



## ORIGINAL

# Protocolo breve de evaluación del suicidio: fiabilidad interexaminadores

Rebeca García-Nieto<sup>a</sup>, Isabel Parra Uribe<sup>b</sup>, Diego Palao<sup>b</sup>, Jorge Lopez-Castroman<sup>a</sup>, Pilar Alejandra Sáiz<sup>c</sup>, María Paz García-Portilla<sup>c</sup>, Jerónimo Saiz Ruiz<sup>d</sup>, Angela Ibañez<sup>d</sup>, Thais Tiana<sup>e</sup>, Santiago Durán Sindreu<sup>e</sup>, Victor Perez Sola<sup>e</sup>, Yolanda de Diego-Otero<sup>f</sup>, Lucía Pérez-Costillas<sup>f</sup>, Rafael Fernández García-Andrade<sup>g</sup>, Dolores Saiz-González<sup>g</sup>, Miguel Angel Jiménez Arriero<sup>h</sup>, Mercedes Navío Acosta<sup>h</sup>, Lucas Giner<sup>i</sup>, Julio Antonio Guija<sup>i,j</sup>, José Luis Escobar<sup>k</sup>, Jorge Antonio Cervilla<sup>k,l</sup>, Marta Quesada<sup>m</sup>, Dolores Braquehais<sup>m</sup>, Hilario Blasco-Fontecilla<sup>a</sup>, Teresa Legido-Gil<sup>a</sup>, Fuensanta Aroca<sup>n</sup>, Enrique Baca-García<sup>a,\*</sup> y Grupo GEICS<sup>◇</sup>

<sup>a</sup> Fundación Jiménez Díaz, Madrid, CIBERSAM, España

<sup>b</sup> Corporación Sanitaria Universitaria Parc Taulí de Sabadell, Barcelona, España

<sup>c</sup> Universidad de Oviedo, CIBERSAM, España

<sup>d</sup> Hospital Universitario Ramón y Cajal, Madrid, IRYCIS, CIBERSAM, España

<sup>e</sup> Hospital de la Santa Creu i Sant Pau, Barcelona, CIBERSAM, España

<sup>f</sup> Hospital Carlos Haya y Fundación IMABIS, Málaga, España

<sup>g</sup> Hospital Universitario Clínico San Carlos, Madrid, CIBERSAM, España

<sup>h</sup> Hospital 12 de Octubre, Madrid, CIBERSAM, España

<sup>i</sup> Universidad de Sevilla, España

<sup>j</sup> Instituto de Medicina Legal, Sevilla, España

<sup>k</sup> Unidad de Salud Mental, Hospital Universitario San Cecilio, Granada, España

<sup>l</sup> Departamento de Psiquiatría, Universidad de Granada, CIBERSAM, España

<sup>m</sup> Hospital Universitario Vall d'Hebron, CIBERSAM, España

<sup>n</sup> Instituto de Matemáticas, Universidad Nacional Autónoma de México, México

Recibido el 7 de septiembre de 2011; aceptado el 17 de octubre de 2011

Disponible en Internet el 16 de noviembre de 2011

### PALABRAS CLAVE

Ensayos clínicos;  
Escala clínicas;  
Estadística;  
Acuerdo  
entre-examinadores;

### Resumen

**Introducción:** El acuerdo entre-examinadores es un aspecto fundamental en la planificación de cualquier trabajo de investigación donde la principal herramienta diagnóstica es la entrevista clínica. El objetivo de este estudio es valorar el acuerdo entre-examinadores de un instrumento de evaluación de la conducta suicida (Protocolo breve de evaluación del suicidio) utilizando las valoraciones de múltiples observadores en una sola sesión.

\* Autor para correspondencia.

Correo electrónico: [ebaca@fjd.es](mailto:ebaca@fjd.es) (E. Baca-García).

◇ En el Anexo 1 se incluye los componentes del grupo.

Psicometría;  
Intentos de suicidio;  
Suicidio

**Método:** Durante la fase piloto de un estudio clínico multicéntrico centrado en la monitorización de intentos de suicidio, 32 examinadores evaluaron el vídeo de la entrevista clínica a un paciente simulado con conducta suicida. Para identificar los ítems en los que existía una mayor discordancia y a los examinadores cuyo criterio se alejaba más del acuerdo general, se utilizó el método Detection Of Multiple Examiners Not In Consensus (DOMENIC).

**Resultado:** El acuerdo interexaminadores osciló entre pobre (<70%) y excelente (90-100%). En la escala de acontecimientos vitales estresantes el nivel de acuerdo osciló entre 48,4 y 97%; en la escala problemas psicosociales del DSM-IV, entre 75,5 y 100%; en la escala de evaluación de la actividad global fue de 82,58%; en la escala de intencionalidad suicida, osciló entre 67,5 y 97%; en la escala de ideación suicida, entre 63,5 y 100% y en la escala de letalidad del intento de suicidio fue de 88,39%. En general, los examinadores mostraron un nivel de acuerdo adecuado tanto en las puntuaciones globales de cada escala como en cada ítem en particular.

**Conclusiones:** El diseño propuesto permite evaluar el acuerdo entre-examinadores de una forma eficiente (en una única sesión). Además, con respecto al Protocolo breve de evaluación del suicidio, el acuerdo entre-examinadores fue apropiado.

© 2011 SEP y SEPB. Publicado por Elsevier España, S.L. Todos los derechos reservados.

## KEYWORDS

Clinical trials;  
Clinical rating scales;  
Statistics;  
Inter-rater  
agreement;  
Psychometrics;  
Suicide attempt;  
Suicide

## Brief Suicide Questionnaire. Inter-rater reliability

### Abstract

**Introduction:** Inter-rater agreement is a crucial aspect in the planning and performance of a clinical trial in which the main assessment tool is the clinical interview. The main objectives of this study are to study the inter-rater agreement of a tool for the assessment of suicidal behavior (Brief Suicide Questionnaire) and to examine whether the inter-examiner agreement when multiple ratings are made on a single subject is an efficient method to assess the reliability of an instrument.

