

## Artículo original

# Validación en Colombia del instrumento para evaluación de la depresión Montgomery-Åsberg Depression Rating Scale (MADRS)<sup>☆</sup>



Juan Fernando Cano<sup>a,\*</sup>, Carlos Gomez Restrepo<sup>b</sup> y Martín Rondón<sup>b</sup>

<sup>a</sup> Centro de Investigaciones del Sistema Nervioso, Grupo CISNE, Bogotá, Colombia

<sup>b</sup> Departamento de Epidemiología Clínica y Bioestadística, Pontificia Universidad Javeriana, Bogotá, Colombia

### INFORMACIÓN DEL ARTÍCULO

#### Historia del artículo:

Recibido el 27 de febrero de 2015

Aceptado el 24 de agosto de 2015

On-line el 9 de octubre de 2015

#### Palabras clave:

Estudios de validación

Depresión

Colombia

### RESUMEN

**Objetivo:** Adaptar y validar la *Montgomery-Åsberg Depression Rating Scale (MADRS)* en el medio colombiano.

**Metodología:** Estudio de tipo observacional para la validación de una escala. Se precisó la validez de criterio determinando los puntos de corte del instrumento a través de los valores de sensibilidad y especificidad de la gravedad de la enfermedad, al contrastarla con los criterios de la Clasificación Internacional de Enfermedades (CIE-10). Se realizó un análisis factorial de la escala; se determinó la consistencia interna del instrumento; se evaluó la reproducibilidad interevaluadores a través de la evaluación de 22 pacientes por cuatro investigadores diferentes, y se estableció la sensibilidad al cambio de la escala en 28 sujetos aplicando el instrumento tras un lapso de 14-28 días.

**Resultados:** El estudio se llevó a cabo en la ciudad de Bogotá, donde se aplicó a 150 pacientes con diagnóstico de depresión mayor. El punto de corte para depresión moderada fue 20 (sensibilidad, 98%; especificidad, 96%) y para grave, 34 (sensibilidad, 98%; especificidad, 92%). Se evidenció que el instrumento es unidimensional y posee buena consistencia interna ( $\alpha = 0,9168$ ). Los hallazgos de las pruebas de confiabilidad interevaluadores demostraron que el instrumento es altamente confiable (coeficiente de correlación intraclase = 0,9833). El instrumento tiene buena sensibilidad al cambio.

**Conclusiones:** La versión colombiana de la escala de evaluación de la depresión Montgomery-Åsberg tiene buenas propiedades psicométricas y se puede utilizar en la práctica clínica y en investigaciones relacionadas con el trastorno depresivo.

© 2015 Asociación Colombiana de Psiquiatría. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

<sup>☆</sup> Trabajo presentado como requisito para el título de magíster en Epidemiología Clínica, obtenido en 2013 en la Pontificia Universidad Javeriana. Trabajo presentado en el LII Congreso Colombiano de Psiquiatría llevado a cabo en la ciudad de Cartagena del 10 al 14 de octubre de 2013.

\* Autor para correspondencia.

Correo electrónico: [juan.cano@grupocisne.org](mailto:juan.cano@grupocisne.org) (J.F. Cano).

<http://dx.doi.org/10.1016/j.rcp.2015.08.006>

0034-7450/© 2015 Asociación Colombiana de Psiquiatría. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

## Validation of the Montgomery-Åsberg Depression Rating Scale (MADRS) in Colombia

### A B S T R A C T

**Keywords:**  
Validation studies  
Depression  
Colombia

**Objective:** To adapt and to validate the Montgomery-Åsberg Depression Rating Scale (MADRS) in Colombia.

**Methods:** Observational study for scale validation. Validity criteria were used to determine the severity cut-off points of the tool. Taking into account sensitivity and specificity values, those cut points were contrasted with ICD-10 criteria for depression severity. A factor analysis was performed. The internal consistency was determined with the same sample of patients used for the validity criteria. Inter-rater reliability was assessed by evaluating the 22 records of the patients that consented to a video interview. Sensitivity to change was established through a second application of the scale in 28 subjects after a lapse of 14 to 28 days.

**Results:** The study was performed in Bogotá, the tool was applied in 150 patients suffering from major depressive disorder. The cut-off point for moderate depression was 20 (sensitivity, 98%; specificity, 96%), and the cut-off point for severe depression was 34 (sensitivity, 98%; specificity, 92%). The tool appears as a unidimensional scale, which possesses a good internal consistency with ( $\alpha=.9168$ ). The findings of inter-rater reliability evaluation showed the scale as highly reliable (intraclass correlation coefficient=.9833). The instrument has a good sensitivity to change.

**Conclusions:** The Colombian version of the Montgomery-Åsberg Depression Rating Scale has good psychometric properties and can be used in clinical practice and in clinical research in the field of depressive disorder.

© 2015 Asociación Colombiana de Psiquiatría. Published by Elsevier España, S.L.U. All rights reserved.

## Introducción

Los instrumentos de medición como las escalas en psiquiatría deben ser ajustados al contexto cultural del grupo al que se va a aplicar<sup>1</sup>. Esto se hace necesario por la variabilidad de las características poblacionales; por tal razón, es importante adaptar los elementos de la escala y validar su comportamiento en un grupo específico de población. Si esto no se realiza, las mediciones obtenidas no reflejan fielmente el estado del evento en estudio<sup>2</sup>. Por eso la validación de una escala para la medición de la gravedad de la depresión significa un paso relevante en la investigación y en la práctica clínica colombianas.

La depresión es la condición que genera mayor discapacidad en el mundo, principalmente por la carga en años de vida laborales perdidos que genera<sup>3</sup>. Los avances que puedan generarse en dicha área siempre tendrán alto impacto.

La investigación en psiquiatría es un campo creciente, algunos de los trabajos de grupos colombianos han alcanzado reconocimiento mundial, pero dichos trabajos no serían posibles si no existieran los instrumentos adecuados para su desarrollo y repetición.

Los estudios clínicos, necesarios para el desarrollo y la implementación de nuevas alternativas terapéuticas, se realizan cada vez con más frecuencia en países latinoamericanos, en gran medida para satisfacer las necesidades de evidencia en diferentes grupos y poblaciones que permitan mayor aplicación de dichas alternativas<sup>4</sup>. Los resultados y la aplicación

de los estudios clínicos en depresión realizados en Colombia se encuentran limitados por la escasez de escalas adaptadas a nuestro medio, lo que genera cierto recelo sobre la aplicación de los resultados obtenidos en esos estudios. Esta situación resalta la necesidad de generar instrumentos adecuados que den mayor solidez a dichos resultados.

