación, tratando a los pacientes como un grupo homogéneo (Jull *et al.*, 2013). Los anteriores autores señalan que se podrían alcanzar mayores ratios de recuperación si el SLC se reconociera definitivamente como una condición heterogénea y se enfocara su valoración y tratamiento desde el modelo biopsicosocial y multidisciplinar.

3.4. Diferencias en la prevalencia/incidencia del Síndrome de Latigazo Cervical

Proponíamos una segunda pregunta previa a la cuestión del alto riesgo de simulación en SLC. ¿Por qué tiene tanta incidencia en unos países y en otros no? Tradicionalmente se ha observado que entre países existen diferencias muy significativas en el número de casos. Ferrari *et al.* (1999) señalaron que estas diferencias no podían ser explicadas por los modelos teóricos de la época y realizaron un análisis de las cifras proporcionadas en diversos países. Por ejemplo, encontraron diferencias muy llamativas entre el número de casos de Nueva Zelanda y el Estado de Victoria en Australia, las cuales poseían vías legales diferentes para los accidentes de circulación. También encontraron diferencias en Australia tras la reforma del sistema legal de 1987, la cual establecía que

la víctima debía informar a la policía sobre el accidente y debía cubrir los gastos iniciales del proceso (317 dólares en aquella época). Se observó que en 1989 había un 45% menos de reclamaciones por latigazo cervical respecto a 1982/1983 y 68% menos que en 1985/1986. También se identificaron diferencias en el número de casos entre países como el Gran Bretaña o Alemania, con una prevalencia muy elevada de SLC, y Grecia o Lituania, donde el SLC apenas existía. Los autores sugieren que el proceso de compensación, no únicamente la existencia del mismo, podía ser una posible explicación a las diferencias observadas entre países. Esta hipótesis ha sido apoyada por el estudio de Represas *et al.* (2008), ya que los autores encontraron que había mucha más incidencia de WAD en Galicia, una comunidad autónoma del noreste de España, que en la zona norte-central de Portugal, a pesar de que las características sociodemográficas de las poblaciones eran muy similares. Consideraron como posible explicación el hecho de que “la legislación española permite la conversión directa de síntomas, fácilmente simulables, a un sistema de puntos legales, resultando en una compensación económica mucho mayor que en Portugal (p.355)”.

Chappuis y Soltermann (2008) añaden que, además de los factores relacionados con la compensación, pueden considerarse también las diferencias en el diagnóstico diferencial de la condición debido a la preparación de los profesionales. Los autores señalan que, por ejemplo, Gran Bretaña e Italia tienen una incidencia de lesiones cervicales ocasionadas por accidente de circulación mucho más elevada que Francia, a pesar de tener características poblacionales similares. La diferencia es que, en las dos primeras localizaciones, los profesionales médicos no recibían preparación especializada para evaluar accidentes de circulación.

En términos generales, esta cuestión no ha recibido tanta atención en las investigaciones sobre la temática y, a día de hoy, disponemos de una cantidad muy

limitada de evidencia. Por ejemplo, podemos observar que estudios muy recientes, como el de Hayashi *et al.* (2019), aún recurren a las cifras expuestas por Ferrari *et al.* (1999) hace ya 21 años. No cabe duda de que el original título que Represas *et al.* (2008) utilizan en su estudio, “*No cash, no whiplash*”, lo que se traduciría como sin dinero no hay latigazo cervical, representa la idea o creencia general sobre la principal explicación a la gran variabilidad de prevalencia/incidencia de la condición. Los hallazgos indican que otros factores, como las actitudes culturales, a la aproximación médica al diagnóstico y tratamiento también pueden jugar un papel importante en el fenómeno, pero aún son necesarias más investigaciones que apoyen estas hipótesis.

3.5. ¿Por qué se simula el Síndrome de Latigazo Cervical?

Una síntesis de las dos anteriores preguntas puede ayudar al/la lector/a a hacerse una idea de porqué el SLC es una de las condiciones más simuladas. En primer lugar, como exponían Represas *et al.* (2008), en algunos países se otorga una compensación económica por sufrir la condición, lo cual actúa como motivación para padecerla, o aparentar padecerla. No se trata de establecer una relación causal entre existencia de compensación y simulación, sino plantear que, en un contexto donde puede obtenerse un beneficio por sufrir una enfermedad o condición, la simulación es una posibilidad y debe realizarse un diagnóstico diferencial de la misma (Arce, 2017). En segundo lugar, entra en juego la facilidad para simular la condición, teniendo en cuenta qué síntomas se pueden simular y cómo puede evaluarse la condición y la simulación. Pongamos por ejemplo una fractura de la tibia o un traumatismo craneoencefálico moderado o severo. Aunque el presunto paciente finja dolor, o cojera, o afirme no poder apoyar la pierna, una radiografía confirmará o descartará la lesión y sacará de dudas al evaluador. Podemos poner otro ejemplo de una condición que no sea objetivable con pruebas de imagen y se diagnostique, en gran medida, en base a las verbalizaciones o relato del paciente, como

el trastorno depresivo mayor o el trastorno adaptativo. El profesional dispondrá de una amplia variedad de instrumentos psicométricos y entrevistas para realizar un diagnóstico diferencial robusto de simulación, como los detallados en el capítulo 2 del presente marco teórico. En el caso del SLC, cuando el impacto ha sido severo y se otorga el tercer o cuarto grado, siendo estos los menos frecuentes (Chappuis y Soltermann, 2008), el experto se encuentra en una situación similar a la de los anteriores ejemplos y dispone de medios para realizar una evaluación objetiva, pero ¿qué ocurre cuando se otorga el primer y segundo grado? Como señala Represas (2017), estos casos tienen una gran complejidad diagnóstica debido a que los síntomas que suelen expresar los pacientes, como contracturas, limitación de la movilidad o rectificaciones cervicales, son comunes a múltiples patologías con alta prevalencia en el contexto, lo que impide el correcto establecimiento de la relación causal entre los síntomas y el accidente. Los profesionales médicos rara vez disponen de información sobre el estado anterior del paciente, por lo que, si se observa, por ejemplo, una contractura, no será posible descartar otras posibles causas. De la misma forma, varios de los citados síntomas son difíciles de objetivar con los métodos diagnósticos disponibles en la actualidad y no pueden ser cuantificados correctamente (Sartori *et al.*, 2003). Por ello, el diagnóstico del SLC de grado 1 ó 2, u ocasionado por alcances a baja velocidad, siendo este el más frecuente, suele realizarse en función de las manifestaciones de los pacientes y sin disponer de una prueba objetiva. Así, nos encontramos con una condición por la cual se otorga un beneficio externo, económico o laboral, es fácil de simular y no dispone de pruebas diagnósticas que objetiven los síntomas.

3.6. ¿Cómo se estudia la simulación en el Síndrome de Latigazo

Cervical?

La anterior cuestión podría solventarse con una respuesta rápida, y es que, a día de hoy, y a pesar de su severo impacto socioeconómico, no existe ningún instrumento, metodología o herramienta que permita realizar un diagnóstico diferencial de simulación en SLC preciso y fiable. La evidencia disponible sobre la temática es especialmente limitada, y una búsqueda en las principales bases de datos (Pubmed, Medline, Psychinfo, Psycharticles y Psycodoc) arroja 59 resultados, de los cuales únicamente 11 son estudios empíricos que tienen el objetivo de validar algún instrumento o técnica para detectar simulación en la condición de interés (Burges y Dworkin, 1993; Capilla Ramírez *et al.*, 2014; Espinosa-Sánchez *et al.*, 2014; Linnell y Easton, 2006; Merten *et al.*, 2007; Sartori *et al.*, 2003, 2007; Schmand *et al.* 1998; Oddsdóttir *et al.*, 2015; Vonk *et al.*, 2010; Wallis y Bogduk, 1996). La revisión de los principales manuales de la temática, como el de González-Ordi *et al.* (2012), Rogers y Benders (2012) o Young (2014) también ofrecen poca información sobre la simulación en SLC, siendo el primero el único que dedica un capítulo completo al mismo. Esta situación también puede encontrarse en los trastornos relacionados con el dolor en general, ya que, como exponen Boskovic *et al.* (2017), tampoco se dispone a día de hoy de recursos válidos y fiables para detectar la simulación en dichos trastornos. La principal diferencia entre ambas es que el número de investigaciones en el campo del dolor sigue creciendo lentamente, y en los últimos años se han explorado nuevas alternativas y posibilidades para tratar de solventar esta problemática, como la adaptación de técnicas propias del campo de la detección de mentiras (Acka *et al.*, 2019; Boskovic *et al.*, 2017, 2018) o el estudio de la capacidad discriminativa de diversas técnicas (Curtis *et al.*, 2019; Merten *et al.*, 2016). En cambio, como puede observarse, las últimas publicaciones sobre simulación en SLC se remontan

a hace 5 años, y en la gran mayoría no se utiliza un diseño de grupo conocido, ni sus resultados han sido replicados, lo que limita significativamente su utilidad en el contexto clínico-forense.

Teniendo en cuenta la alta prevalencia/incidencia del SLC, su gran coste económico y el hecho de que sea una de las condiciones más simuladas, consideramos de gran importancia que se retome la línea de investigación y se trate de desarrollar una metodología válida y eficaz para detectar simulación que pueda ser utilizada en el contexto clínico-forense. No planteamos esta necesidad con la única finalidad de “cazar” a los simuladores, sino también para proporcionar a los profesionales médicos medios que les permita hacer un diagnóstico diferencial de simulación con seguridad. Esto permitiría reducir la posible desconfianza que exista hacia la condición y, además, evitará que se clasifique erróneamente a pacientes genuinos como simuladores, con las severas consecuencias que eso conlleva (Cernovsky *et al.*, 2019a, 2019b).

PARTE EMPÍRICA

4. Objetivos generales y específicos

El fin último que se persigue con la presente tesis doctoral es la elaboración de un protocolo que permita valorar el riesgo de simulación en pacientes diagnosticados con SLC. Para alcanzar adecuadamente dicha finalidad, se han llevado a cabo 6 estudios divididos en tres bloques. Cada bloque corresponde con una fase de la preparación del protocolo, y tiene un objetivo general que se desglosa en objetivos específicos. Dichos objetivos específicos corresponden con cada uno de los citados 6 estudios.

El desarrollo del protocolo se ha realizado teniendo en consideración las recomendaciones de Rogers y Bender (2018) sobre la necesidad de entender la simulación como un constructo multidimensional que requiere de diversas fuentes de información para su análisis y valoración. Cabe la posibilidad de que un simulador sea capaz de fingir correctamente una variable de una de las dimensiones de la condición simulada (por ejemplo, la severidad del dolor), pero le resultará más complejo emular correctamente un perfil compuesto por una mayor cantidad de variables (Bianchini *et al.*, 2005).

El **primer bloque** comprende los pasos iniciales que se han seguido para la construcción del protocolo. Consideramos que el primer paso fundamental era planificar qué profesionales debían participar en la evaluación y qué procedimientos e instrumentos se aplicarían. Resulta de gran importancia que dichos instrumentos hayan sido correctamente validados y presenten unas propiedades psicométricas adecuadas. De la misma forma, debe seguirse una metodología de revisión bibliográfica sistemática y estructurada que garantice que no se ha pasado por alto ningún estudio. Por ello, el primer objetivo general fue:

- **OBG1.** Estudiar las propuestas realizadas hasta el momento para evaluar la simulación en el SLC, seleccionando aquellas más eficaces.

Para su consecución se plantean los dos primeros estudios:

Estudio 1. Criterios para el establecimiento del nexo causal en SLC: Un estudio de concordancia inter-observadores.

- **Objetivo específico 1.1.** Valorar la fiabilidad interjueces del juicio de cuatro expertos en medicina legal que establecen el nexo causal con y sin estudio de la biomecánica, en una muestra de pacientes que han sufrido SLC a partir de un accidente de circulación.

Estudio 2. Valoración de la simulación en SLC: Una revisión sistemática.

- **Objetivo específico 1.2.** Identificar aquellas medidas de resultado utilizadas para evaluar simulación en las tres dimensiones (biológica, psicológica y social) del SLC puedan ser empleadas como indicadores de simulación

El segundo y el tercer bloque comprenden la administración del protocolo. En **el segundo bloque** se utiliza para realizar un análisis pormenorizado del perfil de los pacientes clínicos incluidos en la presente tesis doctoral. El objetivo que se perseguía era:

- **OBG2.** Estudiar el perfil biopsicosocial y médico-legal de los pacientes de SLC.

Para su consecución se ha desarrollado el tercer estudio:

Estudio 3. Perfiles biopsicosociales de percepción del dolor de pacientes con SLC.

- **Objetivo específico 2.1a.** Obtener de forma empírica subgrupos basados en características psicosociales utilizando un análisis de cluster k-medias con una muestra de pacientes litigantes con latigazo cervical.

- **Objetivo específico 2.1b.** Investigar diferencias en los subgrupos obtenidos en variables de interés que no han sido utilizadas para formar los clústeres (severidad del dolor, percepción de discapacidad, variables sociodemográficas y variables médico-legales).

Finalmente, en el **tercer bloque** se analizó la capacidad discriminativa de los principales componentes del protocolo. El objetivo general planteado para este bloque fue:

- **OBG3.** Elaborar y validar un protocolo para valorar la presencia de simulación en Síndrome de Latigazo Cervical.

Los estudios 4, 5 y 6 son los desarrollados para la consecución del objetivo:

Estudio 4. Capacidad discriminativa de un protocolo biopsicosocial para la detección de simulación en SLC.

- **Objetivo específico 3.1.** Evaluar la capacidad de una serie de escalas breves para discriminar entre pacientes clínicos con sintomatología genuina y simuladores instruidos.

Estudio 5. *Checklist* de indicadores de simulación en SLC.

- **Objetivo específico 3.2.** Comprobar la capacidad de una serie de marcadores de simulación, extraídos de bibliografía relevante en la materia, para discriminar entre pacientes clínicos con sintomatología genuina y simuladores instruidos.

Estudio 6. Capacidad discriminativa del Inventario Estructurado de Simulación de Síntomas (SIMS).

- **Objetivo específico 3.3.** Valorar la capacidad discriminativa del Inventario Estructurado de Simulación de Síntomas (SIMS) en el contexto médico-legal.

Cada uno de los estudios incluidos sigue el formato propuesto por APA séptima edición (APA, 2020) para la presentación de estudios científicos, con la excepción del Estudio 2 que, al ser una revisión sistemática, sigue las directrices de la guía PRISMA (Moher *et al.*, 2009).

Por otro lado, el análisis de los resultados de la presente tesis se realiza a dos niveles, individual y general. Para el primero se ha incluido en los citados estudios un apartado de “Discusión” donde se valora de forma específica los resultados de cada uno de ellos. Para el segundo se ha añadido un apartado llamado “Discusión general” donde se realizará un análisis en conjunto tanto del desarrollo del trabajo como de los hallazgos de los estudios.

Finalmente, aclarar al /la lector/a que en la presente tesis doctoral se utilizará el término “simulación” en lugar de “fingimiento” debido a que este último apenas ha sido utilizado en la bibliografía española. No obstante, para el lector internacional, o en el caso de traducir al inglés fragmentos de este documento, “simulación” debe ser entendido y traducido como “feigning”, y no como “malingering”.

5. Bloque 1 - Estudio 1. Criterios para el establecimiento del nexo causal en SLC: Un estudio de concordancia inter-observadores

En España, la valoración médico-legal en el caso de accidentes de circulación, principal causa del SLC, se ha ocupado de identificar el daño que debe ser indemnizado mediante el uso de baremos y tablas de cálculo que estiman los daños personales (Represas, 2017). Como expone el anterior autor, la entrada en vigor de la Ley 35/2015 hace que en la citada valoración cobre especial protagonismo el estudio de la relación causal entre la lesión/daño padecido y el accidente sufrido, convirtiéndose en un elemento de gran importancia en casos de SLC. En un protocolo como el planteado en la presente tesis, los criterios para el establecimiento de la relación causal aportan información adicional que puede ser de gran utilidad para ayudar al/la profesional a tomar una decisión. No obstante, resulta necesario analizar dicha metodología, ya que en las colisiones a baja velocidad el estudio del nexo causal puede ser complejo, y a día de hoy no se dispone de evidencia científica de que los criterios de la Ley 35/2015 realmente permitan alcanzar el fin propuesto (Represas, 2017).

De la misma forma, los hallazgos en la materia indican que un estudio multidisciplinar del SLC que incluya una reconstrucción biomecánica de las dinámicas del evento facilitará en gran medida el estudio de la causalidad (Ferrara *et al.*, 2016). En ese sentido, la biomecánica de las lesiones permitirá conocer el mecanismo de producción de una lesión y distinguir entre las lesiones que son producto de un traumatismo directo, las que son producto de daño inercial o las que es posible que se produzcan como consecuencia de ambos mecanismos (Arregui-Dalmases *et al.*, 2011).

Por todo lo anteriormente expuesto, consideramos que tanto los criterios de la ley 35/2015 como el estudio de la biomecánica pueden ser de gran utilidad para ser incluidos

como fuente de información adicional en la presente tesis, pero creemos necesario que se realice un análisis previo de los mismos. De la misma forma, un estudio como el planteado permitirá a los autores familiarizarse con la metodología seguida habitualmente en el contexto médico-legal para estudiar la condición de interés.

Para alcanzar el fin propuesto se plantea el siguiente objetivo:

- **Objetivo específico 1.1.** Valorar la fiabilidad interjueces del juicio de cuatro expertos en medicina legal que establecen el nexo causal con y sin estudio de la biomecánica, en una muestra de pacientes que han sufrido SLC a partir de un accidente de circulación.

5.1. Metodología

5.1.1. Participantes

Para la elaboración del presente estudio, se utilizó una muestra de 300 pacientes clínicos (55,1% hombres) evaluados en una clínica especializada en la evaluación del daño corporal en A Coruña (España), de enero a septiembre de 2017. La edad media de los pacientes fue de 32,93 ($DT = 15,24$) con un rango de 22-89 años. Todos los casos fueron a la clínica para someterse a la evaluación de una lesión producida tras un accidente de tráfico. Se desconoce si alguno de los pacientes se encontraba en estado de litigio, pero todos tenían la posibilidad de obtener una ganancia secundaria en forma de incentivo económico.

Los pacientes fueron enviados a la clínica por sus respectivas compañías de seguros y presentaron toda la información relevante del caso, incluido un informe biomecánico de la lesión. Los criterios de inclusión fueron: (a) tener diagnóstico de SLC, (b) ser mayor de 18 años, (c) haber firmado el consentimiento informado, y (d) no

presentar otra enfermedad, anomalía o alteración médica que pudiera distorsionar las respuestas de la metodología aplicada. No cumplir dichos criterios suponía la exclusión del estudio.

5.1.2. Instrumentos

Criterios de causalidad. Para el estudio de la causalidad de los pacientes se utilizaron los cuatro criterios que establece el artículo 135 de la ley 35/2015 del baremo de accidentes de tráfico:

- **Criterio a. De exclusión**, que consiste en que no medie otra causa que justifique totalmente la patología.
- **Criterio b. Cronológico**, que consiste en que la sintomatología aparezca en tiempo médicamente explicable. En particular, tiene especial relevancia a efectos de este criterio que se hayan manifestado los síntomas dentro de las setenta y dos horas posteriores al accidente o que el lesionado haya sido objeto de atención médica en este plazo.
- **Criterio c. Topográfico**, que consiste en que haya una relación entre la zona corporal afectada por el accidente y la lesión sufrida, salvo que una explicación patogénica justifique lo contrario.
- **Criterio d. De intensidad**, que consiste en la adecuación entre la lesión sufrida y el mecanismo de su producción, teniendo en cuenta la intensidad del accidente y las demás variables que afectan a la probabilidad de su existencia.

Debido a que los expertos consideraban que algunos de los criterios eran poco específicos, se operativizaron con la inclusión de los siguientes subcriterios:

- **Criterio de exclusión.**

a1. Estudios objetivos específicos. Presencia de evidencia objetiva que permita establecer indudablemente la existencia de una causa alternativa que explique la sintomatología observada.

- **Criterio Topográfico.**

c1. Plausibilidad. La relación entre la zona afectada y el mecanismo de producción es coherente y tiene lógica biológica (adaptado del criterio de plausibilidad científica de Hill, 1965).

c2. Especificidad de síntomas y signos. La relación entre los síntomas y signos observados y/o expresados y el mecanismo de producción es coherente y tiene lógica biológica (adaptado del criterio de especificidad de Hill, 1965).

- **Criterio de Intensidad.**

d1. Eficacia biomecánica. La dinámica del lesionado es coincidente con el mecanismo de producción de la lesión (Represas, 2017).

d2. Compatibilidad biomecánica. Los parámetros físicos del accidente son compatibles con los umbrales poblacionales definidos para la lesión sufrida (Represas, 2017).

No se incluyó ningún subcriterio para el criterio temporal debido a que se consideraba suficientemente específico. Cada criterio, así como sus respectivos subcriterios, se valoraba con “Sí”, “No”, o “Dudoso”. Para clasificar los criterios topográfico y de intensidad como “Sí”, se tenían que valorar de forma positiva los dos

subcriterios. Si uno de ellos resultaba negativo se clasificaba como “No” y si resultaba “Dudoso” se clasificaba como tal. Se valoró el nexo causal como “Sí” cuando los cuatro criterios eran valorados de forma positiva, como "dudoso" cuando al menos 2 de estos criterios se valoraron como dudosos, pero no hubo ningún criterio respondido como "No", y como "No" cuando al menos 3 criterios se valoraron como "No", a excepción del criterio temporal, que era exclusivo.

Abbreviated Injury Scale, AIS (Gennarelli y Wodzin, 2006). Es un sistema de codificación anatómico utilizado para clasificar y describir la severidad de las lesiones, con la finalidad de medir una amenaza vital inmediata. Se otorga un código de 7 dígitos donde el primer número es la región corporal, el segundo el tipo de estructura anatómica, el tercero y el cuarto la estructura anatómica específica, el quinto y el sexto el nivel (fractura, ruptura, laceración, etc.) y el séptimo la gravedad. Para el código de gravedad se utiliza un sistema de clasificación con un rango del 1 (lesión de severidad baja) al 6 (lesión fatal), incluyendo además el valor 9 para lesiones sobre las que se tiene información limitada. En el presente estudio se expone la gravedad de las lesiones.

Baremo español de lesiones en accidentes de tráfico de la Ley 35/2015. Se trata de un artículo único que reforma el sistema para la valoración de los daños y perjuicios causados a las personas en accidentes de circulación en España. El principal objetivo que persigue es conseguir que las indemnizaciones sean más coherentes con la situación de la víctima, e incorpora en el cálculo variables personales, laborales, familiares y profesionales. En el baremo se contemplan el perjuicio personal básico (días no improductivos e improductivos), el perjuicio personal particular (impedimento o limitación que las lesiones o su tratamiento produzcan en la anatomía o desarrollo personal) y perjuicio patrimonial (daño emergente y lucro cesante). En el presente estudio, además de los criterios de causalidad del artículo 135 detallados con anterioridad, se exponen los

días no impeditivos e impeditivos y el perjuicio personal particular, dividido en perjuicio personal particular estético y perjuicio personal particular funcional.

Escala Visual Analógica (EVA). Para la valoración del dolor de los pacientes se utilizó una escala visual analógica de elaboración ad-hoc con puntuaciones del 1 (nada de dolor) al 10 (dolor extremo). Dicha escala incorporaba una guía de colores para facilitar su comprensión.

5.1.3. Procedimiento

Se utilizó un diseño analítico-descriptivo de corte transversal. Para el análisis de la fiabilidad, cuatro expertos en medicina legal y forense, especializados en la evaluación de lesiones corporales y con más de 30 años de experiencia en el campo, participaron como examinadores. Los expertos eran ciegos a los objetivos del estudio, que se presentó como una investigación de las características generales de SLC y los criterios de causalidad del Artículo 135.

Un investigador se encargó de contactar a los pacientes en la clínica donde fueron evaluados. Se les explicó el procedimiento que se iba seguir y, después de obtener su acuerdo verbal, se les solicitó la firma del consentimiento informado y se solventó cualquier posible duda que pudieran tener. A continuación, se remitió al paciente al primer examinador, quien realizó una evaluación clínica completa, analizó el historial clínico y evaluó el nexo causal con los criterios mencionados anteriormente (Clasificación 1). Toda la información recopilada durante la evaluación, excepto la Clasificación 1, se proporcionó a los otros tres evaluadores por separado para que aplicaran los criterios de causalidad y emitieran su clasificación. Una vez realizada la primera clasificación, se les dio el estudio de la biomecánica y se les pidió que reevaluaran el nexo causal, utilizando esta nueva información (Clasificación 2). El informe

biomecánico de la lesión lo realizaron expertos de una consultoría externa con amplia experiencia en el análisis de daños de vehículos. Los resultados del informe se plasmaban en dos valores: el indicador de magnitud de una colisión en ausencia de intrusión del habitáculo (Δv) y la aceleración media (a). Dichos valores eran interpretados por los evaluadores utilizando los umbrales propuestos por Represas *et al.* (2016).

5.1.4. Análisis de datos

El análisis estadístico se realizó con el programa SPSS, versión 24. Se calcularon los estadísticos descriptivos para las variables sociodemográficas y médico-legales recopiladas, incluida la media y la desviación estándar de las variables continuas y las frecuencias de las categóricas. Las diferencias entre las dos clasificaciones de cada uno de los cuatro expertos se estudiaron con la prueba Chi^2 . Para el análisis de la fiabilidad entre evaluadores, se calculó, para las dos clasificaciones, el coeficiente de correlación intraclassa (ICC, Modelo 2.1) y el intervalo de confianza del 95%. Los resultados del ICC se valoraron de la siguiente manera: 1 acuerdo absoluto, .80-.99 acuerdo excelente, .60-.79 acuerdo adecuado, .20-.59 acuerdo moderado y $<.19$ acuerdo mínimo.

De forma alternativa, el grado de acuerdo entre evaluadores también se valoró utilizando el índice de concordancia [$IC = \text{acuerdos} / (\text{acuerdos} + \text{desacuerdos})$], que es más restrictivo que el ICC (13,14). La decisión de utilizar un criterio de evaluación adicional se basa en la alta responsabilidad exigida en el contexto forense, en el que resulta prioritario minimizar el margen de error de las herramientas (Arce y Fariña, 2015).

5.2. Resultados

En relación con las características médico-legales de los pacientes (Tablas 3 y 4), un 78.2% de los mismos presentaron SLC de Grado I, mientras que el 21.8% presentaba Grado II. El 100% de los mismos manifestaba padecer dolor en el área cervical, con una media de severidad de 5.87 ($DT=1.21$) en EVA. Un 41% presentaba limitaciones leves en el rango de movimientos cervicales, un 15% moderadas, un 4% severas y el 40% restante no presentaba limitaciones. De la misma forma, se observaron contracturas leves en el 47.4% de los pacientes, moderadas en el 13.3% y severas en el 3.3%. El 36% restante no tenía ninguna contractura en el momento de la evaluación.

De los 300 pacientes incluidos, el 75.3% utilizó un modelo de asistencia pública y el 24.7% utilizó un modelo privado o mixto. El 28% de los pacientes tuvieron que ser trasladados en ambulancia y de ese total, el 14.28% sufrió pérdida de conciencia. Por otro lado, el 16.7% de esos pacientes tuvo que ser hospitalizado por un periodo superior a 24 horas, con un tiempo medio de hospitalización de 4.35 días ($DT=16.03$), el resto fue dado de alta en un tiempo inferior al citado ($M=12.24$ horas, $DT=4.35$). De dicho 16.7% de pacientes, un 12.5% fue hospitalizado por lesiones musculares leves en los brazos y el 4.2% restante debido a que referían padecer mareos y vértigos graves.

En lo relativo a las variables del Baremo español de lesiones en accidentes de tráfico, los pacientes tuvieron una media de 43.85 ($DT=67.48$) de días impeditivos, 87.86 ($DT=77.06$) de días no impeditivos, 4.22 ($DT=7.47$) de puntos de perjuicio personal particular funcional y 1.09 ($DT=3.03$) de puntos de perjuicio personal particular estético. Un 88.7% de los pacientes habían recibido rehabilitación en el momento de la evaluación, mientras que el 11.3% restante no la había necesitado. Por otro lado, únicamente se tenía información sobre el estado anterior de un 12.7% de los pacientes. Un 21% de los mismos

tenían un diagnóstico psiquiátrico previo al accidente, tratándose de trastornos de tipo depresivo en todos los casos.

Finalmente, los resultados del informe biomecánico indicaron unos valores medios de Delta v (Δv) de 8.75 Km/h ($DT=5.41$) y de aceleración media (a) de 7.11 g ($DT=4.21$).

Tabla 3. *Media y desviación típica de las características médico-legales de los pacientes*

	M	DT
Severidad del dolor	5.87	1.21
Abbreviated injury scale (AIS)	1.26	.60
Días de hospitalización	4.35	16.03
Días impeditivos	43.85	67.48
Días no impeditivos	87.86	77.09
Puntos de alteración funcional	4.22	7.47
Puntos perjuicio estético	1.09	3.03
Delta v (Δv)	8.75	5.41
Aceleración media (a)	7.11	4.21

Tabla 4. *Frecuencias de las características médico-legales de los pacientes*

	Frecuencia	%
Cinturón de seguridad abrochado	268	89.3
Sí	22	7.3
No	10	3.3
No aplicable		
Modelo asistencial		
Público	293	97.7
Privado/mixto	74	24.7
Rehabilitación		
Sí	263	88.7
No	34	11.3
Accidente laboral		
Sí	31	1.3
No	144	48.0
No aplicable	125	41.7
Baja médica		
Sí	124	41.3
No	51	17.0
No aplicable	125	41.7
Estado anterior		
Sí	38	12.7
No	262	87.3
Diagnóstico psiquiátrico previo		
Sí	8	2.7
No	292	97.3

Tabla 4. Frecuencias de las características médico-legales de los pacientes (continuación).

	Frecuencia	%
Grado del SLC		
Grado I	235	78.2
Grado II	65	21.8
Limitación en el rango movimientos		
Ninguna	120	40
Leve	123	41
Moderada	45	15
Severa	12	4
Severidad de la contractura		
Ninguna	108	36
Leve	142	47.4
Moderada	40	13.3
Severa	10	3.3
Pérdida de consciencia		
Si	12	4.0
No	288	96.0
Transferencia en ambulancia		
Si	84	28.0
No	216	72.0
Ingreso en UCI		
Si	2	.7
No	298	99.3
Hospitalización > 24h		
Si	50	16.7
No	250	83.3
Colisión compleja		
Si	49	16.3
No	251	83.7

En lo relativo a la clasificación realizada por cada uno de los evaluadores (Tabla 5), en términos generales en aproximadamente la mitad de los pacientes se ha considerado que sí existe nexo causal, tanto en el análisis sin estudio de la biomecánica (Evaluador 1: 56.7%; Evaluador 2: 84%; Evaluador 3: 69.7%; Evaluador 4: 49%) como con estudio

(Evaluador 1: 52.3%; Evaluador 2: 32%; Evaluador 3: 60.3%; Evaluador 4: 61.3%). De la misma forma, el análisis indica que existen diferencias significativas entre dichos análisis en los cuatro evaluadores (Evaluador 1: $X^2(4)=525.23$, $p<.001$; Evaluador 2: $X^2(4)=12.68$, $p=.013$; Evaluador 3: $X^2(4)=304.55$, $p<.001$; Evaluador 4: $X^2(4)=180.68$, $p<.001$).

Tabla 5. Clasificación de pacientes de cada uno de los evaluadores

	Si		No		Dudosos		X^2	gl	p
	n	%	n	%	n	%			
S.B.									
Evaluador 1	170	56.7	62	20.7	68	22.7			
Evaluador 2	252	84	5	1.7	43	14.3			
Evaluador 3	209	69.7	45	15	46	15.3			
Evaluador 4	147	49	102	34	51	17			
C.B.									
Evaluador 1	157	52.3	67	22.3	76	25.3			
Evaluador 2	96	32	162	54	42	14			
Evaluador 3	181	60.3	70	23.3	49	16.3			
Evaluador 4	184	61.3	65	21.7	51	17.0			
Comparación									
S.B. y C.B.									
Evaluador 1							525.23	4	< .001
Evaluador 2							12.68	4	.013
Evaluador 3							304.55	4	< .001
Evaluador 4							180.68	4	< .001

Abreviaturas: S.B. sin estudio de la biomecánica; C.B., con estudio de la biomecánica., X^2 , prueba Chi cuadrado para las diferencias entre clasificaciones de cada experto; gl, grados de libertad; p, nivel de significación.

Los índices de concordancia (IC) y los coeficientes de fiabilidad de los juicios con su correspondiente intervalo de confianza del 95%, así como el análisis de varianza (ANOVA), se pueden encontrar en la Tabla 6. Los resultados tanto del coeficiente de correlación intraclase como del índice de concordancia del juicio antes del estudio biomecánico de la lesión arrojaron valores adecuados (ICC = .70, CI = .60).

Asimismo, los valores observados en el juicio con el estudio de la biomecánica también indicaron un grado de acuerdo adecuado (ICC = .78, CI = .67). En lo relativo a la clasificación criterio a criterio, en el análisis sin estudio de la biomecánica los criterios a (exclusión) y b (temporal) han obtenido valores elevados (ICC=1, CI=1; ICC=.96, CI=.81 respectivamente), el criterio c (topográfico) valores moderados (ICC=.71, CI=.57) y el criterio d (de intensidad) valores bajos (ICC=.14, CI=.11). Tras el estudio de la biomecánica, los 3 primeros criterios permanecen iguales, pero el criterio d alcanza valores moderados-bajos (ICC=.46, CI= .31).

Tabla 6. Coeficiente de correlación intraclase (ICC), Índice de concordancia y ANOVA

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

Abreviaturas: S.B. sin estudio de la biomecánica; C.B., con estudio de la biomecánica; ICC, Coeficiente de correlación intraclase; CI, índice de concordancia; 95% IC, intervalo de confianza al 95% del ICC; F, F de Fisher del ANOVA; p, nivel de significación.

Discusión

El objetivo de esta investigación era evaluar el grado de acuerdo de cuatro expertos en medicina legal en el establecimiento del nexo causal con y sin el estudio de la biomecánica. Nuestros resultados indican que el grado de concordancia entre las calificaciones de los cuatro expertos fue moderado-alto tanto para la calificación 1, realizada sin el estudio de la biomecánica de la lesión, como para la calificación 2, realizada después de dicho estudio. La calificación 2 alcanzó una mayor fiabilidad entre evaluadores, por lo que se confirma nuestra hipótesis de investigación, y el estudio de la biomecánica genera un incremento, leve en nuestro caso, de la concordancia general. De manera específica, nuestros hallazgos indican que la inclusión del estudio biomecánico no afectó a la clasificación dada por los expertos en los primeros tres criterios, pero tuvo un impacto en el criterio de intensidad, mejorando significativamente los valores tanto del criterio general como de los dos subcriterios, en particular el criterio d2 (compatibilidad biomecánica). A pesar de ello, resulta importante señalar que el estudio de la biomecánica no ha ayudado a disminuir la proporción de casos dudosos, ya que apenas hay cambios entre las dos clasificaciones, sino que ha servido, en términos generales, para que los evaluadores ajusten su tasa de casos no causados.

En este sentido, se esperaría que, al operativizar el criterio y ofrecer a los expertos umbrales para valorarlo, se hubieran alcanzado niveles de concordancia más elevados que los obtenidos, y que hubieran disminuido el número de casos dudosos. Creemos que una explicación a estos resultados puede ser que la operativización propuesta no ha sido lo suficientemente objetiva, y es posible que los umbrales de Represas *et al.* (2016) no alcancen el nivel de especificidad necesario para lograr una interpretación uniforme. En este sentido, deben tenerse en consideración una serie de variables que pueden influir en la posibilidad de aparición de una lesión, como los factores de riesgo, los sistemas de

seguridad, etc. Esto puede hacer que aquellos casos situados muy cerca de los puntos de corte, y que se vean afectados por las variables citadas, generen distintas opiniones al tener el experto que hacer una valoración subjetiva sobre la posible influencia de las mismas. A pesar de ello, creemos que es necesario resaltar que el criterio de intensidad presenta un carácter subjetivo y, de cara a futuras publicaciones sería recomendable explorar nuevas formas de operativizarlo. Los umbrales utilizados eran una propuesta teórica que sintetizaba la evidencia disponible en la materia y, si bien no han permitido alcanzar unos niveles de concordancia elevados, si han generado cierta mejora. Por ello, consideramos que pueden resultar de utilidad para la valoración del mencionado criterio y, en futuras investigaciones, sería de interés desarrollar una clasificación más amplia, concreta y con unos umbrales más específicos.

En relación con la proporción de excluidos, consideramos llamativos los valores observados (entre un 20% y un 30%), ya que es posible que el contexto de la evaluación pueda influir en los mismos. De cara a futuras investigaciones creemos que puede ser de interés estudiar si existen diferencias entre las clasificaciones realizadas por peritos de parte y expertos de los institutos de medicina legal.

Los resultados obtenidos en los otros tres criterios (topográfico, temporal y de exclusión) indican una elevada convergencia en los mismos, sin cambios al introducir el estudio de la biomecánica, lo que sugiere que son independientes de otros posibles criterios y útiles por sí mismos. En nuestro estudio, el criterio temporal no generó ningún desacuerdo, y el criterio de exclusión apenas generó, pero el criterio topográfico obtuvo valores moderados-altos que son de interés. El criterio topográfico se operativizó en dos subcriterios debido a que los evaluadores consideraron que resultaría de interés valorar tanto el área corporal afectada por la lesión (criterio c1) como la coherencia de los síntomas expresados/observados (criterio c2). Si se realiza una valoración exclusivamente

topográfica mediante la aplicación del criterio c1 se obtienen valores de concordancia elevados, ya que en la gran mayoría de los casos la localización de la lesión está clara y solo se producía falta de consenso en aquellos en los que el dolor se irradiaba de forma infrecuente. Pero, al incluir el criterio c2, el nivel de acuerdo sufre una disminución. Creemos que una posible explicación a esos resultados es la amplia heterogeneidad sintomática de la condición estudiada y la dificultad para objetivar parte de los mismos.

Es posible que, al no valorar variables de corte psicosocial siguiendo un procedimiento estandarizado, los expertos interpretaran la coherencia de algunos síntomas o signos basándose únicamente en la propia experiencia. Por ejemplo, uno de los profesionales podría haber considerado que un posible cuadro ansioso-depresivo -que puede haberse originado por las circunstancias y características interpersonales del paciente y no por la magnitud del accidente- no era coherente con la intensidad de la colisión y no necesitaba un análisis más exhaustivo, mientras que, otro de los expertos, podría considerar que es un síntoma válido tras realizar una exploración más profunda. De cara a futuras investigaciones, creemos que puede resultar de interés analizar la coherencia de la sintomatología expresada por los pacientes (criterio c2), pero sería recomendable incluir un listado de todas las variables que puede presentar el SLC, incluidas aquellas de corte psicosocial.

Por otro lado, consideramos de interés las diferencias en las clasificaciones realizadas por los expertos. Puede observarse que, a pesar de alcanzar valores de concordancia moderados-altos, dichas clasificaciones difieren entre sí, en especial las del evaluador 2 que, tras el estudio de la biomecánica, modifica significativamente su clasificación y aumenta dramáticamente la clasificación de casos no causados. Creemos que las diferencias observadas podrían indicar, como ya se ha señalado con anterioridad, que el criterio C2 y el criterio d -los que han presentado menos concordancia- han

conservado un carácter subjetivo y la operativización propuesta no ha sido del todo efectiva para que los evaluadores alcancen consenso. Cabe recordar que, en términos generales, la metodología aplicada para la valoración pericial es eminentemente personal y los resultados obtenidos rara vez son sometidos a una revisión crítica. Ante esta situación, si el profesional no cuenta con un procedimiento estandarizado, la valoración se basará en la propia experiencia profesional, la cual puede diferir significativamente entre los expertos.

Con el objetivo de valorar los resultados obtenidos, la utilidad y suficiencia de los criterios utilizados debería ser considerada. Los valores de concordancia obtenidos indicarían que los criterios a-c (considerando el criterio c1), están operativizados correctamente y son específicos, por lo que pueden resultar de utilidad para establecer el nexo causal. A pesar de ello, creemos que su suficiencia debe ser valorada con precaución. Consideramos que los criterios utilizados son deterministas. En la gran mayoría de los casos, los accidentes de circulación se verán influenciados por variables externas que provoquen cambios en los resultados. Si bien es posible predecir la evolución y el riesgo de complicaciones, estas variables incrementarán significativamente las posibilidades, alejándose de dicha naturaleza determinista (Luna, 2017). Tomando como ejemplo el criterio temporal no es posible asegurar que ningún caso sea excluido erróneamente por la influencia de dichas variables (resistencia al dolor, disponibilidad de dispositivos asistenciales, circunstancias personales y familiares, etc.). Por ello, los criterios utilizados pueden no ser suficientes, y sería aconsejable ampliar la lista con factores validados científicamente. En este sentido, resultaría interesante considerar propuestas como la de Repesas (2017), que incluye tres criterios con sus respectivos criterios secundarios, uno biomecánico, uno epidemiológico y uno del estado previo del paciente. Considerar las particularidades de la condición estudiada puede ser otro enfoque para desarrollar nuevos

criterios. Como se mencionaba en la introducción, el SLC es una condición con una elevada complejidad diagnóstica debido a la ausencia de indicadores fiables (Represas, 2017). El procedimiento de evaluación de la lesión aplicado habitualmente se centra en la búsqueda de divergencias estructurales o anatómicas mediante prácticas diagnósticas que no logran alcanzar una eficacia adecuada (Oddsdóttir *et al.*, 2015; Valera *et al.*, 2001; Waeyaert *et al.*, 2016). La evidencia sugiere que el SLC debe ser evaluado siguiendo un modelo biopsicosocial (Phillips *et al.*, 2010; Walton y Elliott, 2017). Este enfoque permitiría al evaluado observar la interacción entre los factores médicos, biomecánicos, sociales y psicológicos y ofrecería una visión mucho más amplia de la lesión que sería de utilidad en el estudio de la causalidad (Turk *et al.*, 2018). De cara a futuras investigaciones debería complementarse la valoración del daño corporal y el estudio de la biomecánica de la lesión con un análisis de las variables psicosociales que influyen en el SLC, como la percepción y las expectativas de recuperación de la enfermedad, la presencia de sintomatología ansioso-depresiva, la percepción de discapacidad o la catastrofización del dolor (Carroll *et al.*, 2009; Falla *et al.*, 2016; Holm *et al.*, 2008).

Finalmente, creemos que los criterios de nexo causal pueden ser de gran utilidad para la formación de los grupos criterioles que se realiza en los diseños de criterio parcial, e incluso en los de grupo conocido si se complementa con los instrumentos pertinentes. En el marco de la presente tesis doctoral se incluirán en el protocolo para tal fin, así como para estudiar las posibles diferencias que puedan existir entre pacientes con el nexo causal establecido y pacientes sin el mismo. Se considera además el estudio de la biomecánica, pero la posibilidad de que se realice deberá ser consultada con los/as profesionales que participen en los subsiguientes estudios.

6. Bloque 1 - Estudio 2. Valoración de la simulación en SLC: Una revisión sistemática.

Como ya comentábamos en el Marco Teórico, el riesgo de simulación en SLC es alto debido a que ofrece pocos marcadores a nivel físico y la evaluación clínica tiende a seguir el modelo biomédico, centrándose en la búsqueda de divergencias estructurales o anatómicas mediante prácticas diagnósticas que cuentan con poca eficacia (Valera-Garrido *et al.*, 2001; Waeyaert *et al.*, 2016). En términos generales, la evaluación de la simulación en SLC se suele enfocarse desde al análisis de síntomas físicos, metodología que no ha demostrado ser del todo productiva, ya que numerosos estudios han aportado evidencia significativa de que este tipo de lesiones están influenciadas por factores psicológicos y sociales, y buscar únicamente indicios biológicos supone una aproximación simplista a la evaluación del fenómeno (Ferrari *et al.*, 2005; Schoenberg *et al.*, 2003).

Por ello, para el desarrollo del protocolo se pretende enfocar la valoración de la simulación desde una perspectiva biopsicosocial (Engel, 1977) que ayude al evaluador a entender la influencia de los factores psicológicos y sociales en la manifestación médica del fenómeno y permita obtener una amplia gama de resultados que engloben la gran variedad de síntomas y factores que componen la condición (Elliot *et al.*, 2015; Ferrari *et al.*, 2005; Valera *et al.*, 2001). Así, se construirá con instrumentos que valoren al menos una de las 3 dimensiones y deberá incluir, como mínimo, un instrumento en cada dimensión.

Consideramos necesario que la elección de los instrumentos que componen el citado protocolo se realice de forma estructurada y sistemática, garantizando así que se valoran todas las aportaciones realizadas por los expertos en la temática y que dichos

instrumentos tienen propiedades psicométricas adecuadas. De la misma forma debe tenerse presente la posibilidad de que, entre el listado de instrumentos resultante de la búsqueda, se encuentren algunos que no han sido validados en una muestra española y no puedan utilizarse. Por ello, se plantea una revisión sistemática, que tiene como objetivo:

- **Objetivo específico 1.2.** Identificar aquellas medidas de resultado utilizadas para evaluar simulación en las tres dimensiones (biológica, psicológica y social) del SLC que puedan ser empleadas como indicadores de simulación

Así, el producto del estudio no solo serán los instrumentos o herramientas, sino sus medidas del resultado (qué mide el test, del inglés *outcome measures*), de manera que, si uno de ellos no puede utilizarse, se considerará su medida de resultado y se tratará de localizar una herramienta válida que persiga la misma finalidad.

6.1. Metodología

Para alcanzar el objetivo del presente estudio se llevó a cabo una revisión sistemática de la evidencia científica, siguiéndose para ello las directrices de la Declaración PRISMA (Moher *et al.*, 2009).

6.1.1. Criterios de selección

Se establecieron los siguientes criterios de inclusión: (a) estudios experimentales o cuasi-experimentales que utilizan uno de los cuatro diseños de investigación en simulación (grupo conocido, simulación, prevalencia diferencial y criterio parcial); (b) la muestra incluye hombres y mujeres mayores de 18 años; (c) los pacientes clínicos incluidos han sido diagnosticados con latigazo cervical, SLC o se incluye una muestra general con discapacidad relacionada con el dolor (*Pain Related Disability*) o dolor

crónico (*Chronic Pain*) que contenga pacientes con daño cervical; (d) el estudio propone una herramienta o metodología que permite detectar simulación.

Como criterio de exclusión, utilizamos: (a) Ausencia de grupo clínico. Los estudios con diseño de simulación originalmente comparaban a un grupo de estudiantes instruidos para responder correctamente a las escalas y otros instruidos para fingir. Para aumentar la validez externa, comenzó a incluirse un grupo clínico de referencia con el que hacer una comparación cuasi-experimental de los resultados de los simuladores. Como la ausencia de un grupo de referencia clínico es un defecto fatal en este diseño (Rogers y Bender, 2012, 2018), decidimos excluir todos aquellos estudios con un Diseño de Simulación que no incluían un grupo de comparación clínica. (b) Falta de criterios de calidad. Utilizamos una herramienta de riesgo de sesgo con la finalidad de estudiar la calidad de los artículos incluidos en esta revisión y cómo puede afectar la misma a los resultados presentados. Si los artículos estudiados no cumplían con ninguno de los 5 primeros criterios de la herramienta de análisis de sesgo, se excluían del estudio.

