

**ARTÍCULOS**

**Algunos problemas con la utilización  
de la adaptación española del  
Inventario Clínico Multiaxial de Millon-II  
(MCMI-II) con fines diagnósticos**

***Some troubles with the use of the  
Spanish adaptation of Millon Clinical  
Multiaxial Inventory-II (MCMI-II) for  
diagnostic purposes***

**JESÚS SANZ\***

**RESUMEN**

*Uno de los objetivos del Inventario Clínico Multiaxial de Millon-II (MCMI-II) es identificar el diagnóstico específico que mejor se aplica a un paciente, para lo cual propone puntos de corte en puntuaciones tasa base (TB) que están avalados por los resultados satisfactorios de un estudio de validación diagnóstica. La adaptación española del MCMI-II presenta varias inconsistencias en los baremos que sirven para transformar las puntuaciones directas a TB (errores tipográficos, altas discrepancias con los baremos americanos, cambios de baremos entre ediciones) y propone puntos de corte TB para el diagnóstico cuya validez no viene avalada por ningún tipo de estudio empírico. El objetivo del presente artículo es analizar esos problemas y sus repercusiones, así como sugerir algunas recomendaciones al respecto, la más importante de las cuales implica no utilizar la adaptación española del MCMI-II con fines diagnósticos hasta que se obtengan datos que sustenten su validez diagnóstica.*

\* Departamento de Persnalidad, Evaluación y Psicología Clínica. Universidad Complutense de Madrid.

### **ABSTRACT**

*One of the aims of Millon Clinical Multiaxial Inventory-II (MCMI-II) is to identify the best diagnosis for patients. In order to achieve this goal, the MCMI-II suggests cutoff points in terms of base rate (BR) scores that are supported by satisfactory results in a diagnostic validity study. The three editions of the Spanish adaptation of the MCMI-II show a number of inconsistencies concerning norms to translate raw scores into BR –misprints, high discrepancies with American norms, change of norms among editions. There is also a suggestion for diagnostic BR cutoff points which validity is not supported by any empirical research. This paper analyzes problems and implications stemming from this situation. Some suggestions and recommendations are also made to overcome those problems –the most important of which is not to use the Spanish adaptation of the MCMI-II for diagnostic purposes until data to support diagnostic validity have been gathered.*

### **PALABRAS CLAVE**

*Diagnóstico, Inventario Clínico Multiaxial de Millon-II, Validez, Adaptación.*

### **KEY WORDS**

*Diagnosis, Millon Clinical Multiaxial Inventory-II, Validity, Adaptation.*

## INTRODUCCIÓN

El Inventario Clínico Multiaxial de Millon (*Millon Clinical Multiaxial Inventory*; MCMI), a lo largo de sus tres versiones (MCMI-I, MCMI-II y MCMI-III; Millon, 1983, 1987; Millon, Millon y Davis, 1994), se ha convertido en uno de los instrumentos más prestigiosos y utilizados internacionalmente para la evaluación de la psicopatología tanto en ámbitos profesionales como de investigación (Butcher y Rouse, 1996; Piotrowski, 1997). Así, una búsqueda en la base de datos *PsycINFO* mediante las palabras "MCMI" o "Millon Clinical Multiaxial Inventory" en los campos título y resumen para el período comprendido entre 1977, fecha de aparición de la primera versión del MCMI, y enero de 2007, dio lugar a 1013 referencias, de las cuales 678 se referían a artículos de revistas en los cuales se había utilizado o analizado el MCMI.

En 1998 se publicó la adaptación española del MCMI-II. Desde entonces, este instrumento también ha ido ganando en popularidad entre los investigadores y profesionales españoles, como lo demuestra el hecho de que en 1999 se publicara una segunda edición de dicha adaptación y en 2002 una tercera. Asimismo, en los últimos años la adaptación española del MCMI-II ha sido utilizada en más de una treintena de estudios realizados por investigadores españoles y publicados

en revistas científicas<sup>1</sup>. En casi la mitad de estos estudios, el MCMI-II fue empleado como instrumento diagnóstico, de forma que se diagnosticó la presencia de determinados trastornos psicológicos utilizando como único criterio puntuaciones superiores a los puntos de corte en alguna de sus escalas.

Desgraciadamente, la adaptación española del MCMI-II presenta algunos problemas con sus baremos y con su validación diagnóstica que afectan a su utilización como instrumento diagnóstico de los trastornos mentales, problemas que pueden conducir a los profesionales e investigadores a realizar inferencias erróneas sobre si sus pacientes presentan o no un trastorno clínico o un trastorno de personalidad. El objetivo del presente artículo es exponer esos problemas y sus repercusiones, así como sugerir algunas recomendaciones al respecto.

## EL MCMI-II COMO INSTRUMENTO DIAGNÓSTICO

Para entender los problemas que presenta la adaptación española del MCMI-II como herramienta diagnóstica parece conveniente revisar los objetivos para los cuales se diseñó el inventario y cuales son sus principales características con relación a esos objetivos. En el manual del instrumento se nos dice que: "el principal objetivo del MCMI-II es propor-

<sup>1</sup> Los lectores interesados pueden solicitar al autor una lista de los estudios españoles que han utilizado la adaptación española del MCMI-II.

cionar información a los clínicos (...) en tareas de evaluación y tratamiento de personas con dificultades emocionales e interpersonales. El perfil de las puntuaciones de las escalas del MCMI-II puede servir para identificar qué grupo de personas con dificultades emocionales y personales puede requerir una evaluación más intensa o atención profesional. [...] Más específicamente, los puntos de corte individuales para cada escala del MCMI-II pueden usarse para tomar decisiones sobre trastornos comportamentales primarios o síndromes clínicos. De igual manera, los niveles de puntuación de los grupos de escalas pueden ofrecer motivos para enjuiciar el deterioro, la gravedad y cronicidad de la patología” (Millon, 1998, p. 16-17).