Los instrumentos más utilizados en la literatura médica para la evaluación de la depresión tanto en registros como en estudios clínicos son la escala de Montgomery-Åsberg (MADRS) y la escala de Hamilton<sup>5</sup>.

La MADRS tiene elementos en común con la escala de Hamilton, son instrumentos heteroaplicados, tienen alta concordancia y estadifican la gravedad de la depresión de manera adecuada; sin embargo, la MADRS tiene menos enunciados, se centra en aspectos cognitivos, afectivos y somáticos sin dar mayor prioridad a este último componente, lo que le proporciona mayor utilidad para pacientes adultos mayores o con enfermedades médicas concomitantes<sup>6</sup>. También ha mostrado mayor precisión para determinar la gravedad del episodio<sup>5</sup> y además se ha resaltado su mayor sensibilidad al cambio, lo que le confiere ventajas sobre la escala de Hamilton en el contexto de los estudios clínicos<sup>7</sup>.

La MADRS fue publicada por los profesores Stuart Montgomery y Marie Åsberg en 1979<sup>8</sup> y surge con el objetivo de superar la escasa diferenciación que se encontraba en los ensayos clínicos de las diversas moléculas de antidepresivos. Ya desde sus orígenes, la razón principal para la construcción y motivación en la MADRS es identificar el instrumento con mayor sensibilidad al cambio en depresión mayor. A partir de un

instrumento publicado por el mismo autor en 1978, que evaluaba de manera ampliada (65 ítems) los síntomas en los pacientes con trastornos mentales<sup>9</sup>, comenzó la elaboración de este instrumento. En su ejercicio inicial de validación, se construyó el instrumento a partir de las evaluaciones de 106 pacientes (54 ingleses y 52 suecos) a los que se aplicó la escala (escala ampliada de psicopatología [CPRS]) y se identificaron los 17 síntomas más comunes en el trastorno depresivo mayor. Posteriormente se evaluaron estos 17 ítems en 64 pacientes de cuatro estudios de antidepresivos, y se identificaron los 10 ítems que en mayor medida y de forma proporcional se asociaron con la respuesta global, exponiendo mejor sensibilidad al cambio.

La escala es un instrumento de 10 ítems que debe administrarse durante una entrevista clínica personal médico o paramédico previamente entrenado en su aplicación, y puede utilizar información de otras fuentes como los familiares del paciente, de manera similar al desarrollo normal del contexto clínico.

Para cada ítem se consideran siete grados de gravedad (de 0 a 6), que anclan los valores pares (0, 2, 4 y 6) a unos enunciados y permite las puntuaciones intermedias entre esos enunciados cuando hay duda sobre cuál de los dos asignar. La puntuación total de la escala se obtiene sumando los valores seleccionados de cada ítem, con un intervalo total entre 0 y 60 puntos<sup>8</sup>. Los puntos de corte no están claramente definidos, dado que el objetivo inicial de la escala era evaluar los cambios en la gravedad de los síntomas depresivos en pacientes con diagnóstico de depresión mayor, y lo relevante es la variación en las puntuaciones de observaciones repetidas. Sin embargo, su uso se ha generalizado a la estadificación de la gravedad del episodio depresivo, y se proponen algunos puntos de corte:  $\geq 35$  puntos, depresión grave; 20-34, depresión moderada; 7-19, depresión leve, y  $\leq 6$ , depresión en recuperación<sup>10</sup>.

La escala ha atravesado por diversos estudios de validación desde la fecha de su publicación, y ha arrojado buenos resultados en la evaluación de confiabilidad, con valores de consistencia interna entre 0,76 y 0,95<sup>5,10-15</sup>, y confiabilidad interevaluadores alta (de  $\alpha = 0,8$  a  $\alpha = 0,95$ <sup>8,12,13,15,16</sup>).

Los estudios de validación se han hecho a través de validez convergente, validez de criterio y análisis factorial. Validez convergente con otros instrumentos para evaluación de la depresión como la escala de Hamilton y la escala de Beck, con buena correlación<sup>8,16,17</sup>. Validez de criterio a través de determinar la sensibilidad y especificidad de los puntos de corte entre depresión moderada y depresión grave al comparar con las gradaciones de gravedad de la escala de impresión clínica global (CGI)<sup>17-19</sup>. El análisis factorial ha arrojado resultados diversos, desde los que indican que la escala es unidimensional hasta reportes de cuatro dimensiones diferentes<sup>6,11,12,14,20-26</sup>.

No se han realizado análisis de ítem diferencial con la escala, pero este tipo de análisis se ha realizado evaluando la depresión con otros instrumentos, y se han encontrado diferencias leves por sexo y lugar de residencia<sup>27</sup>.

La escala fue traducida al español por Conde y Franch en España en 1984<sup>28</sup>; el primer proceso de validación en español lo llevó a cabo un grupo diferente 7 años después<sup>17</sup>) y el grupo del hospital clínico de Zaragoza realizó un segundo proceso en 2002<sup>13</sup>. En España se utilizan cuatro versiones diferentes del

instrumento, y todavía no hay consenso sobre cuál es el instrumento más apropiado; la experiencia en nuestro medio con las escalas psiquiátricas traducidas y adaptadas en España no siempre ha dado buenos resultados, y se ha identificado gran variabilidad entre el contexto cultural de los países latinoamericanos y España; en Colombia solo se ha realizado este proceso con la escala de Zung<sup>29,30</sup>, instrumento autoaplicable de criba, razón que limita su utilidad, dado que en pacientes con trastornos afectivos la evaluación del observador es más objetiva que la del propio individuo, en especial en los momentos de mayor intensidad del episodio.

Se busca con este estudio adaptar y validar la MADRS en el medio colombiano.

---

## Metodología

### *Traducción, adaptación y validez de apariencia*

Se obtuvo autorización escrita del autor para realizar la validación de la escala, con la única condición de que se citara permanentemente el artículo original y se respetaran siempre los derechos de autor sobre el instrumento.