6.1.2. Estrategia de Búsqueda

La búsqueda se realizó en las siguientes bases de datos: Medline, Cochrane, Academic Search Premier, PsycINFO y PsycARTICLES. Los principales descriptores utilizados fueron “Malingering (Malinge*)” “Feign” (Feign*), “Whiplash”, “Neck pain”, “Chronic pain”, “MPRD” (malingered pain-related disability), “Spine injury”, “WAD” (Whiplash associated disorders), “Physical symptoms” y “Psychological symptoms”, siendo la última fecha de búsqueda el 14/12/2019.

Debido a que no se localizó ninguna revisión sistemática previa sobre simulación en SLC no se estableció ninguna restricción de fechas.

Una búsqueda preliminar utilizando únicamente “Whiplash” o “Neck pain” y “Malinge*” reveló estudios en los que, a pesar de que en el título no se especificaba explícitamente, se incluían una gran cantidad de pacientes con sintomatología propia del SLC bajo el título genérico de “daño cervical”. Para evitar perder instrumentos que pudieran ser de utilidad, se realizó una revisión conjunta de estudios que incluyeran pacientes con daño cervical, latigazo cervical o SLC.

6.1.3. Selección de los estudios

Teniendo en cuenta los criterios de selección mencionados, la extracción de estudios se realizó en dos fases:

- En la primera fase, se desarrolló una hoja de extracción de datos basada en las recomendaciones de la declaración PRISMA (Moher *et al.*, 2009) y se probó con cinco de los estudios incluidos, seleccionados al azar. Posteriormente, dos revisores independientes, preseleccionaron los artículos basándose en el título y el resumen. A continuación, ambos acordaron la lista resultante, resolviéndose los desacuerdos mediante consenso.
- En la segunda fase, los dos revisores leyeron de forma independiente el texto completo de los estudios preseleccionados, desarrollando una nueva lista de artículos potencialmente relevantes, resolviendo nuevamente las discrepancias a través de consenso. Si no se llegaba a un acuerdo, se pedía a un tercer revisor que determinara si el estudio cumplía con los criterios de selección. Los textos completos de los artículos aceptados finalmente fueron cuidadosamente leídos, y sus referencias bibliográficas revisadas a fin de identificar posibles artículos pertinentes no localizados en la búsqueda inicial.

Con el fin de reducir la duplicación no planificada de comentarios y proporcionar transparencia en el proceso de revisión, así como de minimizar el sesgo de notificación, este estudio fue registrado en PROSPERO (International Prospective Register of Ongoing Systematic Reviews, <http://www.crd.york.ac.uk/prospero>) (Booth *et al.*, 2011), con número de registro: CRD42017070746.

6.1.4. Evaluación de Riesgo de Sesgo

Debido a que ésta parece ser la primera revisión sistemática de simulación en SLC realizada hasta la fecha que sigue a la declaración PRISMA, no pudimos encontrar un instrumento o método para evaluar el riesgo de sesgo y la validez de los resultados de este trabajo. Consideramos que una herramienta genérica para evaluar el riesgo de sesgo en estudios experimentales o cuasi-experimentales no cumpliría adecuadamente con su propósito, ya que ignoraría las características específicas de los diseños utilizados en la investigación de simulación. Por esta razón, desarrollamos una herramienta de evaluación del riesgo de sesgo siguiendo los trabajos de Rogers y Bender (2012, 2018), González-Ordi *et al.* (2012) y Singh (2013). Está compuesta por 26 criterios y se divide en cinco apartados (Anexo 1). El primero se llamó “Calidad” y contiene cinco criterios de eliminación basados en el *Programa de Habilidades de Lectura Crítica* (CASP; Al-garadi *et al.*, 2016; Singh, 2013). Estos criterios comprenden aspectos metodológicos básicos y el evaluador puede tomar la decisión de excluir el estudio si alguno de estos criterios no se cumple. El segundo apartado se llamó “Análisis de resultados” y se basa en la propuesta de González-Ordi *et al.* (2012) y aportaciones de Rogers y Bender (2018). Evalúa la estrategia de análisis de datos seguida por los autores. El tercer apartado está basado en la guía CASP y se llamó "Aplicabilidad". Analiza si los resultados obtenidos en los estudios son aplicables al contexto clínico, al entorno del lector y si coinciden con la evidencia disponible. El cuarto apartado se tituló “Diseño” y está basado en el trabajo

de Rogers y Bender (2012). Tiene dos grupos de criterios diferentes, “a” ó “b”, y el autor debe elegir uno u otro en función del diseño que esté utilizando, “a” para diseño de simulación y “b” para diseño de grupo conocido, prevalencia diferencial y criterio parcial. El último apartado se denominó “Otras fuentes de sesgo” y sirve para recopilar todas las posibles fuentes de sesgo identificadas que pueden alterar significativamente los resultados y que no se han incluido en los otros criterios. Debido a que la Declaración PRISMA indica que una valoración numérica total puede resultar confusa, cada uno de los criterios incluidos se valora con “sí”, “no” o “no sé”.

Dos revisores independientes, autores de este artículo, aplicaron esta herramienta a cada uno de los artículos aceptados, alcanzándose una alta fiabilidad entre jueces (coeficiente Kappa de Cohen, $k = .89$).

6.1.5. Tabulación y Análisis de Datos

Los estudios elegidos fueron codificados en una tabla donde se recogieron los siguientes resultados: datos de la investigación (autores y año de publicación); principal objetivo del estudio; características de los participantes (tamaño muestral y subgrupos), lugar de realización del estudio, escalas utilizadas, diseño del estudio, dimensión de las medidas del resultado, medidas del resultado evaluadas, indicadores de simulación y resultados significativos.

La columna “indicadores de simulación” se incluyó con la intención de ofrecer una visión práctica de las medidas del resultado. En dicha columna se expone cómo se manifiesta esta medida en el grupo de simuladores. El signo delante del indicador expresa si se manifiesta en mayor (+) o en menor medida (-) en el grupo de simuladores. De manera que, si la principal medida de resultado era “intensidad o severidad del dolor”, si el grupo de simuladores ofrece una puntuación mucho más elevada que el grupo de

pacientes clínicos, el indicador será “(+) Gravedad del dolor”. De esta forma, si por ejemplo se observa que la severidad del dolor es una medida de resultado/indicador útil para la valoración de la simulación, pero el instrumento utilizado para estudiarlo no está disponible en castellano, se buscará un instrumento alternativo en castellano que mida la variable de interés.

Para facilitar la exposición de los resultados, las medidas de resultado se agrupan en dimensiones según el modelo biopsicosocial (Engel, 1977):

1. *Dimensión Biológica.* Contempla aquellas medidas del resultado de origen físico. Se incluye el número e intensidad de los síntomas somáticos, la localización e intensidad de dolor, movimientos del cuello y signos no-orgánicos.
2. *Dimensión Psicológica.* Esta categoría aúna aquellas medidas del resultado de origen psicológico, entre las que se incluye el rendimiento o grado de alteración cognitiva y el perfil psicopatológico/sintomatología psicológica del evaluado.
3. *Dimensión Social.* Categoría que incluye el grado de alteración en la calidad de vida del evaluado.
4. *Elementos de apoyo a la evaluación.* En este apartado se han añadido aquellas herramientas que obtienen como medida de resultado riesgo o presencia de simulación y que pueden ser utilizadas como herramientas complementarias.

6.2. Resultados

En la Figura 5 se resumen los resultados de las distintas etapas del proceso de revisión. Aplicando la estrategia de búsqueda descrita, se identificaron un total de 523 publicaciones, a las que se sumaron 36 artículos tras realizar búsqueda secundaria en listas de referencias. Se eliminaron 128 trabajos por encontrarse duplicados. De los 395

artículos restantes se realizó una lectura de título y resumen y se aplicaron los criterios de selección establecidos, excluyendo 333 artículos y dejando 62 publicaciones para su lectura a texto completo. Tras realizar dicha lectura, se excluyeron 34 artículos por no cumplir los criterios de selección, quedando 28 artículos que fueron sometidos al análisis de riesgo de sesgo, de los cuales 5 estudios no cumplían los criterios de calidad definidos. El proceso concluyó con la selección de los 23 artículos que forman parte de la presente revisión.

El tamaño muestral total es de 7303 adultos. De dicho total, 4841 adultos corresponden a la categoría de “Pacientes clínicos” y 1528 a la categoría de “Simuladores”, cuya composición varía en función al diseño de los estudios. El resto de los participantes se incluyen dentro de la categoría de “Controles” ($n = 622$) e “Indiferenciados” ($n = 312$).

Quince estudios utilizan un diseño de grupo conocido, con varias categorías de simuladores, habitualmente “posible, probable y definitiva simulación”. Siete de estos estudios utilizan criterios de simulación, como los de Bianchini *et al.* (2005), para formar el grupo conocido. El resto de los estudios utilizan un diseño de simulación, excepto por Fishbain *et al.* (2002) y Schmand *et al.* (1998) que utilizan un diseño de prevalencia diferencial.

Aclarar que, cuando en nuestro texto hablamos de “simuladores”, nos referimos a los *análogos* y a los pacientes clasificados como “probable y definitiva simulación”. Los pacientes clínicos son aquellos que han demostrado tener sintomatología genuina.

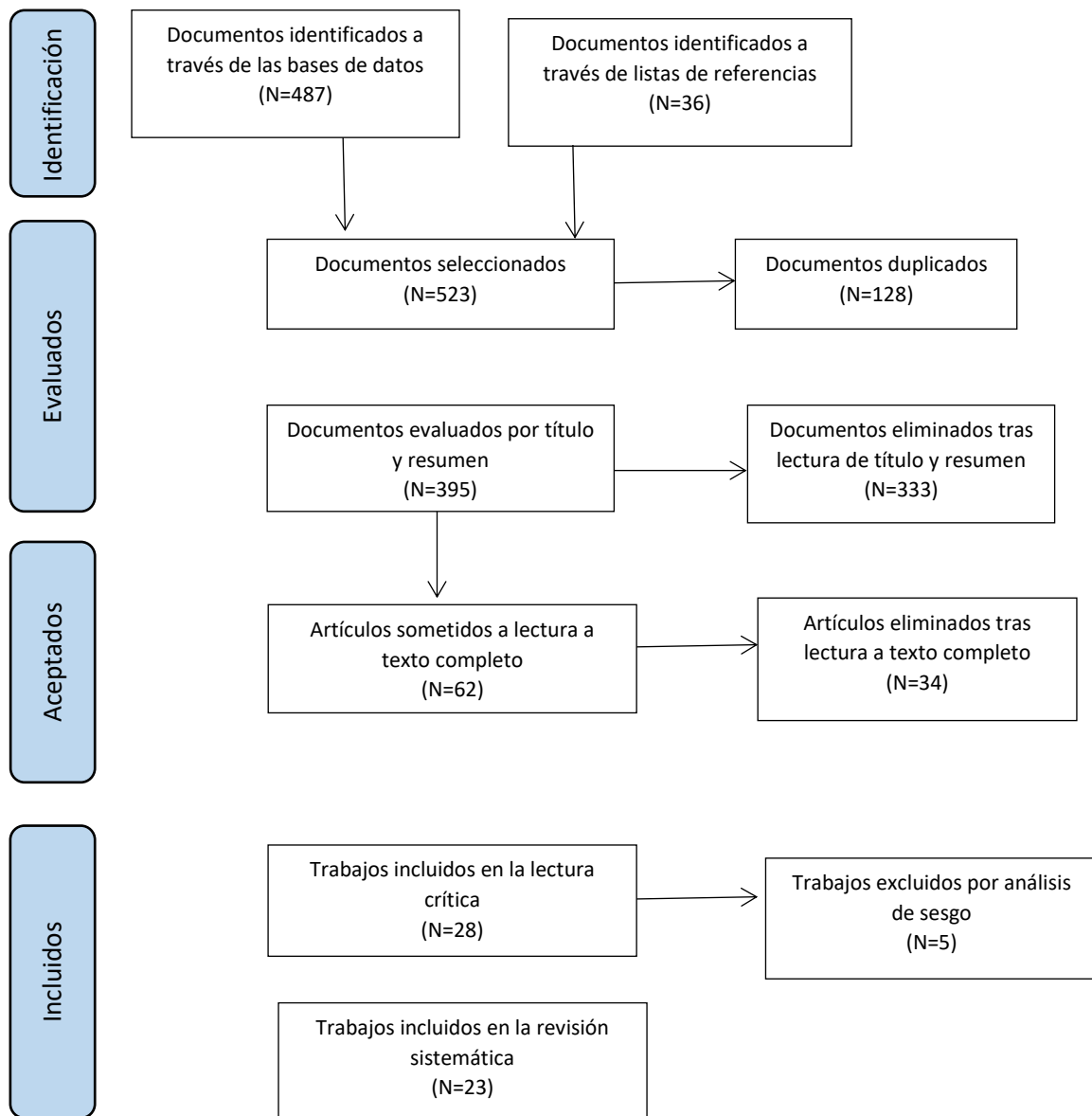


Figura. 5. El proceso de selección (siguiendo la guía PRISMA, Moher *et al.*, 2009).

Los principales resultados de la presente revisión sistemática se muestran en la Tabla 7. Se han identificado 13 medidas de resultado (manifestación de síntomas somáticos, percepción del impacto en la calidad de vida, perfil psicopatológico y sintomatología psicológica, riesgo de simulación, actitud ante la situación de enfermedad, precisión de movimientos del cuello, percepción de grado discapacidad, grado de dolor percibido, inconsistencias durante la evaluación física del paciente, rendimiento cognitivo, grado de deterioro cognitivo, grado de alteración emocional y número de

síntomas), las cuales se exponen a continuación agrupadas en función de la dimensión a la que pertenecen.

6.2.1. Dimensión biológica

Dentro de esta categoría se ha incluido un total de 12 artículos que valoran una herramienta y/o indicador de tipo biológico (Bianchini *et al.*, 2014; Crighton *et al.*, 2014; Hopwood *et al.*, 2010; Oddsdottir *et al.*, 2015; McGuire *et al.*, 2001; Larrabee, 2003; Linnell y Easton, 2006; McGuire y Shores, 2001; Sobel *et al.*, 2000; McGuire y Shores, 2001b; Wallis y Bogduk, 1996; Wygant *et al.*, 2017).

- Dos de estos estudios se centran en el uso de los signos no orgánicos para **buscar inconsistencias durante la evaluación física del paciente**³ (Sobel *et al.*, 2000; Wygant *et al.*, 2017). Para evaluar dichas inconsistencias, Wygant *et al.* (2017) utilizaron una versión actualizada de los signos de Sobel (Sobel *et al.*, 1980), obteniendo una correlación positiva entre los signos detectados y un sesgo en el estilo de respuesta de los síntomas reportados. Observaron que puntuaciones superiores o iguales a 2 en los signos (rango de 0 a 3), se asociaron con un mayor riesgo de que el paciente estuviera magnificando la cantidad de síntomas padecidos. Por otro lado, Sobel *et al.* (2000) adaptaron los Signos evaluados en el estudio anterior, creando una nueva herramienta llamada “Non-Organic Cervical Signs”. Obtuvieron resultados similares al primer estudio mencionado, destacando el signo de “Pérdida sensorial” por su capacidad discriminativa. Por otro lado, recomiendan, desechar los signos “Simulación mientras el paciente está sentado” y, dentro del apartado “Sobrerreacción”, “Tensión muscular y

³ Las medidas de resultado se resaltan en negrita para facilitar la lectura.

“Verbalización desproporcionada” por ser difíciles de ejecutar y no aportar resultados suficientemente robustos.

- El estudio de Oddsdottir *et al.* (2015) usa la herramienta llamada “Fly Test” para comparar la **precisión de los movimientos del cuello**. Se observó que los pacientes clínicos tenían un patrón de movimiento más limitado, menos preciso y pasaban menos tiempo en el objetivo (25% frente al 40% de los simuladores). Las medidas del resultado empleadas arrojaron resultados positivos, con una sensibilidad y especificidad media del 81% y del 66%, respectivamente.
- Los nueve estudios restantes evalúan mediante autoinforme las **inconsistencias en la manifestación de los síntomas somáticos** de los pacientes (Bianchini *et al.*, 2014; Crighton *et al.*, 2014; Hopwood *et al.*, 2010; Larrabee, 2003; Linnell y Easton, 2006; McGuire *et al.*, 2001; McGuire y Shores, 2001a; McGuire y Shores, 2001b; Wallis y Bogduk, 1996), concluyendo que los simuladores obtienen puntuaciones más altas y puntúan en más subescalas que los pacientes clínicos, traduciéndose esta diferencia en que los primeros magnifican la intensidad y el número de síntomas producidos, tienden a sobrestimar la gravedad de la patología, y tienen una peor identidad de síntomas (Bianchini *et al.*, 2014; Crighton *et al.*, 2014, Larrabee, 2003; McGuire *et al.*, 2001; McGuire y Shores, 2001; McGuire y Shores, 2001b; Wallis y Bogduk, 1996). Las herramientas utilizadas para contrastar dichas inconsistencias se han incluido en la Tabla 7. Mencionamos aparte el estudio de Hopwood *et al.* (2010), el cual utilizó el Personality Assessment Inventory (PAI; Morey, 1991; Morey, 2007), concluyendo que dicha escala no alcanza suficiente sensibilidad como para ser útil en el contexto clínico-forense.

Por otro lado, de estos nueve estudios, cinco estudian la utilidad de la **intensidad o grado del dolor** (Larrabee, 2003; Linnell y Easton, 2006; McGuire *et al.*, 2001; McGuire y Shores, 2001b; Wallis y Bogduk, 1996), obteniendo evidencias que indican que los simuladores generan una percepción de dolor más intensa que los pacientes clínicos. Para valorar tales evidencias, Wallis y Bogduk (1996) utilizan la escala McGill Pain Questionnaire (MPQ; Melzack, 1975) y el Symptom Check List SCL-90-Revised (Derogatis, 1983), encontrando diferencias significativas únicamente en esta última escala. Larrabee (2003) sí obtuvo diferencias significativas entre grupos en el MPQ, pero advierte que el tamaño del efecto es bajo. Linnell y Easton (2006), utilizando una batería que incluye el Pain Disability Index (PDI; Pollard, 1984) y el Illness Perception Questionnaire (IPQ; Weinman *et al.*, 1996), añaden que los pacientes de dolor crónico puntúan aún más bajo que los pacientes agudos. Por último, McGuire *et al.* (2001) y McGuire y Shores (2001) utilizaron el Pain Perception Questionnaire (P3; Tollison y Langley, 1995) con dos grupos de simuladores, uno instruido y otro no. La escala fue capaz de detectar la magnificación de la intensidad del dolor en los dos grupos, pero fue más eficaz al identificar el grupo de instruidos gracias a que su escala de validez detectó sesgo en el estilo de respuesta.

6.2.2. Dimensión psicológica

La utilidad del rendimiento cognitivo, del grado de alteración cognitiva y del perfil psicopatológico/sintomatología psicológica del evaluado han sido factores estudiados por once de los estudios incluidos (Eherton *et al.*, 2006; Eherton *et al.*, 2006; Greve *et al.*, 2010; Greve *et al.*, 2009; Greve *et al.*, 2009; Greve *et al.*, 2009; McGuire *et al.*, 2001; McGuire y Shores, 2001; Schadman *et al.*, 1998; McGuire y Shores, 2001b; Wallis y Bogduk, 1996).

- La validez del **grado de alteración cognitiva** como medida discriminativa entre pacientes clínicos que afirman padecer déficits cognitivos por el accidente y simuladores ha sido evaluada por siete de estos estudios (Etherton *et al.*, 2006; Etherton *et al.*, 2006; Greve *et al.*, 2010; Greve *et al.*, 2009; Greve *et al.*, 2009; Greve *et al.*, 2009; Schadman *et al.*, 1998). Cinco indican que los simuladores obtienen puntuaciones más bajas en las escalas empleadas, observándose una tendencia a magnificar la gravedad de los síntomas padecidos y a puntuar por debajo de sus capacidades (**rendimiento subóptimo**) (Etherton *et al.*, 2006; Greve *et al.*, 2010; Greve *et al.*, 2009; Greve *et al.*, 2009; Greve *et al.*, 2009). Schadman *et al.* (1998) confirma dichas tendencias, encontrando rendimiento subóptimo en simuladores, pero también en paciente clínicos que están en proceso de litigio. La principal diferencia observada es que los simuladores obtienen puntuaciones más extremas, similares a la obtenida por el grupo de daño cerebral grave, dato que corrobora el estudio de Etherton *et al.* (2006) con el Processing Speed Index (PSI) del WAIS-III.

De las herramientas utilizadas en los cinco estudios, las cuales se pueden observar en la Tabla 7, destacamos el California Verbal Learning Test (CVLT; Delis *et al.*, 1987), ya que resultó más preciso en términos generales, siendo el Linear Shrinkage Model más efectivo a la hora de detectar simulación en pacientes con dolor crónico, y el Reliable Digit Span Test (RDS; Greiffenstein *et al.*, 1994) por su alta precisión diagnóstica (1% de falsos positivos) con un punto de corte de 6.

- Los cuatro estudios restantes evalúan las diferencias en el **perfil psicológico/sintomatología psicológica**, entre simuladores y pacientes clínicos (McGuire *et al.*, 2001; McGuire y Shores, 2001; McGuire y Shores, 2001b; Wallis y Bogduk, 1996). Wallis y Bogduk (1996) encontraron que los pacientes clínicos

tenían un perfil muy concreto en el SCL-90 que los simuladores eran incapaces de recrear, puntuando de forma moderada en Obsesivo-compulsivo, Somatización y Depresión. De la misma forma, los dos estudios con el P3 arrojaron un perfil en pacientes clínicos con Depresión, Ansiedad y Somatización moderada. En los 4 estudios mencionados, los grupos de simuladores obtienen puntuaciones superiores a los pacientes clínicos.

- Linnell y Easton (2006) utilizaron además el IPQ y el Cuestionario de Salud General (GHQ-12; Goldberg, 1992) para buscar diferencias en la **actitud ante la situación de enfermedad** entre simuladores y pacientes clínicos agudos y crónicos. Los resultados mostraron que los simuladores reportaban menos control sobre la enfermedad y más consecuencias adversas sufridas que los pacientes agudos. Sin embargo, al comparar los simuladores con los pacientes crónicos, se observó que estos últimos reportaban más síntomas, más consecuencias severas y esperaban un mayor tiempo de recuperación, pero consideraban que tenían un mejor control sobre sus síntomas.

6.2.3. Dimensión social

De los estudios incluidos, cuatro evalúan las inconsistencias en la percepción del grado de discapacidad provocado por la condición padecida, así como su impacto en la calidad de vida (Bianchini *et al.*, 2014; Crighton *et al.*, 2014; Larrabee, 2003; Linnell y Easton, 2006). Larrabee (2003) comprobó que el PDI tenía una sensibilidad moderada a la exageración de síntomas físicos con especificidad del 0.90, pero que disminuía drásticamente (por debajo de 0.40) con niveles más altos de especificidad. Siguiendo los pasos de este estudio, Bianchini *et al.* (2014) encontraron que el PDI es capaz de diferenciar entre simuladores y pacientes clínicos con un alto grado de eficacia, pero

requiere que las puntuaciones obtenidas en la escala sean altas o bajas (mayor o igual que 62 indican riesgo de simulación y menor de 55 indican que hay riesgo mínimo de simulación). Cuando las puntuaciones se encuentran en un rango intermedio (entre 55 y 61) se considera que la exageración reflejada puede estar provocada por la influencia de diversas fuentes y no son lo suficientemente específicas como para ser un indicador fiable de exageración. Crighton *et al.* (2014) evaluaron el PDI en el contexto clínico y forense, llegando a la conclusión de que, si bien su especificidad no es excelente, es una escala que puede ser utilizada como un método complementario de detección de simuladores, tanto en el contexto clínico como en el forense.

6.2.4. Elementos de apoyo a la evaluación

Tres artículos evalúan la utilidad de diferentes herramientas como elemento de apoyo de la evaluación, cuya medida de resultado hemos llamado “riesgo de simulación” (Agueverre *et al.*, 2008; Bianchini *et al.*, 2008; Tearnan y An Ross, 2012). Dos tienen como objetivo validar el Inventario Multifásico de la Personalidad de Minnesota en su segunda versión (MMPI-2; Butcher *et al.*, 1989) (Agueverre *et al.*, 2008; Bianchini *et al.*, 2008). Bianchini *et al.* (2008) observaron que los grupos de simuladores obtenían puntuaciones más altas en todas las subescalas, asociando puntuaciones extremas a mayor sensibilidad en todas las variables menos en la escala L. Advierten, sin embargo, que las escalas clínicas (Hs y Hy), si bien resultan sensibles, no son útiles hasta que las puntuaciones son muy altas ($T > 90$). Posteriormente, Agueverre *et al.* (2008) utilizaron una versión reducida del Índice de Meyers para el MMPI-2 (Meyers Index; Meyers *et al.*, 2002) para diferenciar simuladores de pacientes con dolor crónico y daño cerebral traumático, obteniendo una especificidad del 95% con una puntuación igual o superior a 4. Por otro lado, un punto de corte de ≥ 3 se asoció con una excelente precisión diagnóstica. Otro estudio, realizado por Tearnan y An Ross (2012) utiliza el Life

Assessment Questionnaire (LAQ) con valores de predicción positivos (entre 0.79 y 0.88), fuerte validez convergente y alta sensibilidad y especificidad (valores de 0.44 a 0.57 y 0.88 a 0.93 respectivamente). Se observa, sin embargo, que el poder predictivo de la escala de síntomas psicológicos era más débil que la de síntomas físicos.

Tabla 7. Características de los estudios incluidos

Autor	Objetivo	Muestra	Escala	Diseño del estudio	Dimensión de medición	Medidas del resultado	Indicadores	Resultados significativos
Hopwood <i>et al.</i> , 2010	Evaluar la capacidad de la escala PAI para discriminar entre simuladores y pacientes clínicos.	n = 469 // PC = 317; SIM = 152	PAI	Simulación	Biológica	Manifestación de síntomas somáticos	(+) Número y (+) Gravedad de los síntomas	-Los resultados indican que puede discriminar entre grupos, pero los autores no consideran que el PAI sea lo suficientemente preciso / sensible para usarse en un entorno clínico.
Linnell y Easton, 2006	Uso del IPQ para explorar las expectativas y creencias sobre el SLC de pacientes clínicos y simuladores	n = 149 // PC = 45; SIM = 104	IPQ / PDI	Simulación	Psicológica y social	Actitud ante la situación de enfermedad // Grado de discapacidad, percibido por el paciente	(-) Control, (+) gravedad, (-) identidad de síntomas físicos y psicológicos. (+) Gravedad del dolor	-La variedad en la cantidad y severidad de los síntomas que sufren los pacientes con SLC es más amplia de lo que un simulador podría imaginar. Los pacientes crónicos desarrollan ideas más extremas que los pacientes agudos y los simuladores, pero creen que tienen un mejor control sobre los síntomas de la enfermedad y no informan niveles tan altos de dolor.
Wallis y Bogduk, 1996	Evaluar la capacidad del SCL-90 para discriminar entre simuladores y pacientes clínicos.	n = 173 // PC = 133a; SIM = 40	SCL-90, MPQ y EVA	Simulación	Biológica y psicológica	Perfil psicopatológico/ y sintomatología psicológica // Grado de dolor percibido	(-) Identidad, (+) gravedad, (+) número de síntomas físicos y psicológicos. (+) Gravedad del dolor.	-El SCL-90 mostró un perfil muy específico en pacientes con SLC, con puntuaciones moderadas en trastorno obsesivo compulsivo, somatización y depresión. El grupo de análogos obtuvo puntuaciones más altas

Tearnan y An Ross, 2012	Validar el LAQ. Un cuestionario elaborado para evaluar la presencia de simulación en adultos que padecen dolor.	n = 400 // Estudio 1 - PC = 84, Estudio 2 - PC = 174; SIM = 59, Estudio 3 - PC = 83	LAQ	Simulación	Apoyo	Riesgo de simulación	(+) Número, (+) Gravedad, (+) Inconsistencias en la expresión de síntomas físicos y psicológicos	distribuidas aleatoriamente en todas las subescalas. -Valores de predicción positivos (entre 0,79 y 0,88), validez convergente, sensibilidad y especificidad adecuadas (valores de 0,44 a 0,57 y 0,88 a 0,93, respectivamente). LAQ es una herramienta válida y efectiva en la clasificación de pacientes que simulan dolor,
Fishbain <i>et al.</i> , 2002	Evaluar la capacidad de la escala IBQ para detectar simulación.	n = 96 // PC = 36	IBQ	Prevalencia diferencial	Psicológica y social	Actitud ante la situación de enfermedad	No sig.	-La hipótesis principal no se cumple. IBQ no es una herramienta válida para discriminar entre pacientes clínicos y simuladores
Bianchini <i>et al.</i> , 2014	Evaluar la capacidad del PDI y el MSPQ para detectar simulación	n = 426 // PC = 328; SIM = 98	PDI y MSPQ	Grupo conocido	Biológico y social	Grado de discapacidad, percibido por el paciente // Manifestación de síntomas somáticos	(+) Percepción de alteración social, (+) Discapacidad, (-) Autonomía, (+) Gravedad de síntomas físicos	-Puntuaciones extremas en ambas escalas indican exageración intencional. Ambos pueden discriminar entre simuladores y pacientes clínicos, pero el MSPQ es generalmente más efectivo.
Oddsottir <i>et al.</i> , 2015	Valorar si el "fly test" puede distinguir entre simuladores y pacientes clínicos	n = 65 // PC = 34; SIM = 31	Fly test	Simulación	Biológico	Precisión de movimientos del cuello	(+) Gravedad de los síntomas. (-) Precisión de movimientos del cuello	-Sensibilidad correcta, los simuladores fingen deliberadamente dificultades en el movimiento del cuello e informan una precisión mucho menor, inconsistente con lo que se espera en la condición.

Schmand <i>et al.</i> , 1998	Estudio de la simulación de síntomas cognitivos (memoria y concentración) en pacientes con SLC	n = 174 // PC = 65; SIM = 43; Ctrl = 66	Batería compuesta por: ASTM como prueba principal; DART, SDMT, TMT, Stroop Test, AVLT y RBMT	Prevalencia diferencial	Psicológico	Riesgo de simulación (déficit de memoria y concentración)	(+) Gravedad de los síntomas neuropsicológicos (+) Déficits en memoria y (-) Capacidad de concentración.	-La muestra de pacientes se dividió en dos contextos, clínico y legal. En el contexto clínico, 1 de cada 3 o 4 pacientes obtuvo puntajes por debajo de sus capacidades, y en el contexto legal, 2 de cada 3 o 4. Pacientes simuladores y pacientes con SLC obtuvieron puntuaciones que indican un rendimiento subóptimo, pero las de los simuladores fueron más extremas.
Crighton <i>et al.</i> , 2014	Evaluar la capacidad del PDI y el MSPQ para detectar simulación.	n = 287b // PC = 154; SIM = 133	PDI y MSPQ	Grupo conocido	Biologico y social	Grado de discapacidad, percibido por el paciente // Manifestación de síntomas somáticos	(+) Percepción de alteración social, (+) Discapacidad, (+) N y (+) Gravedad de los síntomas	-Los resultados sugieren que ambas son útiles para detectar simulación en el contexto forense. los simuladores tienden a mostrar más síntomas de lo esperado y a exagerar su intensidad. MSPQ obtuvo tamaños de efecto más grandes.
Larrabee, 2003	Evaluar la capacidad de las escalas MPQ, PDI y MSPQ para discriminar entre simuladores y pacientes clínicos	n = 977 // PC = 948 (MPQ PC = 295; PDI PC = 401; MSPQ PC = 252); SIM = 29.	MPQ, PDI, MSPQ	Grupo conocido	Biologico y social	Grado de dolor percibido // Grado de discapacidad, percibido por el paciente // Manifestación de síntomas somáticos	(+) Gravedad del dolor, (+) Percepción de la alteración social, (+) Discapacidad, (+) Número y (+) Gravedad de los síntomas	-Los simuladores obtuvieron puntuaciones significativamente más altas que los pacientes clínicos en las tres escalas. La sensibilidad de las tres herramientas difiere. MPQ tiene poca sensibilidad, PDI moderada y MSPQ alta.

Wygant <i>et al.</i> , 2017	Utilidad de los signos no orgánicos de Sobell para detectar simulación.	n = 216b // PC = 169; SIM = 47	Signos no orgánicos de Sobell	Grupo conocido	Biológico	Inconsistencias durante la evaluación física del paciente (screening para “ <i>abnormal illness behaviour</i> ”)	(+) Inconsistencias en la expresión, (-) Identidad y (+) Gravedad de los síntomas, (+) Sensibilidad	-Presenta una serie de elementos que podrían resultar útiles para la evaluación del riesgo de simulación, como sensación de debilidad en áreas que no deberían verse afectadas. Las puntuaciones en los signos de 2 y 3 se asocian con una mayor probabilidad de que el paciente exagerara la cantidad de síntomas sufridos a niveles observados en simuladores.
Etherton <i>et al.</i> , 2006	Evaluar la capacidad de la subescala WMI del WAIS-III para detectar simulación	n = 220 - Estudio 1 - n = 64 (SIM) - Estudio 2 - n = 156 // PC = 49; SIM = 32; Ctrl = 75	WAIS-III (Índice de memoria de trabajo)	S1 Simulación y S2 Grupo conocido	Psicológico	Rendimiento cognitivo (memoria y concentración a corto plazo)	Non sig.	-No se encontraron diferencias significativas entre los grupos de pacientes clínicos y simuladores. No hay diferencias por edad o sexo.
Etherton <i>et al.</i> , 2006	Evaluar la capacidad de la subescala PSI del WAIS-III para detectar simulación	n = 234 - Estudio 1 - n = 82 (SIM) - Estudio 2 - n = 152 PC = 48; SIM = 32; Ctrl = 72	WAIS-III (Índice de velocidad de procesamiento)	S1 Simulación y S2 Grupo conocido	Psicológico	Rendimiento cognitivo (velocidad de procesamiento de la información)	(+) Gravedad de los síntomas (-) Velocidad de procesamiento	-Los resultados de los estudios indicaron que algunos pacientes mostraron alteraciones en la velocidad del procesamiento cognitivo, a menudo a niveles que indican daño cerebral traumático. Se considera que puntuaciones de 70 o menos no pueden explicarse por problemas relacionados con el dolor.

Greve <i>et al.</i> , 2010	Evaluar la capacidad del RDS para discriminar entre simuladores y pacientes clínicos.	n = 641b // PC = 612; SIM = 29	RDS	Grupo conocido	Psicológico	Grado de deterioro cognitivo	(+) Gravedad de los síntomas neuropsicológicos . (-) Actuación	-Alta precisión para diferenciar a los simuladores de pacientes clínicos con un punto de corte de 6.
Greve <i>et al.</i> , 2009	Evaluar la capacidad de la escala TOMM para detectar simulación	n = 795 // PC = 604; SIM = 20 (67d); Ctrl = 104	TOMM	Grupo conocido	Psicológico	Grado de deterioro cognitivo	(+) Gravedad de los síntomas neuropsicológicos . (+) Déficits en la memoria	-Moderada validez y fiabilidad. Los datos indican que una puntuación más alta de lo normal puede interpretarse como simulación de deterioro cognitivo. Sin embargo, se recomienda una aplicación conjunta con otra técnica de evaluación porque la capacidad de discriminación del TOMM es limitada.
McGuire y Shores, 2001	Evaluar la capacidad de la escala P3 para discriminar entre simuladores y pacientes clínicos.	n = 60 // PC = 40; SIM = 20	P3	Grupo conocido	Biológico y físico	Grado de alteración somática y emocional (depresión y ansiedad)	(+) Severidad, (+) Número de síntomas físicos y psicológicos. (+) Gravedad del dolor; (+) Sintomatología depresiva.	-El índice de validez no ayudó a encontrar diferencias entre los dos grupos. El resto de las subescalas, especialmente depresión, fueron sensibles. Se encontraron puntuaciones altas en el grupo de simuladores.
McGuire <i>et al.</i> , 2001	Evaluar la capacidad de la escala P3 para discriminar entre simuladores y pacientes clínicos.	n = 122 // PC = 62; SIM = 60	P3	Grupo conocido	Biológico y físico	Grado de alteración somática y emocional (depresión y ansiedad)	(+) Severidad, (+) Número de síntomas físicos y psicológicos. (+) Gravedad del dolor; (+) Sintomatología depresiva.	-Los simuladores se dividieron en grupos de dolor y rehabilitación. Los dos grupos de simuladores obtuvieron puntuaciones más altas en las subescalas.

McGuire <i>et al.</i> , 2001b	Evaluar la capacidad del SCL-90 para detectar simulación	n = 70 // PC = 50; SIM = 20	SCL-90	Grupo conocido	Biológico y físico	Número de síntomas, grado de alteración psicopatológica y somática	(-) Identidad, (+) Severidad, (+) Número de síntomas psicológicos	-Simuladores obtienen puntuaciones más altas que los pacientes clínicos en prácticamente todas las subescalas.
Greve <i>et al.</i> , 2009	Evaluar la capacidad de la Prueba Verbal de California para detectar simulación	n = 820; PC = 314; SIM = 194; In = 312	CVLT-I y CVLT-II	Grupo conocido	Psicológico	Rendimiento cognitivo	(+) Gravedad de los síntomas neuropsicológicos . (-) Rendimiento cognitivo	-La precisión de las dos herramientas es similar, pero los puntos de corte no son intercambiables. El uso del punto de corte del CVLT-I en el CVLT-II conduce a una alta tasa de falsos positivos.
Agueverre <i>et al.</i> , 2009	Desarrollar un método para detectar simulación utilizando 7 subescalas del MMPI-2 (Escala de validez de Meyers)	n = 300 // PC = 57; SIM = 58; Ctrl = 185	MMPI-2 (MVSCI)	Grupo conocido	Apoyo	Perfil psicopatológico/ y sintomatología psicológica	(-) Identidad, (+) Severidad, (+) Número de síntomas psicológicos	-Se utilizó una versión abreviada del método propuesto por Meyers <i>et al.</i> (2002), que incluye un índice con cinco escalas de validez. Se obtuvo una alta precisión, comparable con la versión original. Especificidad del 95% y sensibilidad en el rango del 50%.
Bianchini <i>et al.</i> , 2008	Valorar la sensibilidad a MMPI-2 para discriminar entre simuladores y pacientes clínicos	n = 115 // PC = 57; SIM = 58	MMPI-2	Grupo conocido	Apoyo	Perfil psicopatológico/ y sintomatología psicológica	(-) Identidad, (+) Severidad, (+) Número de síntomas psicológicos	-La escala puede diferenciar la exageración intencional, las alteraciones psicológicas relacionadas con el dolor y el estrés asociado con el contexto médico-legal. Para todas las escalas, excepto L, las puntuaciones extremas se asociaron con una mayor especificidad.

Greve <i>et al.</i> , 2009	Evaluar la capacidad de la escala PDRT para discriminar entre simuladores y pacientes clínicos.	n = 468 // PC = 318; SIM = 30; Ctrl = 120	PDRT	Grupo conocido	Psicológico	Grado de deterioro cognitivo	(+) Gravedad de los síntomas neuropsicológicos . (+) Déficits en la memoria	-Moderada validez y fiabilidad. Los datos indican que una puntuación más alta de lo normal puede interpretarse como simulación de deterioro cognitivo. Sin embargo, se recomienda una aplicación conjunta con otra técnica de evaluación porque la capacidad de discriminación del PDRT es limitada.
Sobel <i>et al.</i> , 2000	Desarrolle una herramienta para detectar simulación de daño cervical.	N = 26 (Mal)	Signos no orgánicos de Sobel adaptados	Simulación	Biológico	Inconsistencias durante la evaluación física del paciente (screening para “ <i>abnormal illness behaviour</i> ”)	(+) Inconsistencias en la expresión, (-) Identidad y (+) Gravedad de los síntomas, (+) Sensibilidad	-Correlación positiva entre los signos y un sesgo en el estilo de respuesta, destacando el signo de "pérdida sensorial" por sus buenos resultados, lo que indica que los pacientes con un estilo de respuesta atípico tienden a informar una disminución de la sensibilidad en un patrón que no se corresponde con ningún trastorno nervioso.

a. Tabla del artículo N=126; b. Muestra final; c. Análogos - Cold-pressor task; d. Etherton et. al (2005); e. Los indicadores físicos y psicológicos se consideran de forma independiente.

Abreviaturas, PC: Paciente clínico; Sim: Simuladores; Ctrl: Paciente control; In: indiferenciado.

6.3. Discusión

A través de la presente revisión sistemática se han identificado 13 medidas de resultado: manifestación de síntomas somáticos, percepción del impacto en la calidad de vida, perfil psicopatológico y sintomatología psicológica, riesgo de simulación, actitud ante la situación de enfermedad, precisión de movimientos del cuello, percepción de grado discapacidad, grado de dolor percibido, inconsistencias durante la evaluación física del paciente, rendimiento cognitivo, grado de deterioro cognitivo, grado de alteración emocional y número de síntomas. Se han localizado medidas y herramientas en cada una de las dimensiones del modelo biopsicosocial, lo que pone en manifiesto el uso enfoque en la valoración de la simulación.

En la gran mayoría de estos estudios se observa que la valoración de la simulación no se realiza mediante una herramienta diseñada específicamente con este fin. En ese sentido, los estudios que utilizaron varias herramientas para valorar la simulación indicaron una mayor eficacia discriminativa cuando eran utilizadas en conjunto, lo que apoya la utilidad del modelo de convergencia de datos (Faust *et al.*, 2012).

Los estudios incluidos en la **dimensión biológica** indican que los simuladores tienden a magnificar significativamente la gravedad de sus síntomas, especialmente el dolor cervical sufrido, independientemente del tiempo que lleven padeciendo la condición. La simulación se observa a través de puntuaciones más elevadas en las herramientas aplicadas de lo que se podría esperar en pacientes que realmente sufran dolor cervical, más elevadas incluso que las de aquellos cuyos síntomas se ven magnificados por variables de corte psicosocial. A pesar de dichas puntuaciones pueden deberse a diversos factores, los simuladores tienden a exagerar a un nivel que no ha sido visto en otras condiciones que no sean las de engaño (Bianchini *et al.*, 2014). Se observa también

que, a diferencia de los pacientes clínicos, los simuladores suelen informar de un mayor número de síntomas, coincidiendo con la conocida tendencia al *over-endorsement* (Rogers, 1997). Dicha tendencia se manifiesta mediante los resultados en los instrumentos psicométricos aplicados (SCL-90R, MMPI-2, MSPQ, etc.), puntuando los simuladores en más subescalas que los clínicos, que suelen presentar un perfil más homogéneo. Como consecuencia de este estilo de respuesta, los primeros generan más inconsistencias en el patrón de síntomas presentado.

Finalmente, encontramos significativos varios de los hallazgos obtenidos por Linnell y Easton (2006). Su investigación evidenció que los simuladores manifiestan una actitud ante la situación de enfermedad más negativa que los pacientes clínicos. Además, se observó que los simuladores subestiman el rango de síntomas padecidos. Esta última afirmación choca con los resultados señalados anteriormente y la conocida tendencia al *symptom overendorsement*. Creemos que la explicación de este resultado puede ser que, habitualmente, el *overendorsement* se valora mediante las puntuaciones de la psicometría y/o los síntomas referidos. Los autores han utilizado tres herramientas que arrojan resultados sobre creencias las de la enfermedad (IPQ), discapacidad (NDI) y percepción de salud (GHQ - 12), pero no han utilizado ninguna escala que registre síntomas o síndromes, papel que cumplía el SCL-90 o el P3 en los otros estudios incluidos. De la misma forma, no se pidió a los participantes que informaran sobre síntomas padecidos, de manera que no se obtuvo información referida directamente por los simuladores. Es posible que para que los simuladores informen de síntomas de forma indiscriminada sea necesario facilitarles un medio para ello. Sería recomendable estudiar en qué condiciones los simuladores expresan padecer más síntomas que los pacientes clínicos.

Los estudios analizados en la **dimensión psicológica** arrojan resultados similares a la biológica en lo referido a la magnificación de síntomas. A pesar de que es posible

que los pacientes en proceso de litigio, y los pacientes agudos ofrezcan puntuaciones elevadas, los simuladores alcanzan resultados incoherentes con la condición sufrida o los síntomas reportados (Etherton, *et al.*, 2006). En el plano neuropsicológico se observó que los simuladores tienden a mostrar rendimiento subóptimo a niveles que sólo serían coherentes en pacientes con una lesión cerebral grave, mientras que los pacientes clínicos, si bien mostraban cierta alteración, no alcanzaba un nivel tan elevado (Greve *et al.*, 2010; Greve *et al.*, 2009; Greve *et al.*, 2009; Greve *et al.*, 2009). En el plano psicopatológico y emocional se observa un perfil muy específico en los pacientes clínicos que los simuladores no son capaces de recrear. Los estudios con el P3 y el SCL-90 revelan que los pacientes clínicos muestran sintomatología ansioso-depresiva y somatización a niveles moderados (McGuire *et al.*, 2001; McGuire y Shores, 2001; McGuire y Shores, 2001b; Wallis y Bogduk, 1996). Dicho perfil corresponde con la presentación psicológica de la patología ofrecida habitualmente en la bibliografía, que recoge síntomas como pensamientos catastrofistas relacionados con el dolor, miedo al movimiento, y sintomatología ansioso-depresiva (Philips *et al.*, 2010; Sterling, 2011).

Debido a que el origen más común de la condición, el accidente de circulación, trae consigo una alta carga emocional, creemos que los resultados resultan coherentes a nivel clínico. La experiencia traumática vivida deriva en una importante fuente de estrés que, por la intensidad del dolor y su impacto en la calidad de vida, se mantiene a lo largo del proceso de recuperación. Pacientes con baja resiliencia o mala salud física/psicológica previa al accidente, corren el riesgo a sufrir una mala adaptación a la enfermedad y desarrollar actitudes y pensamientos negativos durante el curso de la condición que magnifiquen y mantengan dicho estrés. Ante esta situación, no es extraño que aflore sintomatología ansioso-depresiva, que configure un cuadro adaptativo (American

Psychiatric Association, 2013) o de estrés postraumático, el cual tiene una incidencia estimada del 25% en pacientes de SLC (Jaspers, 1998).

Consideramos que la dimensión psicológica puede ser especialmente útil para valorar la simulación en la condición de interés y destacamos las variables que componen el concepto “conducta ante la situación de enfermedad”, creemos que analizar cómo interactúan las diferentes dimensiones puede ser útil para discriminar entre pacientes clínicos y simuladores. La limitada evidencia encontrada indica que los simuladores ofrecen una actitud más negativa y resulta coherente con la presentación que se hace habitualmente en la bibliografía de los deshonestos (Vrij, 2008). Podrían creer que aparentar mejoría reduciría su compensación económica, por lo que siempre intentarían dar una imagen muy negativa de su situación de enfermedad. Sería interesante evaluar las diferencias en las conductas ante la situación de enfermedad entre pacientes clínicos y simuladores de cara a futuras investigaciones.