Para cumplir con estos objetivos clínicos, el MCMI-II no transforma las puntuaciones directas (PD) de las distintas escalas del cuestionario en puntuaciones típicas, sino en puntuaciones de tasa base (TB). Esta transformación relaciona las PD de las escalas del MCMI-II con las prevalencias de los trastornos en cuestión en una muestra clínica representativa, de forma que se pueda establecer un punto de corte diagnóstico para cada escala que reproduzca el mismo porcentaje de casos que se encontraría realmente en la población de pacientes psicopatológicos. Así, por ejemplo, si la prevalencia de la depresión entre los pacientes es de 25%, el punto de corte diagnóstico para la escala de depresión mayor sólo podría ser alcanzado y superado por el 25%

de los pacientes de una muestra representativa que completen el cuestionario y, por tanto, dicho punto de corte diagnóstico reproduciría a la postre el 25% de pacientes con diagnóstico de depresión de la población total de pacientes con trastornos mentales.

Por tanto, el uso de transformaciones de TB pretende perfeccionar la eficiencia diagnóstica del MCMI-II ya que, para Millon, “el propósito principal de un instrumento clínico no es ubicar la posición relativa de un paciente en una distribución de frecuencias, sino calcular la probabilidad que el paciente tiene de formar parte de una entidad diagnóstica particular” (Millon, 1998, p. 73). En la misma dirección, Millon afirma: “los resultados del MCMI-II pueden emplearse de distintas maneras con propósitos clínicos. El uso principal es la interpretación de perfil mediante un proceso de inferencia de los aspectos más importantes de la configuración de las 22 escalas clínicas. Otro uso importante es la toma de decisiones para identificar el diagnóstico específico que mejor se aplica al paciente” (Millon, 1998, p. 160).

En resumen, uno de los objetivos más importante del MCMI-II es el diagnóstico clínico y, para cumplir ese objetivo, el MCMI-II utiliza puntuaciones TB que *a priori* permiten obtener mejores puntos de corte diagnóstico.

Para transformar las PD a TB, Millon y su grupo de investigación analizaron los datos de 1.292

pacientes evaluados con el MCMI-II por más de 600 clínicos. Estos clínicos habían asignado a sus pacientes diagnósticos en el Eje I y en el Eje II según los criterios del DSM-III-R. Los pacientes podían haber recibido más de un diagnóstico, pero si este era el caso, los clínicos indicaron cuál era el diagnóstico principal. A partir de estos datos, se estimó la prevalencia de cada uno de los trastornos que pretenden medir las escalas del MCMI-II, se calcularon las PD del MCMI-II y se seleccionaron dos números arbitrarios para designar los puntos de corte de TB que correspondían a las PD de los grupos diagnósticos. Así, en todas las escalas se tomó la puntuación de TB 74 como la línea de corte superior que correspondería al porcentaje de pacientes que, según sus clínicos, presentaban un trastorno clínico o un trastorno de personalidad, y la puntuación de TB de 84 como la línea de corte superior que correspondería al porcentaje de pacientes en los cuales, según sus clínicos, dichos trastornos clínicos o trastornos de personalidad eran los principales o más sobresalientes. Así, por ejemplo, si el 48% de las mujeres de la muestra clínica sufría un trastorno distímico y en el 29% el trastorno distímico era el diagnóstico principal, entonces la PD de la escala de Distimia (D) del MCMI-II que sólo es alcanzada o superada por el 48% de las mujeres de la muestra de pacientes se transformaría en TB 75 y la PD de dicha escala que sólo es alcanzada o superada por el 29% de las mujeres de la muestra de pacientes se transformaría en

TB 85. De forma similar, se seleccionó arbitrariamente una puntuación TB de 60 para representar la mediana de las PD de todos los pacientes de la muestra normativa clínica del MCMI-II (este valor se aplicó a todas las escalas exceptuando la escala de Hipomanía, en la que la puntuación directa de la mediana de los pacientes era menor que la puntuación directa de la mediana de una muestra no clínica o "normal" de personas), y una puntuación TB de 35 para representar la mediana o puntuación del centil 50 de una muestra no clínica de personas en todas las escalas de personalidad patológica ("S", "C", "P") y en las escalas de trastornos clínicos (desde "A" hasta "PP") (Millon, 1998, pp. 77-78).

Como era lógico suponer, en la adaptación española se siguieron esas mismas directrices: "para designar las líneas de corte de las tasas base en cada subgrupo de trastornos de personalidad y síndromes clínicos, se tomaron dichos números [los utilizados por Millon] de la siguiente manera: TB 74 designa la presencia de este tipo de personalidad o síndrome; TB 84 indica el tipo de personalidad o síndrome más alto o sobresaliente; TB 60 representa la puntuación de la mediana de todos los pacientes; y TB 35 es la puntuación media que se correspondería con los sujetos no clínicos" (Millon, 1998, pp. 174-175).

Aunque la transformación de PD a TB puede constituir *a priori* una buena estrategia para obtener los mejores puntos de corte diagnósti-

co, esta es una cuestión empírica que debe confirmarse con los estudios de validación adecuados. En este sentido, las directrices de la Comisión Internacional de Tests (ITC, 2000) afirman que se deberían "utilizar puntos de corte en la interpretación de las puntuaciones sólo cuando se disponga de datos empíricos sobre su validez" (Directriz 2.7.9.; ITC, 2000).