Se utilizó el método de traducción y retrotraducción<sup>31,32</sup>. Se formó un comité de revisión de la escala compuesto por dos psiquiatras con experiencia clínica (Marcela Alzate García, Juan Fernando Cano), una psicóloga con maestría en neurociencias y salud mental (Ana Fernanda Olarte) y un filólogo (Sergi Casals), todos ellos bilingües. Tres psiquiatras (Juan Diego Barrera, Juan Pablo Rosas, Julian Ortigón Valencia), cuya lengua materna es el español y que conocían el instrumento con sus objetivos, realizaron traducciones del inglés al español de toda la escala. Las versiones en español fueron evaluadas por el comité de revisión, que mediante consenso elaboró una versión unificada de estas tres traducciones verificando que la versión final fuera comprensible y reflejara equivalencia transcultural entre la versión original de la escala y la nueva traducida al español. Un traductor oficial cuya lengua materna es inglés británico y que no tenía conocimiento de la versión original de la escala retradujo a su idioma original (inglés) la nueva versión en español. Esta nueva versión en inglés fue evaluada por el comité de revisión, que no encontró grandes discrepancias con el instrumento original.

Cinco psiquiatras y una psicóloga aplicaron la escala a 12 pacientes con características similares a las de la población objeto de estudio, para evaluar la comprensibilidad de los ítems, el tiempo necesario para aplicarlo, la frecuencia de respuesta y la restricción de gama de respuestas y la facilidad de calificación.

### *Población de estudio y procedimientos para evaluación de los pacientes*

En el estudio se incluyó a pacientes con diagnóstico de depresión mayor confirmado a través de la Entrevista Clínica Estructurada para los Trastornos del eje I del DSM-IV (SCID-I) y distintos grados de gravedad según los criterios del CIE-10, que asistieron a la clínica psiquiátrica Campo Abierto, el Hospital Simón Bolívar y la Unidad de Salud Mental Floralia de la ciudad de Bogotá (en los servicios de consulta externa,

hospitalización y urgencias y en el consultorio particular de los psiquiatras entrenados para el desarrollo del protocolo) que aceptaran participar en el estudio y que ser evaluados por un clínico entrenado en la versión colombiana de la MADRS. Se excluyó a los pacientes que tuvieran cualquier otro diagnóstico psiquiátrico o antecedentes de trastorno afectivo bipolar de familiares en primer grado de consanguinidad.

Se realizaron reuniones de entrenamiento en la aplicación de la entrevista SCID, módulo para trastornos afectivos, y los criterios de gravedad del CIE-10, y en la aplicación de la escala empleando como método de calibración la evaluación de un paciente grabado en formato de video y la evaluación y discusión en grupo de las puntuaciones de la escala propuestos por el comité evaluador del instrumento. Por facilidad en el proceso de evaluación de los pacientes, todos los evaluadores fueron entrenados en ambos instrumentos, y se les indicó que en la evaluación individual de cada paciente solo podrían aplicar una de las dos escalas, mientras que la otra evaluación la realizaría otro evaluador entrenado.

Quienes participaron en el presente trabajo en condición de sujetos de observación lo hicieron de manera voluntaria. Para que un paciente fuera incluido en el estudio era necesario que otorgara consentimiento informado por escrito, previa información de las características de la investigación.

El protocolo, la versión final de la escala, el consentimiento informado, el módulo SCID para trastornos afectivos, junto con los criterios de gravedad de un episodio depresivo según CIE-10 y el instrumento de recolección fueron evaluados y aprobados por el Comité de Ética en Investigación Campo Abierto.

#### Validez de criterio

Se calcularon la sensibilidad y la especificidad y la curva ROC del instrumento, al compararlo con la clasificación de gravedad con los estándares del CIE-10.

Para el cálculo del tamaño de la muestra se empleó la fórmula propuesta por McCullagh y Nelder en 1981, con la adaptación para varianza de la curva ROC propuesta por Hanlen y McNeil en 1982<sup>33</sup>:  $m = (Z_{1-\alpha}/2)^2 V(\bar{A})/L^2$ , donde  $Z_{1-\alpha}/2$  es el percentil de la distribución normal, fijado en 1,96;  $\alpha$  es el nivel de confianza, fijado en 0,05;  $V(\bar{A})$  es la varianza de la curva ROC, y  $L$  es la amplitud deseada de medio intervalo de confianza, fijado en 0,1.

La varianza de la curva:  $V(\bar{A}) = Q1/k + Q2 - A2(1/k + 1)$ , donde  $A$  es la estimación esperada para el área bajo la curva, que se fijó en 0,75.

$$Q1 = A / (2-A) \quad Q2 = 2A^2 / (1+A)$$

$k$  es la proporción entre pacientes sanos y enfermos, que para el protocolo se definió como la proporción entre pacientes con depresión moderada y pacientes con depresión grave, que para el caso se fijó en 1, y teniendo en cuenta que se intentó definir los puntos de corte entre depresión moderada y grave, al igual que entre depresión leve y moderada.

Se estimó  $m$  con esos parámetros en 45,28; se requieren  $m(1+k)$  pacientes. Dado que se estudiarán dos curvas diferentes, se propone incluir a 46 pacientes en cada grupo de gravedad para realizar el cálculo.

Se realizó un muestreo no probabilístico entre los pacientes hospitalizados o que asistían a la consulta externa con diagnóstico de depresión mayor y se aplicó el instrumento diagnóstico a 150 pacientes (50 en cada grupo de gravedad), y en un plazo < 24 h un evaluador diferente y entrenado, que no conocía la clasificación de gravedad de la depresión del sujeto en estudio, aplicó el instrumento. Se tuvo en cuenta que en las 12 h previas a la aplicación los pacientes no hubieran recibido fármacos psicotrópicos con efecto hipnótico o sedante que afectaran al resultado de la prueba. La puntuación de la escala se mantuvo en reserva y no se informó de ella al médico tratante, con el objeto de limitar posibles modificaciones de la práctica clínica secundarias al resultado de la prueba.

#### Validez de contenido y constructo

Dado que los resultados de los análisis factoriales realizados a la escala han sido disímiles, pues se ha encontrado entre un único factor y cuatro factores, se determinó que el mejor método de evaluación del contenido del instrumento es un análisis factorial exploratorio para determinar los factores componentes de la escala. Antes de realizar dicho análisis, se verificó que se cumpliera el supuesto de correlación mediante la prueba global de esfericidad de Bartlett (en la que la hipótesis resultó nula: la correlación es igual a la matriz de identidad, es decir, no hay correlación), y el índice de adecuación o medida de adecuación de la muestra de Kaiser Meyer Olkin (KMO), que identifica correlación entre las variables descomponiendo la correlación entre correlación final y correlación parcial.

Como método de extracción de factores, se utilizó el análisis de factores principales. Para identificar el número de factores, se empleó el criterio gráfico y el de Kaiser (los que tuvieran valores propios > 1). Para estos análisis se utilizó a los mismos sujetos que para la validez de criterio, dado que el número de sujetos era 15 por cada ítem que se considera apropiado<sup>34</sup>.