La dimensión social mantiene la utilidad de la magnificación como indicador de simulación, sin embargo, la cantidad de estudios encontrados para esta dimensión ha sido más limitada. En este caso, los estudios realizados con el PDI indicaron que los simuladores tienden a exagerar la alteración que la patología tiene en su calidad de vida, reportando una autonomía reducida y una percepción de discapacidad muy elevada, frente a la percepción de discapacidad moderada de los pacientes clínicos (Bianchini *et al.*, 2014; Crighton *et al.*, 2014; Larrabee, 2003). Encontramos llamativa la poca cantidad de estudios encontrados que validen herramientas que miden el impacto de la patología en la dimensión social. Consideramos que escalas como el Cuestionario de salud SF-36 (SF-36; Ware, 2005) pueden ser de utilidad en este sentido.

Por otro lado, los estudios incluidos en el último apartado, llamado “elementos de apoyo a la evaluación”, indican que el MMPI-2 y el LAQ son capaces de discriminar entre pacientes clínicos y simuladores (Agueverre *et al.*, 2008; Bianchini *et al.*, 2008; Tearnan y An Ross, 2012). Encontramos especialmente útil el MMPI-2, que tiene la ventaja adicional de que ofrece un completo perfil psicopatológico que cubriría la dimensión psicológica, además de habilitar la posibilidad de aplicar índices preparados para la detectar la simulación mediante una selección concreta de ítems, como el Índice de Meyers (Meyers Index; Meyers *et al.*, 2002) o la Fake-Bad Scale (Lees-Haley *et al.*, 1991).

Creemos que, de cara a la presente tesis doctoral, las medidas del resultado descritas en los párrafos anteriores pueden ser de utilidad para la formación del protocolo. Debido a que buscamos construir una metodología basada en el modelo multidimensional y de convergencia de datos, nuestro objetivo es disponer de una amplia variedad de fuentes de información que ayuden al/la profesional a tomar una decisión fundada. Así, los resultados de este estudio pueden traducirse en cinco fuentes de información diferentes: *variables de las dimensiones biológicas, psicológicas y sociales, existencia de inconsistencias y existencia de exageración u overreporting*⁴ (Tabla 8).

⁴ Este término hace referencia a la tendencia a expresar muchos síntomas y exagerar su gravedad. Debido a que en español no existe un término similar, se ha decidido conservar la expresión inglesa.

Tabla 8. *Protocolo y fuentes de información – primera versión*

Fuente de información	Medida de resultado	Instrumento
Protocolo biopsicosocial	Severidad/grado de dolor	Inventario breve del dolor (BPI) y/o Cuestionario del dolor de McGill.
	Inconsistencias síntomas físicos y somáticos	Valoración clínica y SCL-90.
	Perfil psicopatológico y sintomatología psicológica	Inventario Multifásico de Personalidad de Minnesota (MMPI-2), Inventarios de depresión y ansiedad de Beck (BDI y BAI).
	Actitud ante la situación de enfermedad	Illness Perception Questionnaire (IPQ).
Social	Percepción de discapacidad	Índice de discapacidad cervical (NDI).
	Calidad de vida relacionada con la salud	Cuestionario de salud SF-36.
Inconsistencias	Incoherencias entre la presentación de la condición del paciente y lo esperado en la misma.	Checklist de inconsistencias
Exageración u overreporting	Presencia de exageración de síntomas	Inventario estructurado de simulación de síntomas (SIMS), Inventario Multifásico de Personalidad de Minnesota (MMPI-2)

Las tres primeras fuentes de información las agruparemos bajo el término “protocolo biopsicosocial”, el cual estará formado por:

Dimensión biológica. Se incluirá la *severidad/grado del dolor*, la *inconsistencia en la presentación de síntomas físicos y el grado de somatización*. Ninguno de los instrumentos identificados para valorar la severidad/grado de dolor ha sido validado con población española, salvo la *Escala Visual Analógica (EVA)*, así que se planteará la inclusión del *Inventario Breve del Dolor* abreviado en su versión española (BPI-SF, de Andrés Arés *et al.*, 2015) y/o el *Cuestionario del Dolor* de McGill (Lázaro *et al.*, 1994).

Por otro lado, la inconsistencia en la presentación de síntomas físicos y somáticos se realizará mediante la entrevista clínica y el SCL-90.

Dimensión psicológica. Nuestros resultados indican que *el rendimiento cognitivo, el perfil psicopatológico / sintomatología psicológica y la actitud ante la situación de enfermedad* son medidas de resultado a considerar para el protocolo. En ese sentido, a pesar de que los estudios sobre el rendimiento cognitivo ofrecen resultados prometedores, el tiempo de aplicación y coste económico que requieren instrumentos como el WAIS-IV hace inviable su inclusión en el citado protocolo, por lo que se utilizará el perfil psicopatológico / sintomatología psicológica, que será evaluado mediante el MMPI-2, el BDI y el BAI, y la actitud ante la situación de enfermedad, que se valorará mediante el IPQ.

Dimensión social. Se utilizarán las medidas del resultado *percepción de discapacidad e impacto en la calidad de vida*. El instrumento valorado en la presente revisión sistemática, el PDI, no está validado para población española, por lo que se utilizará el *Índice de Discapacidad Cervical* en su versión española (NDI; Andrade Ortega *et al.* 2010), el cual persigue una finalidad similar (percepción de discapacidad producida por el daño cervical, en lugar de por el dolor), presenta propiedades psicométricas adecuadas (Alpha de Cronbach =.93) y se recomienda su uso en SLC (Sterling, 2014), y el cuestionario de salud SF-36 (adaptado al castellano por Alonso *et al.*,1995), que evalúa calidad de vida relacionada con la salud (HRQoL) y es uno de los cuestionarios más utilizados en el contexto (Vilagut *et al.* 2005).

Para la valoración de posibles **inconsistencias en la valoración clínica** se tendrá en consideración la publicación de Akehurst *et al.* (2015), la cual no fue incluida en la presente revisión por no tener un grupo de pacientes clínicos de referencia, pero que

plantea una interesante herramienta para emitir juicios de credibilidad en el contexto médico-legal. Creemos que un instrumento como el propuesto por los anteriores autores, de rápida aplicación y que puede integrarse con facilidad en la valoración clínica, puede ser de gran utilidad al ofrecer un medio para detectar y cuantificar las inconsistencias que puedan observarse en el transcurso de la valoración.

Por último, para analizar la **existencia de exageración u overreporting** se contemplará el MMPI-2 y las subescalas propuestas en los estudios detectados para detectar dicho riesgo. De la misma forma, aunque no se ha incluido en la presente revisión ningún estudio que valore su validez en el contexto, el SIMS (González y Santamaría, 2009) es una herramienta diseñada específicamente para medir riesgo de simulación y ha demostrado tener excelentes propiedades psicométricas en múltiples estudios (van Impelen *et al.*, 2014).

Los instrumentos que no se han mencionado en los párrafos anteriores, como el P3, LAQ, o MSPQ, no se consideran debido a que no han sido adaptados a población española. Por otro lado, el *Flytest* tampoco se considera debido a que, a pesar de resultar prometedor, sería muy difícil disponer de las herramientas necesarias para su ejecución y su uso en el contexto médico-legal sería realmente complejo.

7. Bloque 2 – Estudio 3. Perfiles biopsicosociales de percepción del dolor de pacientes con SLC.

Los estudios 3 y 4 comprenden la aplicación del protocolo para su valoración. Si bien ambos estudios siguen la misma metodología y utilizan la misma muestra, sus objetivos y estrategia de análisis de datos son diferentes y se consideró que plasmarlos como un único estudio resultaría confuso y poco operativo.

El presente Estudio 3 se plantea para dar respuesta a una necesidad observada durante la elaboración de los dos anteriores estudios, alcanzar una mayor comprensión del perfil de paciente con SLC y aportar evidencia que apoye el uso del modelo biopsicosocial, ya no solo para la detección de la simulación, sino también para la valoración clínica general. Esta perspectiva ofrecería una visión más completa de la condición, permitiendo al/la profesional conocer y entender la interacción entre los factores médicos, biomecánicos, sociales y psicológicos (Turk *et al.*, 2018).

Para dar respuesta a dicha necesidad, en el estudio 3 se plantean los dos siguientes objetivos:

- **Objetivo específico 1.** Obtener de forma empírica subgrupos basados en características psicosociales utilizando un análisis de cluster k-medias con una muestra de pacientes litigantes con latigazo cervical.
- **Objetivo específico 2.** Investigar diferencias en los subgrupos obtenidos en variables de interés que no han sido utilizadas para formar los clústeres (severidad del dolor, percepción de discapacidad, variables sociodemográficas y variables médico-legales).

7.1. Metodología

7.1.1. Participantes

La muestra inicial para el presente estudio estaba compuesta por 247 pacientes diagnosticados con SLC tras sufrir un accidente de circulación. Los criterios de inclusión utilizados fueron: Criterio a. Necesidad de firmar el consentimiento informado, Criterio b2. La persona estaba siendo evaluada para valorar el daño físico y psicológico producido tras un accidente de circulación, Criterio c2. Era mayor de edad (≥ 18 años), Criterio d2. Se catalogado como paciente clínico. El incumplimiento de los criterios en cada grupo suponía la exclusión del estudio. Tras la aplicación de los citados criterios, se descartaron 18 participantes por no cumplir el criterio d2 (16 fueron catalogados como probables simuladores y 2 como dudosos). Así, la muestra final se compuso por 229 pacientes, con una media de edad de 37.21 ($DT=10.65$) y un rango de 18 a 60. De dicho total, 119 eran mujeres (52%) y 110 hombres (48%).

7.1.2. Procedimiento

Construcción del protocolo

La primera versión del protocolo aplicado en el presente estudio se formó con las herramientas descritas en la Tabla 8 del Estudio 2, así como una portada donde se recogían las principales variables sociodemográficas y médico-legales de interés, incluido el nexo causal. El protocolo sin el checklist de inconsistencias se aplicó primero con 5 alumnos voluntarios de la Universidad de Murcia, con la finalidad de conocer el tiempo estimado de administración y posibles dificultades. El tiempo medio de aplicación fue de 187 minutos ($DT=52.21$), y las principales quejas emitidas por los alumnos fueron el factor cansancio (100%) y el excesivo tiempo requerido para cumplimentar todas las escalas (100%).

Para la segunda versión del protocolo, se decidió suprimir el MMPI-2, debido a su largo tiempo de cumplimentación y complejidad, y el Cuestionario de dolor de McGill, el cual resultaba redundante con el BPI. Dicha segunda versión volvió a ser sometida a valoración con otros 5 alumnos voluntarios, diferentes a los de la primera versión, obteniéndose un tiempo de aplicación de 53 minutos ($DT=18.87$) y dos quejas (40%) sobre dicho tiempo. A continuación, la propuesta de protocolo fue presentada a un catedrático de medicina legal y forense con amplia experiencia en la temática, el cual colaboró en todo el desarrollo de la presente tesis doctoral y actuó como asesor en aspectos médicos, especialmente aquellos relacionados con la valoración del daño corporal y el SLC. Dicho catedrático consideró que el tiempo de aplicación del protocolo biopsicosocial excedía significativamente el invertido en el proceso de valoración del daño corporal en España (entre 20 y 30 minutos), lo que dificultaría la obtención de muestra y haría prácticamente inviable su aplicación.

Por ello, para la tercera y última versión del citado protocolo biopsicosocial se decidió suprimir el SCL-90, ya que suponía el 79% del tiempo, y se sustituyó el IPQ por su versión abreviada (Brief IPQ; Weinman *et al.*, 1996). La prueba realizada a la tercera versión del protocolo, con las mismas condiciones que las dos anteriores, indicó que el tiempo medio de aplicación era de 22 minutos ($DT=4.24$) y no registró ninguna queja. Dicha tercera versión se muestra en la Tabla 9.

El checklist de indicadores fue analizado aparte debido a que se diseñó para ser integrado en una valoración clínica. Se les facilitó a los/as profesionales que participaron en los estudios y éstos lo evaluaron durante tres días con 5 pacientes cada uno (15 en total). No se formuló ninguna queja ni se consideró que fuera necesaria alguna modificación. No se pudo calcular el tiempo estimado de aplicación, ya que lo rellenaron

durante la valoración clínica, pero los tres afirmaron que era sencillo, rápido de cumplimentar y no entorpecía el citado proceso de valoración.

Tabla 9. *Componentes del protocolo definitivo*

Fuente de información	Medida de resultado	Instrumento	
Protocolo biopsicosocial	Biológica	Severidad/grado de dolor	Inventario breve del dolor (BPI).
		Inconsistencias síntomas físicos	Valoración clínica.
	Psicológica	Perfil psicopatológico y sintomatología psicológica	Inventarios de depresión y ansiedad de Beck (BDI-II y BAI).
		Actitud ante la situación de enfermedad	Brief Illness Perception Questionnaire (BIPQ).
	Social	Percepción de discapacidad	Índice de discapacidad cervical (NDI).
	Calidad de vida relacionada con la salud	Cuestionario de salud SF-36.	
Inconsistencias	Incoherencias entre la presentación de la condición del paciente y lo esperado en la misma.	Checklist de inconsistencias	
Exageración u overreporting	Presencia de exageración de síntomas	Inventario estructurado de simulación de síntomas (SIMS).	

Así, una vez ajustados los componentes del Procolo se organizó de la siguiente forma (Anexo 2):

1- Página de datos. La primera página recogía el primer nombre del paciente y el código de codificación, la fecha de la valoración, la fecha del accidente, el sexo, la edad, el nivel socioeconómico y socioeducativo, la zona de residencia, el estado laboral, la existencia de información sobre el estado anterior, del nexo causal y de un posible incentivo externo material-legal, si había recibido rehabilitación, los síntomas referidos, un resumen de la valoración clínica del experto y variables relacionadas con el accidente (momento de la

colisión, asiento ocupado, localización del impacto, tipo de carretera, cinturón, posición de la cabeza y estado del coche).

2 – Consentimiento informado. Completamente anónimo, donde se detallaba el objetivo que se perseguía en la investigación y los instrumentos aplicados.

3- Checklist de indicadores de simulación.

4 – Protocolo biopsicosocial y SIMS. Se siguió el mismo orden de administración que figura en la Tabla 9.

5 – Última página. Reservada para incluir observaciones y otra información que pueda resultar de interés.

Diseño y aplicación del protocolo

Se utilizó un diseño transversal analítico-descriptivo con participantes reclutados en un centro médico multidisciplinar especializado de la Región de Murcia (España), durante los años 2017, 2018 y 2019. Todos los pacientes acudían a la clínica para ser valorados tras sufrir un accidente de circulación.

Se contactó con el citado centro médico y se concertó una cita donde se expuso a la dirección los objetivos y el procedimiento a seguir para la aplicación del protocolo. Una vez recibido el visto bueno por parte de dirección, se concertó una segunda cita, esta vez con los tres médicos que aceptaron participar, donde se volvió a explicar en detalle el estudio planificado y se organizó un plan de trabajo. El protocolo ideado fue valorado positivamente por los/as profesionales y únicamente solicitaron que no se realizara una exploración sobre la motivación del posible incentivo material-legal, ya que consideraron que podía resultar problemático para los pacientes. Además, recomendaron que se suprimiera el estudio de la biomecánica, ya que se realizaba en muy pocas ocasiones. Se

llevo a cabo una aplicación preliminar con 5 pacientes no incluidos en el estudio. El tiempo de aplicación medio fue de 25 minutos ($DT=1.24$), lo que se consideró, tanto por parte de la gerencia del centro, excesivamente elevado, debido a que entorpecía la actividad rutinaria de la clínica (los pacientes esperaban fuera de la sala, donde el espacio era reducido y dificultaba el tránsito del personal) y los participantes se cansaban de esperar y se marchaban. Por este motivo, se decidió dividir la aplicación del protocolo en dos grupos, el grupo A, en el que se aplicaban todos los instrumentos a excepción del SIMS y el grupo B, donde únicamente se administraba el SIMS. Los citados pacientes eran asignados aleatoriamente a uno de los dos grupos. Para el presente estudio y para el Estudio 4, se utilizó la muestra del grupo A y para el Estudio 6 la del grupo B.

El procedimiento seguido para la recogida de muestra se dividió en dos fases. En la primera, desarrollada por el médico valorador, los pacientes eran sometidos a una valoración clínica completa que ellos, o sus respectivas aseguradoras, habían solicitado. Una vez aplicada dicha valoración, el médico invitaba al paciente a participar en el estudio, proporcionándole una explicación detallada de la finalidad y procedimiento del mismo, haciendo hincapié en su carácter anónimo y voluntario. Los pacientes que aceptaban participar firmaban el consentimiento y eran valorados por un psicólogo forense, el cual aplicaba la batería de escalas preparada. Toda la información médico-legal necesaria (nexo causal, detalles del impacto, inconsistencia con los síntomas) era solicitada posteriormente al/la médico, habitualmente aprovechando alguna pausa entre pacientes. Para evitar la inclusión de posibles simuladores en el grupo de pacientes clínicos (criterio de inclusión d2), se utilizó como método de clasificación la existencia de evidencia contradictoria en la evaluación médico-legal. Para la clasificación, los tres expertos en medicina legal que participaron en el estudio analizaron cada caso y evaluaron si la manifestación de su sintomatología del paciente era consistente/coherente con los

indicadores anatómico-estructurales observados o esperados en la condición (severidad del dolor, localización del dolor, rango de movimiento cervical, balance articular activo / pasivo, etc.). Una vez evaluados, los pacientes se catalogaron como: 1. Paciente clínico (sin evidencia de fingimiento), 2. Paciente sobredimensionado (con evidencia de fingimiento y sin incentivo externo), 3. Probable simulador (con evidencia de fingimiento y con incentivo externo) o 4. Dudoso. Los desacuerdos fueron resueltos por consenso.

Para el desarrollo de este estudio se siguieron las consideraciones éticas propuestas por la American Psychological Association (2002, 2010) y se contó con el informe favorable del Consejo de coordinación del Programa de doctorado en ciencias forenses de la Universidad de Murcia.

7.1.3. Instrumentos y variables

Para el presente estudio las medidas de resultado del protocolo (variables de ahora en adelante) se dividieron en dos grupos: variables de clustering, que son aquellas que se utilizan como *input* en el análisis *cluster* para formar los subgrupos, y variables de perfilado, con las que se comparan los subgrupos resultantes del citado análisis. A continuación, se exponen las variables, así como una descripción de los instrumentos que las miden:

Variables de clustering

- Calidad de vida relacionada con la salud (HRQoL), medida mediante el 36-Item *Short Form Survey* (SF-36). El SF-36 (Ware, 2000). Los ítems del instrumento proporcionan resultados que se dividen en las siguientes subescalas: función física (capacidad para realizar tareas físicas), rol físico (capacidad para cumplir el rol físico), dolor corporal, salud general, vitalidad (energía / fatiga), función social (capacidad para realizar actividades y tareas sociales), rol emocional (limitaciones de

rol debido a problemas emocionales) y salud mental. Cada subescala produce una puntuación de 0 a 100, con puntuaciones más altas que indican la mejor calidad de vida del paciente. El Alfa Cronbach general es de $\alpha = .85$ y $\alpha = .75$ para todas las dimensiones excepto para el funcionamiento social ($\alpha = .45$).

- Presencia y severidad de depresión, medida mediante el *Inventario de Depresión de Beck* (BDI) en su segunda versión. El BDI-II (Beck et al., 1996) es un inventario autoinformado que mide la presencia y severidad de la depresión con un rango de puntuaciones de 0 (depresión mínima / sin depresión) a 63 (depresión severa). Está compuesto por 21 ítems de respuesta múltiple en el que el/la evaluado/a elige la alternativa, ordenadas de menor a mayor severidad, la frase que mejor describe su situación durante las dos últimas semanas. El Alfa Cronbach obtenida por los autores es de $\alpha = .86$. Se utilizó la versión española de Sanz et al. (2003).
- Presencia y severidad de ansiedad, medida mediante el *Inventario de Ansiedad de Beck* (BAI). El BAI (Beck et al., 1988) es un inventario autoinformado que mide la presencia y severidad de ansiedad con un rango de puntuaciones de 0 (depresión mínima / sin depresión) a 63 (depresión severa). Está compuesto por 21 ítems de respuesta múltiple en el que el/la evaluado/a elige la alternativa, ordenadas de menor a mayor severidad, la frase que mejor describe su situación durante las dos últimas semanas. El Alfa Cronbach obtenida por los autores es de $\alpha = .93$.
- Conducta ante la situación de enfermedad, medida mediante el *Brief Illness Perception Questionnaire* (Brief IPQ). El BIPQ es una versión breve del *Illness Perception Questionnaire* (Weinman et al., 1996), diseñado para evaluar las representaciones cognitivas y emocionales de la enfermedad. Está compuesto por 9 ítems, con una puntuación del 1 al 10, que miden consecuencias, duración, control

personal y control del tratamiento, representaciones emocionales, coherencia y causas. La consistencia interna no se evaluó en el estudio utilizado para diseñar la escala, pero mostró una alta validez convergente con el IPQ (Broadbent et al., 2006).

Variables de perfilado

- Severidad del dolor, medida mediante el *Brief Pain Inventory* (BPI). El BPI (Cleeland, 1989, 1990, 1991; Cleeland y Ryan, 1994) es un autoinforme de 11 ítems desarrollado para evaluar la severidad del dolor clínico padecido, así como el grado de alteración social causado por el mismo. Cada uno de dichos ítems se puntúa de 0 a 10 en una escala visual (Cleeland, 1991). Ofrece dos índices, severidad del dolor e interferencia del dolor. El Alfa de Cronbach es de $\alpha = .91$. De este instrumento también se obtuvo información sobre el dolor máximo y el mínimo.
- Percepción de discapacidad – *Índice de Discapacidad Cervical* (NDI). El NDI (Vernon y Mior, 1991) es un autoinforme de 10 ítems que se desarrolló para evaluar la discapacidad percibida por el paciente, con una puntuación que va de 0 (sin discapacidad) a 34 (discapacidad completa). A diferencia de la dimensión de interferencia del BPI, que busca evaluar la alteración provocada por el dolor en general, el NDI está diseñado para medir específicamente la interferencia en las diferentes áreas sociales provocada por el dolor en la región cervical. Tiene una alta consistencia interna con un Cronbach de $\alpha = .92$.
- Otras variables de interés. Se recogieron las variables sociodemográficas de los participantes (sexo y edad) y de interés medico-legal (momento del día en el que ocurrió el accidente, asiento ocupado, localización del impacto, tipo de carretera, cinturón de seguridad abrochado, posición de la cabeza, estado del coche, tiempo desde el accidente y sintomatología descrita). Se comprobó además que los pacientes

presentaran un nexo causal entre la condición diagnosticada y el presunto origen de la misma. Los criterios utilizados para establecer dicho nexo son los recogidos en el Artículo 135 de la ley 35/2015, explicados con detalle en el Estudio 1 (página 84).

7.1.4. Análisis de datos

Se obtuvieron los estadísticos descriptivos y de frecuencias del total de la muestra y se estudió la existencia de diferencias significativas en las variables de *clustering* y perfilado en función del sexo mediante la prueba *t* de Student.

Para el análisis *cluster* de las variables detalladas con anterioridad, se realizaron tres pasos. El primero fue determinar el número óptimo de *clusters*. Para ello se siguió el procedimiento de Aguerrevere *et al.* (2018) aplicando el análisis two step con la solución autoclustering del SPSS, la cual selecciona el número óptimo de clusters tomando el valor más alto de la *ratio of distance measures* (RDM) y el valor más bajo del criterio bayesiano de Schwarz's (BIC). El segundo paso fue el análisis *cluster* en sí mismo. Se utilizó el procedimiento K-medias con las variables de *clustering* y el número óptimo de *clusters* indicado por el análisis anterior. El tercer y último paso fue comprobar la validez de los subgrupos formados mediante un procedimiento de validación cruzada. Para ello se examinaron las diferencias en las variables de perfilado con un análisis de la varianza de un factor (ANOVA) con Tukey HSD *post hoc tests* para las variables continuas y Chi² para las variables categóricas. A continuación, se realizó un análisis discriminante, con el método por pasos, con las variables utilizadas para formar los *clusters*, evaluando así la capacidad de las mismas para predecir la pertenencia a los clusters. Debido a las diferencias entre el formato de los resultados de las escalas, las variables de *clustering* fueron normalizadas en puntuaciones Z. Después de realizar estos análisis, las

puntuaciones Z se convirtieron a los valores originales de cada escala para facilitar la presentación y la comprensión de los resultados.

7.2. Resultados

7.2.1. Clasificación de los grupos y características de la muestra

Los resultados del análisis del grado de acuerdo entre los tres evaluadores para la clasificación de los grupos muestran un ICC (2,1) de .98. Se encontraron 3 desacuerdos, resolviéndose todos mediante consenso. En los 7 casos, dos de los expertos consideraban que era un probable simulador, mientras que el tercero que era un caso dudoso.

En lo relativo a los pacientes clínicos, la valoración médico-legal fue realizada tras una media de 34.12 días ($DT= 7.33$) desde el accidente, con un rango de 25 a 57 días. Todos afirmaban padecer dolor cervical, un 96.1% sensibilidad excesiva en la región cervical, un 20.6% afirmaban padecer mareos, y un 4.4% indicó otro síntoma además de los mencionados, siendo en todos dolor en el área lumbar. En ninguno de los casos los profesionales otorgaron un grado al SLC, no obstante, ninguno de los pacientes padecía signos neurológicos severos, ni se observó ninguna fractura ni lesión musculoesquelética de gravedad. A todos los pacientes incluidos en el estudio se les realizó un estudio del nexo causal y, en todos los casos, se consideró que existía relación causal entre el accidente sufrido y la sintomatología descrita/observada.

Por otro lado, únicamente 12 pacientes (4.82%) presentaron información sobre el estado previo al accidente (estado anterior), no afectando en ninguno de ellos al establecimiento del citado nexo causal. Finalmente, no se encontraron diferencias significativas en función del sexo en ninguna de las variables incluidas en el estudio. En la Tabla 10 se muestran las variables de interés médico legal.

Tabla 10. Factores relacionados con el accidente

	n	%
Momento del día en el que ocurrió el accidente		
Durante el día	166	72.4
Durante la noche	55	24.0
Al amanecer	5	2.2
Al anochecer	3	1.4
Asiento ocupado		
Conductor	50	21.8
Copiloto	77	33.7
Asiento trasero	102	44.5
Localización del impacto		
Frontal	55	24.0
Trasero	153	66.8
Lateral	21	9.2
Tipo de carretera		
Rural	21	9.1
Ciudad	166	72.5
Autovía/autopista	42	18.4
Asiento de seguridad abrochado		
No	15	6.6
Si	214	93.4
Posición de la cabeza		
Mirada al frente	190	82.9
Girada hacia la izquierda	21	9.3
Girada hacia la derecha	18	7.8
Estado del coche		
Daño irreparable	7	3.0
Daño serio	22	9.6
Daño moderado	170	74.3
Daños menores/leves	30	13.1

7.2.2. Análisis cluster

Determinación del número óptimo de clusters. Se aplicó el análisis two step con la opción de autoclustering con toda la muestra. Los resultados obtenidos indicaron que la solución de tres clusters era la más indicada, (BIC=916.42, RDM=1.88), ya que la solución de dos y cuatro cluster obtenían un mayor BIC y un menor RDM (BIC=932.63, RDM=1.80 y BIC=950.19, RDM=1.78 respectivamente). La medida de silueta de cohesión y separación consideró la calidad de los cluster como “buena”. La aplicación de los dos criterios explicados indicó que la solución de tres grupos era la óptima.

Comparación de los grupos en función de las variables de clustering. Se realizó un análisis cluster k-medias con las variables mostradas en la Tabla 11. La distribución de los pacientes con la solución de 3 clusters fue: Cluster 1 con 64 participantes (27.95%), Cluster 2 con 92 (40.17%), y Cluster 3 con 73 (31.88%).

El grupo que forma el Cluster 3 (C3 de ahora en adelante) se caracteriza por tener puntuaciones más elevadas en las variables estudiadas que los otros dos clusters, con la excepción de aquellas en las que no se han encontrado diferencias significativas. Los resultados mostraron que este grupo presentaba sintomatología ansioso-depresiva (BAI, $M=18.17$, $DT=3.41$; BDI, $M=22.17$, $DT=3.84$), alteración moderada-alta de las variables de HRQoL medidas con el SF-36 (Rol emocional, $M=37.22$, $DT=17.42$; Vitalidad, $M=27.11$, $DT=15.02$; Salud mental, $M=53.92$, $DT=19.12$; Función social, $M=33.87$, $DT=12.71$; Salud general, $M=43.05$, $DT=14.95$; Función física, $M=26.46$, $DT=11.22$; Rol físico, $M=55.17$, $DT=18.40$), peor conocimiento de la condición padecida (Coherencia, BIPQ, $M=5.11$, $DT=2.37$) y peor percepción de pronóstico (Duración, BIPQ, $M=6.57$, $DT=2.98$).

Por otro lado, el Cluster 2 (C2 de ahora en adelante) se caracteriza por puntuaciones moderadas – bajas en general, con una alteración moderada-baja en las variables de la HRQoL (Rol emocional, $M=58.68$, $DT=15.42$; Vitalidad, $M=37.72$, $DT=15.24$; Salud mental, $M=84.20$, $DT=15.41$; Función social, $M=51.12$, $DT=14.27$; Salud general, $M=57.19$, $DT=15.35$; Función física, $M=39.94$, $DT=11.49$), significativamente más bajas que el Cluster 1 (C1 de ahora en adelante), excepto por Rol físico ($M=72.94$, $DT=12.82$), y ausencia de sintomatología ansioso-depresiva (BAI, $M=3.75$, $DT=2.89$; BDI, $M=4.78$, $DT=3.54$).

El grupo C1 incluye pacientes que han obtenidos puntuaciones bajas en general, con ausencia de sintomatología ansioso-depresiva BAI, $M=2.74$, $DT=3.88$; BDI, $M=1.92$, $DT=2.35$) y baja alteración de las variables de la HRQoL (Rol emocional, $M=78.77$, $DT=12.97$; Vitalidad, $M=67.95$, $DT=16.89$; Salud mental, $M=96.35$, $DT=7.10$; Salud general, $M=77.82$, $DT=12.71$; Rol físico, $M=73.91$, $DT=12.42$) a excepción de la Función social y física ($M=53.13$, $DT=13.91$ and $M=69.98$, $DT=15.88$ respectivamente).

Salvo las variables indicadas en el grupo C3 (duración y coherencia), los tres grupos tienen un perfil similar en las variables del BIPQ, con una capacidad de control media-alta (tanto para tratamiento como personal), baja preocupación sobre la condición y buena identidad de síntomas.

Tabla 11. *Media, SD y ANOVA del análisis cluster.*

	Cluster 1 (N=64)	Cluster 2 (N=92)	Cluster 3 (N=73)	$F(2,226)$	p	Diferencias intercluster significativas
Depresión (BDI)	1.92(2.35)	4.78(3.54)	22.17(3.84)	773.07	.00	3>2, 3>1, 2>1
Ansiedad (BAI)	2.74(3.88)	3.75(2.89)	18.17(3.41)	487.46	.00	3>2, 3>1
Rol emocional (SF-36)	78.77(12.97)	58.68(15.42)	37.22(17.42)	123.49	.00	3>2, 3>1, 2>1
Vitalidad (SF-36)	67.95(16.89)	37.22(15.24)	27.11(15.02)	125.15	.00	3>2, 3>1, 2>1
Salud mental (SF-36)	96.35(7.10)	84.20(15.41)	53.92(19.12)	149.05	.00	3>2, 3>1, 2>1
Función social (SF-36)	69.98(15.88)	51.12(14.27)	33.87(12.71)	109.13	.00	3>2, 3>1, 2>1
Salud general (SF-36)	77.82(12.71)	57.19(15.35)	43.05(14.95)	98.22	.00	3>2, 3>1, 2>1
Función física (SF-36)	53.13(13.91)	39.94(11.49)	26.46(11.22)	82.58	.00	3>2, 3>1, 2>1
Rol físico (SF-36)	73.91(12.42)	72.94(12.82)	55.17(18.40)	37.89	.00	3>2, 3>1
Duración (BIPQ)	3.85(2.01)	4.32(2.25)	6.57(2.98)	25.45	.00	3>2, 3>1
Control personal (BIPQ)	6.55(1.42)	6.71(1.55)	6.15(2.10)	2.23	.10	n.s.
Control del tratamiento (BIPQ)	6.22(1.34)	6.47(1.28)	6.12(1.83)	1.21	.29	n.s.
Identidad (BIPQ)	2.31(1.47)	2.54(1.61)	2.45(1.98)	.34	.70	n.s.
Coherencia (BIPQ)	6.91(1.55)	6.55(1.77)	5.11(2.37)	17.50	.00	3>2, 3>1
Consecuencias (BIPQ)	3.09(2.32)	2.98(1.81)	3.44(2.35)	.98	.37	n.s.

Abreviaturas: n.s.: no significativo.

Diferencias en los subgrupos en las variables de perfilado. Se analizaron los subgrupos producidos por el análisis cluster en función de las variables de interés médico-legal y sociodemográficas expuestas en la Tabla 12. En lo relativo al dolor, se encontraron diferencias significativas en función de la severidad del dolor [$F(2,226)= 111.96, p=.00$], dolor máximo [$F(2,226)= 76.91, p=.00$] y dolor mínimo [$F(2,226)= 22.92, p=.00$],

percepción de discapacidad [$F(2,226)= 60.71, p=.00$] y tiempo transcurrido desde el accidente [$F(2,226)= 14.74, p=.00$]. No se encontraron diferencias significativas en función de la edad [$F(2,226)= .45, p=.63$].

Tabla 12. *Diferencias intercluster en las variables de perfilado.*

	Cluster 1 (C1)	Cluster 2 (C2)	Cluster 3 (C3)	$F(2,226)$	p	Diferencias intercluster significativas
Severidad del dolor (BPI)	3.97(1.37)	4.63(1.07)	6.70(.98)	111.96	.00	3>2, 3>1 ,2>1
Dolor máximo (BPI)	6.27(1.11)	7.04(1.29)	8.71(1.13)	76.91	.00	3>2, 3>1 ,2>1
Dolor mínimo (BPI)	2.14(1.39)	2.62(1.37)	3.66(1.32)	22.92	.00	3>2, 3>1 ,2>1
Percepción de discapacidad (NDI)	13.62(5.00)	17.39(4.67)	22.09(3.82)	60.71	.00	3>2, 3>1 ,2>1
Tiempo desde la lesión	59.00(22.01)	41.82(23.89)	38.91(23.79)	14.74	.00	3<1, 2<1
Edad	36.84(9.79)	37.93(12.21)	36.27(11.50)	.45	.63	n.s.

n.s., no significativo

Los subgrupos también fueron analizados en función de variables relacionadas con el accidente y el sexo, no encontrándose diferencias significativas en ninguna de ellas (Momento del día en el que ocurrió la colisión: $X^2=6.25$, Coeficiente Phi= .20, $p=.39$; Asiento ocupado: $X^2=8.27$, Coeficiente Phi= .23, $p=.08$; Localización del impacto: $X^2=.77$, Coeficiente Phi= .07, $p=.92$; Tipo de carretera: $X^2=4.39$, Coeficiente Phi= .35, $p=.17$; Cinturón de seguridad abrochado: $X^2=.32$, Coeficiente Phi= .04, $p=.84$; Posición de la cabeza: $X^2=5.52$, Coeficiente Phi= .19, $p=.23$; Estado del coche: $X^2=4.40$, Coeficiente Phi= .17, $p=.62$; Sexo: $X^2=2.65$, Coeficiente Phi= .13, $p=.26$).

7.2.3. Análisis discriminante (validación cruzada)

Para validar la solución de cluster obtenida con la metodología expuesta en el apartado anterior, se realizó un análisis discriminante con las variables que el análisis cluster había utilizado para formar los subgrupos: depresión (Wilk $\lambda = .66$, $p = .000$), ansiedad (Wilk $\lambda = .83$, $p = .000$), rol emocional (Wilk $\lambda = .47$, $p = .000$), vitalidad (Wilk $\lambda = .55$, $p = .000$), salud mental (Wilk $\lambda = .51$, $p = .000$), función social (Wilk $\lambda = .55$, $p = .000$), salud general (Wilk $\lambda = .75$, $p = .000$), función física (Wilk $\lambda = .60$, $p = .000$), rol físico (Wilk $\lambda = .55$, $p = .000$), *timeline* (Wilk $\lambda = .42$, $p = .004$) y coherencia (Wilk $\lambda = .46$, $p = .006$).

Los resultados del análisis mostraron dos funciones con resultados significativos para las variables en las funciones 1 al 2 ($X^2 = 280.32$, $p = .00$, Wilk $\lambda = .14$) y en la función de prueba 2 ($X^2 = 21.64$, $p = .001$, Wilk $\lambda = .86$). Estos resultados indicaron que las variables utilizadas eran capaces de explicar la relación entre las variables y los subgrupos creados por el cluster con un 96.8% (Función 1 y 2, correlación canónica = .91) y un 3.2% (Función 2, correlación canónica = .37). Las relaciones más fuertes para la función 1 se observaron en Rol emocional, Salud mental y BDI. Por otro lado, para la función 2 en Rol Físico y Función social. Finalmente, la validación cruzada de la clasificación mostró que el modelo era capaz de clasificar correctamente a un 96.1% de los 3 subgrupos (97.2% del cluster 1, 97.3% del cluster 2 y 93.2% del cluster 3).

7.3. Discusión

Los objetivos que se perseguían con el presente estudio eran (1) obtener de forma empírica subgrupos basados en características psicosociales utilizando un análisis de cluster k-medias con una muestra de pacientes litigantes con latigazo cervical y (2) investigar diferencias en los subgrupos obtenidos en variables de interés que no han sido

utilizadas para formar los clústeres (severidad del dolor, percepción de discapacidad, variables sociodemográficas y variables médico-legales).

En términos generales, nuestros resultados muestran un perfil del dolor en SLC coherente con lo expuesto con anterioridad en la bibliografía (Åhman y Stålnacke, 2008; Beltran-Alacreu *et al.*, 2018; Sullivan *et al.*, 2002). Se trata de una condición con dolor general de grado intermedio-alto, que puede alcanzar valores muy elevados en situaciones específicas, como hacer un movimiento brusco o un sobreesfuerzo. A pesar de ello, los pacientes son capaces de controlar la severidad del dolor en cierta medida, principalmente mediante el uso de la medicación, como se puede observar en las puntuaciones de dolor mínimo registradas del BPI y control del tratamiento del BIPQ. Por otro lado, la presencia de síntomas psicológicos también es coherente con los resultados de otras investigaciones. La síntesis de evidencia de Carroll *et al.* (2009) mostraba que un alto porcentaje de pacientes de SLC desarrollaban sintomatología ansioso-depresiva a las 6 semanas y manifestaban tener una peor calidad de vida. Nuestros resultados sobre la alteración en la HRQoL también son consistentes con los obtenidos por otros autores (Leaver *et al.*, 2013; Pedisic *et al.*, 2013). Como Manolopoulos *et al.* (2008) explican, las lesiones musculoesqueléticas ubicadas en la zona cervical, como las observadas en el SLC, generan alteraciones físicas que disminuyen el rango de movimientos que la persona puede realizar. Estas alteraciones producen limitaciones funcionales en el correcto desempeño de las actividades diarias, lo que se traduce en una alteración de la calidad de vida de los pacientes.

En lo relativo a los resultados del análisis cluster, se encontraron tres grupos diferentes (C1, C2 and C3). El grupo C3 está compuesto por pacientes que muestran perfil general de severidad moderada-alta. Se observa sintomatología ansioso-depresiva y alteración moderada-elevada de la HQoL y de la representación cognitiva de la

enfermedad. Dicha alteración de la HQoL se manifiesta mediante la disminución de la capacidad para desarrollar adecuadamente tareas físicas y de desempeñar correctamente actividades sociales, pérdida de vitalidad y sensación de fatiga, límites en el desempeño del rol habitual por problemas físicos y emocionales, y percepción de alteración general de la salud. De la misma forma, estos pacientes han desarrollado una representación cognitiva de la enfermedad desadaptativa, caracterizada por una peor comprensión de la condición sufrida, así como la creencia de que se extenderá durante un largo periodo de tiempo. Por otro lado, en las variables de perfilado, este grupo tiene valores significativamente más elevados de severidad del dolor general, mostrando un grado de dolor percibido moderado-alto. De la misma forma, también presentan un mayor dolor máximo y mínimo, lo que indicaría que tienen más dificultades para controlarlo, y una mayor percepción de discapacidad.

El grupo C2 está compuesto por pacientes que presentan una severidad general moderada, sin sintomatología ansioso depresiva y afectación de la HQoL/representación cognitiva de la enfermedad moderada-baja. Muestran alteraciones en la función física y social, la vitalidad y el rol emocional, pero tienen más capacidad que el grupo C3 para desempeñar las actividades que componen su rol físico. De la misma forma, muestran una representación cognitiva de la enfermedad más adaptativa que los otros dos grupos. Puede observarse también una severidad del dolor y una percepción de discapacidad intermedia, menor que el grupo C3 pero significativamente mayor que el tercer y último grupo (C1). Por otro lado, dicho grupo C1 está compuesto por pacientes con puntuaciones bajas en las variables expuestas, pero que aún presentan cierta alteración en la HQoL, especialmente en la vitalidad y las funciones sociales y físicas, y padecen dolor y alteración en la percepción de discapacidad leves.

También se han observado diferencias significativas en función del tiempo transcurrido desde el accidente. El grupo C1, que muestra síntomas leves, tenía una media de tiempo significativamente más elevada que los otros dos. Esto sugeriría que la severidad de los síntomas es menor una vez transcurridos dos o tres meses desde el origen de los mismos, pero algunos, como el dolor, persisten. Estos resultados coinciden con la síntesis de Carroll *et al.* (2009), donde se registraban, mayoritariamente, tiempos de recuperación de entre 1 y 6 meses, siendo el dolor el síntoma que más tiempo tarda en desaparecer.

Consideramos de interés la ausencia de diferencias significativas entre los grupos C2 y C3 en esta variable. Estos dos grupos muestran un tiempo desde el accidente similar -algo más de 1 mes-, pero tienen un perfil de severidad diferente. Ya que no se han observado diferencias en las variables relativas a la colisión, las cuales podrían justificar las diferencias en estos grupos, creemos que una posible explicación es la interacción entre las variables psicológicas y el dolor. Es posible que la presencia de sintomatología psicológica, como la ansioso-depresiva observada en el grupo C3, provoque una mayor percepción del dolor en el paciente por el sesgo atencional que puede ocasionar (Arola *et al.*, 2010; Ho *et al.*, 2018). La depresión y la ansiedad son mediadores importantes en la catastrofización, la hipervigilancia y la evitación, variables asociadas con una exacerbación de la percepción del dolor. Los pacientes con sintomatología ansioso-depresiva tienden a catastrofizar como estrategia de adaptación, produciendo un estado de hipervigilancia o un estado emocional negativo en el que se presta más atención al dolor, lo que se traduce en un incremento de la gravedad percibida (Aparicio *et al.*, 2013; Lumley *et al.* 2011). También es posible que la interacción ocurra en la dirección opuesta, ya que la evidencia de la dirección de la relación entre el dolor y estas variables en pacientes con SLC no es concluyente (Woo, 2010). Así, una severidad del dolor elevada

puede provocar una experiencia de la condición traumática que genera síntomas psicológicos (Phillips *et al.*, 2010). Dado que la presencia de sintomatología ansioso-depresiva es un potencial factor de riesgo en el proceso de cronificación del dolor, de cara a futuras investigaciones, sería conveniente comprender la dirección de la relación entre dicha sintomatología y la gravedad del dolor. Esto ayudaría a predecir el pronóstico de la enfermedad y proporcionaría una intervención más específica y eficaz (Pool *et al.*, 2010).

Las diferencias entre los grupos en la percepción de discapacidad también pueden explicarse desde la influencia de las variables psicológicas. La presencia de sintomatología ansioso-depresiva se asocia con la evitación de actividad y una mayor percepción de alta discapacidad (Zale y Ditre, 2015). Es posible que aquellos pacientes que sufren mayor dolor restrinjan sus movimientos y actividad para evitar la aparición del dolor, lo que conduciría a una mayor percepción de discapacidad (Vlaeyen y Linton, 2000).

En lo relativo al resto de variables estudiadas en la comparación de los subgrupos, no se han encontrado diferencias significativas en los factores relacionados con el accidente (momento del día, asiento ocupado, localización del impacto, tipo de carretera, cinturón de seguridad abrochado, posición de la cabeza y estado del coche) lo indicaría que no son variables que influyan en la gravedad de los síntomas. Estos hallazgos son coherentes con los resultados obtenidos por otros autores, donde se ha observado que son variables que no están asociadas con la prognosis de la condición (Sterling y Kenardy, 2011). Por otro lado, tampoco se han encontrado diferencias significativas en función de la edad o el sexo en ninguna de las variables estudiadas. Estos resultados coinciden con el reciente estudio de Ickmans *et al.* (2017), donde no se encontró ninguna diferencia en función de factores personales, y con el de Malfliet *et al.* (2015), donde no se hallaron

diferencias en función del sexo en variables psicológicas como depresión, miedo, somatización, función física, dolor físico o salud general.

Consideramos que los tres grupos obtenidos pueden resultar de interés para el contexto clínico-forense, ya que aportan evidencia sobre los diferentes perfiles de paciente litigante de SLC que puede encontrar el profesional. Como puede observarse, la clasificación realizada por el análisis cluster indica que hay pacientes que presentan un patrón de síntomas de severidad elevada que podría considerarse atípico, que no se explica por el tiempo transcurrido desde el accidente o la influencia de factores relativos a la colisión. Como se exponía anteriormente, esta magnificación de los síntomas suele interpretarse a menudo como un signo de simulación y crea un efecto de falsa alarma en el que se considera que el paciente litigante exagerará sus síntomas para obtener una mayor compensación económica, con las consecuencias psicológicas y sociales que dicho efecto conlleva (Thompson *et al.*, 2018). Como afirman Merckelbach *et al.* (2019), no solo debe considerarse qué tipo de síntomas presenta el paciente, sino también cuándo y cómo, por lo que el profesional debe analizar y valorar adecuadamente todas las posibles hipótesis. En este sentido, anteriormente exponíamos que la simulación o la interacción cíclica entre el dolor y la sintomatología psicosocial son dos posibles explicaciones a los síntomas de severidad inusual, pero también pueden considerarse otros factores como la secuencia de los ítems y los test que se apliquen, confusión o falta de información sobre los síntomas o el proceso de enfermedad, estilo de respuesta desatento o rasgos de personalidad (Merckelbach *et al.*, 2019).

8. Bloque 3 - Estudio 4. Capacidad discriminativa de un protocolo biopsicosocial para la detección de simulación en SLC.

El Estudio 4 comprende el análisis y estudio de la capacidad del protocolo biopsicosocial para discriminar entre simuladores, instruidos para fingir SLC, y pacientes clínicos que han sido diagnosticados con dicha condición.

El objetivo que se persigue es:

- **Objetivo específico 3.1.** Evaluar la capacidad de una serie de escalas breves para discriminar entre pacientes clínicos con sintomatología genuina y simuladores instruidos.