En consecuencia, la utilización de la TB 75 para indicar la presencia de un determinado tipo de trastorno de personalidad o de trastorno clínico, y la utilización de la TB 85 para indicar el tipo de trastorno de personalidad o de trastorno clínico más sobresaliente, debería contar con los datos apropiados que respaldaran la validez de dichas indicaciones o interpretaciones. Precisamente, este es uno de los puntos fuertes del MCMI-II. En el manual original del instrumento se justifica empíricamente la utilización de dichos puntos de corte analizando (1) los perfiles en el MCMI-II de distintos grupos diagnósticos, tanto en la muestra clínica normativa como en una muestra adicional de validación cruzada compuesta por 703 pacientes (véanse las pp. 109-137 de Millon, 1998, 1999 y 2002) y (2) la eficacia diagnóstica de los puntos de corte de las escalas tomando como criterio los diagnósticos realizados por los clínicos en esa muestra de validación cruzada (véanse la p. 129 y las pp. 138-143 de Millon, 1998, 1999 y 2002).

En general, los resultados de todos estos análisis apoyan de forma

satisfactoria la utilización del MCMI-II como instrumento diagnóstico a partir de las PD seleccionadas como TB 75 y TB 85 y, en especial, los resultados de los análisis de la eficacia diagnóstica de los puntos de corte, ya que, sin duda, estos datos de validez diagnóstica son los que mejor pueden justificar esa utilización del MCMI-II. Así, por ejemplo, en términos del valor o poder predictivo positivo (probabilidad de que un paciente identificado por el MCMI-II como padeciendo un determinado trastorno tenga realmente ese trastorno), el éxito diagnóstico de las TB 75 oscila entre 58% y 80% para las 13 escalas de trastornos de personalidad, con un promedio de 69%, y entre 50% y 92% para las 9 escalas de trastornos clínicos, con un promedio de 70% (véanse las Tablas 3.14 y 3.16 de Millon, 1998, 1999 y 2002). De forma parecida, los índices de sensibilidad (capacidad del MCMI-II para identificar correctamente a los pacientes que tienen un trastorno en particular) y de especificidad (capacidad del MCMI-II para identificar correctamente los pacientes que no tienen un trastorno en particular) son también satisfactorios. Para las escalas de trastornos de personalidad, la sensibilidad de las TB 75 varía entre 50% y 78%, con una media de 69%, y su especificidad oscila entre 91% y 99%, con una media de 95%, mientras que para las escalas de trastornos clínicos, la sensibilidad de las TB 75 fluctúa entre 34% y 87%, con una media de 62%, y su especificidad varía entre 83% y 99%, con una media de 95% (véanse las Tablas 3.14 y

3.16 de Millon, 1998, 1999 y 2002). Es habitual considerar como buenos índices de sensibilidad y especificidad valores de 70% o más, como aceptables índices entre 50% y 69% y como pobres índices menores de 50% (West y Finch, 1997), por lo que esos datos de eficacia diagnóstica, que pueden considerarse entre buenos y aceptables, no sólo corroboran la utilidad del MCMI-II para la evaluación psicopatológica, sino que son el apoyo empírico más necesario y sólido para emplear el MCMI-II como instrumento diagnóstico en un contexto clínico.

### **LOS PROBLEMAS DE LA ADAPTACIÓN ESPAÑOLA DEL MCMI-II**

Desgraciadamente, en su utilización como instrumento diagnóstico, la adaptación española del MCMI-II presenta varios problemas que tienen que ver, por un lado, con los baremos para la transformación de PD a puntuaciones TB y, por otro lado, con los índices de validez sobre las bondades diagnósticas de los puntos de corte TB para el diagnóstico.

#### **Problemas con los baremos de la adaptación española del MCMI-II**

Existen al menos cuatro problemas con los baremos de la adaptación española del MCMI-II que, tomados en conjunto, aconsejan ser cautos en su utilización con propósitos diagnósticos.

#### **Discrepancias entre baremos y tablas en la 1.<sup>a</sup> y 2.<sup>a</sup> edición del manual: errores tipográficos?**

Un primer problema con los baremos de la adaptación española del MCMI-II afecta a los usuarios de la 1.<sup>a</sup> y la 2.<sup>a</sup> edición de dicha adaptación, y tiene que ver con una discrepancia importante entre la información que aparece en una parte del manual (los baremos de los Anexos 3 y 4) y la información que aparece en otra parte de dicho manual (las Tabla 5.3 y 5.4). En concreto, en las dos primeras ediciones del manual español del MCMI-II, las PD que se corresponden con las TB de 35, 60, 75 y 85 en los baremos españoles que aparecen recogidos en los Anexos 3 y 4 (Millon, 1998, p. 185 y 186) *no coinciden*, respectivamente, con: (a) la mediana en PD de la muestra de española de personas normales que se recoge en la Tabla 5.4 del manual (Millon, 1998, p. 174), (b) la mediana en PD de la muestra clínica española de baremación que también se recoge en la Tabla 5.4, (c) la línea de corte en PD correspondiente a la TB 75 que se recoge en la Tabla 5.3 del manual (Millon, 1998, p. 173), y (c) la línea de corte en PD correspondiente a la TB 85 que también se recoge en la Tabla 5.3, cuando tales valores *deberían coincidir* según las directrices anteriormente mencionadas.