Para evaluar la validez de constructo, también se realizó análisis de ítem diferencial poniendo en consideración cuatro hipótesis sobre el comportamiento de algunos ítems en grupos específicos.

**Hipótesis 1.** Los pacientes con depresión grave puntuarán con mayor frecuencia en el ítem 10 que los clasificados como depresión moderada o leve.

**Hipótesis 2.** Las mujeres puntuarán en el ítem 1 («tristeza aparente») con mayor frecuencia que los hombres.

**Hipótesis 3.** Las mujeres puntuarán en el ítem 6 («dificultades en concentración») con mayor frecuencia que los hombres.

**Hipótesis 4.** Los hombres puntuarán en el ítem 8 («incapacidad para sentir») con mayor frecuencia que las mujeres.

#### Confiabilidad

Se determinó la consistencia interna de la prueba, que es una medida indirecta de la confiabilidad del instrumento,

determinando la homogeneidad de la escala, o grado de correlación de cada uno de los ítems de esta con su objetivo global, para evaluar si existen ítems redundantes en la escala.

Se evaluó la consistencia interna, con la misma muestra utilizada para la evaluación de la validez de contenido (150 sujetos), mediante el alfa de Cronbach. Los valores aceptables deben encontrarse entre 0,7 y 0,9. Los cálculos se encontraron a partir de la siguiente fórmula:

$$\alpha = (N/N-1) [(St^2 - \sum S_i^2) / S_t^2]$$

donde  $n$  es el número de ítems en el instrumento;  $S_t^2$ , la varianza total de las puntuaciones de la prueba, y  $\sum S_i^2$ , la suma de las varianzas de las puntuaciones de los ítems. Se determinó el grado de similitud en los puntuaciones asignadas a un mismo sujeto por cuatro evaluadores diferentes. Todos los evaluadores fueron entrenados en la aplicación del instrumento.

El instrumento se aplicó a 22 pacientes, a quienes, además del consentimiento informado, se les solicitó permiso para grabar la entrevista en formato de video; cada entrevista fue enviada a otros tres evaluadores, quienes puntuaron la escala de manera independiente, con lo cual se obtuvieron las cuatro evaluaciones. El grado de concordancia se evaluó mediante el coeficiente de correlación intraclase. El tamaño de la muestra para la evaluación de la reproducibilidad con el coeficiente de correlación intraclase se calculó con el comando «sampicc» del programa STATA, con los siguientes parámetros: un error tipo I de 0,05, un poder del 80%, un coeficiente de correlación intraclase mínimo aceptable ( $\rho_0 = 0,6$ ), un coeficiente de correlación intraclase esperado ( $\rho_1 = 0,8$ ) y cuatro evaluadores. El tamaño de muestra calculado fue de 22 pacientes.

Se consideró aceptable un coeficiente de correlación intraclase  $> 0,6$  y bueno el  $> 0,8$ .

No se consideró necesario realizar en este instrumento evaluación de la confiabilidad test-retest, por lo muy cambiante del fenómeno, que llevaría a valores de esta prueba necesariamente bajos.

### Sensibilidad al cambio

Dado que es el objetivo principal sobre el cual se elaboró el instrumento, es este uno de los puntos al que se dio mayor relevancia dentro del proceso de validación. Se realizó evaluando a un grupo de sujetos a los que se aplicó el instrumento antes de realizarse cualquier intervención farmacoterapéutica y se volvió a evaluarlos al término de 14-28 días. Mediante la prueba de signos de Wilcoxon, se compararon las puntuaciones obtenidas en la primera aplicación con las de la segunda tras el alta.

El tamaño de la muestra se calculó con el programa G\*Power 3.0 con la función de prueba de signos de Wilcoxon, con un tamaño de efecto fijado en 1, un valor  $\alpha = 0,005$  y un poder del 95%. El tamaño de muestra calculado fue de 25 sujetos, se adicionó un 10% para prever posibles pérdidas, para hacer un total de 28 sujetos.

### Procesamiento de los datos y análisis estadístico

Se construyó una base de datos en el paquete estadístico STATA, y los procedimientos que se describen a continuación se efectuaron en el mismo software.

- Estadísticas descriptivas de la población utilizada para el estudio: se utilizaron medidas de tendencia central, de dispersión y de frecuencia para caracterizar a los grupos de pacientes utilizados.
- Para evaluar la validez de criterio, se emplearon pruebas de hipótesis estableciendo que el área bajo la curva del instrumento fuera  $\geq 0,75$ .
- Para evaluar la validez de contenido: con base en todos los ítems de la escala, se hizo un análisis factorial exploratorio de factores principales para determinar las dimensiones (dominios o factores) de la escala. Para la selección de los dominios, se tuvo en cuenta el criterio gráfico y el de Kaiser (valor propio  $> 1$ ). Se exploró hacer una rotación oblicua promax, dado el supuesto de que los dominios estarían correlacionados, pero no se encontraron modificaciones significativas al hacer la rotación.
- Para la evaluación de las hipótesis del análisis de ítem diferencial, se utilizaron pruebas de la  $\chi^2$ .
- Alfa de Cronbach para la evaluación de consistencia interna.
- Coeficiente de correlación intraclase para evaluación de reproducibilidad interevaluadores.
- Para evaluar la sensibilidad al cambio, se comparó la puntuación obtenida en la primera aplicación con el de la segunda tras el alta mediante la prueba de signos de Wilcoxon.

### Resultados

Luego de cumplir con las etapas requeridas en la traducción y adaptación de la escala, el comité evaluador determinó que en apariencia el instrumento evaluaba de manera adecuada el constructo depresión. No fue necesario contactar al autor de la escala, pues no se encontraron divergencias significativas en el proceso de retrotraducción.

Luego de la prueba piloto, se identificó un punto de anclaje en el ítem 6 de la escala que generaba dificultades en su aplicación por emplear una doble negación, y fue modificado por recomendación del comité evaluador; además se identificó que la comprensión de los ítems por los evaluadores era adecuada, pero que se requería entrenamiento para utilizarlo correctamente. El tiempo promedio de aplicación de la escala fue de 15-20 min. Luego de la prueba piloto, se aprobó la versión final de la escala, que fue sometida al comité de ética para emplearla en las demás fases de la validación.

Se evaluó a 150 personas con diagnóstico confirmado de depresión mayor mediante entrevista estructurada SCID; en su mayoría eran mujeres, laboralmente activas y con edades entre los 19 y los 85 (mediana, 50) años (tabla 1).