8.1. Metodología

8.1.1. Participantes

Para la presente investigación se contó con una muestra inicial de 651 participantes divididos en tres grupos: población general, simuladores instruidos y pacientes clínicos. Se utilizaron los siguientes criterios de inclusión para cada grupo: (1) Población general y simuladores: Criterio a1. Necesidad de firmar el consentimiento informado, Criterio b1. Superar un manipulation check. Criterio c1. No haber estado envuelto en un proceso de compensación médico-legal y Criterio d1. No presentar ninguna alteración psicológica o médica que pueda alterar las respuestas de los instrumentos. (2) Pacientes clínicos: Criterio a. Necesidad de firmar el consentimiento informado, Criterio b2. La persona estaba siendo evaluada para valorar el daño físico y psicológico producido tras un accidente de circulación, Criterio c2. Era mayor de edad (≥ 18 años), Criterio d2. Se catalogaba como paciente clínico. El incumplimiento de los criterios en cada grupo suponía la exclusión del estudio. Tras la aplicación de los citados

criterios, se descartaron 18 pacientes clínicos por no cumplir el criterio d2 (16 fueron catalogados como probables simuladores y 2 como dudosos) y 3 simuladores instruidos por no cumplir el criterio c1.

La muestra final se compuso por un total de 630 personas, con una media de edad de 31.17 ($DT=10.17$) y un rango de 18 a 65. La división por grupos fue: *población general*, formado por 200 participantes, 84 hombres (42%) y 116 mujeres (58%), con una media de edad de 24.55 ($DT=7.81$); *simuladores instruidos* (de ahora en adelante simuladores), compuesto por 201 participantes, 95 hombres (47.26%) y 106 mujeres (52.74%), con una media de edad de 23.92 ($DT=7.51$) y *población clínica relevante*, pacientes clínicos de ahora en adelante, que incluyó un total de 229 sujetos (pertenecientes al grupo A descrito en el estudio 3) diagnosticados con latigazo cervical tras un accidente de, con una media de edad de 37.21 ($DT=10.65$) y un rango de 18 a 60. De dicho total, 119 eran mujeres (52%) y 110 hombres (48%).

8.1.2. Instrumentos

Se utilizó el protocolo biopsicosocial expuesto con anterioridad. Los instrumentos que lo componen han sido descritos en el estudio anterior (páginas 151-153).

Manipulation check.

Para asegurarnos de que los participantes que formaba el grupo experimental (simuladores instruidos y población general) habían comprendido y recordado adecuadamente sus instrucciones, se utilizó un manipulation check con cinco preguntas de respuesta abierta en las que se evaluaba: *recuerdo* (¿Cuáles eran tus instrucciones?), *comprensión* (Dime con tus propias palabras qué se te pidió que hicieras), *conformidad con las instrucciones* (Describe como te has sentido de motivado al hacer esta investigación), *esfuerzo* (Del 1 al 10, ¿Cuánto te has esforzado?), *motivación interna* (¿Te

ha motivado el desafío de “vencer” al test?), y *motivación externa* (¿Te has tomado más en serio tu rol por el miedo a perder la recompensa?). Estas instrucciones eran valoradas posteriormente por uno de los investigadores, otorgándoles una puntuación del 1 al 5 (muy mala, mala, regular, buena y muy buena) en función de la adecuación de la respuesta a una plantilla de corrección que se le facilitó. Este manipulation check se aplicaba a ambos grupos, con la única diferencia de que el que se les daba a los simuladores también incluía dos casillas donde se preguntaba cómo había preparado su rol (“¿Qué fuente y/o material has utilizado para este experimento y cómo la has utilizado para preparar tu rol? Por favor, da detalles sobre el procedimiento”). Estas dos preguntas no eran valorables, especificándose este hecho tras las mismas (“Puedes responder con total sinceridad, estas dos cuestiones no serán utilizadas para valorar tu rendimiento en el experimento”).

8.1.3. Procedimiento

Se utilizó un diseño de simulación (transversal) con los tres grupos expuestos con anterioridad. Los participantes asignados a los grupos de pacientes clínicos, así como el procedimiento para reclutarlos, fueron los mismos que los expuestos en el anterior estudio 3 (páginas 149-151).

Para el grupo experimental, los participantes fueron reclutados en la Universidad de Murcia durante los años 2018, 2019 y enero/febrero de 2020. Se inscribieron por correo electrónico tras la presentación del estudio y fueron asignados de forma aleatoria a las condiciones planteadas: simuladores y población general. Posteriormente, se les facilitaron unas instrucciones acordes a su rol y se les citó en las dependencias de la Universidad.

A los simuladores se les solicitó que fingieran padecer SLC, ocasionado tras un supuesto accidente de circulación, con el objetivo de obtener una compensación

económica. Como motivación se les ofreció una recompensa interna (ser capaz de engañar al test) y externa (puntuación extra en la nota final de una asignatura). Siguiendo la metodología de Rogers y Bender (2018), se penalizó el fallo, de manera que solo aquellos que cumplimentaran la escala acorde al rol asignado, recibirían el bono. Las instrucciones que recibieron fueron diseñadas acorde a los requisitos propuestos por Rogers y Cruise (1998) de claridad, especificidad, contextualización y motivación. Los participantes fueron sometidos a un *manipulation check* donde se valoró, con una puntuación del 1 al 5 el recuerdo, la comprensión, la conformidad con las instrucciones, el esfuerzo, la motivación interna y externa. Aquellos participantes que no alcanzaron una puntuación ≥ 4 en todas las variables de dicho *manipulation check* fueron excluidos del experimento (criterio b2). Para evitar el problema del coaching (Van Impelen *et al.*, 2014), no se ofreció ningún tipo de información o entrenamiento sobre cómo simular, únicamente se hizo hincapié en la importancia de que fuera coherente con lo esperable en la condición seleccionada. Por otro lado, a los participantes asignados a la condición de población general se les solicitó que respondieran las escalas con instrucciones estándar (responde sinceramente). Se aplicaron las mismas condiciones de recompensa externa y *manipulation check* que al anterior grupo (simuladores).

Las instrucciones proporcionadas a los dos grupos experimentales pueden consultarse en el Anexo 3.

8.1.4. Análisis de datos

Se calcularon las medias, las desviaciones típicas y las frecuencias para cada una de las variables estudiadas. Para decidir qué estrategia de análisis de datos utilizar, estudiamos la normalidad de la distribución con el test Shapiro-Wilk y la homocedasticidad con el test de Levene. Debido a que se cumplían los dos principios, se

utilizó estadística paramétrica, analizando las diferencias entre los grupos con el análisis de la varianza de un factor (ANOVA). La magnitud del efecto fue calculada utilizando el estadístico d de Cohen. Debido a que los valores originales propuestos por Cohen (1988) son orientativos y deben ser adaptados al contexto estudiado (Ato et al., 2013), se utilizó la categorización que Rogers et al. (2003) propusieron específicamente para la valoración de la simulación (Moderado $\geq .75$; Grande ≥ 1.25 ; Muy grande ≥ 1.50). Para calcular la capacidad discriminativa de los instrumentos, se calculó la curva característica operativa del receptor (curva COR), obteniendo además el área bajo la curva (AUC) y el error estándar de dicha área (SE_{AUC}). Una vez obtenida la curva COR, se utilizó el índice de Youden (J) para determinar el punto de corte óptimo de cada instrumento, calculando para el mismo la sensibilidad, la especificidad, razón de verosimilitud positiva y negativa ($L+$ y $L-$) y el poder predictivo positivo/negativo (PPP y PPN) para una prevalencia del 10%, 30% y 50%. Los análisis anteriormente descritos también se realizaron a los dos puntos de corte siguientes al óptimo, con la finalidad de ofrecer un rango más amplio de decisión.

Debido a que el interés del estudio es aplicar las escalas en conjunto y conocer el rendimiento general del protocolo, se calcularon las razones de verosimilitud encadenadas (chained likelihood ratios, Larrabee, 2008) con un 30% de prevalencia para uno, dos y tres resultados positivos (resultado igual o superior al punto de corte óptimo). Se siguieron las recomendaciones de Bender y Frederick (2018) y en el cálculo se incluyeron las razones de verosimilitud negativa ($L-$), de manera que se utilizó la siguiente fórmula base:

$$[\text{Odds pretest (prevalencia/1-prevalencia)} * LR+ * LR- * LR- * LR- * LR- = \text{Odds posttest} \\ = \text{probabilidad de simulación (odds posttest/1+oddpotest)}]$$

El L+ es la razón de verosimilitud positiva de la escala que se prueba como resultado positivo, el L- es la razón de verosimilitud negativa del resto de instrumentos. Dicha fórmula corresponde a la prueba para un resultado positivo, para calcular con dos resultados positivos se incluyeron 2 L+ y 3 L- y para calcular tres resultados positivos se incluyeron 3 L+ y 2 L-. Se calcularon además los resultados utilizando la fórmula original propuesta por Larrabee (2008) con finalidad comparativa (Odds pretest * L+).

8.2. Resultados

8.2.1. Características de los grupos

Las características del grupo de pacientes clínicos son las mismas que las recogidas en el estudio anterior. Pueden consultarse en la página 155.

Respecto al grupo de simuladores y población general, para la preparación del rol la mayoría (64%) utilizó páginas no científicas de internet (p.e. de grupos de abogados o de aseguradoras) como fuente de información, localizadas a través de un buscador utilizando términos como “latigazo cervical, esguince cervical o síndromes asociados al latigazo cervical”, “accidente de tráfico e indemnización”, “daños comunes de accidentes de circulación”. Del 36% restante, un 30% utilizó una modalidad combinada de buscar en internet las principales consecuencias y extender la información con bibliografía especializada, y el 6% afirmó haber consultado únicamente bibliografía especializada. En todos los casos, la citada bibliografía especializada consistió en manuales disponibles en la biblioteca de la Universidad. No se especificó la forma acceso a los mismos.

8.2.2. Puntuaciones de las escalas y comparación entre grupos

Como puede observarse en la Tabla 1, el ANOVA de un factor indicó la existencia de diferencias significativas entre grupos en todas las escalas utilizadas. Las puntuaciones

más elevadas se registraron en el grupo de simuladores instruidos (BPI, $M= 67.23$, $DT=18.12$; NDI, $M= 27.50$, $DT=9.93$; BDI, $M= 21.70$, $DT=10.77$; BAI, $M= 14.65$, $DT=5.98$; SF-36 PH, $M= 38.15$, $DT=14.65$; SF-36 MH, $M= 47.48$, $DT=15.58$; BIPQ, $M= 5.26$, $DT=2.51$), superando la media de puntuaciones de dicho grupo el punto de corte óptimo (Tabla 2), con la excepción de NDI y BIPQ.

En lo relativo a la magnitud del efecto de las diferencias entre grupos, la comparación entre pacientes clínicos y simuladores (d_3) indicó que había diferencias muy grandes en las puntuaciones del SF-36 PH ($d=1.60$) y SF-36 MH ($d=1.53$), y grandes en BPI ($d=1.39$). En el resto de escalas se observan magnitudes moderadas (NDI, $d=1.15$; BDI, $d=1.18$) y bajas (BAI, $d=.65$; BIPQ, $d=.26$).

Tabla 13. Media, desviación típica, ANOVA y *d* de Cohen para las puntuaciones de las escalas de los grupos.

	Gen.		Pac.		Sim.		F (2,627)	<i>d</i> ₁ [95% CI]	<i>d</i> ₂ [95% CI]	<i>d</i> ₃ [95% CI]
	M	DT	M	DT	M	DT				
BPI	11.15	8.75	45.54	13.21	67.23	18.12	848.90**	3.01 [2.73-3.27]	3.94 [3.60-4.27]	1.39 [1.18 – 1.60]
NDI	8.72	4.52	18.50	5.65	27.50	9.93	358.51**	1.89 [1.66-2.11]	2.44 [2.17-2.69]	1.15 [.94 – 1.34]
BDI	4.15	2.85	11.33	6.74	21.70	10.77	283.02**	1.34 [1.13-1.54]	2.23 [1.98-2.47]	1.18 [.98 – 1.38]
BAI	6.34	2.41	17.74	7.14	14.66	5.98	225.30**	2.05 [1.82-2.28]	1.83 [1.59-2.06]	.45 [.27 – .65]
SF-36 PH	85.28	11.02	60.20	13.07	38.15	14.65	784.81**	2.88 [2.61-3.13]	4.41 [4.06-4.76]	1.60 [1.38 – 1.81]
SF-36 MH	87.27	13.87	70.29	14.41	47.48	15.58	397.82**	1.68 [1.46-1.89]	3.17 [2.87-3.46]	1.53 [1.31-1.73]
BIPQ	2.14	.84	4.74	1.48	5.26	2.51	180.29**	2.10 [1.87-2.33]	1.67 [1.24-1.89]	.26 [.07 - .45]

Abreviaturas: Gen: población general; Pac: paciente clínico; Sim: Simulador; *d*₁: magnitud del efecto de la comparación entre población general y pacientes clínicos; *d*₂: magnitud del efecto de la comparación entre población general y simuladores; *d*₃: magnitud del efecto de la comparación entre pacientes clínicos y simuladores.

Punto de corte óptimo para discriminar entre pacientes clínicos y simuladores. Con la finalidad de obtener un posible punto de corte óptimo para comparar entre simuladores y pacientes clínicos, se realizó una curva COR (Figura 6) y se calculó el índice Youden (Tabla 14). Para el BPI, el punto de corte óptimo es de 64 ($J=.68$), con una sensibilidad del 74.39% y una especificidad del 94.12% (tasa de falsos positivos del 5.88); para el NDI es 29 ($J=.69$), con una sensibilidad del 71.95% y una especificidad del 98.04% (tasa de falsos positivos del 1.96%), para el BDI es 11 ($J=.60$), con una sensibilidad del 87.37% y una especificidad del 75.16% (tasa de falsos positivos del 24.84%); para el BAI es 13 ($J=.37$), con una sensibilidad del 43.90% y una especificidad del 93.46% (tasa de falsos positivos del 6.54%); para el SF-36 de 34 ($J=.61$), con una sensibilidad del 70.73% y una especificidad del 90.85% (tasa de falsos positivos del 9.15%); y para el BIPQ de 6 ($J=.28$), con una sensibilidad del 46.34% y una especificidad del 81.70% (tasa de falsos positivos del 18.30%).

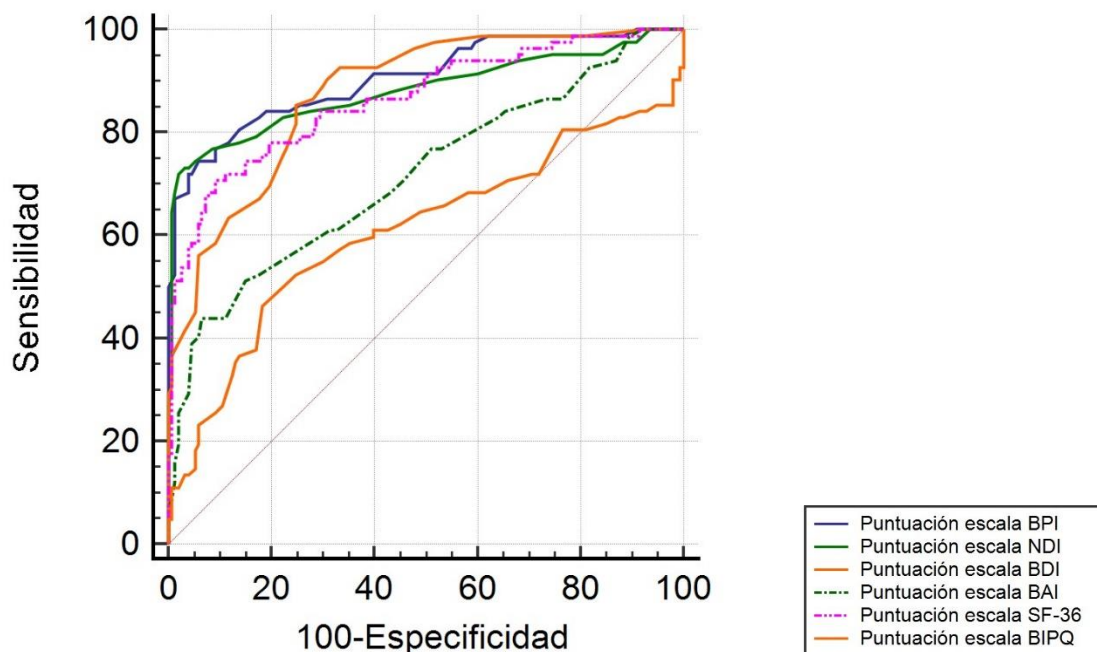


Figura 6. Curvas COR para los instrumentos del protocolo.

Tabla 14. *Capacidad discriminativa del punto de corte óptimo*

Escala	P.C. óptimo	J	SEN	SPEC
BPI	>64	.68	74.39	94.12
NDI	>29	.69	71.95	98.04
BDI	>11	.60	87.37	75.16
BAI	>13	.37	43.90	93.46
SF-36	<34	.61	70.73	90.85
BIPQ	>6	.28	46.34	81.70

Abreviaturas: P.C.: punto de corte; J: Índice de Youden; SEN: sensibilidad; SPEC: Especificidad.

Precisión de la clasificación. La precisión de la clasificación de las escalas utilizadas puede verse en la Tabla 15. Además del punto de corte óptimo expuesto en el apartado anterior, se han incluido dos puntos de corte adicionales para contar con un mayor rango de decisión. El área bajo la curva presenta un rango de .60 a .90, siendo el valor más elevado el del BPI (.90), seguido del NDI (.88), y el más bajo el del BIPQ (.60), seguido del BAI (.71).

Por otro lado, el BPI, el NDI y el BDI han sido los instrumentos que mayor sensibilidad han mostrado. En el BPI, con unos puntos de corte de 64, 65 y 66, se alcanzó una sensibilidad de 74.39, 71.95 y 71.95 respectivamente; en el NDI, con 29, 31 y 32 como punto de corte, la sensibilidad alcanzada fue de 71.41, 68.29 y 64.63, respectivamente; y para el BDI, con puntos de corte de 11, 12 y 13, los valores fueron de 85.37, 81.71, 76.83, respectivamente.

En cuanto a la especificidad, el BPI y el NDI son los instrumentos con valores más elevados, alcanzando el primero, con los puntos de corte anteriormente expuestos, una especificidad de 94.12, 95.42 y 96.08, y el segundo de 98.04, 98.69 y 99.35. El BDI

ha sido el instrumento que ha registrado las cifras más bajas, con valores inferiores al 80% (75.16, 75.16 y 77.12).

En lo relativo a la razón de verosimilitud positiva (L+), el BPI y el NDI también alcanzaron los valores más elevados (12.65, 15.73 y 18.35 el primero, y 36.7, 52.24 y 98.89 el segundo), lo que indicarían que puntuaciones iguales o superiores al punto de corte óptimo en el BPI eran entre 12 y 18 veces más probables en el grupo de simuladores instruidos, y en el NDI entre 36 y 98 veces más probable. El resto de instrumentos obtuvieron cifras que oscilan entre 2.53 y 10.37.

Los PPP y NPP para una prevalencia del 10%, 30% 50% pueden observarse en la tabla X. En términos generales, asumiendo una ratio base del 30%, el más probable de ser encontrado en la práctica clínica (Curtis *et al.*, 2019), el BPI, NDI y SF-36 alcanzaron un PPP superior al 70% (84.4, 87.18 y 88.7 el primero; 94, 95.7 y 97.77 el segundo; y 73.2, 76.2 y 81.6 el tercero), y un NPP superior al 85% (89.6, 88.8 y 88.9 el primero; 89.1, 87.9 y 86.8 el segundo; y 87.6, 87 y 84.9 el tercero). En el caso del NPP, el BAI fue el instrumento que mostró cifras más elevadas (92.3, 90.6 y 88.6).

Tabla 15. *Capacidad discriminativa de los instrumentos*

Escala (P.C.)	AUC [95% IC]	SEAUC	SEN (%)	[95% IC]	SPEC (%)	[95% IC]	L+	[95% IC]	L-	[95% IC]
BPI (64)	.90 [.86 - .94]	.02	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	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	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	71.41	60.7 - 81.1	98.04	94.4 - 99.6	36.7	11.9 - 113.4	0.29	0.2 - 0.4
NDI (31)	.88 [.83 - .92]	.02	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	64.63	53.3 - 74.9	99.35	96.4 - 100.0	98.89	13.9 - 702.1	0.36	0.3 - 0.5
BDI (11)	.87 [.82 - .92]	.02	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	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	76.83	66.2 - 85.4	77.12	69.6 - 83.5	3.36	2.5 - 4.6	0.3	0.2 - 0.4
BAI (13)	.71 [.63 - .77]	.03	43.9	33.0 - 55.3	93.46	88.3 - 96.8	6.72	3.5 - 12.8	0.6	0.5 - 0.7
BAI (14)	.71 [.63 - .77]	.03	43.9	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	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)	.86 [.81 - .91]	.02	70.73	59.6 - 80.3	90.85	85.1 - 94.9	7.73	4.6 - 13.0	0.32	0.2 - 0.5
SF-36 (33)	.86 [.81 - .91]	.02	68.29	57.1 - 78.1	91.5	85.9 - 95.4	8.04	4.7 - 13.8	0.35	0.3 - 0.5
SF-36 (28)	.86 [.81 - .91]	.02	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	46.34	35.3 - 57.7	81.7	74.6 - 87.5	2.53	1.7 - 3.8	0.66	0.5 - 0.8
BIPQ (7)	.60 [.54 - .67]	.04	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	18.29	10.6 - 28.4	94.77	90.0 - 97.7	3.5	1.5 - 7.9	0.86	0.8 - 1.0

Abreviaturas: P.C.: Punto de corte; 95% IC: Intervalo de confianza al 95%; AUC: Área bajo la curva; SEAUC: Error estándar del área bajo la curva; SEN: sensibilidad; SPEC: especificidad; L+ y L-: razón de verosimilitud positiva y negativa.

Tabla 16. Poder predictivo positivo y negativo con un 10%, 30% y 50% de prevalencia

Escala (P.C.)	10% PPP	[95% IC]	10% PPN	[95% IC]	30% PPP	[95% IC]	30% PPN	[95% IC]	50% PPP	[95% IC]	50% PPN	[95% IC]
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	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	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	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	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	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	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)	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)	47.2	34.2 - 60.5	96.3	95.0 - 97.3	76.2	65.5 - 84.3	87	82.9 - 90.2	88.2	81.6 - 92.6	74.1	67.5 - 79.8
SF-36 (28)	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	15.8 - 29.7	93.2	91.7 - 94.4	52	41.9 - 62.0	78	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	14.7 - 46.8	91.3	90.3 - 92.1	60	39.9 - 77.2	73	70.8 - 75.1	77.8	60.8 - 88.8	53.7	51.0 - 56.4

Abreviaturas: P.C.: Punto de corte; 95% IC: Intervalo de confianza al 95%; PPP: poder predictivo positivo; PPN: poder predictivo negativo.

8.2.3. Razón de verosimilitud encadenada

En las tablas 17, 18 y 19 se presenta la razón de verosimilitud encadenada (L+ y L-) de los instrumentos utilizados. En las columnas PoTO1 y PP1 se ha utilizado el método original, propuesto por Larrabee (2008). En la primera se expone el odd ratio post-test resultante de dicha operación y en la segunda la probabilidad de simular con una (tabla 17), dos (tabla 18) y tres (tabla 19) respuestas positivas (igual o superior al punto de corte óptimo). En las columnas PoTO2 y PP2 se han calculado los OR post-test y la probabilidad de simulación siguiendo el método corregido que propusieron Bender y Frederick (2018), en el cual se incluye en la operación la razón de verosimilitud negativa. Las fórmulas empleadas pueden consultarse en “Análisis de datos” del apartado “Metodología”.

Siguiendo el método de Larrabee (2008), con una respuesta positiva en alguna de las escalas (Tabla 17) se alcanza entre un 51% y 93% de probabilidad de simulación, siendo el BPI y el NDI los que más probabilidad muestran (84% y 93% respectivamente) y BDI, BAI y BIPQ los que menos (59%, 58% y 51% respectivamente). A partir de dos resultados positivos, la probabilidad se incrementa a entre un 78% y un 98%, siendo BDI/BAI, BDI/BIPQ y BAI/BIPQ las que se sitúan por debajo del 85% (82%, 78% y 78% respectivamente). Con tres resultados positivos la probabilidad asciende a más del 90%, con un rango de 92% a 99%.

Por otro lado, al aplicar el método de Bender y Frederick (2018), las probabilidades disminuyen significativamente, aunque se mantienen los valores elevados en BPI y NDI, intermedios en el SF-36 y bajos en BDI, BAI e IPQ. Con una respuesta positiva y cinco negativas, las probabilidades se sitúan entre el 1% y el 8%, con dos respuestas entre el 3% y el 80% y con tres entre el 39% y el 99%. En los dos últimos

casos, la combinación del BPI y el NDI fue la que mayor probabilidad obtuvo, mientras que las que menos obtuvieron fueron aquellas en las que estuvo el BIPQ.

Tabla 17. Razón de verosimilitud encadenada con una prueba positiva (RB 30%)

Escala	SEN	SPEC	L+	L-	PrTO	PoTO1	PP1	PoTO2	PP2
BPI	74.39	94.12	12.65	0.27	.42	5.31	0,84	0,03	0,03
NDI	71.95	98.04	36.7	0.29	.42	15.41	0,93	0,09	0,08
BDI	87.37	75.16	3.44	0.19	.42	1.44	0,59	0,01	0,01
BAI	43.90	93.46	3.36	0.30	.42	1.41	0,58	0,08	0,07
SF-36	70.73	90.85	7.73	0.32	.42	3.24	0,76	0,01	0,01
BIPQ	46.34	81.70	2.53	0.66	.42	1.06	0,51	0,002	0,002

Abreviaturas: P.C.: pacientes clínicos; 95% IC: Intervalo de confianza al 95%; PPP: poder predictivo positivo; PPN: poder predictivo negativo.

Tabla 18. Razón de verosimilitud encadenada con dos pruebas positivas (RB 30%)

Escala	PrTO	L+	PoTO1	PP1	PoTO2	PP2
BPI/NDI	5.31	36.7	194.87	0,99	4,24	0,80
BPI/BDI	5.31	3.44	18.26	0,94	0,69	0,41
BPI/BAI	5.31	3.36	17.84	0,94	0,37	0,27
BPI/SF-36	5.31	7.73	41.04	0,97	0,79	0,44
BPI/BIPQ	5.31	2.53	13.43	0,93	0,13	0,11
NDI/BDI	15.41	3.44	53.01	0,98	1,91	0,66
NDI/BAI	15.41	3.36	51.77	0,98	1,02	0,51
NDI/SF-36	15.41	7.73	119.11	0,99	2,19	0,69
NDI/BIPQ	15.41	2.53	38.98	0,97	0,35	0,26
BDI/BAI	1.44	3.36	4.83	0,82	0,17	0,14
BDI/SF-36	1.44	7.73	11.13	0,91	0,36	0,26
BDI/BIPQ	1.44	2.53	3.64	0,78	0,06	0,05
BAI/SF-36	1.41	7.73	10.89	0,91	0,19	0,16
BAI/BIPQ	1.41	2.53	3.56	0,78	0,03	0,03
SF-36/BIPQ	3.24	2.53	8,19	0,89	0,07	0,06

Abreviaturas: PrTO: Odd ratio pre-test; L+: razón de verosimilitud positiva; PoTO1: Odd ratio post-test método Larrabee (2008); PP1: Probabilidad de simular método Larrabee (2008); PoTO2: Odd ratio post-test método Bender y Frederick (2018); PP2: Probabilidad de simular Bender y Frederick (2018).

Tabla 19. Razón de verosimilitud encadenada con tres pruebas positivas (RB 30%)

Escala	PrTO	L+	PoTO1	PP1	PoTO2	PP2
BPI/NDI/BDI	194,87	3.44	670,37	0,99	88.92	0.99
BPI/NDI/BAI	194,87	3.36	654,78	0,99	47.51	0.98
BPI/NDI/SF-36	194,87	7.73	1506,39	0,99	101.93	0.99
BPI/NDI/BIPQ	194,87	2.53	493,03	0,99	16.38	0.94
BPI/BDI/BAI	18,26	3.36	61,37	0,98	7.75	0.89
BPI/BDI/SF-36	18,26	7.73	141,19	0,99	16.63	0.94
BPI/BDI/BIPQ	18,26	2.53	46,21	0,97	2.67	0.73
BPI/BAI/SF-36	17,84	7.73	137,91	0,99	8.88	0.90
BPI/BAI/BIPQ	17,84	2.53	45,13	0,97	1.43	0.59
BPI/SF-36/BIPQ	41,04	2.53	103,84	0,99	3.06	0.75
NDI/BDI/BAI	53,01	3.36	178,11	0,99	21.39	0.96
NDI/BDI/SF-36	53,01	7.73	409,77	0,99	45.89	0.98
NDI/BDI/BIPQ	53,01	2.53	134,11	0,99	7.37	0.88
NDI/BAI/SF-36	51,77	7.73	400,24	0,99	24.52	0.96
NDI/BAI/BIPQ	51,77	2.53	130,99	0,99	3.94	0.80
NDI/SF-36/BIPQ	51,77	2.53	130,99	0,99	8.45	0.89
BDI/BAI/SF-36	4,83	7.73	37,40	0,97	4.00	0.80
BDI/BAI/BIPQ	4,83	2.53	12,24	0,92	0.64	0.39
BDI/SF-36/BIPQ	11,13	2.53	28,16	0,96	1.38	0.58
BAI/SF-36/BIPQ	10,89	2.53	27,57	0,96	0.74	0.42

Abreviaturas: PrTO: Odd ratio pre-test; L+: razón de verosimilitud positiva; PoTO1: Odd ratio post-test método Larrabee (2008); PP1: Probabilidad de simular método Larrabee (2008); PoTO2: Odd ratio post-test método Bender y Frederick (2018); PP2: Probabilidad de simular Bender y Frederick (2018).

8.3. Discusión

El principal objetivo que se perseguía con el presente estudio era comprobar la capacidad de una serie de escalas breves para discriminar entre simuladores instruidos y pacientes clínicos que han sido diagnosticados con SLC. En la comparación entre grupos se pueden observar diferencias significativas en todas las escalas. Como es esperable, la comparación entre simuladores y población general es la que mayor magnitud del efecto presenta. Por otro lado, en la comparación entre simuladores y pacientes clínicos, se observan magnitudes del efecto mayoritariamente moderadas-grandes, siendo el SF-36 el único que ha alcanzado una magnitud muy grande. Los simuladores han alcanzado puntuaciones más elevadas que los otros dos grupos, lo que resulta coherente con lo expuesto en la bibliografía, donde se afirma que dicho grupo tiende a ofrecer una visión de la condición simulada más severa de lo esperable (Rogers y Bender, 2018). De la misma forma, puede observarse que los pacientes clínicos han ofrecido un perfil de gravedad intermedia, con un dolor y una alteración de la calidad de vida moderados, ligera percepción de discapacidad, sintomatología ansioso-depresiva leve y una actitud ante la situación de enfermedad en un punto intermedio entre la visión negativa y positiva. Este perfil es coherente con la presentación de paciente de SLC ofrecida habitualmente en la bibliografía de la temática (Åhman y Stålnacke, 2008; Beltran-Alacreu *et al.*, 2018; Sullivan *et al.*, 2002).

Encontramos una excepción a lo anteriormente descrito sobre los simuladores instruidos, ya que la puntuación registrada en el BAI es significativamente superior a la de los pacientes clínicos. Creemos que una posible explicación a estas diferencias es que los simuladores instruidos consideraron que la sintomatología depresiva era más coherente con la condición a simular, como puede observarse en las puntuaciones del BDI, e infraestimaron el papel de la ansiedad en el SLC. En este sentido, debido al

carácter eminentemente biomédico de la condición, así como la formación académica del grupo (medicina y enfermería), es posible que las fuentes consultadas para preparar el rol siguieran un enfoque principalmente anatómico o biológico y no se hiciera mención expresa a sintomatología de corte ansioso-depresivo. De cara a futuras investigaciones, sería recomendable considerar la inclusión de varios grupos de simuladores para estudiar el posible efecto de la formación previa en la preparación del rol.

En lo relativo a la capacidad discriminativa, los instrumentos han mostrado, en los puntos de corte óptimos indicados por el índice Youden, una especificidad elevada, siendo el BPI, el NDI, el BAI y el SF-36 los que menor tasa de falsos positivos han obtenido. En cuanto a la sensibilidad, el BDI ha sido el que ha alcanzado un valor más elevado, pero su especificidad ha sido significativamente más baja que la del resto de instrumentos, con una tasa de falsos positivos de casi el 25%. Los resultados obtenidos en el SF-36 son similares a los obtenidos por Capilla Ramírez *et al.* (2014), los cuales obtuvieron un punto de corte óptimo y una sensibilidad muy parecida a la nuestra (35 y 73), aunque una especificidad un 10% menor (80%).

Estos resultados indican que, a partir de los puntos de corte óptimos propuestos, el/la evaluador/a debe contemplar la posibilidad de simulación. No obstante, es importante considerar con cuantos test positivos (puntuación igual o mayor al punto de corte óptimo, menor en el caso del SF-36) del presente protocolo el/la profesional tiene que considerar dicha posibilidad. Aplicando el método de Bender y Frederick (2018), con un único resultado positivo, y cinco negativos (por debajo del punto de corte óptimo o por encima, en el caso del SF-36), la probabilidad de simulación es baja. Por ejemplo, con el BPI o el NDI, se alcanza una probabilidad del 3% y 8% respectivamente, mientras que con el BIPQ resulta inferior al 1%. Con dos resultados positivos, y cuatro negativos, la probabilidad se incrementa significativamente, en especial si dichos resultados

positivos son del BPI y el NDI (80%), aunque no alcanza niveles aceptables, ya que en la gran mayoría aún se sitúa por debajo del 50%. Por otro lado, con 3 resultados positivos y 3 negativos, la mayoría de combinaciones de instrumentos alcanza valores iguales o superiores al 90%, con la excepción de aquellas en las que uno de los resultados positivos pertenece al BIPQ.

En vista de estos resultados, podemos señalar una serie de criterios para la toma de decisiones en la práctica clínico-forense. Se puede considerar riesgo de simulación cuando se obtienen tres (o más) resultados positivos. La probabilidad más elevada se alcanza cuando dos de esos tres resultados positivos corresponden al BPI y al NDI (94%-99%), lo que resulta coherente con sus valores predictivos (PPP y PPN) y razones de verosimilitud, ya que son los más elevados de entre todos los instrumentos que componen el protocolo. Cuando se obtienen 3 resultados positivos y únicamente está presente el BPI o el NDI, la probabilidad de simulación también es elevada (por encima del 89% si uno de los positivos no es el BIPQ), aunque es ligeramente superior cuando está el NDI. En ausencia del BPI y el NDI, la probabilidad disminuye al 80% si uno de los positivos no es el BIPQ, y al 39% si dicho instrumento es uno de los positivos. También se puede considerar la posibilidad de simulación, de manera excepcional, con dos resultados positivos, pero dichos resultados deben ser el BPI y el NDI. A continuación, se ofrece una tabla que sintetiza lo anteriormente expuesto, utilizando la nomenclatura de Bianchini *et al.* (2005) de posible, probable y altamente probable (en lugar de definitivo).

Tabla 20. *Criterios para determinar sospecha de simulación en el protocolo*

Riesgo de simulación	Resultado positivo* en	
Posible (80%-89%)	BPI + NDI	NDI + BAI + BIPQ
	BPI + BDI + BAI	NDI + SF-36 + BIPQ
	NDI + BDI + BIPQ	BDI + BAI + SF-36
Probable (90%-96%)	BPI + NDI + BIPQ	NDI + BDI + BAI
	BPI + BDI + SF-36	NDI + BAI + SF-36
Altamente probable (98%-99%)	BPI + NDI + BDI	BPI + NDI + SF-36
	BPI + NDI + BAI	NDI + BDI + SF-36

*Se considera resultado positivo una puntuación igual o superior/menor al punto de corte óptimo de la Tabla 15.

Todas aquellas combinaciones de 3 instrumentos que no se hayan incluido en la tabla anterior no obtienen una probabilidad lo suficientemente elevada como para que el profesional pueda sospechar que hay presencia de simulación. Dichas combinaciones tienen como elemento común la presencia del BIPQ, el cual presenta la peor capacidad discriminativa de todos los instrumentos incluidos en el protocolo, con una sensibilidad por debajo del 50%, la cual cae drásticamente cuando se sobre pasa el punto de corte óptimo. Consideramos que, de cara a futuras investigaciones, el BIPQ puede ser eliminado del protocolo y sustituido por algún instrumento que muestre una capacidad discriminativa más elevada. En términos generales, los resultados obtenidos son prometedores y creemos que el protocolo puede ser de utilidad para la práctica clínico-forense, pero estimamos necesario replicar los resultados del presente estudio utilizando una muestra de simuladores clínicos, mediante un diseño de grupo conocido.

Para finalizar, mediante los hallazgos del Estudio 1 considerábamos que el establecimiento del nexo causal podría ser de utilidad en la presente tesis, bien como un criterio adicional o bien como un método de clasificación de grupos (por ejemplo, valorar los resultados de las escalas en pacientes sin nexo vs con nexo), y se incluyó como

variable de interés. Ninguno de los análisis planificados ha sido posible debido a que en todos los participantes se estableció el nexo causal, y únicamente se tuvo constancia de tres pacientes que no cumplieron dicho criterio, no siendo posible incluirlos en el estudio debido a que la gerencia de la clínica consideró que podría resultar conflictivo al estar envueltos en una situación avanzada de litigio.

9. Bloque 3 - Estudio 5. Checklist de inconsistencias e indicadores de simulación en SLC

A lo largo del presente trabajo, se ha expuesto en múltiples ocasiones que la evidencia sugiere que la valoración de la simulación debe seguir un enfoque multidimensional, o multimétodo, y no limitarse a una sola fuente de información (Rogers y Bender, 2018). En SLC, y dolor en general, ha sido investigada mediante una perspectiva psicométrica basada principalmente en la aplicación de Symptom Validity Test (SVT), pero en condiciones como la descrita su eficacia resulta limitada, y se recomienda el uso de nuevas aproximaciones o enfoques (Akca *et al.*, 2020; Tuck *et al.*, 2019). Ante la importancia que tiene controlar adecuadamente el riesgo de simulación en el contexto forense, desarrollar nuevas técnicas o enfoques que alcancen este objetivo se convierte en una necesidad (Arce *et al.*, 2015; Boskovic *et al.*, 2017). Para tratar de solventar la mencionada problemática, una línea de investigación reciente ha tratado de aplicar herramientas propias del campo de la detección de mentiras en contexto médico-legal. Akehurst *et al.* (2015) diseñaron un checklist de configurado mediante indicadores conductuales y verbales extraídos de la literatura relevante de la mentira y simulación, y a pesar de que los resultados de esta herramienta fueron positivos (aproximadamente un 80% de precisión), muchos de los indicadores incluidos en el estudio no fueron de utilidad para detectar simulación en el ámbito médico-legal. Por otro lado, Boskovic *et al.* (2017) evaluaron la eficacia de la *Verifiability Approach* (Nahari *et al.*, 2012) para valorar la simulación de síntomas, obteniendo resultados significativos, pero encontrando el mismo problema que los anteriores autores, las herramientas utilizadas no son del todo eficaces debido a que están preparadas para evaluar testimonios, o sujetos, con unas características muy concretas que no suelen ser comunes en el contexto médico-legal. Otras investigaciones han intentado validar indicadores exclusivamente conductuales,

conocidos como signos no orgánicos, que puedan ser utilizados para identificar simulación de síntomas físicos como el dolor (Wygant *et al.*, 2017), pero los estudios presentan diversas limitaciones teóricas y empíricas que comprometen la utilidad de sus resultados, por lo que hasta la fecha no se ha logrado asociar unos indicadores conductuales o verbales concretos a simulación (Fishbain *et al.*, 2003).

En vista de la limitada evidencia existente y ante la utilidad que tendría en el contexto médico-legal una herramienta como la propuesta por Akehurst *et al.* (2015), surge la necesidad de identificar y validar indicadores de simulación que sean específicos para dicho contexto. Una herramienta como la descrita, además de aportar un enfoque y una fuente de información adicional, permitiría al/la profesional disponer de una forma de recoger de forma objetiva inconsistencias convincentes identificadas durante la valoración clínico-forense. Por todo lo anteriormente expuesto, mediante el Estudio 5 se plantea el siguiente objetivo:

- **Objetivo específico 3.2.** Comprobar la capacidad de una serie de marcadores de simulación, extraídos de bibliografía relevante en la materia, para discriminar entre pacientes clínicos con sintomatología genuina y simuladores instruidos

9.1. Metodología

9.1.1. Participantes

Para la elaboración de este estudio se ha empleado una muestra de 273 participantes divididos en dos grupos:

Pacientes clínicos

Se utilizó una muestra de 153 pacientes clínicos (de los grupos A y B) evaluados en un centro médico multidisciplinar de la Región de Murcia (España), en los años 2017 y 2018. Todos acudían al centro para ser sometidos a una evaluación del daño corporal a raíz de un accidente de tráfico. Del citado total, 78 eran hombres (51%) y 75 mujeres (49%, con una media de edad de 39.10 años (DT= 11.69) y un rango de edad de 18 a 60 años. En lo que se refiere al nivel económico, predominó el nivel medio (entre 15.000 y 20.000 euros anuales) con un 46.4%, seguido de un 28.1% con nivel económico bajo (>15.000 euros). Más de la mitad de la muestra residen en Murcia capital (56.9%). Finalmente, un 83.5% de los participantes tenían empleo remunerado en el momento de la evaluación, el resto (16.5%) se encontraba en situación de desempleo. En cuanto al nivel educativo, el 48,4% tenía estudios universitarios, el 13,1% tenía bachillerato o formación profesional y el 23,5% tenía educación secundaria.

Se utilizaron los siguientes criterios de selección: Criterio a. Necesidad de firmar el consentimiento informado, Criterio b2. La persona estaba siendo evaluada para valorar el daño físico y psicológico producido tras un accidente de circulación, Criterio c2. Era mayor de edad (≥ 18 años), Criterio d2. Se catalogado como paciente clínico. El incumplimiento de los criterios en cada grupo suponía la exclusión del estudio. Tras la aplicación de los citados criterios, se descartaron 10 pacientes clínicos por no cumplir el criterio d2 (9 fueron catalogados como probables simuladores y 1 como dudoso).

Simuladores instruidos

Compuesta por 120 controles sanos con instrucciones de simulación, alumnos de la Universidad de Murcia. Se les presentó el estudio en una asignatura impartida por un profesor del equipo de investigación y se inscribieron por correo electrónico. El rango de edad fue 19 a 50 años, con una edad media de 23.77 ($DT=7.14$). La mayor parte de los participantes eran mujeres (66.9%), estaban solteras (87.6%) y residían en Murcia capital (60.3%). Ninguno de los sujetos seleccionados había padecido o padecía SLC. Como criterios de selección se utilizaron: Criterio a1. Necesidad de firmar el consentimiento informado, Criterio b1. Superar un manipulation check. Criterio c1. No haber estado envuelto en un proceso de compensación médico-legal y Criterio d1. No presentar ninguna alteración psicológica o médica que pueda alterar las respuestas de los instrumentos.

9.1.2. Instrumentos

Checklist de Detección de la Simulación (CDS)

Los marcadores incluidos se seleccionaron entre aquellos que mayor capacidad discriminativa demostraron en el estudio de Akehurst *et al.* (2015), entre las dimensiones cognitivas de Leventhal (CS-SRM; Leventhal *et al.*, 1980, 2007; Linnell y Easton, 2006), y de las publicaciones de Ferrari (2005); McDermott y Feldman (2007); Resnick (1997) y Vrij (2008), entre otras. El listado completo de indicadores y su fuente de origen se incluyen en el Anexo 4.

Para que los investigadores pudieran evaluar los indicadores incluidos en el estudio se desarrolló un checklist (CDS) compuesto por 22 criterios con 4 niveles de respuesta siendo el primer valor “nada presente” y el cuarto “fuertemente presente”, uno para cada indicador, divididos en 5 dimensiones (Detalles inusuales y pensamientos sobre

el evento, Actitud ante la situación de enfermedad, Actitud ante la evaluación, Causa e Inconsistencias). Cada ítem de las dimensiones 2 a la 5 se planteó de manera que una puntuación elevada indicara un mayor riesgo de simulación. Por otro lado, la dimensión 1 recoge dos criterios del CBCA (Steller y Köhnken, 1989) cuya presencia indica honestidad, por lo que se valoraron a la inversa para que se ajustara al diseño del resto de dimensiones. En el Anexo 5 se muestra el checklist completo.

9.1.3. Procedimiento

El procedimiento seguido en el presente estudio para la valoración de los pacientes clínicos es el mismo que el expuesto en el Estudio 3. En este caso, el CDS lo aplicaba el médico valorador, y no el autor, una vez había finalizado la valoración clínica completa con el paciente. Dichos médicos recibieron una formación previa donde se les explicó detenidamente y con ejemplos prácticos cómo cumplimentar el checklist.

Los simuladores análogos fueron evaluados en las instalaciones de la universidad mediante un procedimiento equivalente al de los pacientes clínicos. Se les indicó que debían simular haber sufrido un accidente que les había ocasionado una cervicalgia o SLC y tenían que convencer al evaluador del estudio (presentado como un Médico Interno Residente, MIR, de traumatología) de que realmente padecían la condición. Como recompensa por participar en el experimento, se les ofreció un bono de 0,5 puntos en la asignatura en la que se había presentado el estudio. Siguiendo el consejo de Rogers (2008) y para lograr que los alumnos recrearan con la mayor fidelidad posible el contexto real, en vez de premiar la participación se penalizó el fallo. Se les advirtió que solo aquellos que lograsen fingir correctamente el SLC y engañar al evaluador recibirían el bono de recompensa. Las instrucciones que recibieron fueron diseñadas siguiendo los requisitos propuestos por Rogers y Cruise (1998) de claridad, especificidad, contextualización y motivación. Una vez elaboradas, fueron evaluadas aplicando los aspectos clave del diseño

de análogos propuestos por Rogers (2008) y operativizados por Gonzalez Ordi et al. (2012).

El investigador encargado de evaluar este grupo era miembro del equipo de investigación del autor, ciego a las condiciones y objetivos del estudio y con experiencia previa en el contexto clínico y forense. Se le explicó que tenía que comprobar una herramienta diseñada por el equipo con personas de la universidad que habían sufrido en un periodo relativamente reciente un accidente de circulación que había ocasionado un SLC. Fue instruido por el mismo doctor que valoró a los pacientes clínicos para aplicar exactamente el mismo procedimiento de valoración del daño corporal. Al finalizar la sesión, se les aplicó el manipulation check detallado con anterioridad.

9.1.4. Análisis de datos

Los datos fueron codificados y analizados en una base de datos con etiquetas genéricas (Grupo 1 en lugar de pacientes clínicos y Grupo 2 en lugar de simuladores) por una persona ajena al estudio, miembro del equipo de investigación, evitando así cualquier sesgo derivado del reconocimiento de los grupos por parte del analista. Todos los análisis estadísticos se realizaron con el paquete estadístico SPSS en su versión 21.