Como muestra de esta discrepancia, en la Tabla 1 se presentan las correspondencias para la escala de Depresión Mayor (CC) en las muestras normativas españolas de mujeres, aunque el problema al que se hace alusión es general y afecta

**Tabla 1. Correspondencia en cuanto a la definición de las puntuaciones de tasa base de la escala de Depresión Mayor (CC) en las muestras normativas de mujeres entre la información que aparece en los baremos de los Anexos del manual de la adaptación española del MCM-II y la información que aparece en las tablas de dicho manual**

	Versión Original - EE.UU.		Versión Española					
			1. <sup>a</sup> y 2. <sup>a</sup> edición				3. <sup>a</sup> edición	
	Muestra clínica de mujeres		Muestra clínica de mujeres		Muestra de mujeres con toxicomanía		Muestra clínica de mujeres	
	Baremos	Tablas	Baremos	Tablas	Baremos	Tablas	Baremos	Tablas
TB	PD	PD	PD	PD	PD	PD	PD	PD
85	44	44	50	71	50	n. e.	35	35
75	36	36	42	63	42	n. e.	31-32	31
60	17	17 <sup>a</sup>	23-31	50	23-31	50	24-25	21
35	5	n. e.	11	8	11	8	13-14	n. e.

Nota.—TB = Puntuaciones tasa base. PD = Puntuaciones directas. n. e. = No se especifican estos valores en el manual de la adaptación española del MCMI-II (1.<sup>a</sup> ed., 2.<sup>a</sup> ed. o 3.<sup>a</sup> ed.) <sup>a</sup>Medida de los pacientes (varones + mujeres) de la muestra de baremación estadounidense.

a todas las escalas del MCMI-II y a todas las muestras de baremación. Por ejemplo, y como puede observarse en la Tabla 1, en la versión española la TB 60 para la escala CC se corresponde en los baremos de pacientes mujeres de los Anexos con una PD de 23-31, mientras que la mediana de la muestra normativa de pacientes mujeres que se recoge en la Tabla 5.4 (y que debería ser la TB 60) es de 50. Por otro lado, la TB 35 para la escala CC se corresponde en los baremos de pacientes

mujeres de los Anexos con una PD de 11, mientras que la mediana de la muestra normal española de mujeres que también se recoge en la Tabla 5.4 (y que debería ser la TB 35) es de 8. Obviamente, tales valores deberían coincidir dada las definiciones de TB 50 y TB 35 que antes se han mencionado. De hecho, en el propio manual español se recogen baremos y tablas de la versión americana del manual, y entre dichos baremos y dichas tablas no hay discrepancia. Así, la TB 60 para



la escala CC en los baremos de las mujeres americanas que se recogen en los Anexos se corresponde con una PD de 17, justo el valor medio<sup>2</sup> de la muestra normativa de pacientes que aparece en la Tabla 2.21 también recogida en el manual español.

Curiosamente, en las dos primeras ediciones de la adaptación española, las TB 85 que aparecen en los baremos de los Anexos se corresponden con las medianas correspondientes en PD de la muestra normativa de pacientes de la Tabla 5.4, cuando dichas medianas deberían ser las TB 60. Por ejemplo, en los baremos de las pacientes mujeres de los Anexos, la TB 85 para la escala CC se corresponde con una PD de 50, justo el valor que aparece en la Tabla 5.4 como mediana (o TB 60) de la muestra normativa de pacientes mujeres para esa escala (véase la Tabla 1).

En definitiva, en las dos primeras ediciones de la adaptación española del MCMI-II hay una discrepancia entre los valores recogidos en los baremos de los Anexos del manual y los valores que aparecen en las tablas que se presentan en las páginas interiores del manual. Posiblemente, estas discrepancias tienen que ver con errores de tipo tipográfico, pero no queda claro si los errores están en las tablas, en los baremos de los Anexos, o en ambos,

aunque dado los baremos estadounidenses y los nuevos baremos españoles que se presentan en la 3.<sup>a</sup> edición de la adaptación española, es muy probable que los errores estén en las tablas. Lógicamente, el mayor problema para los profesionales e investigadores que han utilizado o utilizan el MCMI-II sería que los errores tipográficos estuvieran en los baremos de los Anexos, dado que entonces las transformaciones de PD a TB que se consiguen utilizando dichos baremos, tal y como dictan las normas de corrección del manual, serían totalmente incorrectas.

#### **Diferencias de los baremos españoles de la 1.<sup>a</sup> y 2.<sup>a</sup> edición del manual con los baremos estadounidenses**

Un segundo problema con los baremos españoles del MCMI-II es que, para algunas escalas, tales baremos difieren de los baremos estadounidenses más de lo que cabría esperar, lo cual arrojaría nuevas dudas sobre su validez. Por supuesto, esas diferencias podrían deberse a las características diferenciales de la población española frente a la estadounidense, características que precisamente justificarían la necesidad de una adaptación española del MCMI-II en lugar de una simple traducción. Sin embargo, tales diferencias son a veces muy grandes en magnitud y no son

---

<sup>2</sup> Aunque en la directriz se habla de mediana, los datos de las medianas americanas no aparecen recogidos en el manual, sino tan sólo las medias.

consistentes con las diferencias que se encuentran con otros instrumentos de evaluación psicopatológica.

Por ejemplo, en los baremos españoles de pacientes mujeres que se encuentran en los Anexos de la 1.<sup>a</sup> y 2.<sup>a</sup> edición del MCMI-II, la PD correspondiente a una TB 75 en la escala de Distimia (D) es 47, 16 puntos más que la PD que se corresponde a una TB 75 en los baremos estadounidenses de pacientes mujeres (PD = 31). Esta diferencia es aún mayor si al final resulta que los baremos españoles de los Anexos están equivocados y la PD que realmente corresponde a la TB 75 en la muestra normativa de pacientes españolas es la que aparece recogida en la Tabla 5.3 de las dos primeras ediciones del manual. Efectivamente, en este caso, la PD correspondiente a una TB 75 en la escala D es 64, 33 puntos más que la correspondiente a la TB 75 de los baremos estadounidenses. A partir de una muestra de 73 pacientes con diferentes trastornos mentales de la Clínica Universitaria de Psicología de la Universidad Complutense de Madrid, se ha obtenido una desviación típica de 17,4 en PD para la escala D, por lo que la discrepancia entre los baremos españoles y los estadounidenses en dicha escala

podría oscilar aproximadamente entre una y dos desviaciones típicas<sup>3</sup>, magnitudes considerables que no concuerdan con las diferencias que se encuentran con otros instrumentos que evalúan sintomatología depresiva. Por ejemplo, la media obtenida en la muestra normativa de pacientes españoles con el Inventario para la Depresión de Beck-II (BDI-II) es prácticamente igual a la media obtenida en muestras de pacientes estadounidenses (22,1 frente a 22,3; véase la Tabla 2), semejanza que es coherente con los resultados que se obtienen con dicho instrumento al comparar muestras españolas y estadounidenses seleccionadas de la población general o de la población de estudiantes universitarios (véase la Tabla 2).