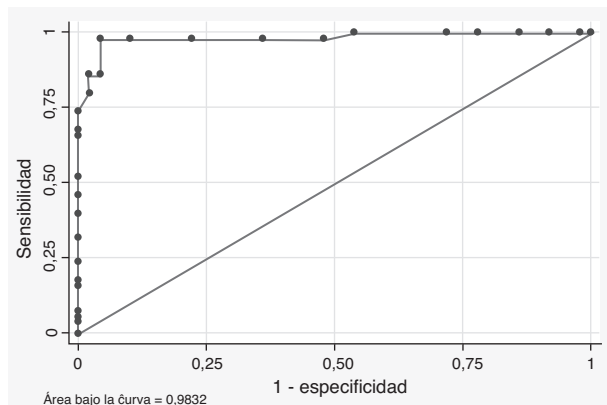
### Validez de criterio

Se identificó que el mejor punto de corte para depresión moderada según la escala es 20, con sensibilidad del 98%

**Tabla 1 – Características sociodemográficas de la muestra**

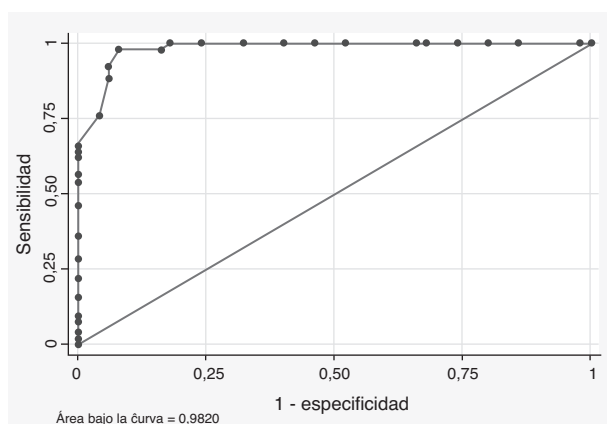
Sexo	
Femenino	98 (65)
Masculino	52 (35)
Ocupación	
Empleado/independiente	67 (45)
Hogar	23 (15)
Estudiante	12 (8)
Jubilado	24 (16)
Cesante	24 (16)
Edad (años cumplidos)	50 [33–61]
Escolaridad (años)	13 [8–16]
Episodios previos, n	2 [1–9]

Salvo otra indicación, los valores expresan n (%) o mediana [intervalo intercuartílico].

**Figura 1 – AUC Punto de Corte Intensidad Moderada.**

y especificidad del 96%, cubriendo un área bajo la curva de 0,9832 (intervalo de confianza del 95% [IC95%] 0,9606-1,00) (fig. 1).

Además, se encontró que el mejor punto de corte para depresión grave según la escala es 34, con sensibilidad del 98% y especificidad del 92%, cubriendo un área bajo la curva de 0,9820 (IC95%, 0,9626-1,00) (fig. 2).

**Figura 2 – AUC Punto de Corte Intensidad Severa.****Tabla 2 – Análisis factorial**

Factor	Valor propio	Diferencia	Proporción	Acumulado
Factor 1	5,47921	4,93967	0,9101	0,9101
Factor 2	0,53954	0,17820	0,0896	0,9997
Factor 3	0,36134	0,25985	0,0600	1,0597
Factor 4	0,10149	0,04754	0,0169	1,0765
Factor 5	0,05395	0,06273	0,0090	1,0855
Factor 6	-0,00879	0,06810	-0,0015	1,0840
Factor 7	-0,07688	0,03043	-0,0128	1,0713
Factor 8	-0,10731	0,03851	-0,0178	1,0534
Factor 9	-0,14582	0,03014	-0,0242	1,0292
Factor 10	-0,17596		-0,0292	1,0000

Log-range test: independiente frente a saturado,  $\chi^2_{(45)} = 1.008,78$ ; Prob  $> \chi^2 < 0,0001$ .

**Tabla 3 – Cargas del factor**

Variable	Factor	Singularidad
Ítem 1	0,8384	0,2971
Ítem 2	0,8794	0,2266
Ítem 3	0,6251	0,6092
Ítem 4	0,5785	0,6653
Ítem 5	0,6555	0,5703
Ítem 6	0,5784	0,6655
Ítem 7	0,7223	0,4783
Ítem 8	0,7972	0,3645
Ítem 9	0,7912	0,3740
Ítem 10	0,8545	0,2699

### Validez de contenido y constructo

Se verificó que se cumplía el supuesto de correlación mediante la prueba global de esfericidad de Bartlett ( $\chi^2 = 1.000,867$ ; grados de libertad, 45;  $p < 0,001$ ), y el índice de adecuación KMO, que arrojó un valor de 0,892, por lo que se rechaza la hipótesis nula de que no hay correlación.

Al realizar el análisis de factores principales, se evidenció un único factor que explica el 91% de la varianza (tabla 2). Al evaluar la carga que aporta cada ítem al factor único, se evidenció que todas las cargas son altas, con un valor mínimo de 0,5780 (tabla 3).

Al evaluar la validez de constructo mediante hipótesis sobre el comportamiento de algunos ítems en grupos específicos a través de diferencias de medianas, pudo evidenciarse que el ítem 10 es marcador de gravedad, pues los pacientes con afección grave puntuaban en dicho ítem con mayor frecuencia que los pacientes con afección leve o moderada.

No se encontró ninguna diferencia significativa por sexo en la puntuación de los ítems de la escala.

### Confiabilidad

#### Consistencia interna

La escala completa mostró adecuada consistencia interna ( $\alpha = 0,9168$ ), por lo que se determinó que no existen ítems redundantes en la escala.

#### Reproducibilidad interevaluadores

Se encontró que la puntuación total de la escala y cada uno de sus ítems tenían excelente reproducibilidad interevaluadores:

**Tabla 4 – Confiabilidad interevaluadores**

Ítem	CCI	IC
1	0,89	0,80-0,94
2	0,97	0,95-0,99
3	0,89	0,80-0,95
4	0,97	0,94-0,99
5	0,97	0,94-0,98
6	0,94	0,89-0,97
7	0,93	0,88-0,97
8	0,94	0,90-0,97
9	0,96	0,92-0,98
10	0,97	0,95-0,99
Total	0,98	0,97-0,99

CCI: coeficiente de correlación intraclase; IC: intervalo de confianza.

coeficiente de correlación intraclase, 0,98 para el total de la escala y 0,88-0,97 para los ítems individuales (tabla 4).