**Method:** In the context of designing a multicenter clinical trial, 32 psychiatrists assessed a videotaped clinical interview of a patient with suicidal behavior. In order to identify those items in which a greater level of discordance existed and detect the examiners whose ratings differed significantly from the average ratings, we used the DOMENIC method (Detection of Multiple Examiners Not in Consensus).

**Results:** Inter-rater agreement was between poor (<70%) to excellent (90-100%). Inter-rater agreement in Brugha's list of threatening experiences ranged from 75.5 and 100%; in the Global Assessment of Functioning (GAF) Scale was 82.58%; in the Beck's Suicidal Intent Scale, ranged from 67.5 and 97%; in the Beck's Scale for Suicide Ideation, ranged from 63.5 and 100%; and in the Lethality Rating Scale was 88.39%. On the whole, the level of agreement among raters, both in general scores and in particular items, was appropriate.

**Conclusion:** The proposed design allows the assessment of the inter-rater agreement in an efficient way (only in one session). In addition, regarding the Brief Suicide Questionnaire, inter-raters agreement was appropriate.

© 2011 SEP y SEPB. Published by Elsevier España, S.L. All rights reserved.

## Introducción

La conducta suicida es causa principal de consumo de recursos sanitarios y mortalidad a nivel mundial, especialmente en jóvenes<sup>1</sup>, y una prioridad de salud pública para la Unión Europea. La conducta suicida (ideación, intentos, suicidio consumado) es heterogénea por la compleja interacción de factores genéticos, biológicos, psicológicos y ambientales<sup>2,28</sup>. La investigación en conducta suicida está limitada por las dificultades en valorar estos aspectos, por ello en muchas ocasiones se estudia subordinada a diagnósticos del eje I (trastornos afectivos y por abuso de sustancia) o eje II (trastorno límite de personalidad límite) sin unos instrumentos específicos de valoración, cuando su impacto clínico y sanitario merece que se trate como una entidad nosológica independiente<sup>3</sup>.

En la actualidad, el *gold standard* para la valoración de la conducta suicida es la evaluación clínica<sup>4</sup>. Sin embargo,

el uso de protocolos y escalas ha demostrado ser muy útil para mejorar la forma en que se documenta la información y aumentar la exhaustividad de la evaluación clínica<sup>5</sup>. Además, el hecho de emplear protocolos y escalas clínicas puede tener valor legal y servir para sustentar la decisión clínica<sup>6,7</sup>. No obstante, algunos estudios recientes han puesto de manifiesto que en nuestro medio la documentación que acompaña a la valoración de la conducta suicida es deficiente<sup>8,9</sup>. El grupo español para la investigación en conducta suicida (GEICS), consciente de esta situación, ha diseñado un Protocolo breve de evaluación del suicidio que incluye las escalas más utilizadas para valorar todo el espectro de la conducta suicida, desde la ideación hasta el intento de suicidio<sup>9</sup>, y examina los factores de riesgo y protectores más importantes (Apéndice).

Para construir el Protocolo breve de evaluación del suicidio se utilizaron las escalas, preferentemente auto-aplicadas, más utilizadas en la literatura de los últimos

40 años en suicidiología, y preguntas que englobasen los factores sociodemográficos con una mayor capacidad descriptiva y predictiva<sup>7</sup>.

Uno de los requisitos esenciales de los instrumentos de valoración es su reproducibilidad<sup>10</sup>. Esta noción se superpone con la noción de acuerdo (*agreement*), que se usa indistintamente para hablar de medidas de consistencia (fiabilidad, reproducibilidad, repetibilidad), que se refieren al acuerdo entre varias medias en las que ninguna es la «correcta», y a medidas de conformidad (validez, exactitud), que se refieren al acuerdo entre una medida y otra que actúa de referencia<sup>11</sup>. El diseño prototípico para poner a prueba la fiabilidad inter-examinador (*interexaminer reliability*) utiliza un pequeño número de examinadores independientes (generalmente 2) que valoran una muestra grande de sujetos (más de 30). La fiabilidad se mide utilizando los coeficientes Kappa, el *weighted* Kappa (Kappa ponderado) o el coeficiente de correlación intraclase, en función de que el tipo de instrumento que se vaya a evaluar sea una escala cualitativa nominal, cualitativa ordinal o cuantitativa<sup>12-16</sup>. El uso de estos índices necesita una muestra mayor de un único sujeto para realizar apropiadamente el estudio de fiabilidad, ya que no es posible calcular el acuerdo por azar (*chance agreement*) con muestras de un único paciente. Su poder estadístico depende tanto del número de examinadores como de sujetos, lo que supone una limitación de recursos muy importante<sup>17</sup>.

Para estimar el acuerdo entre-examinadores del instrumento de valoración de la conducta suicida que se presenta, se empleó la estrategia de un solo caso valorado por múltiples investigadores. Para ello, utilizamos el método propuesto por Cicchetti et al<sup>17</sup>, que permite generar índices, interpretables clínica y estadísticamente, que permitan evaluar el acuerdo global de los examinadores en cada uno de los ítems de las escalas. Por otra parte, permite identificar a los examinadores que diverjan del acuerdo global (entendido como la puntuación media, ya que no se asume un patrón de referencia previo).

## Método

### Examinadores

En este estudio participaron 32 investigadores –psiquiatras y psicólogos clínicos con al menos 2 años de formación- que evaluaron una entrevista clínica grabada en vídeo de un caso prototípico, reproducida en una misma sesión. Previa a la entrevista se les explicó brevemente en qué consistía el instrumento y cada una de las escalas que se incluían. Este soporte audiovisual se ha usado en la valoración de la fiabilidad de instrumentos de evaluación en psiquiatría<sup>18</sup> y, aunque por lo general presenta un menor acuerdo que las historias clínicas, es más cercano a la realidad y resulta más económico que utilizar múltiples entrevistas repetidas individualmente<sup>19</sup>.