### **Nuevos baremos en la 3.<sup>a</sup> edición del manual respecto a las dos ediciones anteriores**

Un tercer problema tiene que ver con la aparición de nuevos baremos para la población de pacientes españoles en los Anexos de la 3.<sup>a</sup> edición de la adaptación española (Millon, 2002) y que sustituyen a los baremos que aparecían en los Anexos de las dos ediciones anteriores.

---

<sup>3</sup> Con este tamaño muestral, el error que se comete al estimar la desviación típica de la escala D en la población de pacientes españoles con trastornos mentales puede ser grande, pero dicha estimación parece conservadora a la hora de considerar la magnitud de la discrepancia entre los baremos españoles y los estadounidenses. Por ejemplo, a partir de tres muestras que totalizaban 978 pacientes españoles con drogodependencias, Herrero (2004) obtuvo desviaciones típicas para la escala D que oscilaban entre 12,4 y 13,6, lo cual indicaría que la discrepancia entre los baremos españoles y los estadounidenses podría llegar a alcanzar más de 2,5 desviaciones típicas.

**Tabla 2. Comparación entre las medias obtenidas con el Inventario para la Depresión de Beck-II (BDI-II) en muestras estadounidenses y en muestras españolas**

	EE.UU.		España	
	<i>Media</i>	<i>N</i>	<i>Media</i>	<i>N</i>
Pacientes ambulatorios con trastornos mentales <sup>a</sup>	22,3	2653	22,1	305
Personas de la población general <sup>b</sup>	7,8	115	9,4	470
Estudiantes universitarios <sup>c</sup>	8,5	818	9,2	590

Nota.— <sup>a</sup>EE.UU. = Media ponderada por el número de participantes de las 8 muestras estadounidenses recogidas en Sanz et al. (2005); España = Sanz et al. (2005).

<sup>b</sup>EE.UU. = Hunt et al. (2003); España = Sanz, Perdigón y Vázquez (2003).

<sup>c</sup>EE.UU. = Media ponderada por el número de participantes de las 3 muestras estadounidenses recogidas en Sanz, Navarro y Vázquez (2003); España = Sanz, Navarro y Vázquez (2003).

Estos nuevos baremos parten de una muestra clínica mayor (710 pacientes frente a los 377 de los baremos antiguos)<sup>4</sup> lo que, en principio, permitiría obtener mejores puntuaciones de corte y, por ende, mejorar las características diagnósticas de la adaptación española del MCMI-II. Por tanto, su aparición obviamente no constituye en sí mismo un problema, sino en todo caso un punto fuerte de la adapta-

ción española del MCMI-II. Es más, en esta tercera edición, al contrario de lo que ocurría en las dos anteriores, no hay ninguna discrepancia entre la información que aparece en una parte del manual (los baremos de los Anexos) y la información que aparece en otra parte de dicho manual (la Tabla 5.3). Así, en esta tercera edición las PD que se corresponden con las TB 60, 75 y 85 en los baremos de los Anexos

<sup>4</sup> Sorprende, no obstante, que en la descripción de las características demográficas y clínicas de la nueva muestra normativa clínica (véase la Tabla 5.2 de la 3.<sup>a</sup> edición), los porcentajes de pacientes que se corresponden con cada una de las categorías de esas características (edad, situación/estado, duración del episodio reciente y problemas importantes del estado de los pacientes) coincidan exactamente, incluso en dos decimales, con los respectivos porcentajes de pacientes de la muestra normativa clínica de las ediciones anteriores (véase la Tabla 5.2 de la 1.<sup>a</sup>-2.<sup>a</sup> edición). Con la única excepción de los dos porcentajes del sexo, la coincidencia exacta abarca a 24 porcentajes, circunstancia muy poco probable por azar y que, probablemente, sea debido de nuevo a algún tipo de error tipográfico.

coinciden con las PD correspondientes en la Tabla 5.3 del manual (véase la Tabla 1 para un ejemplo).

No obstante, las diferencias entre ambos baremos son notables y por tanto los resultados obtenidos con ellos pueden diferir considerablemente. Por ejemplo, para la escala de Depresión Mayor (CC), la PD que se corresponde con la TB 75 en las dos primeras ediciones es 42, mientras que en la tercera edición es 11-10 puntos menor (31-32; véase la Tabla 1). En la muestra de pacientes de la Clínica Universitaria de Psicología anteriormente mencionada, se obtuvo una desviación típica de 13,1 en PD para la escala CC, por lo que la discrepancia entre los baremos de la tercera edición y los baremos de las ediciones anteriores podría alcanzar casi una desviación típica<sup>5</sup>, magnitud ciertamente considerable que podría suponer que con los antiguos baremos el trastorno depresivo mayor no fuera diagnosticado en un buen número de pacientes, pero sí lo fuera en esos mismos pacientes con los nuevos baremos de la 3.<sup>a</sup> edición.