#### *Sensibilidad al cambio*

Los resultados fueron estadísticamente significativos tanto para cada uno de los ítems como para la puntuación total de la escala ( $p \leq 0,0005$ ), por lo que se evidencia que la escala muestra buena sensibilidad al cambio.

## Discusión

Con el presente trabajo queda validada la escala MADRS en pacientes con trastornos depresivos en el medio colombiano. Se alcanzó una versión clara y concisa con utilidad clínica. El personal con formación en salud mental puede aplicar la escala de manera sencilla, práctica y ágil siempre y cuando tenga un entrenamiento previo. El instrumento tiene además adecuada validez y alta confiabilidad, con un desempeño adecuado frente al patrón de referencia de evaluación del episodio depresivo otorgando al clínico una fuente adicional para esa medición.

Los puntos de corte encontrados en esta validación (20 para depresión moderada y 34 para depresión grave) son similares a resultados previos<sup>18,19</sup>; sin embargo, en este estudio no se buscó identificar un punto de corte entre depresión leve y sujetos sin depresión, lo que actualmente es tema de debate<sup>35</sup>.

La evaluación de la validez de la escala MADRS ha sido objeto de múltiples estudios, y se encuentran dificultades con el disímil número de factores reportados en estos trabajos, en los que varían desde modelos unidimensionales hasta la identificación de cuatro factores<sup>5,11-14,20-26</sup>. Por el hecho de encontrar resultados dispares entre el número de factores y los ítems relacionados con cada factor, no es posible realizar un análisis confirmatorio de la escala y se requiere continuar con análisis exploratorios hasta que se dupliquen modelos similares en diferentes partes del mundo.

Sin embargo, la escala se construyó a partir de un instrumento ampliado de 65 ítems para identificar los ítems que tuvieran mayor sensibilidad al cambio tras una intervención farmacológica, por lo que se puede dilucidar que el constructo original era unidimensional. En la publicación inicial, el autor no realizó análisis factorial de la escala, sino que presentó un análisis de validez convergente con la escala de Hamilton<sup>8</sup>. También se ha observado que el resultado disímil

entre los diversos análisis factoriales está condicionado por la población objetivo de la escala, lo que podría dar cuenta de la especificidad del instrumento para sujetos con trastorno depresivo mayor y su variación cuando se evalúan otras enfermedades. En el presente proceso de validación, se seleccionó un grupo de sujetos homogéneo, con características similares a los empleados por el autor en su publicación original y en los trabajos de Galinowski et al<sup>11</sup> y Uher et al<sup>14</sup>, que solo tomaron a pacientes con trastorno depresivo mayor y se duplicó un modelo dimensional en el análisis factorial. Al excluir a los pacientes con antecedentes familiares de trastorno bipolar en el presente trabajo, se homogenizaba más nuestra población, por lo que no es de extrañar que el resultado del análisis exploratorio arrojara que el instrumento era unidimensional. Tal vez el instrumento se comporte de manera diferente si se evalúan otras afecciones, pero ese aspecto se sale del objetivo planteado por el proyecto. Se considera una posible limitación de este estudio no haber realizado un análisis factorial confirmatorio empleando un modelo unidimensional que evaluara qué tanto se ajustan los datos a esa presunción, y sería prudente realizar dicho análisis en futuras apreciaciones del instrumento.

Se han reportado diferencias en las características clínicas del episodio depresivo entre hombres y mujeres, como en el estudio de Cassano et al<sup>27</sup>; en el presente estudio, la escala no mostró diferencias significativas en este aspecto, lo que puede explicarse por no haber realizado el cálculo del tamaño de muestra necesario teniendo en cuenta el sexo como subgrupo, que podría ser objeto de un estudio posterior.

La escala mostró buena consistencia interna, como han reportado previamente la mayoría de los estudios que han evaluado el instrumento, y se encuentra un desempeño superior al de escalas como la de Hamilton o la de Beck<sup>14</sup>.

La reproducibilidad interevaluadores también fue buena, lo cual da cuenta de la facilidad de aplicación de la escala luego de un proceso de entrenamiento y calibración, así como de la utilidad de la escala particularmente en investigación.

Como se esperaba, dado que fue la característica principal desde su concepción, el instrumento mostró buena sensibilidad al cambio, de acuerdo con los resultados en las validaciones predecesoras. Para este estudio, se determinó que el método más sencillo y de más fácil interpretación para la evaluación de la sensibilidad al cambio era la prueba de Wilcoxon, pero se considera que se podría haber evaluado este ítem con estadísticos que dieran cuenta de la dirección y la magnitud del cambio esperado, por lo que se propone que sea un punto a evaluar en investigaciones posteriores con el instrumento.

## Conclusiones

La versión colombiana de la MADRS posee buenas propiedades psicométricas, es un instrumento unidimensional que clasifica adecuadamente a los pacientes por la intensidad del episodio depresivo, con un punto de corte de 20 para depresión moderada y 34 para depresión grave. Luego de un proceso breve de entrenamiento, puede emplearse en la práctica clínica y la investigación.

## 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 han seguido los protocolos de su centro de trabajo sobre la publicación de datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores han obtenido el consentimiento informado de los pacientes y/o sujetos referidos en el artículo. Este documento obra en poder del autor de correspondencia.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Agradecimientos

Agradecemos a los profesionales Marcela Alzate García, Juan Diego Barrera, Sergi Casals, Ana Fernanda Olarte, Julián Ortigón, y Juan Pablo Rosas, así como a todos los compañeros de trabajo de la Clínica Campo Abierto y el Centro de Investigaciones del Sistema Nervioso-Grupo CISNE en la ciudad de Bogotá, por su disposición y dedicación para que este proyecto se haya podido llevar a cabo.

## Anexo. Versión final de la escala de evaluación en depresión de Montgomery-Åsberg

La evaluación debe basarse en una entrevista clínica que inicie con preguntas sobre síntomas formuladas de manera abierta hacia otras más detalladas que permitan precisión en la calificación de gravedad. El entrevistador debe definir si la puntuación se sitúa en los puntos de la escala definidos (0, 2, 4,6) o entre ellos (1, 3, 5).

Es importante recordar que solo en raras ocasiones se encuentra a pacientes deprimidos que no se pueda calificar en un punto particular de la escala. Si no se puede derivar respuestas definitivas del paciente, todos los datos clínicos relevantes, así como la información de otras fuentes, deben usarse como punto de partida para la calificación en concordancia con la práctica clínica habitual.

La escala se puede usar para cualquier intervalo de tiempo entre calificaciones, que puede ser semanal u otro, pero debe estar especificado.