La entrevista fue realizada por dos de los participantes en el estudio (LG y JAG). Con este sistema se trataron de minimizar los factores relacionados con la entrevista y con el paciente que afectan a cualquier estudio de fiabilidad, ya que al tratarse de una muestra de un solo paciente, esta fuente de variabilidad desaparece<sup>20</sup>. Por otra parte,

identificar los factores relacionados con los investigadores era uno de los objetivos del estudio.

### Instrumentos de medida

Los distintos grupos investigadores diseñamos un protocolo de evaluación que examinase las siguientes variables relacionadas con la conducta suicida: desencadenantes (acontecimientos vitales estresantes, problemas psicosociales), funcionalidad (nivel de actividad previo), circunstancias objetivas relacionadas con el intento de suicidio, características de la ideación suicida y letalidad del intento de suicidio. Además de examinar datos sociodemográficos y clínicos, nuestro protocolo de evaluación (Apéndice) incluía los siguientes instrumentos, todos ellos traducidos al castellano<sup>6</sup>:

#### Acontecimientos vitales estresantes (List of Threatening Experiences, LTE)<sup>21</sup>

Se trata de un inventario que examina los acontecimientos vitales experimentados por el paciente en los últimos 6 meses. Consta de 12 ítems dicotómicos, que solo admiten dos respuestas (presente/ausente)

#### DSM-IV-TR. Problemas psicosociales<sup>22</sup>

Mediante este instrumento se recogió la información relativa a los problemas psicosociales y ambientales que habían estado presentes durante los 6 meses anteriores, tal y como están descritos en el DSM-IV (APA, 2000).

#### Escala de evaluación de la actividad global (EEAG)

La EEAG o *Global Assessment Scale* (GAS) es un instrumento heteroaplicado propuesto por el DSM-III-R (APA, 1987)<sup>23</sup>, que valora el nivel general de actividad del sujeto en la esfera psicosocial, social y laboral. La puntuación de esta escala oscila entre 0 y 100, en tramos de 10 puntos. La escala se puntúa de acuerdo a la actividad global previa al intento de suicidio.

#### Escala de intencionalidad suicida de Beck (Suicide Intent Scale, SIS)<sup>24</sup>

Escala heteroaplicada para evaluar las características de la tentativa suicida (TS) que consta de dos subescalas. En la primera se recogen las circunstancias objetivas en las que se llevó a cabo el intento de suicidio; en la segunda se valora la actitud del paciente ante la vida y la muerte y la valoración que hace el paciente de su tentativa. En este trabajo, utilizamos la primera sección, que examina las circunstancias objetivas relacionadas con el intento de suicidio<sup>25</sup>. Esta sección se compone de 15 ítems valorados de 0 a 2. En los trabajos realizados para la validación de la escala, la media de las puntuaciones en TS de elevada gravedad fue de 16,3; en TS de gravedad media, 10,1 y 6,7 para las de baja gravedad<sup>25</sup>. En un trabajo posterior realizado por Baca-García et al se estableció un punto de corte de 11 para distinguir a los pacientes que, tras la TS, precisaban

un ingreso en unidad psiquiátrica de los que no precisaban dicho ingreso<sup>7</sup>.

### Escala de ideación suicida (Scale for Suicidal Ideation, SSI)<sup>26</sup>

Se trata de una escala que cuantifica y evalúa la gravedad de los pensamientos suicidas, o grado de seriedad e intensidad con el que alguien está pensando en suicidarse. Es una escala de 19 ítems que debe ser completada por el examinador en el contexto de una entrevista clínica semiestructurada. Se divide en cuatro secciones que recogen una serie de características relativas a la actitud hacia la vida/muerte; pensamientos o deseos suicidas; plan de intento de suicidio y realización del intento planeado. En la última sección se examinan los intentos de suicidio previos. Para cada ítem hay tres alternativas de respuesta que indican un grado creciente de gravedad y/o intensidad de la intencionalidad suicida.

### Letalidad del intento de suicidio (Lethality Rating Scale)<sup>27</sup>

El método del intento de suicidio empleado fue codificado de acuerdo con el Lethality Rating Scale and Method Attempt Coding (LRS), que valora los diversos métodos empleados y examina también las consecuencias médicas del intento.

### Análisis estadístico

El proceso que se ha seguido para el análisis estadístico se basa en el método propuesto por Cicchetti et al<sup>16</sup>. Según este método, el acuerdo global se define en función de los niveles de acuerdo parciales (las puntuaciones presentan mayor acuerdo conforme menor distancia hay entre ellas). Específicamente, se calcularon los siguientes índices:

Promedio global del nivel de acuerdo inter-examinador (*Average overall level of interexaminer agreement*). Esta media indica el acuerdo global de todos los examinadores. Los valores de referencia para su interpretación son los siguientes: acuerdo excelente (valor comprendido entre 90 y 100), bueno (80-89), débil (70-79) y pobre (menor de 70).

El nivel de acuerdo de cada examinador individualmente. Para ello, se agrupan aquellos investigadores con el mismo grado de acuerdo y se calcula la valoración clínica y estadística del nivel de acuerdo de cada uno de los examinadores utilizando los índices de acuerdo, *Z score* o valor Z (que indica la desviación de cada examinador con respecto al valor de consenso, en este caso la media de las puntuaciones).

Para identificar los ítems en los que existía una mayor discordancia y a los examinadores con una baja fiabilidad entre-examinadores, se utilizó el método Detection Of Multiple Examiners Not In Consensus (DOMENIC)<sup>17</sup>.

## Resultados

### Acontecimientos vitales estresantes

El promedio global de acuerdo entre examinadores en cada uno de los ítems oscila entre el 48,4 y el 97% (tabla 1). En

su mayoría el nivel de acuerdo está entre bueno (80-89%) y excelente (90-100%), salvo en los ítems 6, 7, 8 y 10 (6. Ha roto una relación estable; 7. Ha tenido un problema grave con algún amigo cercano, vecino o familiar; 8. Se ha quedado sin empleo o ha buscado empleo durante más de un mes sin éxito, y 10. Ha tenido una crisis económica grave) donde existe una divergencia importante (tabla 2).