De nuevo, el hecho de que los baremos difieran no es un problema en sí (siempre y cuando se pueda demostrar finalmente que los nuevos baremos mejoran la eficacia diagnóstica de la adaptación espa-

ñola del MCMI-II), pero se puede convertir en un problema si los investigadores o los profesionales no informan sobre qué edición del MCMI-II han aplicado y, por tanto, no es posible saber qué baremos han utilizado para convertir las PD en TB, o si los investigadores o los profesionales no son conscientes de la existencia de estos nuevos baremos y, por ejemplo, comparan erróneamente sus resultados con los obtenidos a partir de baremos distintos. Me temo que, lamentablemente, estas dos posibilidades pueden ser más frecuentes de lo que uno quisiera, y así lo sugiere el hecho de que de los 8 estudios españoles localizados que han utilizado el MCMI-II y que han sido publicados al menos dos años después de la publicación de la 3.<sup>a</sup> edición (entre 2004 y 2006), 3 de ellos no mencionaban la edición española que en concreto habían utilizado (tan sólo citaban la referencia al trabajo original americano).

#### **Diferencias de los baremos españoles de la 3.<sup>a</sup> edición del manual con los baremos estadounidenses**

Un cuarto problema con los baremos españoles del MCMI-II es que también los nuevos baremos de la 3.<sup>a</sup> edición difieren de los baremos

---

<sup>5</sup> De nuevo, esta estimación de la desviación típica de la escala CC debe tomarse con las debidas precauciones; sin embargo, dicha estimación posiblemente sea conservadora a la hora de considerar la magnitud de la discrepancia entre los baremos de la 3.<sup>a</sup> edición y los baremos de la 1.<sup>a</sup>-2.<sup>a</sup> edición, ya que Herrero (2004) obtuvo desviaciones típicas para la escala CC que oscilaban entre 9,8 y 10,7, lo cual indicaría que dicha discrepancia podría sobrepasar una desviación típica.

estadounidenses. Es cierto que dicha diferencia es menor que la que presentan los baremos españoles de las dos ediciones anteriores, pero, no obstante, en algunos casos la diferencia entre los nuevos baremos españoles y los estadounidenses parece mayor de lo que cabría esperar. Por supuesto cabe la posibilidad, tal y como se comentaba antes, de que esas diferencias se deban a las características diferenciales de la población española frente a la estadounidense y que, por tanto, dichas diferencias no sean un problema, sino una virtud de la adaptación. Pero, de nuevo, en algunas escalas las diferencias son notables en magnitud y no son consistentes con las ausencias de diferencias que se encuentran con otros instrumentos de evaluación psicopatológica. Por ejemplo, en los baremos españoles de pacientes mujeres que se encuentran en los Anexos de la 3.<sup>a</sup> edición del MCMI-II, la PD correspondiente a una TB 75 en la escala de Distimia es 40-41, lo que supone 9-10 puntos más que la PD que se corresponde a una TB 75 en los baremos estadounidenses de pacientes mujeres (PD = 31) y, por tanto, representaría una discrepancia de más de media desviación típica entre ambos baremos que no es consistente con la ausencia de diferencias entre muestras españolas y estadounidenses que presenta, por ejemplo, el BDI-II.

En cualquier caso, las diferencias entre los baremos estadounidenses del MCMI-II original y los baremos españoles de cualquiera de las ediciones españolas del MCMI-II

cobran sobre todo relevancia en el contexto de los problemas de validación diagnóstica de la propia adaptación española del MCMI-II y a los que me referiré a continuación.

### **Problemas con los índices de validez diagnóstica de la adaptación española del MCMI-II**

Como se comentaba líneas atrás, la utilización de puntos de corte en la interpretación de las puntuaciones de un test requiere la existencia de datos sobre la validez de dichos puntos de corte (ITC, 2000). Por tanto, la utilización con propósitos diagnósticos de las TB 75 y TB 85 obtenidas a partir de los baremos españoles de cualquiera de las ediciones de la adaptación española del MCMI-II requeriría la existencia de estudios de validez sobre la eficacia diagnóstica de dichos puntos de corte.

Lamentablemente, ninguna de las ediciones de la adaptación española del MCMI-II presenta estudios o datos de validez diagnóstica de los puntos de corte obtenidos a partir de los baremos españoles. De hecho, las dos primeras ediciones de la adaptación española del MCMI-II (Millon 1998, 1999) no incluyen ningún dato o estudio empírico sobre la validez de dicha adaptación en cualquiera de los aspectos en que la validez puede examinarse (p. ej., validez factorial, validez convergente, validez discriminante, validez de contenido, validez de criterio, validez diagnóstica), limitándose tan sólo a presentar los

índices de fiabilidad de consistencia interna de las escalas (los cuales, en general, son adecuados; véase la Tabla 5.5 de Millon 1998, 1999).

La 3.<sup>a</sup> edición de la adaptación española amplia la información sobre las propiedades psicométricas de la misma incluyendo datos sobre su validez factorial y sobre su validez convergente y discriminante respecto a las escalas de la adaptación española del MMPI-2 (véanse las Tablas 5.8 y 5.9 de Millon, 2002), pero tales índices, aunque relevantes para otros propósitos del MCMI-II, no lo son para sustentar la utilización del MCMI-II como instrumento diagnóstico. Para ello, la adaptación española debería incluir información parecida a la que presenta el MCMI-II original, especialmente un análisis, en muestras de pacientes españoles, de la eficacia diagnóstica (p. ej., sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo) de las TB 75 o de las TB 85 obtenidas a partir de los baremos españoles. La ausencia de este tipo de información empírica no permite sustentar con un mínimo de seguridad la utilización del MCMI-II como instrumento diagnóstico a partir de los puntos de corte definidos por las TB 75 y TB 85. Quizás tales datos, de existir, apoyarían las bondades psicométricas de la adaptación española del MCMI-II en cuanto a su validez diagnóstica, pero este extremo es una cuestión empírica que requiere precisamente la realización de tales estudios y, en su ausencia, lo más prudente es no realizar ningún diagnóstico basado únicamente en una puntuación por

encima de la TB 75 o la TB 85 en alguna de las escalas de la adaptación española del MCMI-II.