Para decidir que test estadístico debía ser empleado para analizar la diferencia de medias entre los dos grupos incluidos se comprobó la normalidad de la distribución con el test Shapiro-Wilk y la homogeneidad de la varianza de las variables utilizadas con el test de Levene. Debido a que la distribución de la muestra era normal, dichas diferencias fueron analizadas mediante *t* de Student. Para examinar en qué medida los simuladores podían ser clasificados en función de aquellos indicadores que arrojaron resultados significativos ($p < .05$) utilizamos una regresión logística binaria con método Wald hacia adelante. Finalmente, para valorar el tamaño del efecto se utilizó la *d* de Cohen, calculada mediante las medias y las desviaciones estándar de los indicadores de los dos grupos. Al

igual que en los estudios anteriores, se utilizó la categorización que Rogers, Sewell, Martin y Vitacco (2003) propusieron específicamente para la valoración de la simulación (Moderado ≥ 0.75 ; Grande ≥ 1.25 ; Muy grande ≥ 1.50).

9.1. Resultados

En la Tabla 21 se presentan los factores del accidente. La mayoría de los accidentes ocurrieron de día (70.6%), en ciudad (68.6%) y se produjeron por un alcance trasero (69.3%) mientras el conductor mantenía la mirada al frente (80.4%). Gran parte de los pacientes afirman que llevaban abrochado el cinturón de seguridad (94.1%). Predominan los pacientes que ocupaban el asiento trasero (43.1%), seguido del delantero (32.7%). Finalmente, la mayoría de los coches sufrieron desperfectos de gravedad media, reparables en su totalidad (71.2%). Un 2.6% sufrió daños irreparables.

La gran mayoría de la muestra no había solicitado la baja laboral (85%), a pesar de referir dolor o incomodidad en el puesto de trabajo. Para valorar si el paciente tenía o no abogado, se tuvo en cuenta si respondía de forma afirmativa (90.2%) o si refería no saberlo con seguridad (19.8%). Prácticamente todos los pacientes incluidos en el estudio llevaban padeciendo la patología entre 4 y 8 semanas (95.6%).

Tabla 21. Factores relacionados con el accidente

	n	%
Momento del día en el que ocurrió el accidente		
Durante el día	108	70.6
Durante la noche	40	26.1
Al amanecer	3	2.0
Al anochecer	2	1.3
Asiento ocupado		
Conductor	37	24.2
Copiloto	50	32.7
Asiento trasero	66	43.1
Localización del impacto		
Frontal	34	22.2
Trasero	106	69.3
Lateral	13	8.5
Tipo de carretera		
Rural	16	10.5
Ciudad	105	68.6
Autovía/autopista	32	20.9
Asiento de seguridad abrochado		
No	9	5.9
Si	144	94.1
Posición de la cabeza		
Mirada al frente	123	80.4
Girada hacia la izquierda	16	10.5
Girada hacia la derecha	14	9.2
Estado del coche		
Daño irreparable	4	2.6
Daño serio	18	11.8
Daño moderado	109	71.2
Daños menores/leves	22	14.4

En segundo lugar, se comprobó si existían diferencias significativas entre grupos. Los resultados, expuestos en la Tabla 22, muestran que dieciocho de los veintidós indicadores incluidos en el estudio son capaces de discriminar entre los dos grupos. En la

citada tabla se puede observar que el grupo de simuladores análogos obtiene puntuaciones más elevadas en los indicadores, por lo que presenta un mayor número de los mismos que el grupo de pacientes clínicos. Solo los indicadores I3 Falsa atribución a síntomas reales, $t(239)=-1.16, p=0.14$; I7 Culpa a tercera persona del accidente, $t(231)=1.56, p=0.08$; I18 Resistencia a la evaluación, $t(176)=1.17, p=0.24$; y I19 Énfasis en los aspectos negativos de los síntomas, $t(242)=1.42, p=0.15$, no han sido capaces de discriminar entre los dos grupos. Los indicadores con mayor capacidad de discriminación (magnitud del efecto alta, Cohen's $d > .80$) han sido I12 Síntomas extremos, $t(176)=15.85, p=0.00, d=1.61$; I16 Incremento exponencial del dolor que no corresponde con el curso habitual de la patología, $t(216)=11.57, p=0.00, d=1.43$; I8 Pronóstico de mejora negativo, $t(154)=11.29, p=0.00, d=1.43$; I17 Incapacidad para controlar los síntomas, $t(207)=11.05, p=0.00, d=1.38$; I20 Ausencia de mejoría, $t(214)=9.32, p=0.00, d=1.16$; I9 Confirmación de muchos síntomas, $t(186)=8.68, p=0.00, d=1.08$; I15 Pérdida sensibilidad dermatoma no afectado, $t(148)=7.51, p=0.00, d=0.95$; IA1 Percepción de situación de enfermedad, $t(213)=7.69, p=0.00, d=0.95$; e I13 Sensibilidad extrema al dolor, $t(121)=9.35, p=0.00, d=0.94$. Cinco de estos nueve indicadores pertenecen a la Dimensión 2 – Actitud ante la situación de enfermedad, que ha sido la que ha obtenido mayor poder de discriminación ($d=1$), el resto pertenece a la Dimensión 5 – Inconsistencias, segunda en capacidad de discriminación ($d=0.85$). La Dimensión 4 – Causa, ha sido la que menor poder discriminativo ha presentado ($d=0.3$). Finalmente, utilizamos un análisis de regresión logística binaria (Tabla 23) para identificar los indicadores que mejor discriminaban entre grupos, incluyendo solo aquellos que habían arrojado resultados significativos. El análisis concluyó en ocho pasos, con una Nagelkerke R^2 de .858, y clasificando correctamente a un 91.4% de la muestra, $\chi^2(1, N = 153) = 4.59, p < .05$.

Tabla 22. Resultados prueba Student's T entre Pacientes clínicos y Simuladores instruidos.

	<i>M (DT)</i> <i>P.C.</i>	<i>M (DT)</i> <i>SIM.</i>	<i>t</i>	<i>gl</i>	<i>d</i>	<i>p</i>
I1 Detalles superfluos e inusuales	1.07(1.33)	.45(.90)	4.49	271	.51	.000
I2 Estado mental propio	1.12(1.38)	.69 (1.03)	3.39	271	.34	.013
I3 Falsa atribución síntomas reales	.12(0.52)	.20(.60)	1.16	271	.15	.244
I4 Falta de cooperación	.07(.24)	.58(.88)	6.20	271	.99	.000
I5 Actitud hostil o defensiva	.09(.29)	.26(.60)	3.41	271	.41	.001
I6 Culpa tercera persona síntomas	.81(1.21)	1.27(1.40)	2.56	271	.36	.011
I7 Culpa tercer persona accidente	1.87(1.09)	1.98(1.32)	1.56	271	.09	.084
I8 Pronóstico de mejora general negativo	.59(.52)	1.53(1.21)	11.29	271	1.18	.000
I9 Confirmación muchos síntomas	.19(.57)	1.04(.95)	8.68	271	1.20	.000
I10 Síntomas improbables	.14(.49)	.75(.99)	6.16	271	.89	.000
I11 Contradicción patrón síntomas	.05(.26)	.52(.82)	6.07	271	.84	.000
I12 Síntomas extremos	.18(.60)	1.60(1.09)	12.85	271	1.81	.000
I13 Sensibilidad extrema al dolor	1.01(.81)	1.89(1.03)	9.35	271	1.00	.000
I14 Alta sensibilidad zona incorrecta	1.12(.47)	1.53(.85)	4.74	271	.67	.000
I15 Perdida sensibilidad dermatoma no afectado	1.10(.39)	1.83(1.01)	7.51	271	1.13	.000
I16 Incremento exponencial dolor	.38(.76)	1.66(1.01)	11.57	271	1.52	.000
I17 Incapacidad control de síntomas	.53(.69)	1.70(.98)	11.05	271	1.48	.000
I18 Resistencia evaluación	1.02(.13)	1.05(.25)	1.17	271	.17	.244
I19 Énfasis aspectos negativos síntomas	1.13(1.12)	1.34(1.26)	1.42	271	.18	.156
I20 Ausencia de mejoría	.76(.73)	1.77(.99)	9.32	271	1.23	.000
IA1 Percepción situación enfermedad	.99(.49)	1.55(.67)	7,69	271	1.01	.000
IA2 Identidad de síntomas	1.60(.57)	1.17(1.02)	4.17	271	.52	.000

Abreviaturas: P.C.: Pacientes clínicos; Sim.: Simuladores; *M*: Media; *DT*: Desviación estándar; *d*: *d* de Cohen; *gl*: Grados de libertad.

Tabla 23. Resultados de la regresión logística binaria.

	<i>B</i>	<i>SE</i>	<i>X² Wald</i>	<i>p</i>	<i>OR</i>	<i>95% CI</i>
I4	2.31	.70	10.73	.001	10.10	2.52-40.32
Falta de cooperación						
I5	1.07	.57	3.51	.061	2.93	.95-9.04
Actitud hostil o defensiva						
I8	1.56	.31	25.21	.000	4.75	2.58-8.74
Pronostico de mejora general negativo						
I9	.61	.35	3.33	.082	1.84	.92-3.66
Confirmación muchos síntomas						
I11	1.50	.685	7.21	.007	4.50	1.50-13.53
Contradicción patrón síntomas						
I12	.93	.26	12.12	.000	2.55	1.50-4.32
Síntomas extremos						
I13	1.63	.61	7.17	.000	5.15	1.55-17.09
Sensibilidad extrema al dolor						
I15	.93	.44	4.32	.037	2.54	1.05-6.11
Perdida sensibilidad dermatoma incorrecto						
Constante	-16.66	2.33	50.99	.00	.00	

9.3. Discusión

El principal objetivo que se perseguía mediante el presente estudio era comprobar la capacidad de una serie de marcadores de simulación, extraídos de bibliografía relevante en la materia, para discriminar entre pacientes clínicos con sintomatología genuina y simuladores instruidos. Mediante nuestros resultados hemos podido observar que la gran mayoría de indicadores han sido capaces de diferenciar entre grupos, siendo los simuladores análogos los que obtenían una puntuación más elevada en los mismos. Este estudio representa una de las pocas investigaciones que evalúa indicadores de simulación en SLC y creemos que los hallazgos obtenidos pueden ser de utilidad. Pasamos a continuación a describir como los indicadores se han manifestado en cada grupo:

La dimensión 2 “Actitud ante la situación de enfermedad” ha sido la que mayor capacidad de discriminación ha presentado. Los resultados en esta dimensión indican que los simuladores expresan una actitud ante la situación de enfermedad mucho más negativa que los pacientes clínicos. La gran mayoría de simuladores reportan una vivencia “muy molesta”, suelen describir su experiencia de la enfermedad como un proceso severo y altamente disruptivo, mientras que los pacientes clínicos reconocen que, si bien su calidad de vida se ha visto alterada de forma significativa, no se trata de una situación excesivamente traumática (IA19). Los resultados obtenidos en esta dimensión también indican que los simuladores manifiestan baja capacidad de control de síntomas (I17), expresando en la gran mayoría de los casos que son incapaces de disminuir el dolor que le produce la lesión, ni siquiera mediante el uso de medicamentos. Afirman además haber experimentado un aumento exponencial del dolor desde el accidente que no corresponde con lo esperado en SLC (I16). Los pacientes clínicos, por el contrario, refieren una percepción de control mucho más positiva. Consideran que pueden paliar la sintomatología y, a no ser que realicen algún movimiento brusco, son capaces de mitigar el dolor a niveles tolerables o hacerlo desaparecer prácticamente en su totalidad.

En la misma línea, el grupo de simuladores reporta una visión mucho más negativa del pronóstico de mejora al afirmar que los síntomas van a durar mucho tiempo (I18). Tienden a expresar una visión catastrofista y pesimista del proceso vivido donde no han experimentado mejoría alguna (I20). Los pacientes clínicos, por su parte, ofrecen una visión mucho más optimista del proceso de recuperación. La gran mayoría ha experimentado una mejoría notable a lo largo de las dos primeras semanas y consideran que los síntomas padecidos desaparecerán relativamente rápido. Ante la pregunta “cómo cree que se encontrará en dos meses” la gran mayoría contesta que “curado” o “completamente curado”.

Se evidenció además que los simuladores tienen un discurso mucho más generalista e inespecífico, tanto con los síntomas como con las consecuencias. En general, los simuladores ofrecen una descripción vaga e imprecisa de la sintomatología padecida (AI2), son capaces de enumerar los síntomas que presuntamente sufren, pero no hacen referencia a lugares o áreas específicas. De la misma forma, refieren que experimentan una vivencia muy negativa, pero no ofrecen una descripción detallada de la misma.

A la vista de estos resultados, entendemos que, bajo el punto de vista del simulador, el hecho de que estos admitan una posible mejoría o un pronóstico positivo podría influir en la valoración del médico, el cual otorgaría una menor gravedad y, por lo tanto, menor indemnización. Los hallazgos obtenidos en esta dimensión son consistentes con los aportados anteriormente por Linnell y Easton (2006) en un estudio similar, donde se observaba que los simuladores reportaban menos control, más tiempo de recuperación y más consecuencias severas.

Los resultados obtenidos en esta dimensión deben ser considerados con precaución. Estudios previos han demostrado que diversos factores emocionales y psicológicos juegan un papel importante en el desarrollo del mismo, afectando, por ejemplo, a la percepción sobre la severidad del dolor o la discapacidad funcional (Carroll *et al.*, 2009; Falla *et al.*, 2016; Holm *et al.*, 2008). Debe considerarse la posibilidad de que una exposición negativa del proceso de enfermedad sea resultado de alguna alteración psicológica o emocional provocada por el accidente o incluso previa al mismo. En este sentido, Ferrari *et al.* (1999) señalaban otros dos elementos a tener en cuenta, uno es el beneficio que supone asumir el rol de enfermo, es posible que el accidente sea una solución para los problemas vitales del paciente y ofrezca de forma no intencional una visión más negativa del proceso de enfermedad, otro es la información sobre la enfermedad que se proporciona al paciente y su influencia en la conducta ante la misma.

Cabe la posibilidad de que los pacientes que muestran una actitud positiva hayan sido informados adecuadamente sobre el SLC o, por el contrario, la ausencia de información sobre el pronóstico de mejora o la gravedad de los síntomas desemboquen en una actitud más negativa.

Con todo ello, consideramos que esta dimensión puede ser de utilidad para la valoración de la simulación en el contexto médico-legal y creemos que sería interesante de cara a futuras investigaciones evaluarla junto a los factores señalados.

La dimensión 5, “Inconsistencias”, ha sido la segunda con mayor capacidad de discriminación. Los resultados obtenidos indican que los simuladores suelen confirmar la existencia de muchos síntomas (I9), lo que se conoce como tendencia al symptom over-endorsement (McDermott y Feldman, 2007). En la gran mayoría de ocasiones los describían formando un patrón incoherente (I11) o afirmaban padecer síntomas que no pertenecían a la patología (I10). Se ha observado también que los simuladores manifiestan que los síntomas padecidos son de extrema gravedad (I12). Refieren padecer un dolor realmente elevado (8-9 en VAS) que no mejora con el paso del tiempo o la rehabilitación (I16), mientras que los pacientes clínicos reportan un dolor de gravedad intermedia (5-6 en VAS) que fluctúa en función del efecto de la medicación.

Se puede observar también que las incoherencias en la expresión de la sensibilidad provocada por la lesión es un indicador eficaz (I13, I14, I15). Los simuladores ofrecen una descripción polarizada, o sensibilidad extrema o pérdida total de la misma, en ocasiones ambas en el mismo dermatoma. Estos resultados coinciden con lo aportado con anterioridad en la bibliografía sobre simulación, donde se señala que la valoración de las inconsistencias (internas y externas) es una herramienta útil y sencilla para detectar el engaño en el contexto médico legal (Rogers y Bender, 2018). Sin embargo, es necesario recordar que no toda inconsistencia es sinónimo de engaño, los evaluadores deben evitar

la “trampa de la inconsistencia” y valorar cuidadosamente si las incoherencias observadas son intencionales o fruto de un descuido o confusión (Rogers y Shuman, 2000).

En los resultados obtenidos en la dimensión 3, “Actitud ante la evaluación”, puede observarse que los simuladores presentan una actitud algo más distante y desconfiada que los pacientes clínicos, similar al perfil del sujeto deshonesto presentado en la bibliografía de credibilidad del testimonio (DePaulo *et al.*, 2003; Vrij, 2008). La falta de cooperación observada en el grupo de simuladores análogos se manifiesta exclusivamente a nivel verbal mediante una narración más desgana y rígida y un trato más seco con el evaluador. Por otro lado, en el grupo de pacientes clínicos se observa un esfuerzo por establecer una relación cordial y cercana con el médico valorador. Consideramos que la posible explicación a estos resultados es la visión que el evaluado tiene del médico. El paciente genuino ve al médico valorador como un médico común que puede ayudarle a solucionar su problemática, por lo que tendrá una actitud más positiva ante la evaluación. Por otro lado, el simulador puede tener una visión del valorador como un “evaluador hostil” que podría destapar su “engaño”. No parece probable que los simuladores manifiesten una actitud completamente hostil, ya que le resultaría perjudicial de cara al informe final, de manera que su actitud ante la evaluación será más bien distante y precavida. Al igual que señalábamos en la dimensión 2 los resultados deben ser tomados con precaución y el evaluador debe considerar que la actitud hostil mostrada por el paciente puede deberse a diferentes variables, como la exposición al contexto médico legal o la ansiedad provocada por el proceso de enfermedad (Easton y Akehurst, 2011).

Finalmente, consideramos que la dimensión 1 “Detalles inusuales y pensamientos sobre el evento” y la dimensión 4 “Causa” son las menos útiles para evaluar simulación en SLC. En la Dimensión 1 se han observado que los pacientes aportan más detalles superfluos e inusuales y hacen más alusiones al estado mental propio, sin embargo, no

creemos que sea de especial utilidad ya que, habitualmente, los pacientes evaluados por whiplash, simuladores o no, han sufrido un accidente y por tanto se van a referir a algo que realmente han vivido. Estos dos indicadores han sido obtenidos del CBCA y, si bien su eficacia en otros contextos es elevada (Amado *et al.*, 2015, 2016), creemos que una invención completa del evento resulta altamente improbable en el contexto médico legal. Es probable que los resultados obtenidos en la dimensión 1 se deben en gran parte al empleo de simuladores análogos, los cuales han tenido que inventar completamente el accidente. De la misma forma, los resultados similares obtenidos en el estudio de Akehurst *et al.* (2015) pueden deberse a que los alumnos simuladores no habían pasado por el *cold pressor procedure* e inventaron completamente la vivencia.

Al sintetizar la evidencia anteriormente expuesta, puede observarse que los simuladores tienen a sobreestimar la gravedad del SLC y ofrecen una visión mucho más negativa y catastrofista de la patología que los pacientes genuinos. Estos resultados son compatibles con el perfil del simulador que se ha expuesto habitualmente en la bibliografía (ver Etherton, 2014; McDermott y Feldman, 2007; Rogers y Bender, 2018).

Creemos que la gran mayoría de los indicadores expuestos en este estudio pueden ser de utilidad para el contexto forense, pero consideramos necesario que, de cara a futuras investigaciones, se repliquen estos resultados con un grupo de simuladores reales para garantizar así su validez. De la misma forma, sería necesario localizar más indicadores que sean capaces de discriminar correctamente entre los dos grupos. En este sentido, creemos que las variables que afectan a la expresión de la severidad del dolor, como la catastrofización del mismo o el miedo al movimiento, pueden resultar de interés. Debido a que el dolor se ve condicionado por dichas variables, conocer la interacción entre las mismas podría ser de interés para tratar de encontrar inconsistencias en la manifestación del SLC del simulador.

Para finalizar, consideramos necesario señalar que el CDS es una herramienta diseñada para ser incorporada en un protocolo de screening formado por más instrumentos. No debe usarse como la única herramienta de evaluación porque “puede ayudar a definir un punto final (es decir, exageración de los síntomas), pero no dice nada sobre los factores externos e intrínsecos que contribuyen a este punto final” (Merckelbach *et al.*, 2019, p. 322). La simulación es una posible explicación de la exageración de los síntomas, pero no es la única (Merckelbach *et al.*, 2019). El profesional debe enfocar la evaluación considerando todas las posibles hipótesis de tal exageración y aplicar un protocolo de evaluación que sea capaz de analizarlas.

10. Bloque 3 - Estudio 6. Capacidad discriminativa del Inventario Estructurado de Simulación de Síntomas (SIMS)

El SIMS es un autoinforme compuesto por 75 preguntas dicotómicas diseñado para detectar simulación de sintomatología psiquiátrica y problemas cognitivos. Fue construido bajo la idea de que los simuladores tienden a expresar una mayor cantidad de síntomas y/o afirman padecer síntomas raros, atípicos o extremos (Smith y Burger, 1997; van Impelen *et al.*, 2014). Los resultados del instrumento se plasman en una escala general de riesgo de simulación y cinco subescalas que tienen como finalidad indicar al evaluador en que área puede estar ocurriendo la simulación: Psicosis, Deterioro neurológico, Trastornos amnésicos, Baja inteligencia y Trastornos afectivos. La publicación original de Smith y Burger (1997) establecía un punto de corte de 14 para la escala general y, posteriormente, se propuso un punto de corte de 16 (Rogers *et al.*, 1996). Diversos trabajos han mostrado que las propiedades psicométricas de la escala son adecuadas (Alpha de Cronbach = .96; sensibilidad = .93; especificidad = .98) y ha resultado ser eficaz diferenciando entre honestos y simuladores (Rogers y Bender, 2018; van Impelen *et al.*, 2014).

El SIMS también se utiliza en el contexto médico-legal por profesionales contratados por las aseguradoras para valorar a personas que han sufrido un accidente de circulación (Motor Vehicle Accident, MVA de ahora en adelante) (Cernovsky *et al.*, 2019a, 2019b). Esta práctica resulta problemática debido a que no hay evidencia científica contrastada que valide el instrumento con dicha población. Como indican los anteriores autores, en estas condiciones se corre el riesgo de clasificar erróneamente a los pacientes genuinos como simuladores (falso positivo) generando un impacto iatrogénico en los mismos. De la misma forma, al desconocer la eficacia del instrumento en la citada

población, puede ocurrir el efecto contrario y no detectar a posibles simuladores (falso negativo), con las consecuencias que esto conlleva.

Resulta fundamental que los instrumentos o técnicas que se utilicen en el contexto legal, en especial en aquellos en los que se otorga compensación económica, tengan respaldo empírico y se conozcan sus propiedades psicométricas, en especial la ratio de falsos positivos (Arce, 2017; Cernovsky *et al*, 2019a).

Para dar respuesta a la necesidad expuesta, en el presente estudio se persigue el siguiente objetivo:

- **Objetivo específico 3.3.** Valorar la capacidad discriminativa del Inventario Estructurado de Simulación de Síntomas (SIMS) en el contexto médico-legal.

10.1. Metodología

10.1.1. Participantes

Para el presente estudio se contó con una muestra preliminar de 653 personas divididas, inicialmente, en 3 grupos (paciente clínico, población general y simuladores instruidos). Como criterios de inclusión se utilizaron inicialmente: Criterio a1. Necesidad de firmar el consentimiento informado, Criterio b1. Superar un manipulation check. Criterio c1. No haber estado envuelto en un proceso de compensación médico-legal y Criterio d1. No presentar ninguna alteración psicológica o médica que pueda alterar las respuestas de los instrumentos. (2) Pacientes clínicos: Criterio a. Necesidad de firmar el consentimiento informado, Criterio b2. La persona estaba siendo evaluada para valorar el daño físico y psicológico producido tras un accidente de circulación, Criterio c2. Era mayor de edad (≥ 18 años), Criterio d2. Se catalogaba como paciente clínico. El incumplimiento de los criterios en cada grupo suponía la exclusión del estudio. Debido a

que el número de pacientes descartados por el criterio d2 resultó significativamente más elevado que los descartados en el grupo A ($n=54$), se decidió crear un cuarto grupo cuyos criterios de inclusión fueron: (3) Probable simulador: Criterio a. Necesidad de firmar el consentimiento informado, Criterio b3. La persona estaba siendo evaluada para valorar el daño físico y psicológico producido tras un accidente de circulación, Criterio c3. Era mayor de edad (≥ 18 años), Criterio d3. Se catalogaba como probable simulador.

Así, para el presente estudio se descartaron únicamente 3 simuladores instruidos por no cumplir el criterio c1, siendo la muestra final de 650 participantes, con una media de edad de 30.29 ($DT=11.27$) y un rango de 18 a 65. La división por grupos fue: población general con 200, 84 hombres (42%) y 116 mujeres (58%), con una media de edad de 24.55 ($DT=7.81$); simuladores instruidos, compuesto por 201 participantes, 95 hombres (47.26%) y 106 mujeres (52.74%), con una media de edad de 23.92 ($DT=7.51$); población clínica, que incluyó 195 pacientes, 102 hombres (52.31%) y 93 mujeres (47.69%), con una edad media de 37.45 ($DT=10.97$); y 54 pacientes sobredimensionados (overreporting patients) 25 hombres (46.3%) y 29 mujeres (53.7%), con una edad media de 35.27 ($DT=11.21$). Los dos últimos grupos estuvieron formados por pacientes pertenecientes al grupo B.

24 pacientes atendidos se negaron a participar tras la explicación inicial de los objetivos del estudio, no siendo posible recoger ningún dato más allá del sexo (10 mujeres y 14 hombres) y el motivo de la no participación (16 de ellos afirmaron tener prisa y 8 no querían someterse a ningún tipo de evaluación que no fuera impuesta por las compañías aseguradoras).

10.1.2. Instrumentos

Structured Inventory of Malingered Symptomatology (SIMS). El Inventario Estructurado de Simulación de Síntomas (SIMS; Smith y Burger, 1997) es un screening autoinformado de 75 ítems tipo verdadero y falso que evalúa la exageración o simulación de síntomas neurológicos y psiquiátricos. Ofrece un resultado que indica el riesgo general de simulación y cuenta además con cinco subescalas (Psicosis, Deterioro Neurológico, Trastornos Amnésicos, Baja Inteligencia y Trastornos Afectivos) que aportan información sobre el área específica en la que se detecta la simulación. Los puntos de corte recomendados son 14 y 16. Se utilizó la versión española de la escala, adaptada por González-Ordi y Santamaría (2009). Los estudios realizados para dicha adaptación en un total de 1.005 sujetos participantes indican que el inventario posee un Alfa de Cronbach de .94 para el total de la prueba. En el presente estudio se obtuvo un Alfa de Cronbach de .86 para el total de la muestra. Utilizamos los puntos de corte recomendados por los autores de las dos versiones, 14 de la inglesa (Smith y Burger, 1997) y 16 de la española (González-Ordi y Santamaría Fernández, 2009).

Se aplicó también el mismo Manipulation check utilizado en los anteriores estudios.

10.1.3. Procedimiento

Se utilizó un diseño de simulación (transversal) con los cuatro grupos expuestos con anterioridad. Los pacientes clínicos incluidos en el presente estudio forman parte del grupo B. El procedimiento seguido para su obtención puede consultarse en las páginas 150 y 151.

Para clasificar a los pacientes clínicos en los dos grupos señalados (población general y simuladores) se utilizó como criterio la existencia de evidencia contradictoria

en la evaluación médico-legal. Para valorar dicho criterio, los tres expertos en medicina legal que participaron en el estudio analizaban cada caso y valoraban si la manifestación verbal del paciente sobre su sintomatología era coherente con los indicadores anatómico-estructurales observados u esperados en la condición (severidad del dolor, localización del dolor, rango de movimiento cervical, balance articular activo/pasivo, etc.). Una vez evaluados, los pacientes se catalogaron como 1. Paciente clínico (sin evidencia de fingimiento), 2. Paciente sobredimensionado (con evidencia de fingimiento y sin incentivo externo), 3. Probable simulador (con evidencia de fingimiento y con incentivo externo) o 4. Dudoso. Los desacuerdos fueron resueltos por consenso.

Para el grupo experimental, los participantes fueron reclutados en la Universidad de Murcia durante los años 2018 y 2019. Se inscribieron por correo electrónico tras la presentación del estudio y fueron asignados de forma aleatoria a las condiciones planteadas: simuladores y población general. Posteriormente, se le facilitaron unas instrucciones acordes a su rol y se les citó en las dependencias de la Universidad.

A los simuladores se les solicitó que fingieran daños físicos y psicológicos sufridos tras un supuesto accidente de circulación, con el objetivo de obtener una compensación económica. Como motivación se les ofreció una recompensa interna (ser capaz de engañar al test) y externa (puntuación extra en la nota final de una asignatura). Siguiendo la metodología de Rogers y Bender (2018), se penalizó el fallo, de manera que solo aquellos que cumplimentaran la escala acorde al rol asignado, recibirían el bono. Las instrucciones que recibieron fueron diseñadas acorde a los requisitos propuestos por Rogers y Cruise (1998) de claridad, especificidad, contextualización y motivación. Los participantes fueron sometidos al manipulation check expuesto con anterioridad. Aquellos participantes que no alcanzaron una puntuación ≥ 4 en todas las variables de dicho manipulation check fueron excluidos del experimento (criterio b2). Para evitar el

problema del coaching (Van Impelen *et al.*, 2014), no se ofreció ningún tipo de información o entrenamiento sobre cómo simular, únicamente se hizo hincapié en la importancia de que fuera coherente con lo esperable en sujetos que han sufrido un MVA.

Por otro lado, a los participantes asignados a la condición de población general se les solicitó que respondieran la escala SIMS con instrucciones estándar (responde sinceramente). Se aplicaron las mismas condiciones de recompensa externa y manipulation check que al anterior grupo (simuladores). Las instrucciones proporcionadas pueden consultarse en el apartado “Anexos”.

10.1.4. Análisis de datos

El grado de acuerdo o consistencia de los expertos en la clasificación de los pacientes se analizó con el Coeficiente de Correlación Intraclase (ICC, Modelo 2.1). Se calcularon las medias, las desviaciones típicas y las frecuencias para cada una de las variables estudiadas. Para decidir que estrategia de análisis de datos utilizar, estudiamos la normalidad de la distribución con el test Shapiro-Wilk y la homocedasticidad con el test de Levene. Debido a que se cumplían los dos principios, se utilizó estadística paramétrica, analizando las diferencias entre los grupos con el análisis de la varianza de un factor (ANOVA). La magnitud del efecto fue calculada utilizando el estadístico d de Cohen. Debido a que los valores originales propuestos por Cohen (1988) son orientativos y deben ser adaptados al contexto estudiado (Ato *et al.*, 2013), se utilizó la categorización que Rogers *et al.* (2003) propusieron específicamente para la valoración de la simulación (Moderado ≥ 0.75 ; Grande ≥ 1.25 ; Muy grande ≥ 1.50). Para calcular la capacidad discriminativa del SIMS, se calculó la curva característica operativa del receptor (curva COR), obteniendo además el área bajo la curva (AUC) y el error estándar de dicha área (SE_{AUC}). Una vez obtenida la curva COR, se calculó para los puntos de corte 14 y 16 la sensibilidad, la especificidad, razón de verosimilitud positiva y negativa ($L+$ y $L-$) y el

poder predictivo positivo/negativo (PPP y PPN) para una prevalencia del 10%, 30% y 50%.

Debido a que la comparación de mayor relevancia en los diseños de simulación es la realizada entre el grupo de pacientes clínicos y el grupo de simuladores Rogers y Bender (2018), se expondrá la curva ROC con variable de estado “simuladores” con la finalidad de tomar una decisión sobre un posible punto de corte óptimo para el presente contexto.

El análisis de la capacidad discriminativa anteriormente descrito se llevó a cabo únicamente con la puntuación general del SIMS ya que, como advierten los autores del instrumento, las subescalas no están diseñadas para detectar simulación y solo sirven para orientar al profesional sobre qué tipo de psicopatología puede estar fingiendo el evaluado una vez que se ha excedido el punto de corte (Smith y Burger, 1997). No obstante, se calcularon las medias, las desviaciones estándar y las diferencias entre grupos de dichas subescalas con finalidad descriptiva.

10.2. Resultados

10.2.1. Clasificación de los grupos y variables médico-legales del grupo de pacientes clínicos y sobredimensionados

Los resultados del análisis del grado de acuerdo entre los tres evaluadores para la clasificación de los grupos muestran un ICC (2,1) de .96. Se encontraron 7 desacuerdos, resolviéndose todos mediante consenso. En los 7 casos dos de los expertos consideraban que era un probable simulador, mientras que el tercero que era un caso dudoso.

Todos los participantes incluidos en los grupos de pacientes clínicos y pacientes sobredimensionados eran sometidos a una valoración clínica completa tras sufrir un

accidente de circulación. La valoración fue realizada tras una media de 38.24 días ($DT=9.25$) desde el accidente, con un rango de 26 a 59 días. No se encontraron diferencias significativas en la gravedad de la sintomatología en función del tiempo transcurrido tras el accidente ($p=.86$). El 100% de los participantes afirmaba padecer dolor cervical, con una media de gravedad de 7.56 ($DT=1.23$) en el grupo de pacientes sobredimensionados y 5.15 ($DT=.92$) en el de pacientes clínicos. En lo relativo al diagnóstico realizado por los profesionales, el 95.5% de los pacientes clínicos fue diagnosticado con SLC, el 4.5% restante también estaba diagnosticado de SLC, pero además referían padecer dolor lumbar. En lo relativo a los pacientes sobredimensionados, el diagnóstico de SLC lo presentaban el 89% y el de SLC con dolor lumbar el 8%. El 3% restante fue diagnosticado desde los servicios de urgencias públicos como “cervicalgia postraumática”, pero el médico que los valoró en la clínica confirmó que el cuadro clínico que presentaban era compatible con SLC. En ninguno de los casos los profesionales otorgaron un grado al whiplash, no obstante, ninguno de los pacientes padecía signos neurológicos, a excepción de 6 pacientes sobredimensionados que afirmaban padecer mareos leves, 3 de ellos al mover la cabeza de forma brusca. De la misma forma, no se observó ninguna fractura ni lesión musculo-esquelética de gravedad.

En lo relativo al grupo de simuladores instruidos, la gran mayoría (64%) utilizó páginas no científicas de internet (p.e. de grupos de abogados o de aseguradoras) como fuente de información, localizadas a través de un buscador utilizando términos como “consecuencias de accidente de circulación”, “accidente de tráfico indemnización”, “daños comunes de accidentes de circulación”. Del 36% restante, un 30% utilizó una modalidad combinada de buscar en internet las principales consecuencias y extender la información con bibliografía especializada, y el 6% afirmó haber consultado únicamente bibliografía especializada. En todos los casos, la citada bibliografía especializada

consistió en manuales disponibles en la biblioteca de la Universidad. No se especificó la forma acceso a los mismos.

10.2.2. Puntuaciones del SIMS y comparación entre grupos

Como puede observarse en la Tabla 24, el ANOVA de un factor indicó que había diferencias significativas entre grupos para todas las subescalas (Baja inteligencia, $F(2,646) = 166.27, p < .001$; Trastornos afectivos, $F(2,646) = 132.91, p < .001$; Deterioro neurológico, $F(2,646) = 148.46, p < .001$; Psicosis, $F(2,646) = 2994.21, p < .001$; Trastornos amnésicos, $F(2,646) = 723.89, p < .001$). En los cuatro grupos, la subescala de trastornos afectivos obtuvo la puntuación más elevada, mientras que la subescala Psicosis obtuvo la más baja, registrando valores de 0 en los grupos de población general, sobredimensionados y pacientes clínicos.

Para la PD de la escala, la puntuación media fue de 4.91 ($DT = 2.83$) en el grupo de población general; 9.01 ($DT = 2.95$) en el de pacientes clínicos; 12.87 ($DT = 4.11$) en el de probables simuladores y 16.74 ($DT = 6.49$) en el de simuladores instruidos. El análisis ANOVA indicó que existían diferencias entre los grupos, $F(2, 646) = 253.09, p < .001$. Los resultados obtenidos mostraron que los simuladores producían puntuaciones más elevadas que los otros tres grupos ($M = 16.74, DT = 6.49$), sobrepasando los puntos de corte propuestos por los autores del instrumento. Las magnitudes del efecto obtenían fueron más elevadas entre los simuladores instruidos/probables y la población general ($d_1 = 2.36$ y $d_3 = 2.23$, respectivamente) entre los simuladores instruidos/probables y los pacientes clínicos ($d_2 = 1.53$ y $d_4 = .1.09$, respectivamente).

Tabla 24. Resultados de la prueba ANOVA y Magnitudes del efecto para la comparación entre grupos.

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

Abreviaturas: Gen: Población general; Pac: Pacientes clínicos; Pro: Probables simuladores; Sim: Simuladores instruidos; * Subescalas del SIMS d_1 : Magnitud del efecto para la comparación simuladores vs población general; d_2 : Magnitud del efecto para la comparación simulador vs paciente clínico; d_3 : Magnitud del efecto para la comparación sobredimensionados vs población general; d_4 : Magnitud del efecto para la comparación sobredimensionados vs paciente clínico.; d_5 : Magnitud del efecto para la comparación paciente clínico vs población general; d_6 : Magnitud del efecto para la comparación sobredimensionados vs simuladores; 95% IC: intervalo de confianza al 95%. ** $p < .001$; n.a.: no aplicable.

10.2.3. Precisión de la clasificación

La precisión de la clasificación del SIMS puede verse en la Tabla 25. Utilizando el punto de corte 14, el SIMS fue capaz de discriminar entre simuladores instruidos y población general con una sensibilidad del 59.70% y una especificidad del 97.50% y entre simuladores instruidos y pacientes clínicos con una sensibilidad del 59.70% y una especificidad del 78.97%. Por otro lado, discriminó entre probables simuladores y población general con una sensibilidad del 55% y una especificidad del 97.50% y entre pacientes sobredimensionados y pacientes clínicos con una sensibilidad del 55% y una especificidad del 78.97%. De la misma forma la probabilidad de obtener una puntuación mayor en el SIMS fue más elevada en los simuladores instruidos/probables simuladores que en los pacientes clínicos (AUC=.94 y .91 respectivamente) y que en población general (AUC=.76 y .73 respectivamente). En lo relativo a la razón de verosimilitud positiva (L+), se observó que puntuaciones de ≥ 14 en el SIMS eran aproximadamente 22 veces más probables en los dos primeros grupos que en el de población general y aproximadamente 3 veces más probables en los dos primeros grupos que en el de pacientes clínicos.

Los PPP, PPN y el *accuracy index* para una prevalencia del 10 al 50% pueden observarse en la tabla 26. En términos generales, asumiendo una ratio base del 30%, el más probable de ser encontrado en la práctica clínica (Curtis *et al.*, 2019), los resultados indicaron que con un punto de corte de 14 hay una probabilidad del 91% aproximadamente de clasificar a los simuladores y a los probables simuladores si se comparaba con el grupo de población general, y del 54% aproximadamente si se comparaba con el grupo de pacientes clínicos.

Tabla 25. *Estimadores de utilidad con un punto de corte de 14*

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

Abreviaturas: Gen: Población general; Pac: Pacientes clínicos; Pro: Probables simuladores; Sim: Simuladores instruidos; AUC: Área Bajo la Curva; 95% CI: Intervalo de confianza al 95%; SEAUC: Error estándar del área bajo la curva; Se: Sensibilidad; Sp: Especificidad; DOR: Odd Ratio diagnóstica; L+: Razón de verosimilitud positiva; L-: Razón de verosimilitud negativa.

Tabla 26. *Poder predictivo positivo y negativo, y Accuracy Index para una prevalencia del 10% al 50% (Punto de corte de 14)*

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

Abreviaturas: Gen: Población general; Pac: Pacientes clínicos; Pro: Probables simuladores; Sim: Simuladores instruidos; PPP, Poder predictivo positivo; PPN, Poder predictivo negativo; ACC, Accuracy Index.

Al utilizar 16 como punto de corte (Tabla 27) la sensibilidad disminuyó, tanto en la comparación entre simuladores/probables simuladores y población general, como entre simuladores/probables simuladores y pacientes clínicos (55.22% y 51.85 respectivamente para ambos casos), pero la especificidad aumentó (98.50% y 83.08% respectivamente). La precisión diagnóstica y los L+ también aumentaron, manteniéndose más elevados entre simuladores/probables simuladores y población general (L+=39.82 y 34.57 respectivamente) que entre simuladores/probables simuladores y población clínica (L+ 3.26 y 3.06 respectivamente).

Finalmente, los PPP, PPN y el *accuracy index*, expuestos en la tabla 28, indicaron que utilizando el punto de corte 16 se incrementó al 94% aproximadamente la probabilidad de clasificar a los simuladores y a los probables simuladores si se comparaba con el grupo de población general, y al 58% aproximadamente si se comparaba con el grupo de pacientes clínicos.

Tabla 27. *Estimadores de utilidad con un punto de corte de 16*

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

Abreviaturas: Gen: Población general; Pac: Pacientes clínicos; Pro: Probables simuladores; Sim: Simuladores instruidos; AUC: Área Bajo la Curva; 95% CI: Intervalo de confianza al 95%; SEAUC: Error estándar del área bajo la curva; Se: Sensibilidad; Sp: Especificidad; DOR: Odd Ratio diagnóstica; L+: Razón de verosimilitud positiva; L-: Razón de verosimilitud negativa.

Tabla 28. *Poder predictivo positivo y negativo, y Accuracy Index para una prevalencia del 10% al 50% (Punto de corte de 16)*

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

Abreviaturas: Gen: Población general; Pac: Pacientes clínicos; Pro: Probables simuladores; Sim: Simuladores instruidos; PPP, Poder predictivo positivo; PPN, Poder predictivo negativo; ACC, Accuracy Index.

Punto de corte óptimo para discriminar entre pacientes clínicos y simuladores instruidos. Con la finalidad de obtener un posible punto de corte óptimo para comparar entre simuladores y pacientes clínicos, en la Figura 7 se expone la curva ROC utilizando como variable de estado “simuladores”. La sensibilidad disminuye significativamente desde el punto de corte 8 (Sen: 90%; Spec: 30%) alcanzando una especificidad intermedia en el punto de corte 10 (Sen: 82%; Spec: 48%). Con valores >16 la sensibilidad se sitúa por debajo del 50%, aunque la especificidad aumenta hasta el 80% (Punto de corte 17, Sen: 42%; Spec: 80%).

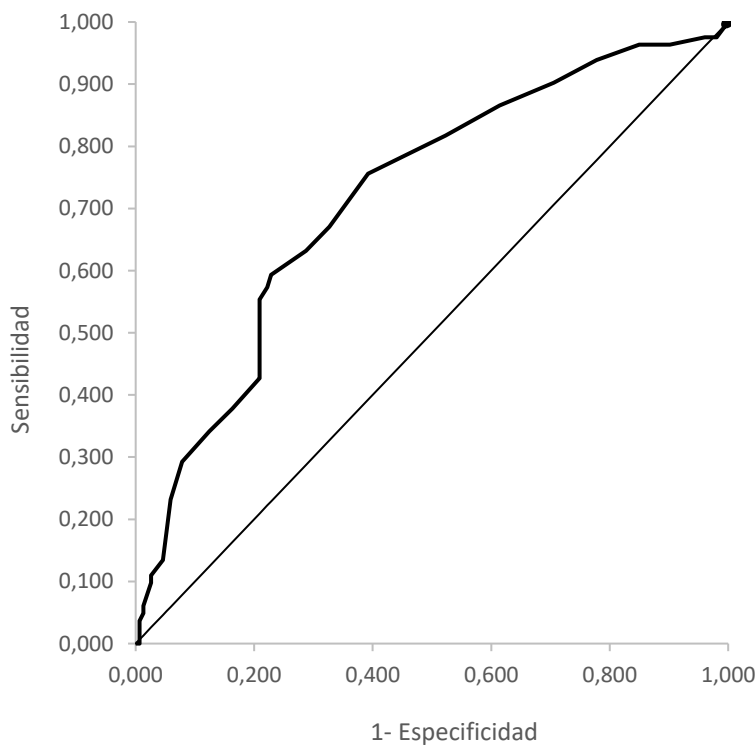


Figura 7. Curva COR para la comparación entre simuladores instruidos y pacientes clínicos.

10.3. Discusión

El principal objetivo que se perseguía mediante el presente estudio era valorar la capacidad discriminativa del Inventario Estructurado de Simulación de Síntomas (SIMS) en el contexto médico-legal. Aplicando los puntos de corte propuestos y teniendo en cuenta la necesidad de reducir al mínimo la tasa de falsos positivos, se obtienen mejores resultados con el punto de corte de 16 que con 14. Los resultados indican que los simuladores obtienen puntuaciones más altas en el SIMS que los otros tres grupos. En términos generales, el SIMS puede discriminar mejor entre simuladores instruidos / probables simuladores y la población general, que entre simuladores instruidos / probables simuladores y pacientes clínicos. De manera similar, el instrumento discrimina mejor a los simuladores que a los probables simuladores.

En cuanto a las puntuaciones del SIMS, la subescala de Trastornos afectivos registró la puntuación más alta en los grupos de pacientes clínicos y sobredimensionados. Estos resultados coinciden con las críticas realizadas a esta subescala, en las que se plantea que puede inducir a error ya que incluye sintomatología depresiva genuina común después de un accidente de circulación (van Impelen *et al.*, 2014; Sarrami *et al.* 2016). Por otro lado, estos hallazgos contradicen la hipótesis de Cernovsky *et al.* (2019b) de que al menos 30 ítems del instrumento podrían inducir a error a los pacientes clínicos porque son consistentes con sintomatología comúnmente observada tras un MVA.

En general, en el análisis entre simuladores / sobredimensionados y pacientes clínicos, la escala muestra una especificidad relativamente alta y una sensibilidad intermedia. Estos resultados indican que el instrumento clasificará erróneamente aproximadamente la mitad de los simuladores (falsos negativos) y tendrá una tasa de falsos positivos de entre el 17% y el 21%, o uno de cada cinco casos. Por otro lado, en la comparación de probables simuladores con pacientes clínicos, se puede observar una

ligera disminución de la sensibilidad respecto al análisis anterior, manteniendo la tasa de falsos positivos entre el 17% y el 21%.

Esto sugiere que el SIMS no es capaz de discriminar correctamente a posibles simuladores con los puntos de corte propuestos por Smith y Burger (1997) y González-Ordi y Santamaría Fernández (2009) (de ahora en adelante nos referiremos a ellos como puntos de corte originales), por lo que cabe la posibilidad de que deba buscarse otro punto de corte que se ajuste a las características de una población para la cual el instrumento no fue diseñado originalmente. En este sentido, en el manual de la versión española del SIMS, los autores recogen el estudio de Capilla Ramírez *et al.* (2008) donde aplican el SIMS a pacientes con dolor crónico del contexto médico-legal. En el citado estudio se sugiere utilizar para este tipo de pacientes un punto de corte más estricto (puntuaciones > 6), que permite discriminar con un 88% de sensibilidad y un 47% de especificidad. Teniendo en consideración este criterio de decisión, para nuestra muestra el punto de corte óptimo se situaría en 10, ya que se alcanzan valores similares a los de los anteriores autores, con una sensibilidad del 82% y una especificidad del 48%. A pesar de ello, creemos que para realizar un screening de simulación con el SIMS en poblaciones como las valoradas en nuestro estudio, y en el de Capilla Ramírez *et al.* (2008), sería necesario contar con instrumentos adicionales que compensen esta limitación.