### DSM-IV. Problemas psicosociales

El promedio global de acuerdo entre examinadores en cada uno de los ítems oscila entre el 75,5 y el 100% (tabla 1). En su mayoría el nivel de acuerdo está entre bueno y excelente, salvo el ítem «Otros problemas psicosociales» donde el acuerdo era débil.

### Escala evaluación actividad global previa (EEAG)

El acuerdo en la puntuación de esta escala fue bueno (82,58%). Solo 2 examinadores (Raters 2 y 3) presentaban un acuerdo significativamente pobre ( $p < 0,001$ ) con respecto a la media de las puntuaciones totales (patrón de referencia).

### Escala de intencionalidad suicida de Beck parte I: circunstancias objetivas relacionadas con el intento de suicidio

El promedio global de acuerdo entre examinadores en cada uno de los ítems oscila entre el 67,5 y el 97% (tabla 1). En la mayoría de los ítems el nivel de acuerdo oscila entre bueno y excelente, salvo en los ítems 3, 4, 6, 8, 13, 15 (3. Precauciones contra el descubrimiento/intervención; 4. Actos para obtener ayuda durante y después del intento; 6. Preparación activa para el intento; 8. Comunicación de la intención antes del intento; 13. Actitud sobre la vida/muerte y 15. Grado de premeditación), donde se detecta una divergencia importante. Los evaluadores cuyas puntuaciones diferían más de las del resto, en cada uno de los ítems, fueron el 5 y el 7 (tabla 3). El acuerdo en la puntuación total de la escala, usando como punto de corte 11, fue bueno (87,5%).

### Escala de ideación suicida

El promedio global de acuerdo entre examinadores en cada uno de los ítems oscila entre el 63,51 y el 100% (tabla 1). En su mayoría el nivel de acuerdo está entre bueno y excelente, salvo en los ítems 2, 6, 7, 16 y 19 (2. Deseo de morir; 6. Duración de la ideación/deseo suicida; 7. Frecuencia de la ideación/deseo suicida; 16. Expectancia/Anticipación del intento actual; 19. Nota suicida) donde hay una divergencia importante. Los observadores cuyas valoraciones diferían más de las del resto en cada uno de los ítems fueron el 12 y el 18 (tabla 4).

### Letalidad del intento de suicidio

El acuerdo alcanzado en la puntuación de esta escala fue bueno (88,39%). Únicamente un examinador (el número 6) presentaban un nivel de acuerdo medio con respecto a la

**Tabla 1** Acuerdo en los diferentes ítems de cada instrumento

Escala	Item	Promedio global de acuerdo entre examinadores (%)	Nivel de significación
Brugha	1. Ha sufrido usted mismo una enfermedad, lesión o agresión grave.	100,00	Excelente
	2. Algún familiar cercano ha sufrido una enfermedad, lesión o agresión grave.	100,00	Excelente
	3. Ha muerto uno de sus padres, hijos o su pareja/cónyuge.	93,55	Excelente
	4. Ha muerto un amigo cercano a la familia o algún otro familiar (tíos, primos, abuelos).	100,00	Excelente
	5. Se ha separado a causa de problemas en su matrimonio.	81,94	Bueno
	6. Ha roto una relación estable.	60,65	Pobre
	7. Ha tenido un problema grave con algún amigo cercano, vecino o familiar.	62,58	Pobre
	8. Se ha quedado sin empleo o ha buscado empleo durante más de un mes sin éxito.	54,84	Pobre
	9. Le han despedido de su trabajo.	87,53	Bueno
	10. Ha tenido una crisis económica grave.	48,39	Pobre
	11. Ha tenido problemas con la policía o ha comparecido ante un tribunal.	81,94	Bueno
	12. Le han robado o ha perdido algún objeto de valor.	93,55	Excelente
DSM-IV	Problemas con el grupo de apoyo primario	93,01	Excelente
	Medio social	100,00	Excelente
	Escuela	87,31	Bueno
	Trabajo	89,35	Bueno
	Vivienda	83,39	Bueno
	Finanzas	81,45	Bueno
	Acceso a servicios de salud	81,29	Bueno
	Sistema legal	83,39	Bueno
EEAG SIS	Otros problemas psicosociales	75,48	Medio
	Escala evaluación actividad global previo	82,58	Bueno
	1. Aislamiento:	100,00	Excelente
	2. Tiempo:	70,62	Medio
	3. Precauciones contra el descubrimiento/intervención:	67,57	Pobre
	4. Actos para obtener ayuda durante y después del intento:	66,88	Pobre
	5. Actos finales, anticipando la muerte (p.ej.: seguros, regalos, testamento):	86,02	Bueno
	6. Preparación activa para el intento:	69,46	Pobre
	7. Nota de suicidio:	81,89	Bueno
	8. Comunicación de la intención antes del intento:	67,20	Pobre
	9. Propósito del intento:	71,78	Medio
	10. Expectativas sobre el resultado mortal:	91,96	Excelente
	11. Conocimiento de la letalidad del método:	89,12	Bueno
	12. Seriedad del intento:	74,84	Medio
	13. Actitud sobre la vida/muerte:	64,13	Pobre
	14. Concepción de la intervención médica:	86,15	Bueno
15. Grado de premeditación:	67,96	Pobre	
Total punto de corte = 11	87,53	Bueno	
SSI	1. Deseo de vivir:	93,76	Excelente
	2. Deseo de morir:	67,89	Pobre
	3. Razones para vivir/morir	100,00	Excelente
	4. Deseo de intentar activamente el suicidio	93,76	Excelente
	5. Intento de suicidio de forma pasiva	77,42	Medio