Podría argumentarse, en apoyo de la utilización diagnóstica de la adaptación española del MCMI-II, que los datos de validación diagnóstica que presenta el MCMI-II original en muestras de pacientes estadounidenses son generalizables a las muestras españolas. Sin embargo, la generalización de los índices de validación no se puede asumir automáticamente, sino que debería confirmarse empíricamente. De hecho, esta asunción es más problemática dado que, como se ha visto, los baremos y, por tanto, los puntos de corte TB 75 y TB 85 de la versión original son diferentes de los de la adaptación española en cualquiera de sus tres ediciones.

Es más, la estrategia de transformar las PD en puntuaciones TB no siempre tiene por qué lograr puntos de corte con índices adecuados de validez diagnóstica, ya que el éxito de dicha estrategia depende de factores tales como el tamaño y representatividad de la muestra de baremación y la forma y el cuidado con que los clínicos realizaron los diagnósticos en dicha muestra (dependiendo, por ejemplo, del tiempo de contacto del clínico con el paciente, de la utilización de entrevistas libres, estructuradas o semiestructuradas, del uso de listados de síntomas o criterios diagnósticos de comprobación, de la verificación posterior de los diagnósticos, de la realización de análisis de fiabilidad interjueces, etc.). A estos factores

también habría que añadir las diferencias culturales, ya que las tasas de prevalencia de los trastornos suelen ser diferentes en distintos países, aunque estos países estén tan cercanos culturalmente como los europeos. Por ejemplo, estudios recientes indican que la prevalencia a los 12 meses de los trastornos depresivos y de los trastornos de ansiedad difiere de forma significativa en España, Holanda, Francia, Italia, Alemania y Bélgica (Comisión Europea, 2004). Estos datos cuestionarían el procedimiento utilizado en la adaptación española del MCMI-II de asumir las tasas de prevalencia estadounidenses para delimitar los puntos de corte TB 75 y TB 85<sup>6</sup>. En conjunto, estos factores, pues, también apuntarían a la necesidad de realizar un estudio de validación diagnóstica con muestras españolas de los puntos de corte TB 75 y TB 85 obtenidos a partir de los baremos, en lugar de asumir la generalización de los índices de validación estadounidenses. De hecho, en Bélgica, la adaptación del MCMI-III ha tenido que prescindir de la estrategia de Millon de transformar las PD en puntuaciones TB para obtener los puntos de corte óptimos para el diagnóstico, ya que dicha estrategia ofrecía índices diagnósticos muy pobres en las muestras de pacientes belgas y, en su lugar, se ha utilizado el análisis

de las curvas ROC (*Receiver Operating Characteristic* o curvas de rendimiento predictivo) que permiten representar gráficamente la sensibilidad y la especificidad para cada una de las PD de las escalas y, en consecuencia, seleccionar como punto de corte para cada escala la PD que mejor rendimiento diagnóstico ofrece (Rossi y Sloore, 2005).

## CONCLUSIONES

El MCMI-II fue diseñado, entre otros objetivos, para llegar a un diagnóstico del trastorno o de los trastornos mentales que puede estar sufriendo un paciente. En su versión original, los puntos de corte que propone el instrumento para realizar esos diagnósticos a partir de la estrategia de transformar las PD en TB están avalados por los resultados satisfactorios de un estudio de validación diagnóstica en una muestra cruzada de pacientes estadounidenses.

La adaptación española del MCMI-II, en sus tres ediciones, ha sido también utilizada por algunos profesionales e investigadores españoles para la realización de diagnósticos. Sin embargo, la adaptación española muestra varias inconsistencias en los baremos para la transformación de PD a puntuaciones TB

<sup>6</sup> El manual de las dos primeras ediciones de la adaptación española explícitamente afirma la asunción de las tasas de prevalencia estadounidenses para la obtención de los puntos de corte (p. 174, Millon, 1998, 1999). Tal afirmación ha desaparecido de la 3.<sup>a</sup> edición de la adaptación, pero tampoco hay ninguna información en dicha edición sobre el procedimiento finalmente utilizado para la obtención de los puntos de corte.

(errores tipográficos, altas discrepancias con los baremos americanos, cambios de baremos entre ediciones) que hacen dudar de la adecuación de tales baremos, pero, sobre todo, la adaptación española propone puntos de corte TB para el diagnóstico cuya validez no viene avalada por ningún tipo de estudio empírico. Hasta que tales inconsistencias en los baremos hayan sido aclaradas y hasta que se obtengan datos que sustenten la validez diagnóstica de los puntos de corte propuestos a partir de esos baremos, lo más prudente es que la adaptación española del MCMI-II *no sea utilizada con fines diagnósticos* o, en todo caso, no sea utilizada como criterio único para llegar a un diagnóstico y que, por tanto, cualquier inferencia diagnóstica que se haga a partir de la presencia de una puntuación que supere el punto de corte establecido sea posteriormente confirmada mediante algún tipo de entrevista diagnóstica realizada por un clínico.

Por supuesto, los problemas señalados en este artículo no afectan a la utilización del MCMI-II como instrumento para detectar síntomas y síndromes psicopatológicos y para evaluar su gravedad. El mismo proceso de construcción del MCMI-II y el propio contenido de sus ítems, en conjunción con los datos de fiabilidad, validez factorial y validez convergente que presenta el manual español, avalan ese uso de la adaptación española del MCMI-II, lo que obviamente es muy importante para evaluar, por ejemplo, el estado general de los pacientes o la efectividad de un proceso terapéutico.