Consideramos de interés el hecho de que nuestros hallazgos aplicando los puntos de corte originales son contrarios a los del meta análisis de van Impelen *et al.* (2014), donde se observa que la sensibilidad del instrumento es bastante elevada y la especificidad es baja, pero resultan similares a los de Capilla Ramírez *et al.* (2014), donde se expone una especificidad elevada y una sensibilidad moderada-baja. De la misma forma, nuestras puntuaciones totales del SIMS en el grupo de simuladores son parecidas a las del estudio

de los últimos autores, mientras que difieren significativamente de las registradas por van Impelen *et al.* (2014).

Creemos que una posible explicación a estas diferencias es que en el citado meta-análisis se han incluido en mayor medida estudios que evalúan simulación en psicopatología -con alguna salvedad como las publicaciones de González-Ordi y Santamaría Fernández (2008) y González-Ordi *et al.* (2010) que evalúan pacientes con fibromialgia y dolor crónico sin causa especificada-, mientras que en nuestro estudio, y en el de Capilla Ramírez *et al.* (2014), se evalúa simulación en patología cervical (whiplash injury). Es posible que estas diferencias en las condiciones afecten a la actuación tanto de los simuladores reales como de los simuladores instruidos, ya que responderían el instrumento adaptándose a la condición a fingir.

El SIMS fue construido siguiendo la teoría del *overendorsement*, según la cual se plantea que los presuntos simuladores mostrarán indiscriminadamente síntomas de cualquier tipo, sin tener en consideración el contexto (Monaro *et al.*, 2018; Rogers y Bender, 2018), pero como exponen Merckelback *et al.* (2009), la influencia de dicho contexto debe tenerse en cuenta ya que modula la forma y la intensidad con la que se simulan los síntomas. En este caso, es posible que los simuladores finjan únicamente síntomas esperables y descarten aquellos que consideren infrecuentes. Un sujeto con daño mínimo tras un accidente de circulación que quiera simular SLC se centrará en síntomas que resulten coherentes con la condición, y considerará que fingir sintomatología llamativa y atípica, como la psicótica, puede restarle credibilidad.

De la misma forma, habitualmente se solicita a los simuladores instruidos que recreen la condición asignada de la forma más realista y creíble posible. Si a un participante análogo se le pide que finja padecer psicopatología grave, encontrará en el SIMS una mayor cantidad de ítems que considerará de interés para recrear su rol, ya que

el instrumento ha sido diseñado específicamente para evaluar dicha sintomatología. Por otro lado, un participante asignado a simular una patología cervical encontrará síntomas atípicos que no considerará por la posibilidad de disminuir su credibilidad. Los resultados observados en las subescalas de nuestro estudio y en el de Capilla Ramírez *et al.* (2014) son coherentes con esta hipótesis, ya que únicamente la subescala de trastornos afectivos, la cual contiene ítems consistentes con algunos de los síntomas que podrían manifestarse en una condición como el SLC, ha obtenido valores relativamente elevados.

Creemos que, si bien la capacidad discriminativa del instrumento para detectar simulación en psicopatología es excelente (Rogers y Bender, 2018; van Impelen *et al.*, 2014), de cara a futuras investigaciones sería recomendable tratar de adaptar el instrumento a las particularidades de aquellos contextos para los que no fue diseñado. En este sentido, podrían incluirse subescalas adicionales que evalúen sintomatología común en dicho contexto, como podría ser el dolor. No solo podrían incorporarse ítems relativos a la severidad del dolor, sino que se podrían incluir factores de corte psicosocial relacionados con la experiencia del mismo, como el pensamiento catastrofista, la actitud ante la situación de enfermedad o la sintomatología ansioso-depresiva. (Aguilera *et al.*, 2019; Casey *et al.*, 2015; Falla *et al.*, 2016; Holm *et al.*, 2008).

11. Discusión general

En la presente tesis doctoral se perseguía, como objetivo general, el diseño de un protocolo para detectar simulación de síntomas en Síndrome de Latigazo cervical. Para alcanzar dicho fin, se formularon 3 objetivos generales, desglosados en 7 específicos, y se desarrollaron 6 estudios. Los hallazgos de cada uno de estos objetivos específicos han sido expuestos con anterioridad en sus respectivos apartados, por lo que en esta discusión general se pretende realizar un análisis global del desarrollo de la tesis, y de los estudios que la conforman, así como una valoración de la viabilidad y adecuación del protocolo aplicado.

En términos generales, consideramos que hemos alcanzado adecuadamente los 3 objetivos planteados. Si bien es cierto que se han encontrado diversas dificultades que han obligado a realizar ajustes al protocolo, como la limitación temporal impuesta o la poca variedad de herramientas validadas en población española, hemos obtenido un producto capaz de discriminar correctamente entre grupos. No obstante, durante la ejecución de los estudios que componen este trabajo se han observado una serie de cuestiones de interés que consideramos oportuno mencionar y analizar antes de valorar nuestros hallazgos.

En primer lugar, mediante nuestra revisión sistemática (Estudio 2) y las múltiples revisiones bibliográficas desarrolladas durante el transcurso del trabajo realizado, hemos podido observar que la disciplina evoluciona lentamente a nivel de *producción de resultados*. Considerando que tanto el SLC como el dolor crónico son dos de las condiciones más simuladas (González-Ordi *et al.*, 2012; Represas *et al.*, 2008; Represas *et al.*, 2020), y teniendo en cuenta el altísimo coste socioeconómico que supone (Crouch *et al.*, 2006; Kamper *et al.*, 2008; Pink *et al.*, 2016), creemos que la evidencia disponible, a día de hoy, para detectar dicha problemática es insuficiente. Mediante la revisión

sistemática realizada se han detectado 23 estudios que evalúan diferentes instrumentos, pero la gran mayoría no han sido replicados. Herramientas como el *Pain Patient Profile* (P3) o el *Life Assessment Questionnaire* (LAQ) resultan prometedoras, pero su uso en la práctica clínico-forense es arriesgado debido a la poca evidencia disponible sobre su funcionamiento y el elevado lapso transcurrido desde su diseño y valoración. Por otro lado, estudios como el del MSPQ y el PDI de Bianchini *et al.* (2014), o el reciente trabajo de Curtis *et al.* (2019) sobre la capacidad discriminativa de escala de *Catastrofización del Dolor* (PCS), evidencian la utilidad de las escalas breves como herramienta de *screening*, pero encontramos el mismo problema de replicabilidad. Además, en varias de estas publicaciones, los autores señalan la necesidad de aplicar el instrumento evaluado junto a otros que compensen una baja sensibilidad o especificidad (Crighton *et al.* 2014), pero no disponemos de evidencia sobre el funcionamiento conjunto de las mismas. Estas limitaciones se hacen más evidentes en la literatura española, donde el número de estudios empíricos y herramientas validadas es significativamente menor, habiéndose localizado únicamente los trabajos desarrollados por los doctores Héctor Ordí y Pablo Santamaría, y la doctora Pilar Capilla, entre otros autores (ej. Capilla Ramírez y González-Ordí, 2009; González-Ordí, *et al.*, 2010; González-Ordí, *et al.*, 2012).

Como expone Santamaría Fernández (2015), si bien es cierto que se existen instrumentos que sí disponen de múltiples estudios que abalan su eficacia para detectar simulación en las condiciones estudiadas, como el *Inventario Multifásico de la Personalidad de Minnesota* (MMPI-2) o la *Escala de Inteligencia de Wechsler para Adultos* (WAIS-IV), su extenso tiempo de aplicación, su alto coste y la formación requerida para su administración y corrección pueden limitar sus posibilidades de implementación en la aplicación práctica. De la misma forma, estos instrumentos evalúan una única variable dentro de una dimensión que puede estar compuesta por más variables

de interés (por ejemplo, rendimiento cognitivo dentro de la dimensión psicológica), por lo que su aplicación no sería suficiente para realizar un diagnóstico diferencial de simulación.

Merten *et al.* (2013), en su revisión sobre el estado de la cuestión (*state of the art*) a nivel europeo, señalaban que en España la simulación en el contexto médico-legal no era una cuestión que se tratara regularmente, y esta carencia de evidencia se traducía en la ausencia de una metodología o procedimiento sistemático y fiable para tratar de detectar la simulación en SLC. El proceder seguido por los profesionales en valoración del daño corporal que han colaborado en este estudio resulta coherente con dicho planteamiento. Para la valoración del SLC se seguía una aproximación eminentemente biomédica, donde la existencia de nexo causal era un factor determinante, de manera que, si no existía causalidad, no se otorgaba compensación económica, aunque existiera un diagnóstico de SLC. Dicho diagnóstico se realizaba únicamente mediante una valoración clínica del daño corporal, sin la aplicación de ningún instrumento validado mediante el cual se pudiera descartar o confirmar objetivamente la presencia de simulación. Únicamente en aquellos casos en los que el evaluado rechazaba la compensación económica propuesta inicialmente por la compañía aseguradora se trataba de evaluar si estaba fingiendo sus síntomas, habitualmente mediante el uso de un detective que trataba de fotografiar movimientos o acciones incoherentes con el presunto daño padecido, mediante un informe biomecánico como los incluidos en el Estudio 1, o mediante la revisión de un traumatólogo que trataba de evidenciar inconsistencias entre la gravedad expresada y los síntomas observados. Entendemos que el segundo recurso puede utilizarse con unos puntos de corte adecuados y el primer y tercer recurso podrían catalogarse como inconsistencias convincentes (del inglés, *compelling inconsistencies*), término utilizado por Slick *et al.* (1999) y Bianchini *et al.* (2005) en sus criterios

diagnósticos. El problema es que, al igual que comentábamos con el uso del MMPI-2 o el WAIS-IV, no aportan toda la información necesaria para realizar un diagnóstico completo y fiable de simulación.

Consideramos de gran importancia tratar de solventar esta situación y avanzar hacia la búsqueda de medidas válidas y protocolos adecuados. Merten *et al.* (2013) ofrecían una visión optimista del desarrollo europeo en el campo de la valoración de la simulación, pero creemos que aún queda mucho camino por recorrer, en especial en España y en el contexto médico-legal.

Por otro lado, durante el desarrollo de la presente tesis también hemos encontrado ciertas dificultades a nivel metodológico, específicamente en los diseños de investigación utilizados en el campo de la simulación. Las principales publicaciones sobre la temática, como Franzen *et al.* (1990), Graham *et al.* (1991) o la primera edición del libro de Rogers (1988) surgen en un contexto eminentemente psicológico, donde el objetivo que se perseguía o se mostraba era, habitualmente, detectar simulación en psicopatología. Gran parte de la investigación en la temática se ha desarrollado alrededor de dicha área, en especial en la neuropsicológica (Rogers y Bender, 2018), y los avances metodológicos, como el diseño de grupo conocido, se han hecho teniendo en consideración los recursos y materiales disponibles en la misma.

Así, nos encontramos diseños que requieren una serie de condiciones que en el campo de la simulación en dolor y SLC difícilmente pueden alcanzarse. Tomando como ejemplo el citado diseño de grupo conocido, donde se comparan grupos de simuladores y pacientes clínicos que han sido formados por una serie de criterios altamente precisos, en psicopatología se dispone de una variedad relativamente amplia de instrumentos que pueden utilizarse para tal fin, como el SIMS o la entrevista estructurada SIRS (Rogers *et al.*, 2010), pero en el campo del dolor dicha variedad es mucho más limitada. En este

sentido, las únicas aportaciones realizadas, a día de hoy, son las del equipo de Kevin Bianchini y Kevin Greve, los cuales han adaptado la metodología utilizada en neuropsicología al campo que nos ocupa, proponiendo los criterios ya señalados en el marco teórico (ver Bianchini *et al.* 2005). Dichos criterios han sido ampliamente utilizados en sus investigaciones como método de clasificación, acompañados de un extenso protocolo de instrumentos psicométricos como el MMPI-2, el TOMM o el WAIS (por ejemplo, ver Tabla A1. del apéndice de Curtis *et al.*, 2019), y representan una aproximación multidimensional completa a la valoración de la simulación. A pesar de ello, como señalan Tuck *et al.*, (2018), los citados criterios presentan la limitación de que requieren una gran cantidad de observaciones y recursos, ya que tendría que realizarse una evaluación psicológica, neuropsicológica, física, de la capacidad funcional y un análisis de la historia clínica, así como obtener información de los miembros de la familia o amigos cercanos, y de algún sistema de vigilancia, como un detective. Además, tendría que sumarse el tiempo de aplicación y coste de los instrumentos de medida que se deseen valorar en la investigación. Por ejemplo, en caso de querer valorar un instrumento con un tiempo de aplicación relativamente elevado, como puede ser el PAI se emplearía con cada paciente un tiempo aproximado de 3 ó 4 horas. Si bien entendemos que esta propuesta representa un modelo multidimensional apropiado y completo para realizar un diagnóstico diferencial de simulación, la elevada cantidad de información necesaria para su valoración hace que no se encuentre dentro de las posibilidades de la gran mayoría de los/as profesionales y limita significativamente la elección de diseño de investigación. Específicamente en el contexto médico-legal español esta propuesta resulta prácticamente inviable por las razones anteriormente descritas.

Ante esta situación, el/la profesional se ve obligado a aplicar únicamente un diseño de simulación que, si bien cuenta con una elevada validez interna, tiene una validez

externa algo más limitada que afecta a la posibilidad de generalizar sus resultados. En condiciones ideales, el citado diseño de simulación se utilizaría como una primera fase de valoración de la idoneidad del instrumento y, posteriormente y en el caso de observar un rendimiento óptimo en el mismo, se validaría con un diseño de grupo conocido o criterio parcial para disponer de unos datos más generalizables. Consideramos que, de cara a futuras investigaciones, la elaboración de una metodología apropiada para la formación de grupos criterios es una prioridad, ya que hará más accesible y fiable la investigación en el campo de la simulación del dolor.

Una vez tratadas las cuestiones de interés expuestas con anterioridad, pasamos a valorar el protocolo objeto de la presente tesis, analizando primero los instrumentos incluidos en el mismo y segundo su validez y rendimiento global.

En lo relativo a los instrumentos que lo componen, para los puntos de corte óptimos identificados, a nivel individual el BPI (punto de corte ≥ 64), el NDI (punto de corte ≥ 29) y el SF-36 (punto de corte ≤ 34) han sido los instrumentos con valores más adecuados, con una sensibilidad por encima del 70%, una especificidad por encima del 80% y poderes predictivos (30% de ratio base) por encima del 80%, salvo el PPP del SF-36 que se sitúa en el 73.2%. El BDI (punto de corte ≥ 11) es el instrumento con mayor sensibilidad (85.37%), pero su especificidad es más limitada, con un 25% de tasa de falsos positivos. Por otro lado, el BAI (punto de corte ≥ 13) y el BIPQ (punto de corte ≥ 6) muestran una sensibilidad por debajo del 50% y una especificidad del 93% y del 81% respectivamente, con valores predictivos positivos entre 59.6% y 52%, y negativos del 92% y del 78%.

En términos generales, creemos que el BPI, el NDI y el SF-36 pueden discriminar entre simuladores y no simuladores, y un sistema o metodología creada para tal fin puede beneficiarse de su inclusión. El BDI tiene una sensibilidad adecuada, pero su

especificidad moderada limita su posible aplicación en el contexto médico-legal. Por otro lado, estimamos que el BAI y el BIPQ no resultan adecuados para diferenciar simuladores de no simuladores, ya que su sensibilidad y poder predictivo positivo indica que sería capaz de discriminar aproximadamente a un 50% de los simuladores.

Varios autores señalan que existe la posibilidad de “compensar” valores inferiores a lo deseable, y la aplicación conjunta de varios instrumentos puede reforzar posibles “puntos débiles” (Crighton *et al.*, 2014; Gonzalez Ordi y Santamaría, 2009), pero consideramos que el aporte real de un instrumento que presente unas características como las descritas con anterioridad es limitado. Los resultados obtenidos mediante el uso de la razón de verosimilitud encadenada apoyan este planteamiento. Dentro de un protocolo de 6 instrumentos, con dos resultados positivos en dos instrumentos que alcanzan una sensibilidad y una especificidad aproximada del 70% y del 95% respectivamente (BPI y NDI), la probabilidad de que dicho resultado positivo haya sido producido por una persona que está simulando sus síntomas es del 80%, pero al combinar uno de estos dos instrumentos con otro que presente unos valores algo más moderados, la probabilidad disminuye significativamente. Por ejemplo, combinando el NDI con el SF-36, el cual tiene una sensibilidad y una especificidad aproximada del 70% y del 90% respectivamente, la probabilidad disminuye al 66%, y al combinarlo con el BPI al 44%. Si por el contrario el BPI o el NDI se empareja con el BIPQ, la probabilidad disminuye al 11% y al 26% respectivamente. Por otro lado, con 3 resultados positivos, la disminución observada es menor si, al igual que en los ejemplos anteriores, dos de los 3 resultados pertenecen a instrumentos con propiedades similares a las del BPI o el NDI. Si el tercer positivo pertenece al BIPQ la probabilidad se reduciría de un 99%, alcanzada con el BDI o el SF-36 como tercer positivo, a un 94%. No obstante, si el segundo positivo pertenece a un instrumento con unos valores inferiores, como el BDI o el SF-36,

observamos disminuciones mucho más marcadas, como en la combinación BPI, BDI y BIPQ, que desciende al 73%, o en NDI, BDI BIPQ que desciende a al 88%. En el caso de que 2 de los 3 positivos sean de instrumentos con una capacidad discriminativa más reducida, como el BAI o el BIPQ en nuestro caso, el PCS en el estudio de Curtis *et al.* (2019) o el MSPQ y PDI en el de Bianchini *et al.* (2014), la probabilidad se reduce drásticamente, a no ser que el tercer instrumento tenga una especificidad muy elevada (probabilidad del 80% si el tercer positivo es el NDI, pero del 59% si es el BPI, 39% si es el BDI y 42% si es el SF-36).

Esto indicaría que la capacidad discriminativa de los instrumentos influye en gran medida en la cantidad de resultados positivos necesarios para obtener una probabilidad elevada de simulación, siendo necesario, en nuestro caso, hasta 4 resultados positivos según los instrumentos combinados. Por ese motivo, consideramos que, aunque se plantee la opción de utilizar la herramienta estudiada como *screening*, y se advierta de la necesidad de utilizar fuentes de información adicionales, sería recomendable analizar cuidadosamente qué valores consideramos “adecuados” en los estimadores de utilidad. En este sentido, cabe destacar que autores como Bianchini *et al.* (2014), Crighton *et al.* (2014) y Curtis *et al.* (2019), consideran que valores similares a los alcanzados en BIPQ y en el BAI son adecuados para su uso como screening. Por ejemplo, en el estudio de Curtis *et al.* (2019) con la Escala de Catastrofismo del Dolor (PCS), con un punto de corte de $T \geq 81$, obtuvieron una sensibilidad del 47%, una especificidad del 93% y un PPP (al 30%) del 74%, afirmando que “es capaz de diferenciar con precisión simuladores de no simuladores” (p.11). Si bien todos los autores citados indican explícitamente, como hemos señalado, que no debería realizarse una valoración de la simulación únicamente con dichos instrumentos, creemos que debería seguirse un criterio de valoración de la “idoneidad” algo más conservador. Comentábamos en la introducción que, para elegir

qué valores considerar, debería tenerse en cuenta el uso que se le pretenda dar al instrumento. Recordemos que el/la evaluador/a dispone de dos tipos de instrumentos: aquellos dirigidas a hacer *screening* de riesgo de simulación, como el SIMS, y aquellos para realizar un diagnóstico diferencial de simulación (Rogers, 2008). Generalmente, se considera que los estimadores de utilidad deben ser lo más elevados posibles, pero que un test de *screening* debe tener una sensibilidad alta, mientras que el test de diagnóstico diferencial que se utilice a continuación debe tener alta especificidad (McNamara y Martin, 2018). Debido a que la detección de la simulación está ligada al contexto forense, contexto donde resulta fundamental reducir al mínimo la tasa de falsos positivos, esta visión general no es compartida y se estima que, independientemente del uso que se le dé a la herramienta, la especificidad debe ser lo más elevada posible (Arce, 2017). Boone (2007) estableció que, para instrumentos destinados a realizar un análisis de la validez de los síntomas, resulta recomendable que, como mínimo, se alcance un 90% de especificidad, pero Vitacco *et al.* (2007) consideraron que, aunque se trate de medidas de *screening*, la cifra debía ser lo más elevada posible y un 10% de tasa de falsos positivos no es aceptable en el contexto que nos ocupa. Creemos que los valores adecuados pueden estimarse en función del número de instrumentos de *screening* que se utilicen, y coincidimos con Vitacco *et al.* (2007) en que, en caso de que se disponga de una única herramienta de *screening*, la especificidad debe ser superior al 95%. No obstante, consideramos que, en caso de que se trate de una batería de herramientas, se puede aceptar un valor ligeramente inferior, ya que con el procedimiento de razón de verosimilitud encadenada se podrá calcular qué número de positivos es necesario para realizar una valoración precisa. A pesar de ello y en base a nuestros resultados, coincidimos con Boone (2007) y no debería ser inferior al 90%.

Una vez analizadas a nivel individual las herramientas que componen el protocolo, podemos considerar que, a nivel global, nuestros resultados apoyan su validez y el uso conjunto de medidas breves es una metodología válida para discriminar entre simuladores y no simuladores. Se ha podido observar que, con 3 resultados positivos, siendo al menos dos de ellos el BPI y el NDI, se alcanzaba una probabilidad superior al 95% de que dicho resultado estuviera producido por una persona que estuviera simulando sus síntomas. No obstante, si bien consideramos que nuestra propuesta es prometedora, sería prematuro recomendarla para su inmediata aplicación en el contexto médico-legal. Creemos que la presente tesis doctoral supone un primer paso hacia la creación de una metodología válida para la detección de la simulación en SLC, pero resulta necesario continuar con el proceso de validación del mismo, incluyendo una serie de modificaciones. En primer lugar, sería conveniente estudiar la inclusión de nuevos instrumentos para sustituir el BAI y el BIPQ. En ese sentido, tanto la ansiedad como la conducta ante la situación de enfermedad son dos variables de interés en el estudio del SLC (Carroll *et al.*, 2009), por lo que puede considerarse sustituir dichas herramientas por otras que tengan la misma finalidad. Por ejemplo, para valorar la ansiedad, el *Cuestionario de Ansiedad Estado-Rasgo* (STAI; Spielberg *et al.* 1970), en la adaptación española de Buela-Casal *et al.* (2011) puede ser una elección adecuada por sus buenas propiedades psicométricas ($\alpha = .90$ para rasgo y $.94$ para estado), aunque su tiempo de aplicación aproximado es de 15 minutos aproximadamente. También puede considerarse un instrumento de aplicación rápida que incluya entre sus medidas del resultado la ansiedad, como el *Listado de Síntomas Breve* (LSB-50; de Rivera y Abuín; 2018), que valora, entre otros, la hostilidad, la somatización y sintomatología ansioso-depresiva y se administra en aproximadamente 10 minutos. Por otro lado, para el estudio de la actitud o conducta ante la situación de enfermedad puede considerarse el *Cuestionario de Conducta de Enfermedad* o *Cuestionario para el*

Comportamiento de Enfermedad (IBQ; Pilowsky *et al.*, 1984) en su adaptación española por Llor-Esteban *et al.* (1991), o la versión revisada del *Cuestionario de Percepción de Enfermedad* (IPQ-R; Broadbent *et al.*, 2006), que es la versión no abreviada del instrumento utilizado en la presente tesis.

También puede plantearse la posibilidad de incluir instrumentos adicionales en el caso de que no se encuentre una limitación temporal estricta, aunque recomendamos que el tiempo de aplicación no sea superior a 30 minutos para poder garantizar su viabilidad. En este sentido, la catastrofización del dolor, variable ya evaluada por Curtis *et al.* (2019) a través del uso del PCS, tiene una influencia significativa en la vivencia de la enfermedad del SLC (Bostick *et al.*, 2013), por lo que creemos que resultaría de interés estudiar el rendimiento del instrumento en una población con dicha condición, o de otra herramienta que estudie la citada variable, como el *Coping Style Questionnaire* (CSQ; Roger *et al.*, 1993), en su adaptación española por Guarino *et al.* (2007). También consideramos de interés el *Whiplash Disability Questionnaire* (WDQ; Pinfold *et al.*, 2004), en su adaptación española por Tejero (2017), el cual tiene como objetivo estudiar la percepción de discapacidad ocasionada por el SLC, tiene un tiempo de administración breve y puede complementar los resultados del NDI.

Otro tipo de herramientas que podría ser de utilidad en un protocolo como el estudiado son aquellas que tienen como objetivo la detección de la simulación. En el Estudio 6 hemos evaluado la capacidad discriminativa del SIMS, y si bien con el punto de corte propuesto (≥ 10) se ha alcanzado una sensibilidad elevada, su especificidad es baja, con una tasa de falsos positivos del 52%, lo que limita significativamente su aplicación en el contexto estudiado. Como hemos expuesto en la discusión del citado estudio, creemos que el SIMS puede ser de gran utilidad para la tarea que nos ocupa, ya que en el campo de la detección de simulación en psicopatología ha demostrado una

capacidad discriminativa muy elevada y su tiempo de administración es bajo (van Impelen *et al.*, 2013). En este sentido, creemos que, de cara a futuras investigaciones, podría evaluarse la posibilidad de adaptar el instrumento al contexto médico-legal sustituyendo subescalas poco relevantes, como la de psicosis, por otras que valoren variables frecuentes en el mismo, como el dolor o la percepción de discapacidad.

El *Self-Report Symptom Inventory* (SRSI; Merten *et al.*, 2016) es otro instrumento que consideramos que puede ser de utilidad. Se trata de un cuestionario de 107 ítems divididos en 10 subescalas, 5 para pseudosíntomas y 5 para síntomas, y dos índices totales, diseñado para evaluar *symptom over-reporting* en dolor y sintomatología ansioso-depresiva, cognitiva y somática no específica. Los estudios preliminares realizados muestran resultados prometedores, con una consistencia interna elevada (entre .81 y .85 para síntomas y entre .67 y .92 para pseudosíntomas) y una capacidad discriminativa alta (sensibilidad y especificidad aproximada del 82% y del 91%, respectivamente, y razón de verosimilitud de 9.31). Si bien su tiempo de aplicación es algo más elevado que el del SIMS, todas las subescalas incluidas valoran variables de interés para el SLC por lo que, de cara a futuras investigaciones, su validación al castellano, y la valoración de su capacidad discriminativa con condiciones como la citada, sería de gran interés.

Con el objetivo de incorporar en el protocolo una fuente de información adicional, además de los instrumentos psicométricos estudiados, también se ha evaluado un *checklist* formado por indicadores de simulación citados habitualmente en los estudios de la materia (Estudio 5). Si bien no se ha podido incluir en el cálculo de la razón de verosimilitud encadenada, debido a que el *checklist* no ha sido construido como una escala, los resultados preliminares obtenidos apoyan su uso e indican que se trata de una herramienta que puede ser de utilidad para el/la profesional. Como ya exponíamos en la discusión del citado estudio, el *checklist* puede ser utilizado como puede ser utilizado

como guía para una valoración sistemática de varias de las estrategias propuestas por Rogers y Bender (2018) expuestas en el marco teórico, como la presencia de inconsistencias en los síntomas informados *versus* los observados. Al igual que el protocolo, los resultados del *checklist* son prometedores, pero algunos de los indicadores utilizados no han resultado de utilidad para la tarea propuesta. Es posible que el diseño de investigación utilizado haya producido estos resultados, ya que algunos de los citados indicadores, como los que valoran hostilidad, pueden ser difíciles de recrear con simuladores instruidos. Si bien en el caso de los instrumentos psicométricos recomendamos su sustitución, en el caso del *checklist* consideramos que, en la medida de lo posible, debería estudiarse su validez mediante un diseño de grupo conocido que incluya simuladores reales antes de considerar la eliminación y/o sustitución de alguno de los citados indicadores. Creemos que herramientas como nuestro *checklist*, el propuesto por Akehurst *et al.* (2015), y los estudios sobre la *Verifiability Approach* en el contexto médico-legal (Akca *et al.*, 2020; Boscovik *et al.*, 2017) son aportaciones de gran utilidad para la detección de la simulación en patologías dolorosas. Como señalaban Akca *et al.* (2020), el uso de instrumentos psicométricos no es suficiente para detectar con eficacia la fabricación de síntomas en la condición señalada, por lo que consideramos que la adaptación de metodologías del campo de la detección de mentiras puede ser un enfoque prometedor para solventar dicha problemática. En este sentido, creemos necesario señalar que las limitaciones del contexto reducen significativamente la variedad de recursos a utilizar en el mismo. Los simuladores no tienen, en la gran mayoría de casos, un relato extenso sobre el que poder aplicar un instrumento como el SVA, o un protocolo como el *Sistema de Evaluación Global* (SEG; Arce y Fariña, 2004, 2005) y, aunque dispusieran de él, se necesitaría un procedimiento de obtención del testimonio, como la *Entrevista Cognitiva Mejorada* (Fisher y Geiselman, 1992), que necesitaría una elevada

cantidad de tiempo. Por ello, creemos que el uso de *checklists* de indicadores de simulación resulta una aportación de utilidad que podría ser abordado en futuras investigaciones.

El principal producto de nuestro trabajo ha sido el protocolo en sí mismo, así como un modelo de estructura que puede utilizarse en futuros trabajos en la temática. Inicialmente, se planteó la opción de desarrollar un listado de criterios de valoración, como los de Bianchini *et al.* (2005), y ajustar el protocolo en función de dichos criterios, pero consideramos que un listado de tales características debía elaborarse una vez se hubiera estudiado la validez del protocolo, para que la elección de cada criterio estuviera basada en hallazgos empíricos. Así, teniendo en consideración nuestros resultados y las propuestas de Bianchini *et al.* (2005), Slick *et al.* (1999), Sherman *et al.* (2020) y Rogers y Bender (2018), proponemos una serie de criterios que podrían seguirse en subsiguientes estudios:

A) Existencia de incentivo externo significativo de tipo material-legal. Como exponen González-Ordi *et al.* (2012), la existencia de un incentivo externo es un punto común en todos los criterios diagnósticos propuestos para la simulación. El más frecuente es el beneficio económico, mediante indemnización o compensación, pero también puede darse el mantenimiento de una baja laboral. Debe considerarse que determinar la presencia de un incentivo externo no es suficiente, ya que su existencia no es sinónimo de simulación y debe examinarse además la percepción del/la examinado/a sobre dicho incentivo y valorar si es lo suficientemente motivador como para justificar la simulación (Rogers y Bender, 2018). Debido a que se trata de un contexto en el que resulta frecuente que los pacientes muestren cierta desconfianza y suspicacia, la exploración necesaria para el análisis de la motivación puede resultar compleja, por lo que debería realizarse en

último lugar para no alterar el resto del protocolo y seguir un enfoque conciliador y cordial, similar al de la entrevista investigativa (ver Bull, 2014).

B) Evidencia de inconsistencias psicométricas. Mediante el uso de un protocolo biopsicosocial como el analizado en la presente tesis doctoral. El/la profesional debe valorar si el perfil obtenido es consistente con lo esperado en la condición y determinar la probabilidad de simulación mediante el uso de la razón de verosimilitud encadenada.

C) Evidencia de fingimiento de síntomas. Además de evidencia de un protocolo biopsicosocial, cuyo producto es un perfil de la condición específico que se compara con el esperado en la misma, resulta de utilidad disponer de un resultado directo sobre la existencia de fingimiento, como el proporcionado por el SIMS o el SRSI (Merten *et al.*, 2016).

D) Inconsistencias clínicas y conductuales convincentes. Considerar, mediante un registro similar al *checklist* valorado en el Estudio 5, inconsistencias entre las manifestaciones verbales del paciente y los indicadores clínicos observados, así como patrones de comportamiento anómalos, atípicos o incoherentes.

E) Se han descartado hipótesis alternativas. El/la profesional deberá considerar otras posibles explicaciones a la exageración de síntomas, entre las que se encuentran la secuencia de los ítems y los tests aplicados, falta de información sobre los síntomas que puede padecerse, un patrón de respuesta a los instrumentos inatento, rasgos de personalidad (Merckelbach *et al.*, 2019), un patrón de actuación defensivo, la discapacidad social y la gestión de la impresión (Rogers y Bender, 2018). De la misma forma, deberá descartar la presencia de un trastorno facticio.

Consideramos oportuno finalizar la presente tesis doctoral reflexionando sobre la necesidad de un cambio de paradigma en la detección de simulación en patologías

dolorosas y SLC en España. En los estudios desarrollados hemos intentado mantener una proyección nacional e internacional, pero inevitablemente, tanto nuestra metodología como el producto obtenido, se han tenido que adaptar a los requisitos del contexto. Nuestros hallazgos son prometedores y creemos que pueden ser de utilidad para futuras investigaciones en el territorio nacional, pero nuestro esfuerzo representa unos pocos pasos en un largo camino por recorrer. Comentábamos con anterioridad que Merten *et al.* (2013) ofrecían una visión optimista del avance europeo, pero la descripción del estado de la cuestión que realizan de España sigue siendo aplicable a día de hoy, y aún quedan pendientes una serie de cuestiones a las que se debería prestar especial atención para alcanzar un desarrollo óptimo. En primer lugar, consideramos que fomentar el uso del modelo biopsicosocial, así como una metodología multidimensional, es fundamental para el crecimiento de la disciplina. Tanto nuestros resultados, en especial los obtenidos en el Estudio 3, como las aportaciones de los expertos en la materia (Sterling, 2004, 2011, 2014), apoyan su uso tanto a nivel clínico como a nivel forense. La evidencia indica que variables psicosociales como la ansiedad, depresión o pensamientos catastrofistas sobre el dolor o el proceso de la enfermedad se asocian con una peor recuperación de la condición (Ritchie y Sterling, 2016). El modelo biopsicosocial permitiría realizar un análisis completo y pormenorizado de dichas variables, lo que haría posible ofrecer un tratamiento adaptado a las características y necesidades de cada paciente y reduciría el riesgo de cronificación. De la misma forma, el/la profesional dispondría de una mayor cantidad de evidencia empírica mediante la cual poder tomar una decisión objetiva y argumentable (González-Ordi *et al.* 2012). En segundo lugar, además de desarrollar y validar nuevas herramientas e instrumentos para población española, resulta fundamental visibilizar y transmitir, en la medida de lo posible, los resultados y avances obtenidos. El procedimiento para la valoración del SLC seguido a día de hoy por los/as profesionales

se ciñe a las peticiones de las partes interesadas, ya que para otorgar la compensación económica se solicitan una serie de criterios y la valoración persigue el objetivo de responder a tales criterios. Presentar una metodología eficaz y con evidencia robusta a dichas partes interesadas, con la cual se podría incrementar la detección de simuladores, podría facilitar un cambio de paradigma. En tercer y último lugar, somos conscientes de que un cambio como el planteado es complejo y resulta poco probable que se haga a corto plazo, por lo que una línea de investigación que consideramos prioritaria es el desarrollo de una metodología eficaz para desarrollar los grupos criteriosales en diseños como el de grupo conocido o de criterio parcial. Consideramos que dichos diseños son de gran importancia para comprobar la validez externa de la investigación desarrollada y garantizar así que están preparadas para su aplicación en la práctica clínico-forense.

12. Limitaciones

Los resultados de la presente tesis doctoral están sujetos a una serie de limitaciones que detallamos a continuación:

En primer lugar, no ha sido posible aplicar un diseño longitudinal mediante un seguimiento y análisis de la evolución de los pacientes clínicos, debido a que gran parte de los pacientes no acudían a una segunda sesión, o eran evaluados en una clínica externa a la que difícilmente se podía tener acceso. Si bien consideramos que el citado diseño habría tenido un impacto relativamente menor en el protocolo, ya que el diseño transversal es el que se aplica habitualmente en la bibliografía en la temática (Rogers y Bender, 2018), habría sido posible aportar resultados adicionales en el análisis *cluster* del Estudio 3 que habrían enriquecido los perfiles obtenidos. En segundo lugar, los simuladores instruidos incluidos eran estudiantes, en gran parte con una edad inferior a 25 años, y procedían de estudios relacionados con las ciencias de la salud (medicina y

enfermería). Como exponíamos con anterioridad, si bien eran alumnos de segundo año, y aún no habían recibido formación específica en la temática, es posible que dispusieran de conocimientos previos que condicionaran la presentación de la condición realizada. De la misma forma, si bien no se encontraron diferencias significativas en los resultados en función de la edad, dicha variable también puede ser un factor de influencia y es posible que la ejecución del rol asignado difiera en función de la misma. De cara a futuras investigaciones, creemos que sería de interés disponer de una variedad más amplia de simuladores instruidos, tanto en lo referido a la edad como a la procedencia. En tercer lugar, debido a las severas limitaciones temporales impuestas, no se ha podido aplicar un protocolo de clasificación de pacientes clínicos como el empleado habitualmente los diseños de grupo conocido (ver, por ejemplo, Curtis *et al.*, 2019). Si bien se utilizó el juicio de varios expertos en la temática, Rogers (2008) indica que no es una metodología de clasificación tan precisa como la empleada en el anterior diseño, por lo que no podemos garantizar una completa exclusión de posibles simuladores. En cuarto lugar, las citadas limitaciones temporales y la imposibilidad de utilizar un diseño de grupo conocido también influyen en la validez externa de nuestros resultados. No obstante, debemos matizar que el uso del diseño de simulación es apropiado para una fase inicial en la que la utilidad de los resultados se valore a nivel de investigación. Creemos que nuestros hallazgos permiten construir una base sobre la que desarrollar futuros estudios en la temática en la que se incluyan presuntos simuladores reales. Mencionar también que, desde nuestra experiencia en el desarrollo de la presente tesis doctoral, creemos que es altamente complejo aplicar en el contexto que nos ocupa la misma metodología de clasificación que utiliza Bianchini y su equipo. Como ya mencionábamos con anterioridad, adaptar el diseño de grupo conocido al contexto médico-legal español es una línea de investigación que consideramos altamente prioritaria. En quinto lugar, nuestra

batería de instrumentos ha sido ideada para la valoración del SLC. Si bien puede considerarse su aplicación en otros trastornos relacionados con el dolor, nuestros resultados no pueden generalizarse a otras condiciones sin una adecuada validación. En sexto y último lugar, no fue posible estudiar la relevancia que tenía la ganancia externa para los pacientes clínicos. Creemos que el análisis de dicha variable podría ofrecer información adicional de interés para la clasificación de grupos, y debería estudiarse su inclusión en criterios como los de Bianchini *et al.* (2005).

13. Conclusiones

En la presente tesis doctoral se perseguía el objetivo de desarrollar un protocolo biopsicosocial que permitiera evaluar el riesgo de simulación en esguince cervical o síndrome de latigazo cervical. Para alcanzar dicho objetivo, se han desarrollado seis estudios cuyas conclusiones exponemos a continuación.

Nuestros hallazgos indican que, en términos generales, el procedimiento empleado en España para la evaluación del SLC sigue un enfoque eminentemente biomédico, basado en la búsqueda de divergencias anatómico-estructurales. No se suele utilizar una metodología específica para la valoración de la simulación y, a día de hoy, se dispone de muy pocas herramientas validadas para realizar dicha valoración. En este sentido, el análisis realizado en el grupo de pacientes clínicos muestra tres perfiles diferentes y una amplia heterogeneidad sintomática, lo que apoya la evaluación del SLC desde el modelo biopsicosocial de la enfermedad y la evaluación del dolor desde el modelo multidimensional. La comparación de los grupos mostró que los pacientes del grupo de alto dolor percibido presentaron un perfil de mayor severidad en las variables psicosociales estudiadas. Teniendo en cuenta estos resultados, consideramos que la evaluación biopsicosocial de SLC, con un abordaje multidisciplinar, podría ser útil para

obtener una visión completa del mismo, asegurando que el tratamiento sigue un abordaje adaptado a las necesidades del paciente. Este enfoque multidimensional y biopsicosocial también puede ser de utilidad para la valoración de la simulación. El protocolo psicométrico diseñado desde los citados enfoques, compuesto por el BPI, NDI, SF-36, BAI, BDI y BIPQ, ha arrojado resultados prometedores. Nuestros hallazgos indican que el BPI, el NDI y el SF-36 han presentado una capacidad discriminativa adecuada y pueden ser de utilidad integrados en un sistema, metodología o batería destinada a realizar *screening* de simulación. El BAI, BDI, BIPQ han mostrado una capacidad más reducida, con una alta tasa de falsos positivos en el caso del BDI y de falsos negativos en las otras dos, por lo que creemos que no son apropiados para la finalidad que se persigue. No obstante, la evidencia indica que las variables medidas por dichas herramientas juegan un papel importante en el desarrollo de la condición estudiada, por lo que sería recomendable, de cara a futuras investigaciones validar otras herramientas que las valoren. En general, nuestros resultados indican que medidas breves como las utilizadas pueden ser de utilidad tanto en el contexto forense, como en el ámbito de la investigación, donde se pueden utilizar con sistemas de evaluación como el de Bianchini *et al.* (2005).

Además del protocolo mencionado, se han evaluado dos herramientas adicionales, un *checklist* de indicadores y el SIMS. En lo relativo al primer instrumento, nuestros resultados muestran que 18 de 22 indicadores podrían discriminar entre grupos adecuadamente, siendo la dimensión 2 “Actitud ante la situación de enfermedad” la que presentó la mayor capacidad de discriminación. Creemos que la mayoría de indicadores examinados pueden ser de utilidad para el contexto forense, pero consideramos necesario replicar estos resultados con un grupo de simuladores genuinos para asegurar su validez.

Por otro lado, en cuanto al SIMS, nuestros hallazgos sugieren que los puntos de corte propuestos por los autores de las versiones inglesa y española (14 y 16) producen

valores de sensibilidad y especificidad moderados, con varias subescalas, como Psicosis, que no ha resultado de utilidad para discriminar entre grupos. Creemos que, si bien las propiedades psicométricas de la herramienta para detectar fingimiento en trastornos mentales son excelentes (Rogers y Bender, 2018; van Impelen *et al.*, 2014), para futuras investigaciones sería recomendable intentar adaptar el instrumento a las particularidades de los contextos para los que no fue diseñado originalmente. En este sentido, podría sustituirse escalas no relevantes para las condiciones por otras que estudien sintomatología común en contexto, como podría ser el dolor. No solo podrían incorporarse ítems relacionados con la severidad del dolor, sino que también se podrían incluir factores de corte psicosocial relacionados con la experiencia del dolor como la catastrofización o el miedo al movimiento.

14. Referencias

- Aguerrevere, L. E., Calamia, M. R., Greve, K. W., Bianchini, K. J., Curtis, K. L. y Ramirez, V. (2018). Clusters of financially incentivized chronic pain patients using the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF). *Psychological Assessment*, 30(5), 634–644. <https://doi.org/10.1037/pas0000509>
- *⁵Aguerrevere, L. E., Greve, K. W., Bianchini, K. J. y Meyers, J. E. (2008). Detecting malingering in traumatic brain injury and chronic pain with an abbreviated version of the Meyers Index for the MMPI-2. *Archives of clinical Neuropsychology*, 23(7-8), 831–838. <https://doi.org/10.1016/j.acn.2008.06.008>
- Aguerrevere, L. E., Greve, K. W., Bianchini, K. J. y Ord, J. S. (2011). Classification accuracy of the Millon Clinical Multiaxial Inventory–III modifier indices in the detection of malingering in traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, 33(5), 497–504. <https://doi.org/10.1080/13803395.2010.535503>
- Aguilera, M., Paz, C., Compañ, V., Medina, J. C. y Feixas, G. (2019). Cognitive rigidity in patients with depression and fibromyalgia. *International Journal of Clinical and Health Psychology*, 19, 160-164. <https://doi.org/10.1016/j.ijchp.2019.02.002>
- Åhman, S. y 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, 1245-1249. <https://doi.org/10.2147/NDT.S4104>

⁵ Las referencias marcadas con “*” son las incluidas en la revisión sistemática.

- Akca, Y. E. A., Slootmaekers, L. y Boskovic, I. (2020). Verifiability and Symptom Endorsement in Genuine, Exaggerated, and Malingered Pain. *Psychological Injury and Law*, 13, 235–245. <https://doi.org/10.1007/s12207-020-09375-w>
- Akehurst, L., Easton, S., Fuller, E., Drane, G., Kuzmin, K. y Litchfield, S. (2015). An evaluation of a new tool to aid judgements of credibility in the medico-legal setting. *Legal and Criminological Psychology*, 22(1), 22-46. <https://doi.org/10.1111/lcrp.12079>
- Al-garadi, M. A., Khan, M. S., Varathan, K. D., Mujtaba, G. y Al-Kabsi, A. M. (2016). Using online social networks to track a pandemic: A systematic review. *Journal of Biomedical Informatics*, 62, 1-11. <https://doi.org/10.1016/j.jbi.2016.05.005>
- Alonso, J., Prieto, L. y Antó, J. M. (1995). La version española del "SF-36 Health Survey" (Cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos. *Medicina Clinica*, 104(20), 771-776.
- Alozai, U. U. y McPherson, P. K. (2020). *Malingering*. Disponible en: <https://www.ncbi.nlm.nih.gov/books/NBK507837/>
- Altman, D. G., Machin, D., Bryant, T. N. y Gardner, M. J. (Eds.). (2000). *Statistics with confidence, 2nd ed.* BMJ Books.
- Amado, B. G., Arce, R. y Fariña, F. (2015). Undeutsch hypothesis and criteria-based content analysis: A meta-analytic review. *The European Journal of Psychology Applied to Legal Context*, 7(1), 3-12. <http://dx.doi.org/10.1016/j.ejpal.2014.11.002>
- Amado, B. G., Arce, R., Fariña, F. y Vilariño, M. (2016). Criteria-based content analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal*

of Clinical and Health Psychology, 16(2), 201-210.

<https://doi.org/10.1016/j.ijchp.2016.01.002>

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders revised (4th ed.)*. Masson.

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

American Psychological Association. (2002). *Ethical principles of psychologists and code of conduct*. Autor.

American Psychological Association. (2010). *Ethical principles for psychologists and code of conduct*. Autor

American Psychological Association. (2020). *Publication manual of the American Psychological Association: the official guide to APA style*. Autor.

Andrade Ortega, J. A., Delgado Martínez, A. D. y Alméjida Ruiz, R. (2010). Validation of the Spanish version of the Neck Disability Index. *Spine*, 35(4), E114–E118.
<https://doi.org/10.1097/BRS.0b013e3181afea5d>

Aparicio, V. A., Ortega, F. B., Carbonell-Baeza, A., Cuevas, A. M., Delgado-Fernández, M. y Rodríguez, R. (2013). Anxiety, depression and fibromyalgia pain and severity, *Behavioral Psychology / Psicología Conductual*, 21, 381-392.