Tabla 1 (Continuación)

Escala	Item	Promedio global de acuerdo entre examinadores (%)	Nivel de significación
	6. Duración de la ideación/deseo suicida	63,51	Pobre
	7. Frecuencia de la ideación/deseo suicida	65,48	Pobre
	8. Actitud hacia la ideación/deseo suicida	87,53	Bueno
	9. Control sobre el acto suicida: acting-out/deseo	87,53	Bueno
	10. Disuasores («frenos») para hacer un intento activo (familia, secuelas si no se consume)	74,80	Medio
	11. Razones para el intento planeado	84,84	Bueno
	12. Método: especificidad/planificación	74,62	Medio
	13. Método: disponibilidad/oportunidad	88,73	Bueno
	14. Sensación de «capacidad para realizar el intento»	84,84	Bueno
	15. Expectancia/Anticipación del intento actual	87,53	Bueno
	16. Preparación real	64,62	Pobre
	17. Nota suicida	88,73	Bueno
	18. Últimos arreglos para preparar la muerte (seguros, testamento, donaciones, etc.)	87,53	Bueno
	19. Engaño/Ocultación de la tentativa planeada	69,38	Pobre
	20. Intentos de suicidio previos	93,76	Excelente
	21. Intención de morir relacionada con el último intento	91,35	Excelente
LRS	Letalidad del intento de suicidio	88,39	Bueno

media de las puntuaciones totales (patrón de referencia) (tabla 5).

## Discusión

El objetivo de este trabajo fue evaluar la fiabilidad de un protocolo de evaluación de la conducta suicida (protocolo breve de evaluación del suicidio) para su uso en la investigación en un proyecto multicéntrico utilizando las valoraciones de múltiples examinadores en una muestra de un solo paciente. Los resultados obtenidos en este estudio permiten afirmar que las escalas clínicas que componen el Protocolo breve de evaluación del suicidio tiene

una fiabilidad buena. Es más, los problemas de fiabilidad observados son atribuibles a examinadores concretos y, en el caso de las escalas con más de un ítem, están relacionados con el hecho de que algunos examinadores han dejado sin responder alguno de los ítems.

No obstante, no hay que olvidar que, al tratarse de un diseño con un solo paciente, los resultados sobre la fiabilidad de los instrumentos no pueden generalizarse a la población de la que el paciente ha sido seleccionado. De cara a construir un instrumento aplicable a la clínica, el planteamiento para estimar su fiabilidad sería diferente. Requeriría la valoración de varios pacientes grabados en video (aproximadamente 10 por cada observador incluido) y el uso de otros parámetros estadísticos como el *weighted*

Tabla 2 Acuerdo entre los diferentes observadores en la Escala evaluación actividad global previo (EEAG)

Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
0	2	0,24	Pobre	-3,53	<0,001	2, 3
1	0					
2	0					
3	0					
4	0					
5	3	0,80	Bueno	-0,54	0,59	12, 18, 23
6	14	0,89	Bueno	0,06	0,95	5, 6, 7, 11, 13, 17, 19, 22, 24, 26, 27, 30, 31, 32
7	12	0,86	Bueno	0,65	0,51	4, 8, 9, 10, 14, 15, 16, 20, 21, 25, 28, 29
8	0					
9	0					

**Tabla 3** Acuerdo entre los diferentes observadores en los ítems de la Escala de tentativa suicida de Beck parte I: Circunstancias objetivas relacionadas con el intento de suicidio

Ítem	Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
SIS1	0	0					
	1	0					
	2	31	100,00	Excelente			2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
SIS2	0	2	47,33	Pobre	-2,65	0,01	4 25
	1	24	83,33	Buena	0,00	1,00	3 5 7 8 10 11 12 13 14 16 17 18 19 20 22 23 24 26 27 28 29 30 31 32
SIS3	2	2	47,33	Pobre	2,65	0,01	2 21
	0	2	55,33	Pobre	-2,37	0,02	23 27
	1	12	66,67	Pobre	-0,75	0,45	3 4 8 9 12 13 15 18 19 21 25 30
SIS4	2	16	74,00	Medio	0,86	0,39	2 5 6 7 10 11 14 16 17 20 24 26 28 29 31 32
	0	0					
	1	19	76,67	Medio	-0,73	0,47	3 4 6 8 9 10 13 14 15 16 19 20 21 22 25 26 28 30 31
SIS5	2	10	61,67	Pobre	1,38	0,17	2 5 7 11 17 23 24 27 29 32
	0	26	91,67	Excelente	-0,44	0,66	2 3 4 5 6 8 10 11 12 14 16 17 18 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	5	56,67	Pobre	2,28	0,02	7 9 13 15 19
SIS6	2	0					
	0	14	70,00	Medio	-1,07	0,29	2 4 8 10 11 12 13 14 16 17 18 19 20 22
	1	16	73,33	Medio	0,94	0,35	3 5 6 7 9 15 21 23 24 25 27 28 29 30 31 32
SIS7	2	0					
	0	23	88,67	Buena	-0,57	0,57	2 3 4 5 6 7 8 10 12 14 16 18 20 21 22 23 25 26 28 29 30 31 32
	1	2	51,67	Pobre	0,69	0,49	13 19
SIS8	2	6	66,00	Pobre	1,94	0,05	9 11 15 17 24 27
	0	15	72,33	Medio	-0,92	0,36	3 4 5 7 10 11 12 13 14 16 17 18 19 20 23
	1	13	68,33	Pobre	0,70	0,48	2 6 21 22 24 25 26 27 28 29 30 31 32
SIS9	2	2	55,00	Pobre	2,33	0,02	9 15
	0	14	72,33	Medio	-1,01	0,31	2 3 4 5 8 11 12 17 18 23 24 26 27 30
	1	15	73,33	Medio	0,64	0,52	6 7 9 10 14 15 16 20 21 22 25 28 29 31 32
SIS9	2	2	56,33	Pobre	2,29	0,02	13 19