Sin embargo, no hay que confundir los niveles de análisis síntoma/síndrome/trastorno. El diagnóstico de un trastorno se efectúa teniendo en cuenta no sólo el tipo y número de síntomas presentes, sino también ciertos criterios que difícilmente puede evaluar el MCMI-II a partir del contenido de sus ítems como, por ejemplo, el grado de interferencia (p. ej., los síntomas deben provocar deterioro social, escolar, o laboral) y la ausencia de ciertas causas posibles o de ciertos diagnósticos concurrentes (p. ej., para el diagnóstico de algunos trastornos, los síntomas no deben considerarse el efecto fisiológico directo de una enfermedad médica o de la ingestión de medicamentos o drogas, o no se deben explicar mejor por la presencia de otro trastorno mental). A pesar de las limitaciones que el MCMI-II y, en general, la mayoría de los cuestionarios tienen para evaluar esos criterios diagnósticos, las puntuaciones que ofrece el MCMI-II podrían servir para estimar la presencia de un trastorno, tal y como Millon pretendía, siempre y cuando el MCMI-II, como cualquier otro cuestionario, pueda demostrar su validez diagnóstica. Por tanto, futuras investigaciones con la adaptación española del MCMI-II deberían confirmar la validez diagnóstica de las puntuaciones de corte propuestas mediante el correspondiente estudio empírico, lo cual fue precisamente lo que el propio Millon realizó con su versión original y lo que también demandan las directrices internacionales sobre el diseño y uso de tests psicológicos (ITC, 2000).



Finalmente, la novedosa utilización en todas las versiones del MCMI de la transformación de PD en puntuaciones TB y la constatación de un cambio en la 3.<sup>a</sup> edición española del MCMI-II en cuanto a los baremos que sirven para obtener dichas TB, sugiere que tanto los profesionales como los investigadores que utilicen la adaptación española del MCMI-II deberían especificar claramente en sus informes: (a) qué tipo de puntuaciones están informando o analizando, las PD o

las TB; (b) qué baremos en concreto están utilizando para transformar las PD a TB, los de la 1.<sup>a</sup>-2.<sup>a</sup> edición o los de la 3.<sup>a</sup> edición, ya que los resultados pueden ser muy diferentes, y (c) qué punto de corte en concreto están utilizando para realizar cualquier inferencia sobre la presencia de un trastorno, la TB 75 o la TB 85, inferencia que, de hacerse, debería por ahora verse acompañada de la oportuna confirmación mediante algún tipo de entrevista diagnóstica realizada por un clínico.

## REFERENCIAS

Butcher, J. N., y Rouse, S. V. (1996). Personality: Individual differences and clinical assessment. *Annual Review of Psychology*, 47, 87-111.

European Commission (2004). *The state of mental health in the European Union*. Bruselas: Comisión Europea.

Herrero, J. (2004). Alteraciones de la personalidad asociadas a las conductas adictivas: influencia de la duración del consumo y sus implicaciones. *Psyche*, 13, 91-100.

Hunt, M., Auriemma, J., y Cashaw, A. C. A. (2003). Self-report bias and underreporting of depression on the BDI-II. *Journal of Personality Assessment*, 80(1), 26-30.

International Test Commission (2000).

*Directrices Internacionales para el Uso de los Tests de la Comisión Internacional de Tests (ITC)*. [Traducción de la Comisión de Tests del Colegio Oficial de Psicólogos]. Documento consultado en Internet en: <http://www.cop.es/vernumero.asp?id=40>.

Millon, T. (1983). *Millon Clinical Multiaxial Inventory manual* (3.<sup>a</sup> ed). Minneapolis. MN: National Computer Systems.

Millon, T. (1987). *Millon Clinical Multiaxial Inventory-II manual*. Minneapolis. MN: National Computer Systems.

Millon, T. (1998). *MCMI-II. Inventario Clínico Multiaxial de Millon II. Manual*. Madrid: TEA.

Millon, T. (1999). *MCMI-II. Inventario Clínico*

co *Multiaxial de Millon II. Manual* (2.<sup>a</sup> ed.). Madrid: TEA.

Millon, T. (2002). *MCMI-II. Inventario Clínico Multiaxial de Millon II. Manual* (3.<sup>a</sup> ed.). Madrid: TEA.

Millon, T., Millon, C., y Davis, R. (1994). *Millon Clinical Multiaxial Inventory-III manual*. Minneapolis, MN: National Computer Systems.

Piotrowski, C. (1997). Use of the Millon Clinical Multiaxial Inventory in clinical practice. *Perceptual and Motor Skills*, 84, 1185-1186.

Rossi, G., y Sloore, H. (2005). International uses of the MCMI: Does interpretation change? En R. Craig (Ed.), *New directions in interpreting the Millon Clinical Inventory-III (MCMI-III)* (pp. 144-163). Hoboken, NJ: Wiley.

Sanz, J., Navarro, M. E., y Vázquez, C.

(2003). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 1. Propiedades psicométricas en estudiantes universitarios. *Análisis y Modificación de Conducta*, 29(124), 239-288.

Sanz, J., Perdigón, L. A., 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.

Sanz, J., García-Vera, M. P., Espinosa, R., Fortún, M., y Vázquez, C. (2005). Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. *Clínica y Salud*, 16(2), 121-142.

West, S. G., y Finch, J. F. (1997). Personality measurement: Reliability and validity issues. En R. Hogan, J. Johnson y S. Briggs (Eds.), *Handbook of personality psychology* (pp. 143-164). San Diego, CA: Academic Press.