### Lista de ítems

1. Tristeza aparente.
2. Tristeza reportada.
3. Tensión interna.
4. Sueño reducido.
5. Apetito reducido.

6. Dificultades de concentración.
7. Apatía (desgano).
8. Incapacidad para sentir.
9. Pensamientos pesimistas.
10. Pensamientos suicidas.

1. **Tristeza aparente:** Manifestada por desaliento, tristeza y desesperanza (más que una sensación transitoria de desánimo), reflejada en el discurso, la expresión facial y la postura. Evalúe de acuerdo con la severidad y la incapacidad para animarse.
  - 0) Sin tristeza.
  - 2) Se ve desanimado, pero se anima sin dificultad.
  - 4) Parece triste e infeliz la mayor parte del tiempo.
  - 6) Se ve desdichado todo el tiempo. Desánimo extremo.
2. **Tristeza reportada:** Manifestada por los reportes del individuo acerca de humor deprimido, sin tener en cuenta si se refleja o no en la apariencia. Incluye desánimo, desaliento o la sensación de estar desvalido y desesperanzado. Evalúe de acuerdo con intensidad, duración y grado en que las circunstancias externas influyen en el ánimo.
  - 0) Tristeza ocasional de acuerdo con las circunstancias.
  - 2) Triste o desanimado, pero se anima sin dificultad.
  - 4) Sentimientos generalizados de tristeza o desánimo. El ánimo aún está influenciado por las circunstancias externas.
  - 6) Tristeza, desdicha o desaliento, continuos e invariables.
3. **Tensión interna:** Manifestada por sentimientos de malestar pobremente definido, nerviosismo, confusión interna, o incertidumbre que puede llegar a transformarse en angustia, terror o pánico. Evalúe de acuerdo con intensidad, frecuencia, duración y grado de consuelo requerido.
  - 0) Plácido. Solo tensión interna pasajera.
  - 2) Sentimientos ocasionales de nerviosismo o malestar pobremente definido.
  - 4) Sentimientos continuos de tensión interna o pánico intermitente que el paciente solo con dificultad puede controlar.
  - 6) Terror o angustia extrema desbordante. Pánico incontenible.
4. **Sueño reducido:** Manifestado por la experiencia de reducción en la duración o profundidad del sueño comparado con el patrón habitual del paciente cuando se encuentra bien.
  - 0) Duerme como de costumbre.
  - 2) Leve dificultad para conciliar el sueño o sueño levemente reducido o ligero o con interrupciones ocasionales.
  - 4) Sueño reducido o interrumpido por al menos dos horas.
  - 6) Menos de dos o tres horas de sueño.
5. **Apetito reducido:** Manifestado por la sensación de pérdida de apetito en comparación con el patrón habitual. Evalúe pérdida de las ganas de comer o la necesidad de forzarse a comer.
  - 0) Apetito normal o aumentado.
  - 2) Apetito ligeramente reducido.



- 4) Sin apetito. La comida no tiene sabor.
  - 6) Requiere que lo obliguen a comer.
6. **Dificultades de concentración:** Manifestadas por dificultades para poner en orden las ideas, que se incrementa hasta la falta de concentración incapacitante. Evalúe de acuerdo con intensidad, frecuencia y grado de incapacidad producida.
- 0) Ninguna dificultad.
  - 2) Dificultad ocasional para ordenar las ideas.
  - 4) Dificultades en la concentración y en el pensamiento que reducen la capacidad para leer o mantener una conversación.
  - 6) Gran dificultad o incapacidad para leer o conversar.
7. **Apatía (desgano):** Manifestada por el enlentecimiento o las dificultades para iniciar o realizar las actividades de la vida diaria.
- 0) Casi ninguna dificultad para iniciar. No hay pereza.
  - 2) Dificultad para iniciar las actividades.
  - 4) Dificultad para iniciar las actividades simples de la rutina diaria, las cuales se llevan a cabo con esfuerzo.
  - 6) Completa apatía. Incapacidad para hacer nada sin ayuda.
8. **Incapacidad para sentir:** Manifestada por la experiencia subjetiva de interés reducido en lo que le rodea, o en las actividades que normalmente le producen placer. La capacidad para reaccionar con la emoción adecuada a las circunstancias o las personas está reducida.
- 0) Interés normal en lo que le rodea y en los demás.
  - 2) Reducción en la capacidad para disfrutar de los intereses habituales.
  - 4) Pérdida de interés por lo que le rodea, por amigos y conocidos.
  - 6) Experiencia de estar paralizado emocionalmente, incapacidad para sentir rabia, pena o placer, y una completa e inclusive dolorosa ausencia de sentimientos por amigos o familiares cercanos.
9. **Pensamientos pesimistas:** Manifestados por pensamientos de culpa, inferioridad, autorreproche, pecado, remordimiento y ruina.
- 0) Sin pensamientos pesimistas.
  - 2) Ideas fluctuantes de haber fallado, autorreproche o desprecio por sí mismo.
  - 4) Autoacusaciones persistentes o categóricas, pero aún racionales de culpa o pecado. Pesimismo incrementado sobre el futuro.
  - 6) Delirios de ruina, remordimiento o pecado imperdonable. Autoacusaciones absurdas e irreductibles.
10. **Pensamientos suicidas:** Manifestados por la sensación que no vale la pena vivir, que la muerte natural sería bienvenida, pensamientos suicidas y planes suicidas.
- 0) Disfruta la vida tal y como es.
  - 2) Cansado de la vida. Únicamente pensamientos suicidas pasajeros.
  - 4) Piensa que estaría mejor muerto. Los pensamientos suicidas son comunes, y el suicidio es considerado como una posible solución, pero sin planes o intenciones específicas.
  - 6) Planes explícitos para suicidarse cuando se presente una oportunidad. Preparación activa para el suicidio.