Tabla 3 (Continuación)

Ítem	Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
SIS10	0	2	61,00	Pobre	-3,58	0,00	3 4
	1	1	50,00	Pobre	-1,63	0,10	29
	2	28	95,67	Excelente	0,31	0,75	2 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 30 31 32
SIS11	0	2	60,67	Pobre	-3,38	0,00	13 19
	1	2	51,67	Pobre	-1,51	0,13	5 24
	2	27	94,00	Excelente	0,36	0,72	2 3 4 6 7 8 9 10 11 12 14 15 16 17 18 20 21 22 23 25 26 27 28 29 30 31 32
SIS12	0	0					
	1	13	70,00	Medio	-1,18	0,24	3 4 8 11 12 13 17 18 19 22 26 28 31
	2	18	78,33	Medio	0,85	0,40	2 5 6 7 9 10 14 15 16 20 21 23 24 25 27 29 30 32
SIS13	0	6	60,67	Pobre	-1,60	0,11	5 11 13 17 19 30
	1	12	66,67	Pobre	-0,27	0,79	2 6 12 14 18 20 21 22 23 25 26 27
	2	12	68,67	Pobre	1,07	0,29	3 4 7 9 10 15 16 24 28 29 31 32
SIS14	0	26	92,00	Excelente	-0,41	0,68	2 3 4 5 6 7 8 10 11 12 13 16 17 18 19 21 22 24 25 26 27 28 29 30 31 32
	1	4	55,00	Pobre	1,72	0,09	9 14 15 20
	2	1	58,67	Pobre	3,85	0,00	23
SIS15	0	0					
	1	9	60,00	Pobre	-1,49	0,14	5 7 8 9 13 15 19 24 25
	2	20	78,33	Medio	0,67	0,50	2 3 4 6 10 11 14 16 17 20 21 22 23 26 27 28 29 30 31 32



**Tabla 4** Acuerdo entre los diferentes observadores en los ítems de la Escala de ideación suicida

Ítem	Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
SSI1	0	29	96,67	Excelente	-0,26	0,79	2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	2	51,67	Pobre	3,81	0,00	12 18
	2	0					
	3	0					
SSI2	0	11	65,67	Pobre	-1,19	0,23	5 7 8 9 10 12 13 15 16 18 19
	1	17	75,00	Medio	0,51	0,61	2 3 4 6 11 14 17 20 21 23 24 25 26 27 29 30 32
	2	2	53,67	Pobre	2,22	0,03	28 31
	3	0					
SSI3	0	31	100,00	Excelente			2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	0					
	2	0					
	3	0					
SSI4	0	29	96,67	Excelente	-0,26	0,79	2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	2	51,67	Pobre	3,81	0,00	12 18
	2	0					
	3	0					
SSI5	0	10	65,00	Pobre	-1,45	0,15	2 3 5 9 12 13 15 18 19 26
	1	21	83,33	Bueno	0,69	0,49	4 6 7 8 10 11 14 16 17 20 21 22 23 24 25 27 28 29 30 31 32
	2	0					
	3	0					
SSI6	0	8	63,67	Pobre	-1,39	0,16	3 4 5 13 19 21 23 25
	1	11	65,00	Pobre	-0,13	0,90	7 9 14 15 20 27 28 29 30 31 32
	2	11	67,67	Pobre	1,14	0,25	6 8 10 11 12 16 17 18 22 24 26
	3	0					
SSI7	0	0					
	1	21	78,33	Medio	-0,58	0,56	3 4 5 6 7 10 11 14 16 17 20 21 23 24 25 27 28 29 30 31 32
	2	7	55,00	Pobre	1,73	0,08	8 12 13 18 19 22 26
	3	0					

**Tabla 4 (Continuación)**

Ítem	Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
SSI8	0	28	93,33	Excelente	-0,27	0,79	3 4 5 6 7 8 9 10 11 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	2	50,00	Pobre	3,74	0,00	12 18
	2	0					
	3	0					
SSI9	0	28	93,33	Excelente	-0,27	0,79	3 4 5 6 7 8 9 10 11 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	2	50,00	Pobre	3,74	0,00	12 18
	2	0					
	3	0					
SSI10	0	22	84,67	Bueno	-0,56	0,57	3 4 6 7 9 10 11 13 15 16 17 19 21 22 23 24 25 26 27 29 30 32
	1	4	53,33	Pobre	0,84	0,40	5 8 14 20
	2	4	60,67	Pobre	2,25	0,02	12 18 28 31
	3	0					
SSI11	0	3	51,67	Pobre	-3,00	0,00	3 4 27
	1	27	91,67	Excelente	0,33	0,74	5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 28 29 30 31 32
	2	0					
	3	0					
SSI12	0	22	83,33	Bueno	-0,60	0,55	7 8 10 12 13 14 16 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	8	60,00	Pobre	1,66	0,10	3 4 5 6 9 11 15 17
	2	0					
	3	0					
SSI13	0	2	59,33	Pobre	-3,74	0,00	23 27
	1	0					
	2	28	94,00	Excelente	0,27	0,79	3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 24 25 26 28 29 30 31 32
	3	0					
SSI14	0	27	91,67	Excelente	-0,33	0,74	3 4 5 6 7 8 10 11 12 13 14 16 17 18 19 20 22 23 24 25 26 27 28 29 30 31 32
	1	3	51,67	Pobre	3,00	0,00	9 15 21
	2	0					
	3	0					

Tabla 4 (Continuación)