Arce, R. (2017). Análisis de contenido de las declaraciones de testigos: Evaluación de la validez científica y judicial de la hipótesis y la prueba. *Acción Psicológica*, 14(2), 171-190. <https://doi.org/10.5944/ap.14.1.21347>

Arce, R. y Fariña, F. (2001). *Construcción y validación de un procedimiento basado en una tarea de conocimiento para la medida de la huella psíquica en víctimas de*

- delitos: la entrevista forense* (Manuscrito inédito). Universidad de Santiago de Compostela.
- Arce, R. y Fariña, F. (2004) Obtención del testimonio y evaluación de la veracidad de las declaraciones. En G. Buela-Casal, D. Bunce y E. Jiménez (Eds.), *Manual de Psicología Forense*. Biblioteca Nueva.
- Arce, R. y Fariña, F. (2005). El Sistema de Evaluación Global (SEG) de la credibilidad del testimonio: Hacia una propuesta integradora. En R. Arce, F. Fariña, y M. Novo (Eds.), *Psicología jurídica*. Consellería de Xustiza, Interior e Administración Local.
- Arce, R. y Fariña, F. (2015). Evaluación psicológico-forense de la credibilidad y daño psíquico mediante el Sistema de Evaluación Global. En P. Rivas y G.L. Barrios (Eds.), *Violencia de género: Perspectiva multidisciplinar y práctica forense*. Thomson Aranzadi.
- Arce, R., Fariña, F., Carballal, A. y Novo, M. (2009). Creación y validación de un protocolo de evaluación forense de las secuelas psicológicas de la violencia de género. *Psicothema*, 21, 241-247.
- Arce, R., Fariña, F. y Quintero, I. (2013). Evaluación forense de la simulación en casos de acoso escolar. *Aula Abierta*, 41(2), 67-74.
- Arce, R., Fariña, F. y Vilariño, M. (2015). Daño psicológico en casos de víctimas de violencia de género: estudio comparativo de las evaluaciones forenses. *Revista Iberoamericana de Psicología y Salud*, 6(2), 72–80.
<https://doi.org/10.1016/j.rips.2015.04.002>
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V. y Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of*

- Personality Assessment*, 87(1), 84–94.
https://doi.org/10.1207/s15327752jpa8701_07
- Arregui-Dalmases, C., Teijeira, R., Rebollo-Soria, M. C., Kerrigan, J. R. y Crandall, J. R. (2011). La biomecánica del impacto: una herramienta para la medicina legal y forense en la investigación del accidente de tráfico. *Revista Española de Medicina Legal*, 37(3), 97–104. [https://doi.org/10.1016/s0377-4732\(11\)70071-2](https://doi.org/10.1016/s0377-4732(11)70071-2)
- Arola, H. M., Nicholls, E., Mallen, C. y Thomas, E. (2010). Self-reported pain interference and symptoms of anxiety and depression in community-dwelling older adults: can a temporal relationship be determined? *European Journal of Pain*, 14, 966-971. <https://doi.org/10.1016/j.ejpain.2010.02.012>
- Atherton, K., Wiles, N. J., Lecky, F. E., Hawes, S. J., Silman, A. J., Macfarlane, G. J. y Jones, G. T. (2006). Predictors of persistent neck pain after whiplash injury. *Emergency Medicine Journal*, 23(3), 195–201. <https://doi.org/10.1136/emj.2005.027102>
- Ato, M., López, J. J. y Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <http://doi.org/10.6018/analesps.29.3.178511>
- Axelrod, B. N., Meyers, J. E. y Davis, J. J. (2014). Finger Tapping Test performance as a measure of performance validity. *The Clinical Neuropsychologist*, 28(5), 876–888. <https://doi.org/10.1080/13854046.2014.907583>
- Bass, C. y Halligan, P. (2014). Factitious disorders and malingering: challenges for clinical assessment and management. *Lancet*, 383(9926), 1422–1432. [https://doi.org/10.1016/S0140-6736\(13\)62186-8](https://doi.org/10.1016/S0140-6736(13)62186-8)

- Beck, A. T., Brown, G., Epstein, N. y Steer, R. A. (1988). An Inventory for Measuring Clinical Anxiety: Psychometric Properties. *Journal of Consulting and Clinical Psychology, 56*, 893-897. <https://doi.org/10.1037/0022-006X.56.6.893>
- Beck, A. T., Steer, R. A., Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.
- Becke, M., Fuermaier, A. B. M., Buehren, J., Weisbrod, M., Aschenbrenner, S., Tucha, O. y Tucha, L. (2019). Utility of the Structured Interview of Reported Symptoms (SIRS-2) in detecting feigned adult attention-deficit/hyperactivity disorder. *Journal of Clinical and Experimental Neuropsychology, 41*(8), 1–17. <https://doi.org/10.1080/13803395.2019.1621268>
- 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. y 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>
- Bender, S. y Frederick, R. (2018). Neuropsychological Models of Feigned Cognitive Deficits. En R. Rogers y S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4ta ed.). Guilford.
- Ben-Porath, Y. S. y Tellegen, A. (2008). *MMPI-2: Restructured Form (MMPI-2-RF) manual for administration*. University of Minnesota Press.
- *Bianchini, K. J., Aguerrevere, L. E., Guise, B. J., Ord, J. S., Etherton, J. L., Meyers, J. E., Soignier, R. D., Greve, K. W., Curtis, K. L. y Buy, J. (2014). Accuracy of the modified somatic perception questionnaire and pain disability index in the detection of malingered pain-related disability in chronic pain. *The Clinical*

Neuropsychologist, 28(8), 1376-1394.

<https://doi.org/10.1080/13854046.2014.986199>

*Bianchini, K. J., Etherton, J. L., Greve, K. W., Heinly, M. T. y Meyers, J. E. (2008).

Classification accuracy of MMPI-2 validity scales for the detection of pain-related malingering: A known groups study. *Assessment*, 15(4), 435–449.

<https://doi.org/10.1177/1073191108317341>

Bianchini, K. J., Greve, K. W. y Glynn, G. (2005). On the diagnosis of malingered pain-

related disability: lessons from cognitive malingering research. *The Spine*

Journal, 5(4), 404-417. <https://doi.org/10.1016/j.spinee.2004.11.016>

Binder, L. M. y Rohling, M. L. (1996). Money matters: a meta-analytic review of the

effects of financial incentives on recovery after closed-head injury. *The American*

Journal of Psychiatry, 153(1), 7–10. <https://doi.org/10.1176/ajp.153.1.7>

Blasco Sáiz, J. L. y Pallardó Durá, L. (2013). Detección de exageración de síntomas

mediante el SIMS y el MMPI-2-RF en pacientes diagnosticados de trastorno mixto ansioso-depresivo y adaptativo en el contexto medicolegal: Un estudio preliminar.

Clínica y Salud, 24(3), 177–183. [https://doi.org/10.1016/S1130-5274\(13\)70019-7](https://doi.org/10.1016/S1130-5274(13)70019-7)

Boone, K. B. (2007). *Assessment of feigned cognitive impairment: A neuropsychological*

perspective. Guilford Press.

Booth, A., Clarke, M., Ghera, D., Moher, D., Petticrew, M. y Stewart, L. (2011). An

international registry of systematic-review protocols. *The Lancet*, 377(9760), 108-

9. [https://doi.org/10.1016/S0140-6736\(10\)60903-8](https://doi.org/10.1016/S0140-6736(10)60903-8)

- Boskovic, I., Bogaard, G., Merckelbach, H., Vrij, A. y Hope, L. (2017). The Verifiability Approach to detection of malingered physical symptoms. *Psychology, Crime & Law*, 23(8), 717–729. <https://doi.org/10.1080/1068316x.2017.1302585>
- Boskovic, I., Gallardo, C. T., Vrij, A., Hope, L. y Merckelbach, H. (2018). Verifiability on the run: An experimental study on the verifiability approach to malingered symptoms. *Psychiatry, Psychology and Law*, 26(1), 1–12. <https://doi.org/10.1080/13218719.2018.1483272>
- Bostick, G. P., Carroll, L. J., Brown, C. A., Harley, D. y Gross, D. P. (2013). Predictive capacity of pain beliefs and catastrophizing in Whiplash Associated Disorder. *Injury*, 44(11), 1465–1471. <https://doi.org/10.1016/j.injury.2012.10.007>
- Broadbent, E., Petrie, K.J., Main, J. y Weinman, J. (2006). The brief illness perception questionnaire. *Journal of Psychosomatic Research*, 60, 631-637. <https://doi.org/10.1016/j.jpsychores.2005.10.020>
- Bruns, D. y Disorbio, J. M. (2003). *Battery for Health Improvement 2 (BHI-2) manual*. Pearson.
- Buela-Casal, G., Guillén-Riquelme, A. y Seisdedos Cubero, N. (2011). *Cuestionario de ansiedad estado-rasgo*. Octava edición. TEA ediciones S.A.U.
- Bull, R. (2014). *Investigative Interviewing*, Springer.
- Burgess, J. A. y Dworkin, S. F. (1993). Litigation and post-traumatic TMD: how patients report treatment outcome. *Journal of the American Dental Association*, 124(6), 105–110. <https://doi.org/10.14219/jada.archive.1993.0145>
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R. y Silver, C. H. (2005). Symptom validity assessment: practice issues and

- medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, 20(4), 419–426. <https://doi.org/10.1016/j.acn.2005.02.002>
- Butcher, J. N., Dahlstrom, W., Graham, J. R., Tellegen, A. y Kaemmer, B. (1989). *Manual for administering and scoring the MMPI-2*. Minneapolis: University of Minnesota Press.
- Caillet, R. (2006). *Whiplash-associated diseases*. American Medical Association.
- Cameron, P. y Gabbe, B. (2009). The effect of compensation claims on outcomes after injury. *Injury*, 40(9), 905–906. <https://doi.org/10.1016/j.injury.2009.06.309>
- Capilla Ramírez, P. y González-Ordi, H. (2009). Protocolo para la detección de la simulación del dolor en la práctica clínica: estudio de casos. *Trauma*, 20, 255–263.
- Capilla Ramírez, P., González-Ordi, H. y Santamaría Fernández (2008). *Detección de simulación en dolor crónico*. Comunicación presentada en el VIII Congreso Nacional de la Sociedad Española de Traumatología Laboral – SETLA. Logroño, 27-29 de noviembre de 2008.
- Capilla Ramírez, P., González-Ordi, H., Santamaría Fernández, P. y Casado Morales, M. I. (2014). Detección de exageración de síntomas en esguince cervical: pacientes clínicos versus sujetos análogos. *Trauma*, 24(1), 4-10.
- Cardenal, V. y Sánchez, M. P. (2007). *MCMI-III. Inventario Clínico Multiaxial de Millon-III. Manual*. TEA Ediciones, S.A.U.
- Carroll L. J. (2011). Beliefs and expectations for recovery, coping, and depression in whiplash-associated disorders: lessening the transition to chronicity. *Spine*, 36(25), S250–S256. <https://doi.org/10.1097/BRS.0b013e31823881a4>

- Carroll, L. J., Hogg-Johnson, S., van der Velde, G., Haldeman, S., Holm, L. W., Carragee, E. J., Hurwitz, E. L., Côté, P., Nordin, M., Peloso, P. M., Guzman, J. y Cassidy, J. D. (2008). Course and prognostic factors for neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine*, 33(4), S75–S82. <https://doi.org/10.1097/BRS.0b013e31816445be>
- Carroll, L. J., Holm, L. W., Ferrari, R., Ozegovic, D. y Cassidy, J. D. (2009). Recovery in whiplash-associated disorders: do you get what you expect? *The Journal of Rheumatology*, 36(5), 1063–1070. <https://doi.org/10.3899/jrheum.080680>
- Casey, P. P., Feyer, A. M. y Cameron, I. D. (2015). Course of recovery for whiplash associated disorders in a compensation setting. *Injury*, 46, 2118-2129. <https://www.doi.org/10.1016/j.injury.2015.08.038>
- Cassidy, J. D., Carroll, L. J., Côté, P., Lemstra, M., Berglund, A. y Nygren, A. (2000). Effect of eliminating compensation for pain and suffering on the outcome of insurance claims for whiplash injury. *The New England journal of Medicine*, 342(16), 1179–1186. <https://doi.org/10.1056/NEJM200004203421606>
- Cerda, J. y Cifuentes, L. (2012). Uso de curvas ROC en investigación clínica: Aspectos teórico-prácticos. *Revista Chilena de Infectología*, 29(2), 138–141. <https://doi.org/10.4067/s0716-10182012000200003>
- Cernovsky, Z., Ferrari, J. y Mendonca, J. (2019a). Deceptive clinical diagnosing of malingering via Structured Inventory of Malingered Symptomatology. *Archives of Psychiatry and Behavioral Sciences*, 2(1), 44-49.
- Cernovsky, Z., Ferrari, J. y Mendonca, J. (2019b). Pseudodiagnoses of malingering of neuropsychological symptoms in survivors of car accidents by the Structured

- Inventory of Malingered Symptomatology. *Archives of Psychiatry and Behavioral Sciences*, 2(1), 55-65.
- Chappuis, G. y Soltermann, B. (2008). Number and cost of claims linked to minor cervical trauma in Europe: results from the comparative study by CEA, AREDOC and CEREDOC. *European Spine Journal*, 17(10), 1350–1357.
<https://doi.org/10.1007/s00586-008-0732-8>
- Cleeland, C. S. (1989). Measurement of pain by subjective report. En C. R. Chapman y J.D. Loeser (Eds.), *Issues in Pain Measurement* (pp. 391-403). Raven Press.
- Cleeland, C. S. (1990). Assessment of pain in cancer: measurement issues. En K. M. Foley, J.J. Bonica y 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. En D. Osoba (Ed.), *Effect of Cancer on Quality of Life* (pp. 293-305). CRC Press, Inc.
- Cleeland, C.S. y Ryan, K.M. (1994). Pain assessment: global use of the Brief Pain Inventory. *Annals of the Academy of Medicine, Singapore*, 23, 129-138.
<https://www.doi.org/10.4103/0973-1075.84531>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Earlbaum Associates.
- Colegio Oficial de Psicólogos de Madrid (2004). *Código deontológico del psicólogo. Reglamento de la comisión deontológica del Colegio Oficial de Psicólogos de Madrid*.

- Correa, A. y Rogers, R. (2012). Entrevistas estructuradas. En H. González-Ordi, P. Santamaría Fernández y P. Capilla Ramírez (Coords.). *Estrategias de detección de la simulación* (pp. 153-184). TEA Ediciones, S.A.U.
- Crighton, A. H., Marek, R. J., Dragon, W. R. y Ben-Porath, Y. S. (2016). Utility of the MMPI-2-RF Validity Scales in Detection of Simulated Underreporting: Implications of Incorporating a Manipulation Check. *Assessment*, 24(7), 853–864. <https://doi.org/10.1177/1073191115627011>
- *Crighton, A. H., Wygant, D. B., Applegate, K. C., Umlauf, R. L. y 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. *The Spine Journal*, 14(9), 2042–50. <https://doi.org/10.1016/j.spinee.2014.04.012>
- Crombez, G., Eccleston, C., Van Damme, S., Vlaeyen, J. W. S. y Karoly, P. (2012). Fear-Avoidance Model of Chronic Pain. *The Clinical Journal of Pain*, 28(6), 475–483. <https://doi.org/10.1097/ajp.0b013e3182385392>
- Crouch, R., Whitewick, R., Clancy, M., Wright, P. y 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. y Nicks, R. C. (2019). Detecting malingered pain-related disability with the Pain Catastrophizing Scale: A criterion groups validation study. *The Clinical Neuropsychologist*, 33(8), 1485-1500. <https://doi.org/10.1080/13854046.2019.1575470>

- Daubert, S. D. y Metzler, A. E. (2000). The detection of fake-bad and fake-good responding on the Millon Clinical Multiaxial Inventory III. *Psychological Assessment*, 12(4), 418–424. <https://doi.org/10.1037/1040-3590.12.4.418>
- De Andrés Ares, J., Cruces Prado, L. M., Canos Verdecho, M. A., Penide Villanueva, L., Del Valle Hoyos, M., Herdman, M., Traseira Lugilde, S. y Velázquez Rivera, I. (2015). Validation of the Short Form of the Brief Pain Inventory (BPI-SF) in Spanish Patients with Non-Cancer-Related Pain. *Pain Practice*, 15(7), 643–653. <https://doi.org/10.1111/papr.12219>
- Delis, D. C., Kramer, J. H., Kaplan, E. y Ober, B. A. (1987). *California Verbal Learning test*. Psychological Corporation.
- De Marchi, B. y Balboni, G. (2018). Detecting malingering mental illness in forensics: Known-Group Comparison and Simulation Design with MMPI-2, SIMS and NIM. *PeerJ*, 6, e5259. <https://doi.org/10.7717/peerj.5259>
- DePaulo, B., Lindsay, J., Malone, B., Muhlenbruck, L., Charlton, K. y Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129(1), 47-118. <https://www.doi.org/10.1037/0033-2909>
- De Rivera, L. y Abuín, M. R. (2018). Listado de Síntomas Breves – LSB-50 (2da edición). Tea Ediciones S.A.U.
- Derogatis, L. R. (1983). *SCL-90-R Administration, Scoring and Procedures Manual-II*. Towson, MD: Clinical Psychometric Research.
- Derogatis, L. R., Lipman, R. S. y Covi, L. (1973). SCL-90: an outpatient psychiatric rating scale--preliminary report. *Psychopharmacology bulletin*, 9(1), 13–28.

- Detullio, D., Messer, S. C., Kennedy, T. D. y Millen, D. H. (2019). A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST). *Psychological Assessment, 31*(11), 1319–1328. <https://doi.org/10.1037/pas0000753>
- Díaz-Bravo, L., Torruco-García, U., Martínez-Hernández, M. y Varela-Ruiz, M. (2013). La entrevista, recurso flexible y dinámico. *Investigación en educación médica, 2*(7), 162-167. [https://doi.org/10.1016/S2007-5057\(13\)72706-6](https://doi.org/10.1016/S2007-5057(13)72706-6)
- Duckworth, M. P. y Iezzi, T. (2018). Motor Vehicle Collisions and Their Consequences—Part II: Predictors of Impairment and Disability. *Psychological Injury and the Law, 11*, 288–306. <https://doi.org/10.1007/s12207-018-9334-x>
- Easton, S. y Akehurst, L. (2011). Tools for the detection of lying and malingering in the medico-legal interview setting. *The Medico-Legal Journal, 79*(3), 103-108.
- Elliot, J. M., Courtney, M. D., Rademaker, A., Pinto, D., Sterling, M. M. y Parrish, T. D. (2015). The Rapid and Progressive Degeneration of the Cervical Multifidus in Whiplash: An MRI Study of Fatty Infiltration. *Spine, 40*(12), 694-700. <https://doi.org/10.1097/BRS.0000000000000891>
- Elliott, J. M., Noteboom, J. T., Flynn, T. W. y 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>
- Engel, G. L. (1977). The need for a new medical model: A challenge for biomedicine. *Science, 196*, 129–136.

- Espinosa-Sánchez, J. M., Conde-Negri, E. y López-Escámez, J. A. (2014). Caracterización de las alteraciones vestibulares en pacientes con síndrome de latigazo cervical. *Trauma*, 25(4), 181-187.
- Esteve Zarazaga, R. y Ramírez Maestre, C. (2013). Modelo de miedo-evitación del dolor. *Revista de Psicología de la Salud*, 1(1), 45-73.
- Etherton, J. (2014). Diagnosing malingering in chronic pain. *Psychological Injury and Law*, 7(4), 362–369. <https://doi.org/10.1007/s12207-014-920>
- *Etherton, J. L., Bianchini, K. J., Ciota, M. A., Heinly, M. T. y Greve, K. W. (2006). Pain, malingering, and the WAIS-III Working Memory Index. *Spine Journal*, 6(1), 61-71. <https://doi.org/10.1016/j.spinee.2005.05.382>
- *Etherton, J. L., Bianchini, K. J., Heinly, M. T. y Greve, K. W. (2006). Pain, malingering, and performance on the WAIS-III Processing Speed Index. *Journal of Clinical and Experimental Neuropsychology*, 28(7), 1218-37. <https://doi.org/10.1080/13803390500346595>
- Falla, D., Peolsson, A., Peterson, G., Ludvigsson, M. L., Soldini, E., Schneebeli, A. y 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>
- Faust, D., Ahern, D. C., Bridges, A. J. y Yonce, L. J. (2012). Assessment of malingering and falsification: Conceptual foundations and sources of error. En C. R. Reynolds y A. M. Horton, Jr. (Eds.), *Detection of malingering during head injury litigation* (p. 1–67). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-0442-2_1

- Fernández-Huerta, J. (1959). Medidas sencillas de lecturabilidad. *Consigna*, 214, 29-32.
- Ferrara, S. D., Ananian, V., Baccino, E., Banczerowski, P., Bordignon, D., Boscolo-Berto, R., Domenici, R., Gorriz Quevedo, J., Graw, M., Hell, W., Hernandez Cueto, C., Knudsen, P. J., Masiero, S., Montisci, M., Norelli, G. A., Pinchi, V., Raudys, R., Raul, J. S., Sterzik, V., Tessitore, E., ... Zoia, R. (2016). Whiplash-Associated Disorders: Clinical and medico-legal guidelines on the methods of ascertainment. *International Journal of Legal Medicine*, 130(1), 13–22. <https://doi.org/10.1007/s00414-015-1243-x>
- Ferrari, R., Kwan, O., Russell, A. S., Pearce, J. M. y Schrader, H. (1999). The best approach to the problem of whiplash? One ticket to Lithuania, please. *Clinical and Experimental Rheumatology*, 17(3), 321–326.
- Ferrari, R., Russell, A. S., Carroll, L. J. y Cassidy, J. D. (2005). A re-examination of the whiplash associated disorders (WAD) as a systemic illness. *Annals of the Rheumatic Diseases*, 64(9), 1337-42. <https://doi.org/10.1136/ard.2004.034447>
- Fishbain, D. A., Cole, B., Cutler, R., Lewis, J., Rosomoff, H. y Rosomoff, R. S. (2003). A structured evidence- based review on the meaning of nonorganic physical signs: Waddell signs. *Pain Medicine*, 4, 141-181. <https://doi.org/10.1046/j.1526-4637.2003.03015.x>
- *Fishbain, D. A., Cutler, R. B., Rosomoff, H. L. y Rosomoff, R. S. (2002). Does the conscious exaggeration scale detect deception within patients with chronic pain alleged to have secondary gain? *Pain Medicine*, 3(1), 39–46. <https://doi.org/10.1046/j.1526-4637.2002.02002.x>
- Fisher, R. P. y Geiselman, R. E. (1992). *Memory-enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Charles C. Thomas.

- Fordyce, W. E. (1976). *Behavioral method for chronic pain and illness*. Mosby.
- Florkowski C. M. (2008). Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *The Clinical Biochemist. Reviews*, 29(Suppl 1), S83–S87.
- Franzen, M. D., Iverson, G. L. y McCracken, L. M. (1990). The detection of malingering in neuropsychological assessment. *Neuropsychology Review*, 1(3), 247–279. <https://doi.org/10.1007/bf01112573>
- Frederick R. I. (2000). Mixed group validation: a method to address the limitations of criterion group validation in research on malingering detection. *Behavioral Sciences & the Law*, 18(6), 693–718. <https://doi.org/10.1002/bsl.432>
- Frederick, R. (2018). Feigned amnesia and memory problems. En R. Rogers y S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4ta edición). Guilford.
- Fritz, J. (2017). Toward improving outcomes in whiplash: implementing new directions of care. *Journal of Orthopaedic & Sports Physical Therapy*, 47(7), 447–448. <https://doi.org/10.2519/jospt.2017.0107>
- Gardner, I. A. y Greiner, M. (2006). Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. *Veterinary Clinical Pathology*, 35(1), 8–17. <https://doi.org/10.1111/j.1939-165x.2006.tb00082.x>
- Gatchel, R. J. (1996). The Diathesis–Stress Model and chronic pain: cause and effect relationships. En R. J. Gatchel y D. C. Turk (Eds.), *Psychological Approaches to Pain Management: A Practitioner’s Handbook*. Guilford.

- Gennarelli, T. A. y Wodzin, E. (2006). AIS 2005: a contemporary injury scale. *Injury*, 37(12), 1083–1091. <https://doi.org/10.1016/j.injury.2006.07.009>
- Giummarra, M. J., Baker, K. S., Ioannou, L., Gwini, S. M., Gibson, S. J., Arnold, C. A., Ponsford, J. y Cameron, P. (2017a). Associations between compensable injury, perceived fault and pain and disability 1 year after injury: a registry-based Australian cohort study. *BMJ Open*, 7(10), e017350. <https://doi.org/10.1136/bmjopen-2017-017350>
- Giummarra, M. J., Cameron, P. A., Ponsford, J., Ioannou, L., Gibson, S. J., Jennings, P. A. y Georgiou-Karistianis, N. (2016a). Return to Work After Traumatic Injury: Increased Work-Related Disability in Injured Persons Receiving Financial Compensation is Mediated by Perceived Injustice. *Journal of Occupational Rehabilitation*, 27(2), 173–185. <https://doi.org/10.1007/s10926-016-9642-5>
- Giummarra, M. J., Casey, S. L., Devlin, A., Ioannou, L. J., Gibson, S. J., Georgiou-Karistianis, N., Jennings, P. A., Cameron, P. A. y Ponsford, J. (2017b). Co-occurrence of posttraumatic stress symptoms, pain, and disability 12 months after traumatic injury. *Pain Reports*, 2(5), e622. <https://doi.org/10.1097/PR9.0000000000000622>
- Giummarra, M. J., Ioannou, L., Ponsford, J., Cameron, P. A., Jennings, P. A., Gibson, S. J. y Georgiou-Karistianis, N. (2016b). Chronic Pain Following Motor Vehicle Collision. *The Clinical Journal of Pain*, 32(9), 817–827. <https://doi.org/10.1097/ajp.0000000000000342>
- Glas, A. S., Lijmer, J. G., Prins, M. H., Bonsel, G. J. y Bossuyt, P. M. M. (2003). The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical*

Epidemiology, 56(11), 1129–1135. [https://doi.org/10.1016/s0895-4356\(03\)00177-](https://doi.org/10.1016/s0895-4356(03)00177-)

x

Goldberg, D. (1992). *General Health Questionnaire (GHQ-12)*. Windsor, UK: NFER-Nelson.

González-Ordi, H., Capilla Ramírez, P., Santamaría, P. y Casado Morales, M. I. (2012). Abordaje multidisciplinar para la detección de la simulación en lumbalgia crónica. *Trauma*, 23, 145–154.

González-Ordi, H. y Santamaría Fernández, P. (2008). Detection of malingering in clinical, medicolegal, and forensic settings. En H. González-Ordi y P. Santamaría Fernández (Eds.), *Inventario Estructurado de Simulación de Síntomas – SIMS Manual* (pp. 60–66). TEA Ediciones, S.A.U.

González-Ordi, H. y Santamaría, P. (2009). *Inventario Estructurado de Simulación de Síntomas – SIMS Manual*. TEA Ediciones, S.A.U.

González-Ordi, H., Santamaría Fernández, P. y Capilla Ramírez, P. (Eds.) (2012). *Estrategias de Detección de la Simulación. Un manual clínico multidisciplinar*. TEA Ediciones, S.A.U.

González-Ordi, H., Santamaría, P. y Fernández Marín, P. (2010). Precisión predictiva del Inventario Estructurado de Simulación de Síntomas – SIMS en el contexto médicolegal. *Edupsykhé*, 9(1), 3-22.

Gopinath, B., Jagnoor, J., Kifley, A., Nicholas, M., Blyth, F., Kenardy, J., Craig, A. y Cameron, I. D. (2019). Differential Predictors of Pain Severity Over 12 Months Following Noncatastrophic Injury Sustained in a Road Traffic Crash. *The Journal of Pain*, 20(6), 676–684. <https://doi.org/10.1016/j.jpain.2018.11.011>

- Gopinath, B., Jagnoor, J., Harris, I.A., Nicholas, M., Casey, P., Blyth, F., Maher, C. G. y Cameron I.D. (2015). Prognostic indicators of social outcomes in persons who sustained an injury in a road traffic crash. *Injury*, 46, 909-917.
<https://www.doi.org/10.1016/j.injury.2015.01.002>
- Gorny, I. y Merten, T. (2006). Symptom Information—Warning—Coaching. *Journal of Forensic Neuropsychology*, 4(4), 71–97.
https://www.doi.org/10.1300/j151v04n04_05
- Graham, J. R., Watts, D. y Timbrook, R. E. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment*, 57, 264–277.
- Greiffenstein, M. F., Baker, W. J. y Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6(3), 218–224.
<https://doi.org/10.1037/1040-3590.6.3.218>
- Greve, K. W., Bianchini, K. J., Black, F. W., Heinly, M. T., Love, J. M., Swift, D. A. y Ciota, M. (2006). Classification accuracy of the Test of Memory Malingering in persons reporting exposure to environmental and industrial toxins: Results of a known-groups analysis. *Archives of Clinical Neuropsychology*, 21(5), 439–448.
<https://doi.org/10.1016/j.acn.2006.06.004>
- *Greve, K. W., Bianchini, K. J., Etherton, J. L., Meyers, J., Curtis, K. y Ord, J. (2010). The Reliable Digit Span test in chronic pain: Classification accuracy in detecting malingered pain-related disability. *The Clinical Neuropsychologist*, 24(1), 137–152. <https://doi.org/10.1080/13854040902927546>
- *Greve, K. W., Bianchini, K. J., Etherton, J. L., Ord, J. y Curtis, K. (2009). Detecting malingered pain-related disability: Classification accuracy of the Portland Digit

- Recognition Test. *The Clinical Neuropsychologist*, 23(5), 850-869.
<https://doi.org/10.1080/13854040802585055>
- *Greve, K. W., Curtis, K. J. Bianchini, K. J. y Ord, J. (2009). Are the original and second edition of the California Verbal Learning Test equally accurate in detecting malingering? *Assessment*, 16(3), 237-248.
<https://doi.org/10.1177/1073191108326227>
- *Greve, K. W., Etherton, J., Ord, J., Bianchini, K. J. y Curtis, K. (2009). Detecting malingered pain-related disability: Classification accuracy of the Test of Memory Malingering. *The Clinical Neuropsychologist*, 23(7), 1250–1271.
<https://doi.org/10.1080/13854040902828272>.
- Greve, K. W., Ord, J. S., Bianchini, K. J. y Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation*, 90(7), 1117-26.
<https://doi.org/10.1016/j.apmr.2009.01.018>
- Guarino, L., Sojo, V. y Bethelmy, L. (2007). Adaptación y validación preliminar de la versión Hispana del cuestionario de estilos de afrontamiento (Coping Style Questionnaire, CSQ). *Psicología Conductual*, 15(2), 173–189.
- Guest, R., Tran, Y., Gopinath, B., Cameron, I. D. y Craig, A. (2018). Prevalence and psychometric screening for the detection of major depressive disorder and post-traumatic stress disorder in adults injured in a motor vehicle crash who are engaged in compensation. *BMC Psychology*, 6(1), 4. <https://doi.org/10.1186/s40359-018-0216-5>
- Guise, B. J., Thompson, M. D., Greve, K. W., Bianchini, K. J. y West, L. (2014). Assessment of performance validity in the Stroop Color and Word Test in mild

- traumatic brain injury patients: A criterion-groups validation design. *Journal of Neuropsychology*, 8(1), 20–33. <https://doi.org/10.1111/jnp.12002>
- Gunner, J. H., Miele, A. S., Lynch, J. K. y McCaffrey, R. J. (2012). The Albany Consistency Index for the Test of Memory Malinger. *Archives of Clinical Neuropsychology*, 27(1), 1– 9. <https://doi.org/10.1093/arclin/acr089>
- Gur, R. C. y Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, 37(2), 147–169. <https://doi.org/10.1037/0022-3514.37.2.147>
- Hajian-Tilaki K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, 4(2), 627–635.
- Harris, I., Mulford, J., Solomon, M., van Gelder, J. M. y Young, J. (2005). Association between compensation status and outcome after surgery: a meta-analysis. *JAMA*, 293(13), 1644–1652. <https://doi.org/10.1001/jama.293.13.1644>
- Hawes, S. y Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment*, 21, 112–124 <https://doi.org/10.1037/a0015036>
- Hayashi, K., Miki, K., Ikemoto, T., Ushida, T. y Shibata, M. (2019). Factors influencing outcomes among patients with whiplash-associated disorder: A population-based study in Japan. *PLOS One*, 14(5), e0216857. <https://doi.org/10.1371/journal.pone.0216857>
- Heikkilä, H., Johansson, M. y Wenngren, B. I. (2000). Effects of acupuncture, cervical manipulation and NSAID therapy on dizziness and impaired head repositioning of

suspected cervical origin: a pilot study. *Manual Therapy*, 5(3), 151–157.
<https://doi.org/10.1054/math.2000.0357>

Hill A. B. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58(5), 295–300.

Ho, S. M. Y., Dai, D. W. T., Mak, C. y Liu, K. W. K. (2018). Cognitive factors associated with depression and anxiety in adolescents: A two-year longitudinal study. *International Journal of Clinical and Health Psychology*, 19, 115-123.
<https://doi.org/10.1016/j.ijchp.2019.01.001>

Holm, L. W., Carroll, L. J., Cassidy, J. D., Skillgate, E. y Ahlbom, A. (2008). Expectations for Recovery Important in the Prognosis of Whiplash Injuries. *PLOS Medicine*, 5(5), e105. <https://doi.org/journal.pmed.0050105>

Holmquist, L. A. y Wanlass, R. L. (2002). A multidimensional approach towards malingering detection. *Archives of Clinical Neuropsychology*, 17, 143-156.
[https://doi.org/10.1016/S0887-6177\(00\)00106-2](https://doi.org/10.1016/S0887-6177(00)00106-2)

*Hopwood, C. J., Orlando, M. J. y Clark, T. (2010). The detection of malingered pain-related disability with the Personality Assessment Inventory. *Rehabilitation Psychology*, 55(3), 307–310. <https://doi.org/10.1037/a0020516>

Horowitz, S. W. (1991). Empirical support for statement validity assessment. *Behavioural Assessment*, 13, 293- 313.

Hosmer, D. W. y Lemeshow, S. (2000). *Applied Logistic Regression, 2nd Ed.* John Wiley and Sons.

Ickmans, K., Malfliet, A., De Kooning, M., Goudman, L., Hubloue, I., Schmitz, T., Goubert, D. y Aguilar-Ferrandiz, M. E. (2017). Lack of Gender and Age

Differences in Pain Measurements Following Exercise in People with Chronic Whiplash-Associated Disorders. *Pain Physician*, 20(6), E829–E840.

Iles, R. A., Davidson, M. y Taylor, N. F. (2008). Psychosocial predictors of failure to return to work in non-chronic non-specific low back pain: a systematic review. *Occupational and Environmental Medicine*, 65(8), 507–517. <https://doi.org/10.1136/oem.2007.036046>

Ingram, P. B. y Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Reestructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, 30(4), 473–496. <https://doi.org/10.1080/13854046.2016.1187769>

Iverson, G. L. (2011). Positive Predictive Power. En J. S. Kreutzer, J. DeLuca y B. Caplan (Eds), *Encyclopedia of Clinical Neuropsychology*. Springer.

Jaspers, J. P. (1998). Whiplash and post-traumatic stress disorder. *Disability and Rehabilitation*, 20(11), 397–404. <https://doi.org/10.3109/09638289809166101>

Jelicic, M., Ceunen, E., Peters, M. J. V. y Merckelbach, H. (2011). Detecting coached feigning using the test of Memory Malingered (TOMM) and the structured inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, 67(9), 850–855. <https://doi.org/10.1002/jclp.20805>

Jull, G. (2016). Whiplash Continues Its Challenge. *The Journal of Orthopaedic and Sports Physical Therapy*, 46(10), 815–817. <https://doi.org/10.2519/jospt.2016.0112>

- Jull, G., Kenardy, J., Hendrikz, J., Cohen, M. y Sterling, M. (2013). Management of acute whiplash: a randomized controlled trial of multidisciplinary stratified treatments. *Pain*, *154*(9), 1798–1806. <https://doi.org/10.1016/j.pain.2013.05.041>
- Kamper, S. J., Rebeck, T. J., Maher, C. G., McAuley, J. H. y Sterling, M. (2008). Course and prognostic factors of whiplash: a systematic review and meta-analysis. *Pain*, *138*(3), 617–629. <https://doi.org/10.1016/j.pain.2008.02.019>
- Keefe, F. J., Rumble, M. E., Scipio, C. D., Giordano, L. A. y Perri, L. M. (2004). Psychological aspects of persistent pain: current state of the science. *The Journal of Pain*, *5*(4), 195–211. <https://doi.org/10.1016/j.jpain.2004.02.576>
- Kelley, K. y Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152. <https://doi.org/10.1037/a0028086>
- Kim, K., Shin, D. G., Park, M. K., Baik, S. H., Kim, T. H., Kim, S. y Lee, S. (2014). Circulating cell-free DNA as a promising biomarker in patients with gastric cancer: diagnostic validity and significant reduction of cfDNA after surgical resection. *Annals of Surgical Treatment and Research*, *86*(3), 136–142. <https://doi.org/10.4174/ast.2014.86.3.136>
- Kincaid, J. P., Fishburne, R. P., Rogers, R. L. y Chissom, B. S. (1975). Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for Navy enlisted personnel. *Research Branch Report*, 8–75.
- Köhnken, G. y Steller, M. (1988). The evaluation of the credibility of child witness statements in German procedural system. En G. Davies y J. Drinkwater (Eds.), *The child witness: Do the courts abuse children? (Issues in Criminological and Legal Psychology)* (13) (pp. 37–45). British Psychological Society.

- Kongsted, A., Bendix, T., Qerama, E., Kasch, H., Bach, F. W., Korsholm, L. y Jensen, T. S. (2008). Acute stress response and recovery after whiplash injuries. A one-year prospective study. *European Journal of Pain*, 12(4), 455–463. <https://doi.org/10.1016/j.ejpain.2007.07.008>
- Kraemer, H. C. (1992). Risk ratios, odds ratio, and the test QROC. En H.C. Kraemer, *Evaluating medical tests: Objective and quantitative guidelines*. SAGE Publications.
- Kropp, P. R. y Rogers, R. (1993). Understanding malingering: Motivation, method, and detection. En M. Lewis y C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 201–216). Guilford Press.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 26(4), 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lamb, S. E., Gates, S., Williams, M. A., Williamson, E. M., Mt-Isa, S., Withers, E. J., Castelnovo, E., Smith, J., Ashby, D., Cooke, M. W., Petrou, S., Underwood, M. R. y Managing Injuries of the Neck Trial (MINT) Study Team (2013). Emergency department treatments and physiotherapy for acute whiplash: a pragmatic, two-step, randomised controlled trial. *Lancet*, 381(9866), 546–556. [https://doi.org/10.1016/S0140-6736\(12\)61304-X](https://doi.org/10.1016/S0140-6736(12)61304-X)
- Lange, R. T. y Lippa, S. M. (2017). Sensitivity and specificity should never be interpreted in isolation without consideration of other clinical utility metrics. *The Clinical neuropsychologist*, 31(6-7), 1015–1028. <https://doi.org/10.1080/13854046.2017.1335438>
- Lanyon, R. I. (2004). Favorable self-presentation on psychological inventories: An analysis. *American Journal of Forensic Psychology*, 22(1), 53–65

- Lanyon, R. I. y Carle, A. C. (2007). Internal and External Validity of Scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement*, 67(5), 859–876. <https://doi.org/10.1177/0013164406299104>
- *Larrabee, G. J. (2003). Exaggerated Pain Report in Litigants With Malingered Neurocognitive Dysfunction. *The Clinical Neuropsychologist*, 17(3), 395-401. <https://doi.org/10.1076/clin.17.3.395.18087>
- Larrabee, G. J. (Ed.). (2007). *Assessment of malingered neuropsychological deficits*. Oxford University Press.
- Larrabee G. J. (2008). Aggregation across multiple indicators improves the detection of malingering: relationship to likelihood ratios. *The Clinical neuropsychologist*, 22(4), 666–679. <https://doi.org/10.1080/13854040701494987>
- Larrabee, G. J., Millis, S. R. y Meyers, J. E. (2009). 40 Plus or Minus 10, a New Magical Number: Reply to Russell. *The Clinical Neuropsychologist*, 23(5), 841–849. <https://doi.org/10.1080/13854040902796735>
- Larrabee, G. J. (2015). The multiple validities of neuropsychological assessment. *American Psychologist*, 70, 779–788. <https://www.doi.org/10.1037/a0039835>
- Lázaro, C., Bosch, F., Torrubia, R. y Baños, J. E. (1994). The development of a Spanish questionnaire for assessing pain: Preliminary data concerning reliability and validity. *European Journal of Psychological Assessment*, 10(2), 145–151.
- Leaver, A. M., Maher, C. G., McAuley, J. H., Jull, G., Latimer, J. y Refshauge, K. M. (2013). People seeking treatment for a new episode of neck pain typically have

- rapid improvement in symptoms: an observational study. *Journal of Physiotherapy*, 59, 31-37. [https://www.doi.org/10.1016/S1836-9553\(13\)70144-9](https://www.doi.org/10.1016/S1836-9553(13)70144-9)
- Leak, A. M., Cooper, J., Dyer, S., Williams, K. A., Turner-Stokes, L. y Frank, A. O. (1994). The Northwick Park Neck Pain Questionnaire, devised to measure neck pain and disability. *Rheumatology*, 33, 469–474. <https://doi.org/10.1093/rheumatology/33.5.469>.
- Leeuw, M., Goossens, M. E., Linton, S. J., Crombez, G., Boersma, K. y Vlaeyen, J. W. (2007). The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. *Journal of Behavioral Medicine*, 30(1), 77–94. <https://doi.org/10.1007/s10865-006-9085-0>
- Lees-Haley, P. R., English, L. T. y Glenn, W. J. (1991). A Fake Bad Scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, 68(1), 203–210. <https://doi.org/10.2466/pr0.1991.68.1.203>
- Ledesma, R., Macbeth, G. y Cortada de Kohan, N. (2008). Effect size: A conceptual review and applications with the ViSta statistical system. *Revista Latinoamericana de Psicología*, 40 (3), 425-439.
- Lethem, J., Slade, P. D., Troup, J. D. y Bentley, G. (1983). Outline of a Fear-Avoidance Model of exaggerated pain perception. *Behaviour Research and Therapy*, 21(4), 401–408. [https://doi.org/10.1016/0005-7967\(83\)90009-8](https://doi.org/10.1016/0005-7967(83)90009-8)
- Leventhal, H., Forster, R. S. y Leventhal, E. (2007). Self-regulation of health threats, affect and the self: lessons from the elderly. En C. Aldwin, C. L. Park, A. L. Spiro (Eds), *Handbook of health psychology and aging* (pp. 341–366). New York, NY: Guilford Publications.

- Leventhal, H., Leventhal, E. A. y Contrada, R. J. (1998). Self-regulation, health, and behavior: A perceptual–cognitive approach. *Psychology & Health, 13*(4), 717–733. <https://doi.org/10.1080/08870449808407425>
- Leventhal, H., Meyer, D. y Nerenz, D. (1980). The commonsense representation of illness danger. In S. Rachman (Ed.), *Contributions to medical psychology* (Vol. 2, pp 7-30). Oxford, UK: Pergamon.
- Lewis, J. L., Simcox, A. M. y Berry, D. T. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the structured inventory of malingered symptomatology. *Psychological Assessment, 14*(2), 170–176. <https://doi.org/10.1037//1040-3590.14.2.170>
- 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 – 84979.
- Lindley, S. E., Carlson, E. B. y Hill, K. R. (2014). Psychotic-like experiences, symptom expression, and cognitive performance in combat veterans with posttraumatic stress disorder. *The Journal of Nervous and Mental Disease, 202*(2), 91–96. <https://doi.org/10.1097/NMD.0000000000000077>
- *Linnell, M. y Easton, S. (2006). Malingering, perceptions of illness, and compensation seeking in whiplash injury: A comparison of illness beliefs between individuals in simulated compensation scenarios and litigation claimants. *Journal of Applied Social Psychology, 36*(11), 2619–2634. <https://doi.org/10.1111/j.0021-9029.2006.00119.x>

- Llor-Esteban, B. (1991). Estudio de la conducta de enfermedad en pacientes crónicos [Tesis Doctoral, Universidad de Murcia].
<https://dialnet.unirioja.es/servlet/tesis?codigo=259584>
- Luna, A. (2017). Nexo de causalidad en patología forense. *Ciencia Forense*, 14, 49-60.
- Lumley, M. A., Cohen, J. L., Borszcz, G. S., Cano, A., Radcliffe, A. M., Porter, L. S., Schubiner, H. y Keefe, F. J. (2011). Pain and emotion: a biopsychosocial review of recent research. *Journal of Clinical Psychology*, 67, 942–968.
<https://www.doi.org/10.1002/jclp.20816>
- Malfliet, A., De Kooning, M., Inghelbrecht, E., Hachimi-Idrissi, S., Willems, B., Bernheim, J. y Nijs, J. (2015). Sex Differences in Patients with Chronic Pain Following Whiplash Injury: The Role of Depression, Fear, Somatization, Social Support, and Personality Traits. *Pain Practice*, 15(8), 757–764.
<https://doi.org/10.1111/papr.12244>
- Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology*, 5(9), 1315–1316. <https://doi.org/10.1097/jto.0b013e3181ec173d>
- Manolopoulos, L., Vlastarakos, P. V., Georgiou, L., Giotakis, I., Loizos, A. y Nikolopoulos, T. P. (2008). Myofascial pain syndromes in the maxillofacial area: a common but underdiagnosed cause of head and neck pain. *International Journal of Oral and Maxillofacial Surgery*, 37, 975-984. <https://www.doi.org/10.1016/j.ijom.2008.04.023>
- Martin, A. L., Halket, E., Asmundson, G. J. G., Flora, D. B. y Katz, J. (2010). Posttraumatic Stress Symptoms and the Diathesis-Stress Model of Chronic

Pain and Disability in Patients Undergoing Major Surgery. *The Clinical Journal of Pain*, 26(6), 518–527. <https://doi.org/10.1097/ajp.0b013e3181e15b98>

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N. y Okut, H. (2019). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, 34(1), 88–119. <https://doi.org/10.1080/13854046.2019.1637027>

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)

McCaffrey, R. J. y Weber, M. (2000). A clinical approach to evaluating malingering in forensic neuropsychological evaluations. *Revista Española de Neuropsicología*, 3(2), 21-36.