## BIBLIOGRAFÍA

1. Sanchez R, Gómez C. Conceptos básicos sobre validación de escalas. *Rev Colomb Psiquiatr.* 1998;27:121-30.
2. Streiner D. A checklist for evaluating the usefulness of rating scales. *Can J Psychiatry.* 1993;38:140-8.
3. Murray CJ, Lopez AD. Global mortality, disability, and the contribution of risk factors: Global Burden of Disease Study. *Lancet.* 1997;349:1436-42.
4. Razzouk D, Zorzetto R, Dubugras MT, Gerolin J, Mari JJ. Leading countries in mental health research in Latin America and the Caribbean. *Rev Bras Psiquiatria.* 2007;29:118-22.
5. Carmody TJ, Rush AJ, Bernstein I, Warden D, Brannan S, Burnham D, et al. The Montgomery-Åsberg and the Hamilton ratings of depression: A comparison of measures. *Eur Neuropsychopharmacol.* 2006;16:601-11.
6. Hammond MF. Rating depression severity in the elderly physically ill patient: reliability and factor structure of the Hamilton and the Montgomery-Åsberg depression ratings scales. *Int J Geriatr Psychiatry.* 1998;13:257-61.
7. Khan A, Khan SR, Shankles EB, Polissar NL. Relative sensitivity of the Montgomery-Åsberg Depression Rating Scales, the Hamilton Depression Rating Scale and the Clinical Global Impressions rating Scale in antidepressant clinical trials. *Int Clin Psychopharmacol.* 2002;17:281-5.
8. Montgomery SA, Åsberg M. A new depression rating scale designed to be sensitive to change. *Br J Psychiatry.* 1979;134:382-9.
9. Åsberg M, Montgomery S, Perris C, Schalling D, Sedvall G. A comprehensive psychopathological psychiatric rating scale (CPRS). *Acta Psych Scand.* 1978;271 Supl:5-27.
10. Snaith RP, Harrop FM, Newby DA, Teale C. Grade scores of the Montgomery-Åsberg Depression and the Clinical Anxiety Scales. *Br J Psychiatry.* 1986;148:599-601.
11. Galinowski A, Lehert P. Structural validity of MADRS During antidepressant treatment. *Int Clin Psychopharmacol.* 1995;10:157-61.
12. Andersson S, Krogstad JM, Finset A. Apathy and depressed mood in acquired brain damage: relationship to lesion localization and psychophysiological reactivity. *Psychol Med.* 1999;29:447-56.
13. Lobo A, Chamorro L, Luque A, Dal-Ré R, Badia X, Baró E, et al. Validación de las versiones en español de la Montgomery-Åsberg Depression Rating Scale y la Hamilton Anxiety Rating Scale para la evaluación de la depresión y de la ansiedad. *Med Clin (Barc).* 2002;118:493-9.
14. Uher R, Farmer A, Maier W, Rietschel M. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med.* 2008;38:289-300.
15. Maier W, Philipp M. Comparative analysis of observer depression scales. *Acta Psych Scand.* 1985;72:239-45.
16. Davidson J, Turnbull CD, Strickland R, Miller R, Graves K. The Montgomery-Åsberg Depression Scale: Reliability and validity. *Acta Psychiatr Scand.* 1986;73:544-8.
17. Martínez-Gallardo R, Bourgeois M, Peyre F, Lobo A. Estudio de validación de la escala de depresión de Montgomery y Åsberg. *Rev Asoc Esp Neuropsiquiatria.* 1991;11:9-14.
18. Müller MJ, Szegedi A, Wetzel H, Benkert O. Moderate and severe depression Gradations for the Montgomery-Åsberg Depression Rating Scale. *J Affect Disord.* 2000;60:137-40.
19. Müller MJ, Himmerich H, Kienzle B, Szegedi A. Differentiating moderate and severe depression using the Montgomery-Åsberg Depression Rating Scale (MADRS). *J Affect Disord.* 2003;77:255-60.
20. Serretti A, Jori MC, Casadei G, Ravizza L, Smeraldi E, Akiskal H. Delineating psychopathologic clusters within dysthymia:

- a study of 512 out-patients without major depression. *J Affect Disord.* 1999;56:17-25.
21. Rocca P, Fonzo V, Ravizza L, Rocca G, Scotta M, Zanalda E, et al. A comparison of paroxetine and amisulpride in the treatment of dysthymic disorder. *J Affect Disord.* 2002;70:313-7.
  22. Benazzi F. Factor analysis of the Montgomery-Åsberg Depression Rating Scale in 251 bipolar II and 306 unipolar depressed outpatients. *Prog Neuro Psychopharmacol Biol Psychiatry.* 2001;25:1369-76.
  23. Parker RD, Flintz EP, Bosworth HB, Pieper CF, Steffens DC. A three-factor analytic model of the MADRS in geriatric depression. *Int J Geriatr Psychiatry.* 2003;18:73-7.
  24. Suzuki A, Aoshima T, Fukasawa T, Yoshida K, Higuchi H, Shimizu T, et al. A three-factor model of the MADRS in major depressive disorder. *Depression Anxiety.* 2005;21:95-7.
  25. Craighead WE. Factor analysis of the Montgomery-Åsberg depression rating scale. *Depression.* 1996;4:31-3.
  26. Williamson D, Brown E, Perlis RH, Ahl J, Baker RW, Tohen M. Clinical relevance of depressive symptom improvement in bipolar I depressed patients. *J Affect Disord.* 2006;92:261-6.
  27. Cassano GB, Benvenuti A, Miniati M, Calugi S, Mula M, Maggi L, et al. The factor structure of lifetime depressive spectrum in patients with unipolar depression. *J Affect Disord.* 2009;115:87-99.
  28. Conde V, Franch JI. Escalas de evaluación comportamental para la cuantificación de la sintomatología de los trastornos angustiosos y depresivos. Madrid: Upjohn; 1984.
  29. Torres Y, Posada J. Resultados de la escala de depresión clínica. Estudio Nacional de Salud Mental y Consumo de Sustancias Psicoactivas, Colombia 1993. Bogotá: Carrera Séptima; 1994. p. 167-77.
  30. Campo-Arias A, Díaz LA, Rueda-Jaimes GE, Barros JA. Validación de la escala de Zung para depresión en universitarias de Bucaramanga, Colombia. *Rev Colomb Psiquiatr.* 2005;34:54-63.
  31. Brislin RW. The wording and translation of research instruments. *Field methods in cross-cultural research*, vol. 8. Thousand Oaks: Sage Publications; 1986. p. 137-64.
  32. Sartorius N, Janca A. Psychiatric assessment instruments developed by the World Health Organization. *Soc Psychiatry Psychiatr Epidemiol.* 1996;31:55-69.
  33. Zhou XH, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine.* New York: Wiley Interscience; 2002. p. 201-6.
  34. Sánchez R, Echeverry J. Validación de escalas de medición en salud. *Rev Salud Pública.* 2004;6:302-18.
  35. Schulte-van Maaren IWM, Carlier IVE, Zitman FG, Van Hemert AM, De Waal MWM, Van der Does AJW, et al. Reference values for major depression questionnaires: The Leiden Routine Outcome Monitoring Study. *J Affect Disord.* 2013;149:342-9.