Ítem	Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
SSI15	0	28	93,33	Excelente	-0,27	0,79	3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 30 31
	1	2		Pobre	3,74	0,00	29 32
	2	0					
	3	0					
SSI16	0	12	68,33	Pobre	-1,07	0,28	3 4 5 6 13 19 21 22 23 25 26 27
	1	13	68,33	Pobre	0,33	0,74	8 9 10 11 12 15 16 17 18 28 29 31 32
	2	5	59,00	Pobre	1,72	0,08	7 14 20 24 30
	3	0					
SSI17	0	28	94,00	Excelente	-0,27	0,79	3 4 5 6 7 8 10 11 12 13 14 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	0					
	2	2	59,33	Pobre	3,74	0,00	9 15
	3	0					
SSI18	0	28	93,33	Excelente	-0,27	0,79	3 4 5 6 7 8 9 11 12 13 14 15 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	1	2	50,00	Pobre	3,74	0,00	10 16
	2	0					
	3	0					
SSI19	0	11	65,33	Pobre	-1,24	0,21	3 4 9 10 11 13 15 16 17 19 25
	1	18	76,67	Medio	0,62	0,54	6 7 8 12 14 18 20 21 22 23 24 26 27 28 29 30 31 32
	2	1	52,00	Pobre	2,48	0,01	5
	3	0					
SSI20	0	2	51,67	Pobre	-3,81	0,00	12 18
	1	29	65,67	Excelente	0,26	0,79	2 3 4 5 6 7 8 9 10 11 13 14 15 16 17 19 20 21 22 23 24 25 26 27 28 29 30 31 32
	2	0					
	3	0					
SSI21	0	1	59,33	Pobre	-4,45	0,00	5
	1	2	51,67	Pobre	-2,07	0,04	7 25
	2	28	95,33	Excelente	0,31	0,76	2 3 4 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 26 27 28 29 30 31 32
	3	0					

**Tabla 5** Acuerdo entre los diferentes observadores en la Letalidad del intento de suicidio (LRS)

Puntuación	n	Promedio del acuerdo entre-examinador	Significación clínica	Valor Z	p	Examinadores que otorgan la misma puntuación
0	0					
1	0					
2	0					
3	15	0,89	Bueno	-0,98	0,33	2, 3, 4, 5, 7, 10, 11, 14, 16, 17, 20, 24, 25, 27, 28
4	15	0,89	Bueno	0,81	0,42	8, 9, 12, 13, 15, 18, 19, 21, 22, 23, 26, 29, 30, 31, 32
5	1		Medio	2,60	0,01	6
6	0	0,7				

Kappa (Kappa ponderado) o el coeficiente de correlación intraclase para escalas ordinales o cuantitativas. Nuestro estudio podría indicar a los desarrolladores de este trabajo las áreas de baja consistencia de estos instrumentos y por tanto que se podrían inicialmente eliminar.

El diseño de un solo caso controla las fuentes de variabilidad relacionadas con el examen y el paciente examinado, de esta forma la variabilidad de la evaluación se reduce únicamente a las dependientes del examinador. De hecho, tal y como se expuso anteriormente, identificar a los evaluadores cuyas valoraciones discrepaban más del conjunto era uno de los objetivos de este trabajo. En la fase de preparación de todo tipo de estudios multicéntricos (incluidos ensayos clínicos), el uso de este tipo de diseños (acuerdo de múltiples evaluadores sobre un solo paciente) ha demostrado ser útil para la detección de áreas de baja consistencia e identificación de evaluadores que discrepan del conjunto<sup>17</sup>. No obstante, cabe destacar que este tipo de diseño es muy infrecuente en la literatura, fundamentalmente debido a la complejidad del tratamiento estadístico que implica<sup>16</sup>. Solventado este problema con los procedimientos propuestos por Cicchetti y Showalter<sup>16</sup>, el procedimiento aquí descrito puede hacer más eficiente la fase de preparación de investigadores de un estudio multicéntrico. Una de las características más importantes de esta preparación es la formación de los examinadores hasta garantizar una fiabilidad entre-examinadores adecuada. La identificación de ítems y examinadores con bajos niveles de fiabilidad, seguida de un entrenamiento específico en estos ítems más conflictivos, podría ayudar a corregir las fuentes potenciales de variabilidad de la evaluación de los participantes en el ensayo clínico. Este hecho, a su vez, contribuye a aumentar el poder del diseño sin tener que incrementar el tamaño muestral del estudio<sup>17</sup>.

En resumen, tal y como ponen de manifiesto los resultados del presente estudio, la técnica desarrollada por Cicchetti et al<sup>16</sup> ayuda a cumplir estos objetivos de forma eficiente, ya que requiere un tamaño muestral muy pequeño (un sujeto), una única sesión de valoración que se puede hacer diferida, y no necesita que todos los investigadores evalúen al sujeto en el mismo momento, puesto que los índices pueden calcularse posteriormente. Además, las nuevas tecnologías, como las videoconferencias, permiten que las evaluaciones tengan lugar al mismo tiempo pero desde diferentes lugares. En cuanto al Protocolo breve de evaluación del suicidio, podemos concluir que presenta un el acuerdo entre-examinadores adecuado para los propósitos de investigación al tiempo que se han identificado las áreas de bajo acuerdo y los examinadores que se alejan del acuerdo global. De esta manera, con objeto de utilizar el instrumento con mayor fiabilidad, se han puesto en marcha medidas de adiestramiento de los investigadores.

### Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que los procedimientos seguidos se conformaron a las normas éticas del comité de experimentación humana responsable y de acuerdo con la Asociación Médica Mundial y la Declaración de Helsinki.

**Confidencialidad de los datos.** Los autores declaran que en este artículo no aparecen datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Anexo 1.

### GEICS (Grupo español para la investigación en conducta suicida):

*Universidad Autónoma de Madrid:* Concepción Vaquero Lorenzo

*Corporación Sanitaria Universitaria Parc Taulí de Sabadell, Barcelona:* Gemma García-Parés, María Giró Batalla, M. Garrido

*Hospital 12 de Octubre, Madrid, CIBERSAM:* M. Aragües  
*Hospital Carlos Haya y Fundación IMABIS, Málaga:* E. Martín, M. Alba, M.I. Gómez, A. González, M. Maté, M. Romero y N. Cantero

*Hospital de la Santa Creu i Sant Pau, Barcelona, CIBERSAM:* J. Hernández y S. Durán Sindreu

*Universidad de Oviedo, CIBERSAM:* María Teresa Bascarán, Julio Bobes, Manuel Bousoño y P. Burón, Luis Jiménez Treviño

## Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.rpsm.2011.10.001](https://doi.org/10.1016/j.rpsm.2011.10.001).