McDermott, B.E. y Feldman, M.D. (2007). Malingering in the medical setting. *Psychiatric Clinics of North America*, 30, 645-662. <https://doi.org/10.1016/j.psc.2007.07.007>

*McGuire, B. E., Harvey, A. G. y Shores, E. A. (2001). Simulated malingering in pain patients: a study with the Pain Patient Profile. *British Journal of Clinical Psychology*, 40(1), 71-79. <https://doi.org/10.1348/014466501163490>

*McGuire, B. E. y Shores, E. A. (2001). Pain Patient Profile and the assessment of malingered pain. *Journal of Clinical Psychology*, 57(3), 401-409. <https://doi.org/10.1002/jclp.1022>

- *McGuire, B. E. y Shores, E. A. (2001b). Simulated pain on the Symptom Checklist 90-Revised. *Journal of Clinical Psychology*, 57(12),1589-96.
<https://doi.org/10.1002/jclp.1121>
- McNamara, L. A. y Martin, S. W. (2018). Principles of Epidemiology and Public Health. En S. S. Long, C. G. Prober y M. Fischer (Eds.), Principles and Practice of Pediatric Infectious Diseases (5ta edición). Elsevier.
- Meeus, M., Nijs, J., Hamers, V., Ickmans, K. y van Oosterwijcks, J. (2012). The efficacy of patient education in whiplash associated disorders: a systematic review. *Pain Physician Journal*, 15, 351-361.
- Melton, G., Petrila, J., Poythress, N. y Slobogin, C. (2007). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers (3rd ed.)*. Guilford.
- Melzack R. (1975). The McGill Pain Questionnaire: major properties and scoring methods. *Pain*, 1(3), 277-99. [https://doi.org/10.1016/0304-3959\(75\)90044-5](https://doi.org/10.1016/0304-3959(75)90044-5)
- Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelacic, M. y Otgaar, H. (2019). When Patients Overreport Symptoms: More Than Just Malingering. *Current Directions in Psychological Science*, 28(3), 321–326.
<https://doi.org/10.1177/0963721419837681>
- Merckelbach, H., Smeets, T. y Jelacic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *The Journal of Forensic Psychiatry & Psychology*, 20(3), 378-386. <https://doi.org/10.1080/14789940802456686>
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmand, B. A., Santamaría Fernández, P. y González-Ordi, H. (2013). Symptom validity assessment in European

- countries: Development and state of the art. *Clínica y Salud*, 24(3), 129–138.
<https://doi.org/10.5093/cl2013a14>
- Merten, T., Friedel, E., Mehren, G. y Stevens, A. (2007). Über die Validität von Persönlichkeitsprofilen in der nervenärztlichen Begutachtung. [Negative response bias and the validity of personality profiles in neuropsychiatric assessment] *Der Nervenarzt*, 78(5), 511–520. <https://doi.org/10.1007/s00115-006-2101-3>
- Merten, T. y Merckelbach, H. (2020). Factious disorders and malingering. En J.R. Geddes, N.C. Andreasen, y G.M. Goodwin (Eds.), *New Oxford Textbook of Psychiatry (3ª edición)*. Oxford University Press.
- Merten, T., Merckelbach, H., Giger, P. y Stevens, A. (2016). The Self-report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*, 9, 102–111.
<https://doi.org/10.1007/s12207-016-9257-3>.
- Meyers, J. E., Miller, R. M., Thompson, L. M., Scalese, A. M., Allred, B. C., Rupp, Z. W., Dupaix, Z. P. y Junghyun Lee, A. (2014). Using likelihood ratios to detect invalid performance with performance validity measures. *Archives of Clinical Neuropsychology*, 29(3), 224–235. <https://doi.org/10.1093/arclin/acu001>
- Meyers, J.E., Millis, S.R. y Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology*, 17(2), 157-169. [https://doi.org/10.1016/S0887-6177\(00\)00107-4](https://doi.org/10.1016/S0887-6177(00)00107-4)
- Miller, H. A. (2001). *Miller forensic assessment of symptoms test: Professional manual*. Psychological Assessment Resources.

- Millon, T. (1994). *The Millon Clinical Multiaxial Inventory-III manual*. National Computer Systems.
- Millon, T., Davis, R. D. y Millon, C. (1997). *The Millon Clinical Multiaxial Inventory-III manual (2ª ed.)*. National Computer Systems.
- Millon, T., Grossman, S. y Millon, C. (2015). *Millon Clinical Multiaxial Inventory-IV*. Pearson Assessments.
- Millon, T., Millon, C., Davis, R. y Grossman, S. (2009). *MCMII-III. Millon Clinical Multiaxial Inventory-III. Manual (4ª ed.)*. Pearson Clinical Assessment.
- Mittenberg, W., Patton, C., Canyock, E. M. y Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1094–1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. y Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLOS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Monaro, M., Toncini, A., Ferracuti, S., Tessari, G., Vaccaro, M. G., De Fazio, P., Pigato, G., Meneghel, T., Scarpazza, C. y Sartori, G. (2018). The detection of malingering: A new tool to identify made-up depression. *Frontiers in Psychiatry*, 9, 249. <https://doi.org/10.3389/fpsyt.2018.00249>
- Montes, O. y Guyton, M. R. (2014). Performance of Hispanic inmates on the Spanish Miller Forensic Assessment of Symptoms Test (M-FAST). *Law and Human Behavior*, 38(5), 428–438. <https://doi.org/10.1037/lhb0000074>

- Morey, L. C. (1991). *The Personality Assessment Inventory professional manual*. Psychological Assessment Resources, Inc.
- Morey, L. C. (1996). *An interpretive guide to the Personality Assessment Inventory (PAI)*. Psychological Assessment Resources.
- Morey, L. C. (2007). *The Personality Assessment Inventory professional manual (2nd ed.)*. Psychological Assessment Resources, Inc.
- Morgan, C. D., Schoenberg, M. R., Dorr, D. y Burke, M. J. (2002). Overreport on the MCMI-III: Concurrent validation with the MMPI-2 using a psychiatric inpatient sample. *Journal of Personality Assessment*, 78(2), 288–300. https://doi.org/10.1207/S15327752JPA7802_05
- Murgatroyd, D. F., Casey, P. P., Cameron, I. D. y Harris, I. A. (2015). The Effect of Financial Compensation on Health Outcomes following Musculoskeletal Injury: Systematic Review. *PLOS One*, 10(2), e0117597. <https://doi.org/10.1371/journal.pone.0117597>
- Nahari, G., Vrij, A. y Fisher, R.P. (2012). Does the truth come out in the writing? Scan as a lie detection tool. *Law and Human Behavior*, 36(1), 68–76. <https://www.doi.org/10.1037/h0093965>
- Nicholson, R. A., Mouton, G. J., Bagby, R. M., Buis, T., Peterson, S. A. y Buigas, R. A. (1997). Utility of MMPI–2 indicators of response distortion: Receiver operating characteristic analysis. *Psychological Assessment*, 9(4), 471–479. <https://doi.org/10.1037/1040-3590.9.4.471>

- Nieto, R., Miró, J. y Huguet, A. (2009). The fear-avoidance model in whiplash injuries. *European Journal of Pain*, 13(5), 518–523. <https://doi.org/10.1016/j.ejpain.2008.06.005>
- Nijs, J. e Ickmans, K. (2014). Chronic whiplash-associated disorders: to exercise or not? *The Lancet*, 384(9938), 109–111. doi:10.1016/s0140-6736(14)60130-6
- *Oddsdóttir, G. L., Kristjansson, E. y Gislason, M. K. (2015). Sincerity of effort versus feigned movement control of the cervical spine in patients with whiplash-associated disorders and asymptomatic persons: a case-control study. *Physiotherapy Theory and Practice*, 31(6), 403–409. <https://doi.org/10.3109/09593985.2015.1024299>
- Olivari Medina, C. y Urra Medina, E. (2007). Autoeficacia y conductas de salud. *Ciencia y Enfermería*, 13(1), 9-15. <https://doi.org/10.4067/s0717-95532007000100002>
- Olmedilla Zafra, A., Ortega Toro, E. y Abenza Cano, L. (2013). Validación de la escala de catastrofismo ante el dolor (Pain Catastrophizing Scale) en deportistas españoles. *Cuadernos de Psicología del Deporte*, 13(1), 83–93. <https://doi.org/10.4321/S1578-84232013000100009>
- Organización Mundial de la Salud (1996). *CIE 10: trastornos mentales y del comportamiento: versión multiaxial para adultos*. Meditor.
- Ortiz-Tallo, M., Cardenal, V., Ferragut, M. y Cerezo, M. V. (2011). Personalidad y síndromes clínicos: un estudio con el MCMI-III basado en una muestra española. *Revista de Psicopatología y Psicología Clínica*, 16, 49-59. <https://doi.org/10.5944/rppc.vol.16.num.1.2011.10350>

- Pastakia, K. y Kumar, S. (2011). Acute whiplash associated disorders (WAD). *Open Access Emergency Medicine*, 29, 29-32. <https://www.doi.org/10.2147/oaem.s17853>
- Patrick, C. J. y Iacono, W. G. (1989). Psychopathy, threat, and polygraph test accuracy. *Journal of Applied Psychology*, 74(2), 347–355. <https://doi.org/10.1037/0021-9010.74.2.347>
- Peace, K. A. y Richards, V. E. (2014). Faking it: incentives and malingered PTSD. *Journal of Criminal Psychology*, 4(1), 19–32. <https://doi.org/10.1108/JCP-09-2013-0023>
- Pedusic, Z., Pranic, S. y Jurakic, D. (2013). Relationship of back and neck pain with quality of life in the Croatian general population. *Journal of Manipulative & Physiological Therapeutics*, 36, 267-275. <https://www.doi.org/10.1016/j.jmpt.2013.05.012>
- Peng, Y. B., Fuchs, P. N. y Gatchel, R. J. (2006). Chronic Pain: the Diathesis-Stress Model. En A.B. Bengt y R. Ekman (Eds.), *Stress in Health and Disease*. Wiley-VCH.
- Phillips, L., Carroll, L., Cassidy, D. y Côté, P. (2010). Whiplash-associated disorders: who gets depressed? Who stays depressed? *European Spine Journal*, 19, 945–956. <https://doi.org/10.1007/s00586-010-1276-2>
- Pilowsky, I., Spence, N., Cobb, J. y Katsikitis, M. (1984). The Illness Behavior Questionnaire as an aid to clinical assessment. *General Hospital Psychiatry*, 6(2), 123–130. [https://doi.org/10.1016/0163-8343\(84\)90070-7](https://doi.org/10.1016/0163-8343(84)90070-7)

- Pincus, T., Smeets, R. J., Simmonds, M. J. y Sullivan, M. J. (2010). The fear avoidance model disentangled: improving the clinical utility of the fear avoidance model. *The Clinical Journal of Pain*, 26(9), 739–746. <https://doi.org/10.1097/AJP.0b013e3181f15d45>
- Pinfold, M., Niere, K. R., Hoving, J., Green, S., O’Leary, E. y Buchbinder, R. (2004). Validity and internal consistency of a whiplash-specific disability measure. *Spine*, 29, 263–268. <https://doi.org/10.1097/01.brs.0000107238.15526.4c>
- Pink, J., Petrou, S., Williamson, E., Williams, M. y Lamb, S.E. (2016). Economic and Health-Related Quality of Life Outcomes of Whiplash Associated Disorders. *Spine*, 4(17), 1378-1386. <https://doi.org/10.1097/BRS.0000000000001512>
- Pollard, C.A. (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>
- Pool, J.J., Ostelo, R.W., Knowl, D., Bouter, L.M. y de Vet, H.C. (2010). Are psychological factors prognostic indicators of outcome in patients with sub-acute neck pain? *Manual Therapy*, 15, 111-116. <https://www.doi.org/10.1016/j.math.2009.08.001>
- Pozzato, I., Craig, A., Gopinath, B., Kifley, A., Tran, Y., Jagnoor, J. y Cameron, I. D. (2020). Outcomes after traffic injury: mental health comorbidity and relationship with pain interference. *BMC Psychiatry*, 20(1), 189. <https://doi.org/10.1186/s12888-020-02601-4>
- Radley, A. y Green, R. (1987). Illness as adjustment: a methodology and conceptual framework. *Sociology of Health and Illness*, 9(2), 179–207. <https://doi.org/10.1111/1467-9566.ep11346972>

- Real Academia Española (2014). *Diccionario de la lengua española*, 23.^a ed. Autor.
- Regal Ramos, R. (2011). Síndrome de latigazo cervical. Características epidemiológicas de los pacientes evaluados en la Unidad Médica de Valoración de Incapacidades de Madrid. *Medicina y Seguridad del Trabajo*, 57(225), 348-360.
- Reiss, S. y McNally, R. J. (1985). Expectancy model of fear. En S. Reiss y R. R. Bootzin (Eds.), *Theoretical issues in behavior therapy*. Academic Press.
- Represas, C. (2017). Nexo de causalidad en accidentes de tráfico. *Ciencia Forense*, 14, 61-76.
- Represas, C., Muñoz, J. I. y Luna, A. (2016). Importancia de la biomecánica del impacto en la valoración pericial del síndrome del latigazo cervical. *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. y Llor-Esteban, B. (2020). Rating criteria of the causal nexus in whiplash associated disorders: A study of interrater reliability. *Revista Española de Medicina Legal*. <https://doi.org/10.1016/j.reml.2020.06.002>
- Represas, C., Vieira, D. N., Magalhães, T., Dias, R., Frazão, S., Suárez-Peñaranda, J. M., Rodríguez-Calvo, M. S., Concheiro, L. y Muñoz, J. I. (2008). No cash no whiplash?: Influence of the legal system on the incidence of whiplash injury. *Journal of Forensic and Legal Medicine*, 15(6), 353–355. <https://doi.org/10.1016/j.jflm.2007.12.016>
- Resnick, P. J. (1997). Malingered psychosis. En R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2da ed., pp. 47–67). Guilford Press.

- Resnick, P. J. y Knoll, J.L. (2008). Malingered psychosis. En R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3ra ed., pp. 34–53). Guilford Press.
- Ritchie, C. y Sterling, M. (2016). Recovery Pathways and Prognosis After Whiplash Injury. *Journal of Orthopaedic & Sports Physical Therapy*, 46(10), 851–861. <https://doi.org/10.2519/jospt.2016.6918>
- Roger, D., Jarvis, G. y Najarian, B. (1993). Detachment and coping: The construction and validation of a new scale for measuring coping strategies. *Personality and Individual Differences*, 15(6), 619-626. [https://doi.org/10.1016/0191-8869\(93\)90003-L](https://doi.org/10.1016/0191-8869(93)90003-L)
- Rogers, R. (1984). Towards an empirical model of malingering and deception. *Behavioral Sciences and the Law*, 2, 93-112.
- Rogers, R. (1986). *Structured interview of reported symptoms (SIRS)*. Clarke Institute of Psychiatry, Escala sin publicar.
- Rogers, R. (1988). *Clinical assessment of malingering deception* (1st ed.). Guilford.
- Rogers, R. (1997). *Clinical assessment of malingering deception* (2nd ed.). Guilford.
- Rogers, R. (2003). Standardizing DSM-IV Diagnoses: The Clinical Applications of Structured Interviews. *Journal of Personality Assessment*, 81(3), 220–225. https://doi.org/10.1207/s15327752jpa8103_04
- Rogers, R. (2008). *Clinical assessment of malingering and deception* (3rd ed.). Guilford.
- Rogers, R. y Bender, S. D. (2003). Evaluation of malingering and deception. En A. M. Goldstein (Ed.), *Comprehensive handbook of psychology: Forensic psychology* (Vol. 11, pp. 212-234). Wiley.

- Rogers, R. y Bender, S. D. (2012). Evaluation of malingering and related response styles. En R. K. Otto, I. B. Weiner, R. K. Otto y I. B. Weiner (Eds.), *Handbook of psychology: Forensic psychology* (pp. 517–540). Wiley.
- Rogers, R. y Bender, S. D. (2018). *Clinical assessment of malingering and deception* (4th ed.). Guilford.
- Rogers, R. y Cavanaugh, J. L. (1983). “Nothing but the Truth”... a Reexamination of Malingering. *The Journal of Psychiatry & Law*, 11(4), 443-459. <https://doi.org/10.1177/009318538301100404>
- Rogers, R. y 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. y Granacher, R. J. (2011). Conceptualization and assessment of malingering. En E. Y. Drogin, F. M. Dattilio, R. L. Sadoff y T. G. Gutheil (Eds.), *Handbook of forensic assessment: Psychological and Psychiatric Perspectives* (pp. 659–678). Wiley.
- Rogers, R. y Gillard, N. D. (2011). Research methods for the assessment of malingering. En B. Rosenfeld y S. D. Penrod (Eds.), *Research methods in forensic psychology* (p. 174–188). John Wiley & Sons, Inc.
- Rogers, R., Harrell, E.H. y Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, 13(3), 255-274. [https://doi.org/10.1016/0272-7358\(93\)90023-F](https://doi.org/10.1016/0272-7358(93)90023-F)

- Rogers, R., Hinds, J. D. y Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS - PB - Routledge. *Journal of Personality Assessment*, 67(2), 244–257. https://doi.org/10.1207/s15327752jpa6702_2
- Rogers, R. y Payne, J. W. (2006). Damages and rewards: assessment of malingered disorders in compensation cases. *Behavioral Sciences & the Law*, 24(5), 645–658. <https://doi.org/10.1002/bsl.687>
- Rogers, R., Robinson, E.V. y Gilliard, N. D. (2014). The SIMS screen for feigned mental disorders: The development of detection-based scales. *Behavioral Sciences & the Law*, 32(4), 455-466. <https://doi.org/10.1002/bsl.2131>
- Rogers, R., Sewell, K. W. y Gillard, N. D. (2010). *Structured interview of reported symptoms 2nd Edition (SIRS-2): Professional manual*. Psychological Assessment Resources.
- Rogers, R., Sewell, K. W., Martin, M. A. y Vitacco, M. J. (2003). Detection of feigned mental disorders: a meta-analysis of the MMPI-2 and malingering. *Assessment*, 10(2), 160–177. <https://doi.org/10.1177/1073191103010002007>
- Rogers, R. y Shuman, D. W. (2000). *Conducting insanity evaluations* (2nd ed.). New York, NY: Guilford Press.
- Rumschik, S. M. y Appel, J. M. (2019). Malingering in the Psychiatric Emergency Department: Prevalence, Predictors, and Outcomes. *Psychiatric Services*, 70(2), 115–122. <https://doi.org/10.1176/appi.ps.201800140>
- Ruscheweyh, R., Nees, F., Marziniak, M., Evers, S., Flor, H. y Knecht, S. (2011). Pain catastrophizing and pain-related emotions: influence of age and type of pain. *The*

Clinical Journal of Pain, 27(7), 578–586.

<https://doi.org/10.1097/AJP.0b013e31820fde1b>

Salech, F., Mery, V., Larrondo, F. y Rada, G. (2008). Estudios que evalúan un test diagnóstico: interpretando sus resultados. *Revista Médica De Chile*, 136(9). 10.4067/S0034-98872008000900018.

Samuel, R. Z. y Mittenberg, W. (2006). Determination of malingering in disability evaluations. *Primary Psychiatry*, 12(12), 60-68.

Samoborec, S., Ruseckaite, R., Ayton, D. y Evans, S. (2018). Biopsychosocial factors associated with non-recovery after a minor transport-related injury: A systematic review. *PLOS One*, 13(6), e0198352. <https://doi.org/10.1371/journal.pone.0198352>

Sánchez, G., Ampudia, A., Jiménez, F. y 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, 51-56. <https://doi.org/10.1016/j.ejpal.2017.03.002>

Santamaría Fernández, P. (2009). *Adaptación española del MMPI-2-RF*. TEA Ediciones S.A.U.

Santamaría Fernández, P. (2015) Utilidad diagnóstica del inventario estructurado de simulación de síntomas (SIMS) en población española. [Tesis Doctoral, Universidad Complutense de Madrid]. <https://eprints.ucm.es/28412/>

Santamaría Fernández, P., Capilla Ramírez, P. y González-Ordi, H. (2013). Prevalencia de simulación en incapacidad temporal: percepción de los profesionales de la

salud. *Clínica y Salud*, 24(3), 139–151.

<https://doi.org/10.1016/10.5093/cl2013a15>

Santamaría Fernández, P. y González-Ordi, H. (2012). Evaluación mediante autoinforme.

En H. González-Ordi, P. Santamaría Fernández y P. Capilla Ramírez (Coords.).

Estrategias de detección de la simulación (pp. 153-184). TEA Ediciones, S.A.U.

Sanz, J., Perdigón, A. L. y Vázquez, C. (2003). Adaptación española del Inventario para

la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población

general. *Clínica y Salud*, 14(3), 249-280.

Sarrami, P., Armstrong, E., Naylor, J. M. y Harris, I. A. (2016). Factors predicting

outcome in whiplash injury: a systematic meta-review of prognostic factors.

Journal of Orthopaedics and Traumatology, 18(1), 9–

16. <https://doi.org/10.1007/s10195-016-0431-x>

Sartori, G., Agosta, S. y Gnoato, F. (2007). *High accuracy detection of malingered*

whiplash syndrome. International Whiplash Trauma Congress.

Sartori, G., Forti, S., Birbaumer, N. y 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-](https://doi.org/10.1016/s0304-3940(03)00233-7)

3940(03)00233-7

*Schmand, B., Lindeboom, J., Schagen, S., Heijjt, R., Koene, T. y Hamburger, J. L.

(1998). Cognitive complaints in patients after whiplash injury: The impact of

malingering. *Journal of Neurology, Neurosurgery & Psychiatry*, 64(3), 339–342.

<https://doi.org/10.1136/jnnp.64.3.339>

- Schoenberg, M. R., Dorr, D. y Morgan, C. D. (2003). The ability of the Millon Clinical Multiaxial Inventory--Third Edition to detect malingering. *Psychological Assessment, 15*(2), 198–204. <https://doi.org/10.1037/1040-3590.15.2.198>
- Scholten-Peeters, G. G., Verhagen, A. P., Bekkering, G. E., van der Windt, D. A. W., Barnsley, L., Oostendorp, R. A. y Hendriks, E. J. (2003). Prognostic factors of whiplash-associated disorders: a systematic review of prospective cohort studies. *Pain, 104*(1), 303–322. [https://doi.org/10.14219/10.1016/s0304-3959\(03\)00050-2](https://doi.org/10.14219/10.1016/s0304-3959(03)00050-2)
- Schroeder, R. W., Buddin, W. H., Jr, Hargrave, D. D., VonDran, E. J., Campbell, E. B., Brockman, C. J., Heinrichs, R. J. y Baade, L. E. (2013). Efficacy of test of memory malingering Trial 1, Trial 2, the Retention Trial, and the Albany Consistency Index in a criterion group forensic neuropsychological sample. *Archives of clinical Neuropsychology, 28*(1), 21–29. <https://doi.org/10.1093/arclin/acs094>
- Schroeder, R. W., Twumasi-Ankrah, P., Baade, L. E. y Marshall, P. S. (2012). Reliable digit span: A systematic review and cross-validation. *Assessment, 19*(1), 21–30. <https://doi.org/10.1177/1073191111428764>
- Seijo, D., Fariña, F. y Vilariño, M. (2014). Procedimientos y técnicas para la evaluación psicológica forense. En S. Souto Camba, A. Souto Gestal, F. Fariña Rivera. *Salud y Bienestar* (pp. 25-54). GEU Editorial.
- Sellbom, M. y Bagby, R. M. (2008). Response styles on multiscale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., 182–206). New York: Guilford Press.
- Sellbom, M., Toomey, J. A., Wygant, D. B., Kucharski, L. T. y Duncan, S. (2010). Utility of the MMPI–2-RF (Restructured Form) validity scales in detecting malingering in

- a criminal forensic setting: A known-groups design. *Psychological Assessment*, 22(1), 22–31. <https://doi.org/10.1037/a0018222>
- Sherman, E., Slick, D. J. y 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/aaa019>
- Silverton, L. y Gruber, C. (1998). *Malingering Probability Scale (MPS)*. Western Psychological Services.
- Singh, J. (2013). Critical appraisal skills programme. *Journal of Pharmacology & Pharmacotherapeutics*, 4(1), 76-77. <https://doi.org/10.4103/0976-500X.107697>
- Slick, D. J. (2006). Psychometrics in neuropsychological assessment. En E. Strauss, E. M. S. Sherman y O. Spreen (Eds.), *A compendium of neuropsychological tests: Administration, norms and commentary* (pp. 3–43). Oxford University Press.
- Slick, D. J., Sherman, E. M. S. y Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13(4), 545–561. [https://doi.org/10.1076/1385-4046\(199911\)13:04;1-Y;FT545](https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545)
- Slick, D. J., Tan, J. E., Sherman, E. M. S. y Strauss, E. (2010). Malingering and related conditions in pediatric populations. En A. S. Davis, (Ed.), *Handbook of pediatric neuropsychology* (pp. 457–470). Springer.
- Smith, G.P. y Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy on Psychiatry and Law*, 25, 180–183.

- *Sobel, J.B., Sollenberger, P., Robinson, R., Polatin, P.B. y Gatchel, R.J. (2000). Cervical nonorganic signs: a new clinical tool to assess abnormal illness behavior in neck pain patients: a pilot study. *Archives of Physical Medicine and Rehabilitation*, 81(2), 170-75. [https://doi.org/10.1016/S0003-9993\(00\)90136-9](https://doi.org/10.1016/S0003-9993(00)90136-9)
- Sokol, M. C., McGuigan, K. A., Verbrugge, R. R. y Epstein, R. S. (2005). Impact of medication adherence on hospitalization risk and healthcare cost. *Medical care*, 43(6), 521–530. <https://doi.org/10.1097/01.mlr.0000163641.86870.af>
- Spearing, N. M. y Connelly, L. B. (2011). Is compensation "bad for health"? A systematic meta-review. *Injury*, 42(1), 15–24. <https://doi.org/10.1016/j.injury.2009.12.009>
- Spearing, N. M., Connelly, L. B., Gargett, S. y Sterling, M. (2012). Does injury compensation lead to worse health after whiplash? A systematic review. *Pain*, 153(6), 1274–1282. <https://doi.org/10.1016/j.pain.2012.03.007>
- Spielberger, C. D., Gorsuch, R. y Lushene, R. (1970). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologist Press.
- Spitzer, W. O., Skovron, M. L., Salmi, L. R., Cassidy, J. D., Duranceau, J., Suissa, S. y Zeiss, E. (1995). Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining "whiplash" and its management. *Spine*, 20(8), 1S–73S.
- Steller, M. y Köhnken, G. (1989). Criteria-based content analysis. En D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). Springer-Verlag.

- Sterling M. (2004). A proposed new classification system for whiplash associated disorders--implications for assessment and management. *Manual therapy*, 9(2), 60–70. <https://doi.org/10.1016/j.math.2004.01.006>
- 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>
- Sterling, M. (2014). Physiotherapy management of whiplash-associated disorders (WAD). *Journal of Physiotherapy*, 60(1), 5–12. <https://doi.org/10.1016/j.jphys.2013.12.004>
- Sterling, M., Jull, G. y Kenardy, J. (2006). Physical and psychological factors maintain long-term predictive capacity post-whiplash injury. *Pain*, 122(1-2), 102–108. <https://doi.org/10.1016/j.pain.2006.01.014>
- Sterling, M. y Kenardy, J. (2011). *Whiplash: evidence base for clinical practice*. Churchill Livingstone
- Sullivan, M. J. L., Bishop, S. R. y Pivik, J. (1995). The Pain Catastrophizing Scale: Development and validation. *Psychological Assessment*, 7(4), 524–532. <https://doi.org/10.1037/1040-3590.7.4.524>
- Sullivan, M.J., Stanish, W., Sullivan, M.W. y Tripp, D. (2002). Differential predictors of pain and disability in patients with whiplash injuries. *Pain Research and Management*, 7, 68-74. <https://doi.org/10.1155/2002/176378>
- Szigriszt Pazos, F. (1993) Sistemas predictivos de legibilidad del mensaje escrito: fórmula de perspicuidad. [Tesis Doctoral inedita, Universidad Complutense de Madrid]. <http://webs.ucm.es/BUCM/tesis//19911996/S/3/S3019601.pdf>

- *Tearman, B.H. y Ross, S.A. (2012). The development and classification accuracy of the life assessment questionnaire in the detection of pain-related malingering. *Behavioral Sciences & the Law*, 30(4), 516-36. <https://doi.org/10.1002/bsl.2028>
- Tejero, B. (2017). Adaptación transcultural y validación de la versión española del whiplash disability questionnaire (WDQ-SP). [Tesis doctoral, Universidad Miguel Hernández]. <http://dspace.umh.es/jspui/handle/11000/4862>
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 44(5), 423–432. <https://doi.org/10.1002/pits.20234>
- Thompson, C. M., Lin, H. y Parsloe, S. (2018). Misrepresenting Health Conditions Through Fabrication and Exaggeration: An Adaptation and Replication of the False Alarm Effect. *Health Communication*, 33(5), 562–575. <https://doi.org/10.1080/10410236.2017.1283563>
- Tollison, D.C. y Langley, J.C. (1995). *Pain Patient Profile manual*. Minneapolis: National Computer Services.
- Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. Multi-Health Systems, Inc.
- Trevethan R. (2017). Sensitivity, Specificity, and Predictive Values: Foundations, Pliabilities, and Pitfalls in Research and Practice. *Frontiers in Public Health*, 5, 307. <https://doi.org/10.3389/fpubh.2017.00307>
- Tuck, N. L., Johnson, M. H. y Bean, D. J. (2018). 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>

- Turk D. C. (2002). A diathesis-stress model of chronic pain and disability following traumatic injury. *Pain Research & Management*, 7(1), 9–19.
<https://doi.org/10.1155/2002/252904>
- Turk, D. C. y Holzman, A. D. (1986). Chronic pain: Interfaces among physical, psychological, and social parameters. En A. D. Holzman y D. C. Turk (Eds.), *Pergamon general psychology series, Vol. 136. Pain management: A handbook of psychological treatment approaches* (p. 1–9). Pergamon Press.
- Turk, D.C., Robinson, J.P. y Duckworth, M.P. (2018). The Impact of Psychosocial and Contextual Factors on Individuals Who Sustain Whiplash-Associated Disorders in Motor Vehicle Collisions. *Psychological Injury and Law* 11, 218-232.
<https://www.doi.org/10.1007/s12207-018-9317-y>
- Vachon-Preseau, E., Roy, M., Martel, M. O., Caron, E., Marin, M. F., Chen, J., Albouy, G., Plante, I., Sullivan, M. J., Lupien, S. J. y Rainville, P. (2013). The stress model of chronic pain: evidence from basal cortisol and hippocampal structure and function in humans. *Brain*, 136(Pt 3), 815–827.
<https://doi.org/10.1093/brain/aws371>
- Valera-Garrido, F., Martínez-Rodríguez, A., Medina-Mirapeix, D., Meseguer-Henarejos, A.B., Milán-Robles, F. y Campillo-Pomata, J.L. (2001). El modelo biopsicosocial en los síndromes de dolor vertebral: Implicaciones para la protocolización. *Revista Iberoamericana de Fisioterapia y Kinesiología*, 4(2), 81-87.
- Van Impelen, A., Merckelbach, H., Jelicic, M. y Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A Systematic Review and Meta-Analysis. *The Clinical Neuropsychologist*, 28(8), 1336–1365.
<https://doi.org/10.1080/13854046.2014.984763>

- Van Impelen, A., Merckelbach, H., Jelicic, M., Niesten, I. y Campo, J. À. (2017). Differentiating Factitious from Malingered Symptomatology: the Development of a Psychometric Approach. *Psychological Injury and Law*, 10(4), 341–357. <https://doi.org/10.1007/s12207-017-9301-y>
- Vernon, H. y Mior, S. (1991). The Neck Disability Index: A study of reliability and validity. *Journal of Manipulative and Physiological Therapeutics*. 14, 409-415.
- Vilagut, G., Ferrer, M., Rajmil, L., Rebollo, P., Permanyer-Miralda, G., Quintana, J. M., ... Alonso, J. (2005). El Cuestionario de Salud SF-36 español: una década de experiencia y nuevos desarrollos. *Gaceta Sanitaria*, 19(2), 135–150. <https://doi.org/10.1157/13074369>
- Vilar López, R., Aparicio, M., Gómez Río, M. y Pérez García, M. (2013). Utilidad de los índices de memoria verbal para detectar simulación en población española. *Clínica y Salud*, 24(3), 169–176. [https://doi.org/10.1016/s1130-5274\(13\)70018-5](https://doi.org/10.1016/s1130-5274(13)70018-5)
- Vilar-López, R., Pérez, M. y Puente, A. E. (2011). *Test de Simulación de Problemas de Memoria (TOMM)*. Manual. TEA Ediciones, S.A.U.
- Vilariño, M., Arce, R. y Fariña, F. (2013). Forensic-clinical interview: Reliability and validity for the evaluation of psychological injury. *European Journal of Psychology Applied to Legal Context*, 5(1), 1-21.
- Vitacco, M. J., Rogers, R., Gabel, J. y Munizza, J. (2007). An evaluation of malingering screens with competency to stand Trial patients: A known-groups comparison. *Law and Human Behavior*, 31(3), 249–260. <https://doi.org/10.1007/s10979-006-9062->

- Vlaeyen, J. W., Kole-Snijders, A. M., Rotteveel, A. M., Ruesink, R. y Heuts, P. H. (1995). The role of fear of movement/(re)injury in pain disability. *Journal of Occupational Rehabilitation*, 5(4), 235–252. <https://doi.org/10.1007/BF02109988>
- Vlaeyen, J.W. y Linton, S.J. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85, 317-332. [https://www.doi.org/10.1016/S0304-3959\(99\)00242-0](https://www.doi.org/10.1016/S0304-3959(99)00242-0)
- Vonk, J., Horlings, C. G. C. y Allum, J. H. J. (2010). Differentiating malingering balance disorder patients from healthy controls, compensated unilateral vestibular loss, and whiplash patients using stance and gait posturography. *Audiology & Neurotology*, 15(4), 261–272. <https://doi.org/10.1159/000258682>
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley.
- Waddell, G., Newton, M., Henderson, I., Somerville, D. y Main, C. J. (1993). A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain*, 52(2), 157–168. [https://doi.org/10.1016/0304-3959\(93\)90127-b](https://doi.org/10.1016/0304-3959(93)90127-b)
- Waeyaert, P., Jansen, D., Bastiaansen, M., Scafoglieri, A., Buyl, R., Schmitt, M. y Cattrysse, E. (2016). Three-dimensional cervical movement characteristics in healthy subjects and subgroups of chronic neck pain patients based on their pain location. *Spine*, 41(15), 908-914. <https://doi.org/10.1097/BRS.0000000000001500>
- *Wallis, B. J. y Bogduk, N. (1996). Faking a profile: can naive subjects simulate whiplash responses? *Pain*, 66(2), 223–227. [https://doi.org/10.1016/0304-3959\(96\)03044-8](https://doi.org/10.1016/0304-3959(96)03044-8)

- Walton, D. M. y Elliott, J. M. (2017). An Integrated Model of Chronic Whiplash-Associated Disorder. *The 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), 3130–39. <https://doi.org/10.1097/00007632-200012150-00008>
- Walczyk, J.J., Sewell, N.R. y Dibenedetto, M.B. (2018). A Review of Approaches to Detecting Malingering in Forensic Contexts and Promising Cognitive Load-Inducing Lie Detection Techniques. *Frontiers in Psychiatry*, 9, 700. <https://doi.org/10.3389/fpsyt.2018.00700>
- Weinman, J., Petrie, K.J., Moss-Morris, R. y 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>
- Wessely, S. (2003). Malingering: Historical perspectives. En P. W. Halligan, C. M. Bass, y D. A. Oakley (Eds.), *Malingering and illness deception* (pp. 31–41). Oxford University Press.
- Whiteside, D. M., Kogan, J., Wardin, L., Phillips, D., Franzwa, M. G., Rice, L., Basso, M. y Roper, B. (2015). Language-based embedded performance validity measures in traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, 37(2), 220–227. <https://doi.org/10.1080/13803395.2014.1002758>
- Widows, M. R. y Smith, G. P. (2005). SIMS: *Structured Inventory of Malingered Symptomatology. Professional manual*. Lutz, FL: Psychological Assessment Resources.

- Wisdom, N. M., Callahan, J. L. y Shaw, T. G. (2010). Diagnostic Utility of the Structured Inventory of Malingered Symptomatology to Detect Malingering in a Forensic Sample. *Archives of Clinical Neuropsychology*, 25(2), 118–125. <https://doi.org/10.1093/arclin/acp110>
- Woo, A.K.W. (2010). Depression and anxiety in pain. *Reviews in Pain*, 4, 8-12. <https://www.doi.org/10.1177/204946371000400103>
- Wooley, C. N. y Rogers, R. (2015). The Effectiveness of the Personality Assessment Inventory With Feigned PTSD: An Initial Investigation of Resnick's Model of Malingering. *Assessment*, 22(4), 449–458. <https://doi.org/10.1177/1073191114552076>
- *Wygant, D.B., Arbisi, P.A., Bianchini, K.J. y Umlauf, R.L. (2017). Waddell non-organic signs: new evidence suggests somatic amplification among outpatient chronic pain patients. *The Spine Journal*, 17(4), 505-510. <https://doi.org/10.1016/j.spinee.2016.10.018>
- Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: Implications for practice and court*. Springer Science.
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15 ± 15 %. *Psychological Injury and Law*, 8(3), 188–199. <https://doi.org/10.1007/s12207-015-9232-4>
- Zale, E.L. y Ditre, J.W. (2015). Pain-Related Fear, Disability, and the Fear-Avoidance Model of Chronic Pain. *Current Opinion in Psychology*, 1, 24-30. <https://www.doi.org/10.1016/j.copsyc.2015.03.014>

Zhou, X.H., Obuchowski, N.A. y McClish, D.K. (2002) *Statistical methods in diagnostic medicine*. Wiley.

ANEXOS

Anexo 1 - Herramienta de análisis de sesgo para estudios de simulación

Sección 1 - Calidad

- 1.1 - *¿El estudio se centra en un tema claramente definido?*
- 1.2 - *¿Los autores han utilizado un método adecuado para responder a la pregunta de investigación?*
- 1.3 - *¿La muestra incluye un espectro adecuado de pacientes?*
- 1.4 - *¿Existe una descripción adecuada de la herramienta de evaluación?*
- 1.5 - *¿Existe un informe selectivo de resultados? Utilice los 3 tipos de Kirham et al. (2010).*

Sección 2 - Análisis profundo de resultados

- 2.1- *¿Incluye análisis de significación estadística?*
- 2.2- *¿Incluye medidas del tamaño del efecto? Si es así, ¿cuál es su magnitud?*
- 2.3- *¿Incluye información para diferentes puntos de corte?*
- 2.4- *¿Incluye al menos 4 estimadores de utilidad? (por ejemplo, sensibilidad, especificidad, poder predictivo positivo y negativo)*
- 2.5- *¿Incluye estudios de curvas ROC?*
- 2.6- *¿Proporciona los valores del área bajo la curva?*

Sección 3 - Aplicabilidad

- 3.1- *¿Son los resultados aplicables / viables en su entorno?*
- 3.2- *¿Son los resultados aplicables en el contexto clínico-forense?*
- 3.3- *¿Los resultados coinciden con la evidencia disponible?*

Sección 4a - Diseño - Simulación

- 4a.1- *¿Son relevantes las comparaciones o se incluye un grupo clínico de referencia?*
- 4a.2- *¿Son claras y comprensibles las instrucciones para las condiciones experimentales?*
- 4a.3- *¿Son las instrucciones para las condiciones experimentales suficientemente específicas para replicar las condiciones reales?*
- 4a.4- *¿Las instrucciones para las condiciones experimentales resaltan la necesidad de hacer una presentación creíble?*
- 4a.5- *¿Las instrucciones para las condiciones experimentales desafían al paciente como método de motivación?*
- 4a.6- *¿Se ofrecen incentivos positivos para una simulación exitosa?*
- 4a.7- *¿Se ofrecen incentivos negativos para la simulación fallida?*
- 4a.8- *¿Se le ha dado al simulador de tiempo suficiente para preparar la condición a simular?*
- 4a.9- *¿Se realiza un cuestionario de seguimiento o entrevista para comprobar si se entendieron correctamente las instrucciones?*

Sección 4b - Diseño - Grupo conocido

- 4b.1- *¿Incluye un grupo de simuladores instruidos? Si es así, ¿la respuesta también incluye el cuarto criterio?*
- 4b.2- *¿Es adecuada la metodología de selección de los grupos criterios?*
- 4b.3- *¿Informa sobre cómo se tratan los casos indeterminados? Si es así, ¿están excluidos?*

Sección 5 - Otras fuentes de sesgo

Anexo 2 – Portada del protocolo

Estudio Protocolo Latigazo Cervical

Código:

Fecha evaluación:

Fecha del accidente:

Duración:

Baja laboral:

Abogado/Litigio:

Evidencia:

Rehab:

Nexo causal:

RSP*	E1: 0 1 2 3
	E2: 0 1 2 3
	E3: 0 1 2 3

0.1 Edad:

0.2 Sexo:

0.3 Estado civil:

0.4 Nivel de educación:

0.5 Nivel socioeconómico:

0.6 Residencia:

0.7 Empleo:

0.8 Esfuerzo laboral:

0.9 Esfuerzo ocio:

0.10 Síntomas referidos:

1.1 Momento de la colisión:

1.2 Localización del accidentado:

1.3 Localización alcance:

1.4 Tipo de carretera:

1.5 Cinturón:

1.6 Dirección mirada:

1.7 Estado del coche:

*Riesgo de simulación percibido

Anexo 3 – Instrucciones para los simuladores instruidos y población general

Simuladores instruidos

“Este estudio trata sobre la simulación en esguince cervical. Te pedimos que tomes el papel de **simulador** y queremos que te imagines que has recibido un impacto en el coche mientras conducías. Aunque te encuentras bien y no te duele demasiado (o nada), has escuchado que la indemnización por sufrir un esguince cervical es muy generosa y quieres hacer todo lo posible por recibirla. Vas a acudir a un centro médico para que un médico valorador te evalúe para dar parte a las aseguradoras de tus supuestos daños. Dicho médico te aplicará un instrumento y tendrás que rellenarlo exagerando o inventando los síntomas que padeces.

Es fundamental tu presentación del esguince sea realista y coherente con lo esperable tras un accidente de circulación. Te pedimos te esfuerces en ser completamente creíble, tienes que tratar de ponerte lo máximo posible en el papel que te hemos asignado y responder el test como si realmente padecieras la condición que has preparado. Por ejemplo, psicopatología grave como delirios o alucinaciones no son consecuencias de este tipo de accidentes. Debes prepararte tu condición antes del seminario que tengas asignado, el método que sigas lo dejamos a tu elección, puedes mirar en internet, en manuales, etc.

El instrumento que te aplicaremos es capaz de detectar si no haces bien el papel que te hemos asignado y, en caso de que te “cace”, no recibirás el bono de la asignatura.

¿Serás capaz de superarlo correctamente? ¡Ánimo y a por ello!”

Población general

“Este estudio trata sobre la simulación en esguince cervical. Necesitamos que respondas a este instrumento de forma **completamente honesta y teniendo en consideración tu estado actual.** No es necesario que inventes o finjas ningún síntoma, sencillamente ten en cuenta cómo te encuentras ahora mismo. Te pedimos te esfuerces en ser completamente creíble, tienes que tratar de ponerte lo máximo posible en el papel asignado. El instrumento que te aplicaremos es capaz de detectar si no haces bien el papel que te hemos asignado y, en caso de que te “cace”, no recibirás el bono de la asignatura.

¿Serás capaz de superarlo correctamente? ¡Ánimo y a por ello!”

Anexo 4 – Indicador, ítem correspondiente del CDS y fuente

<i>Ítem e indicador</i>	<i>Fuente</i>
I1 Detalles superfluos e inusuales	CBCA (Steller y Köhnken, 1989)
I2 Estado mental propio	CBCA (Steller y Köhnken, 1989)
I3 Falsa atribución síntomas reales	Resnick (1997)
I4 Falta de cooperación	Resnick (1997), Vrij (2008)
I5 Actitud hostil o defensiva	Resnick (1997), Vrij (2008)
I6 Culpa tercera persona síntomas	Ferrari <i>et al.</i> (1999)
I7 Culpa tercer persona accidente	Ferrari <i>et al.</i> (1999)
I8 Pronostico de mejora general negativo	Linell y Easton (2006), Leventhal et al. (1998)
I9 Confirmación muchos síntomas	Rogers y Bender (2018)
I10 Síntomas improbables	Rogers y Bender (2018)
I11 Contradicción patrón síntomas	Rogers y Bender (2018)
I12 Síntomas extremos	Ferrari (2005), McDermott y Feldman (2007)
I13 Sensibilidad extrema al dolor	Ferrari (2005), McDermott y Feldman (2007)
I14 Alta sensibilidad zona incorrecta (Transferencia)	Ferrari (2005)
I15 Perdida sensibilidad dermatoma no afectado (Transferencia)	Ferrari (2005)
I16 Incremento exponencial dolor	Ferrari (2005), Rogers y Bender (2018)
I17 Incapacidad control de síntomas	Linell y Easton (2006), Leventhal et al. (1998)
I18 Resistencia evaluación	Resnick (1997), Vrij (2008)
I19 Énfasis aspectos negativos síntomas	Rogers y Bender (2018)
I20 Ausencia de mejoría (Pronóstico)	Linell y Easton (2006), Leventhal et al. (1998)
IA1 Percepción situación enfermedad	Linell y Easton (2006), Leventhal et al. (1998)
IA2 Identidad de síntomas	Linell y Easton (2006), Leventhal et al. (1998)

Anexo 5 – Checklist CDS

Dimensión 1 – Detalles inusuales y pensamientos sobre el evento

- I1. Cuando el paciente habla del accidente y las experiencias posteriores.
- ¿Describe detalles inusuales o complicaciones inesperadas?
- I2. Cuando el paciente habla del accidente y las experiencias posteriores. ¿Comenta su propio estado mental?

Dimensión 2 – Actitud ante situación de enfermedad

- I8. ¿Espera el paciente que su sintomatología dure un largo periodo de tiempo?
- I16. ¿Se queja el paciente de un incremento exponencial del dolor desde el accidente que no corresponde con su patología?
- I17. ¿Se queja el paciente de que es incapaz de controlar sus síntomas? (mediante medicación)
- I19. ¿Da el paciente detalles muy específicos de la lesión y sus secuelas, enfatizando los aspectos negativos y discapacitantes de las mismas?
- I20. Si se le pregunta sobre su estado. ¿Refiere que no ha mejorado?
- *A1. Percepción de la situación de enfermedad.
- *A2. Identidad de los síntomas.

Dimensión 3 – Actitud ante la evaluación

- I4. ¿Muestra el paciente falta de cooperación en la entrevista?
- I5. ¿Muestra el paciente una actitud defensiva, desconfiada u hostil?
- I18. ¿Trata el paciente de evitar las situaciones de evaluación/exploración a no ser que sean necesarias para poder obtener un beneficio económico?

Dimensión 4 – Causa

- I6. ¿El paciente culpa a alguien de sus síntomas?
- I7. ¿Cree el paciente que otra persona se comportó de forma descuidada y es ese comportamiento lo que ha causado su problemática?

Dimensión 5 - Inconsistencias

- I3. ¿Está el paciente atribuyendo falsamente síntomas reales al latigazo?
 - I9. ¿Tiene el paciente a confirmar la existencia de muchos síntomas?
 - I10. ¿Hay síntomas que no pertenecen al latigazo cervical?
 - I11. ¿Hay contradicciones en el patrón de síntomas?
 - I12. ¿Ha informado de síntomas extremos o con un nivel improbable de severidad?
 - I13. ¿Se queja el paciente de dolor al rozarlo o pellizcarlo ligeramente?
 - I14. ¿Se queja el paciente de alta sensibilidad en otra zona que no sea la cervical o la región torácica superior?
 - I15. ¿Se queja el paciente de pérdida de sensibilidad en regiones que no corresponden al dermatoma de la lesión?
-