## Bibliografía

- Oquendo MA, Baca-García E, Mann JJ, Giner J. Issues for DSM-V: suicidal behavior as a separate diagnosis on a separate axis. *Am J Psychiatry*. 2008;165:1383-4.
- Mann JJ, Apter A, Bertolote J, Beautrais A, Currier D, Haas A, et al. Suicide prevention strategies: a systematic review. *JAMA*. 2005;294:2064-74.
- Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10-24 years: a systematic analysis. *Lancet*. 2011;377:2093-102.
- Links PS, Hoffman B. Preventing suicidal behaviour in a general hospital psychiatric service: priorities for programming. *Can J Psychiatry*. 2005;50:490-6.
- Oquendo MA, Currier D, Posner K. Reconceptualización de la nosología psiquiátrica: el caso de la conducta suicida. *Rev Psiquiatr Salud Ment (Barc)*. 2009;2:63-5.
- Baca-García E, Diaz-Sastre C, García Resa E, Blasco H, Braqueis Conesa D, Saiz-Ruiz J, et al. Variables associated with hospitalization decision by emergency psychiatrists after a patient's suicide attempt. *Psychiatr Serv*. 2004;55:792-7.
- Baca-García E, Perez-Rodriguez MM, Basurte-Villamor I, Saiz-Ruiz J, Leiva-Murillo JM, De-Prado-Cumplido M, et al. Using Data Mining to Explore Complex Clinical Decisions: A Study of Hospitalization after a Suicide Attempt. *J Clin Psych*. 2006;67:1124-33.
- Miret M, Nuevo R, Morant C, Sainz-Cortón E, Jiménez-Arriero MA, López-Ibor JJ, et al. Calidad de los informes médicos sobre personas que han intentado suicidarse. *Rev Psiquiatr Salud Ment (Barc)*. 2010;3:13-8.
- García-Portilla MP, Bascarán MT, Sáiz PA, Bousoño M, Parellada M, Bobes J. Banco de instrumentos básicos para la práctica de la psiquiatría clínica. 6ª ed. Madrid: Comunicación y Ediciones Sanitarias, SL. Psiquiatría Editores; 2011.
- Brundage MD, Pater JL, Zee B. Assessing the reliability of two toxicity scales: implications for interpreting toxicity data. *J Natl Cancer Inst*. 1993;85:1138-48.
- Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med*. 1994;13:2465-76.
- Andersen J, Korner A, Larsen JK, Schultz V, Nielsen BM, Behnke K, et al. Agreement in psychiatric assessment. *Acta Psychiatr Scand*. 1993;87:128-32.
- Bartko JJ, Carpiniello B. On the methods and theory of reliability. *J Nerv Ment Dis*. 1976;163:307-17.
- Spitzer RL, Fleiss JL. A re-analysis of the reliability of psychiatric diagnosis. *Br J Psychiatry*. 1974;125:341-7.
- Shorout PE, Spitzer RL, Fleiss JL. Quantification of agreement in psychiatric diagnosis revisited. *Arch Gen Psychiatry*. 1987;44:172-7.
- Cicchetti DV, Showalter D, Rosenheck RA. A new method for assessing interexaminer agreement when multiple ratings are made on a single subject: applications to the assessment of neuropsychiatric symptomatology. *Psychiatr Res J*. 1997;72:51-63.
- Baca-García E, Blanco C, Sáiz-Ruiz J, Rico F, Diaz-Sastre C, Cicchetti DV. Assessment of reliability in the clinical evaluation of depressive symptoms among multiple investigators in a multicenter clinical trial. *J Psychiatr Res*. 2001;102:163-73.
- Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307-10.
- Altman EG, Hedeker DR, Janicak PG, Peterson JL, Davis JM. The Clinician-Administered Rating Scale for Mania (CARS-M): development, reliability, and validity. *Biol Psychiatry*. 1994;36:124-34.
- Lobo A, Huyse FJ, Herzog T, Maltz UF, The ECLW collaborative study: II. Patient registration form (PRF) instrument, training and reliability. *J Psychosom Res*. 1996;40:143-56.
- Brugha TS, Cragg D. The list of threatening experiences: The reliability and validity of a brief life events questionnaire. *Acta Psychiatr Scand*. 1990;82:77-81.
- American Psychiatric Association. DSM-IV-TR. Manual diagnóstico y estadístico de los trastornos mentales. En: Texto revisado. Barcelona: Masson; 2000.
- American Psychiatric Association. DSM-III-R. En: Diagnostic and Statistical Manual of Mental Disorders. Washington DC: APA; 1987.
- Beck AT, Schuyler D, Herman I. Development of suicidal intent scales. En: Beck AT, Resnick, Lettieri DJ, editores. The prediction of suicide. Bowie: Charles Press; 1974.
- Diaz FJ, Baca-García E, Diaz-Sastre C, García Resa E, Blasco H, Braquehais Conesa D, et al. Dimensions of suicidal behavior according to patient reports. *Eur Arch Psychiatry Clin Neurosci*. 2003;253:197-202.
- Beck AT, Kovacs M, Weissman A. Assessment of suicidal intention: the Scale for Suicide Ideation. *J Consult Clin Psychol*. 1979;47:343-52.
- Beck AT, Resnik HLP, Lettieri DJ. The prediction of suicide. Charles Press Publishers; 1974.
- Ayuso-Mateos JL, Baca-García E, Bobes J, Giner J, Giner L, Pérez V, Saiz P, Saiz J, Grupo RECOMS. Recomendaciones preventivas y manejo del comportamiento suicida en España. *Rev Psiquiatr Salud Ment (Barc)*. 2012;5:8-23.