

# El error estándar de medida: ¿una alternativa a la diferencia mínimamente importante para evaluar cambios en las medidas de salud autopercebida?

J. REJAS, M. RUIZ<sup>1</sup>, A. PARDO<sup>1</sup>

*Departamento de Investigación de Resultados en Salud. Unidad Médica Pfizer España. Alcobendas, Madrid. <sup>1</sup>Departamento de Metodología. Facultad de Psicología. Universidad Autónoma. Madrid*

## RESUMEN

**Objetivo:** Valorar el grado de concordancia existente entre los criterios error estándar de medida (EEM) y diferencia mínimamente importante (DMI) para evaluar la magnitud del cambio producido por una intervención médica en el estado de salud autopercebido.

**Método:** Se han utilizado dos escalas de salud autopercebida por el paciente [una medida de impacto de la hiperplasia prostática benigna (BPH-PIM) y una escala de satisfacción con la vida (LISAT-8)] en dos estudios independientes realizados para adaptar ambas escalas a la población española. Estos estudios incluían 129 y 537 varones con hiperplasia benigna de próstata (HBP) y disfunción eréctil (DE), respectivamente. Tras calcular los correspondientes valores EEM y DMI se ha clasificado a los pacientes con cada criterio en tres grupos: mejoría, sin cambio y empeoramiento. Para calcular el grado de concordancia entre ambos criterios se ha utilizado el índice de acuerdo kappa de Cohen y el coeficiente de correlación lineal tau-b de Kendall.

**Resultados:** El grado de concordancia encontrado entre los criterios EEM y DIM ha sido alto en ambas escalas: 0,81 y 0,76 para índice de acuerdo kappa, y 0,87 y 0,79 para el coeficiente de correlación tau-b.

**Conclusión:** El grado de acuerdo encontrado permite afirmar que el criterio EEM representa una alternativa apropiada al criterio DMI para evaluar la magnitud del cambio producido en la salud autopercebida de los pacientes.

**PALABRAS CLAVE:** Diferencia mínimamente importante. Error estándar de medida. Medidas de salud autopercebida. Relevancia clínica. Magnitud del cambio.

*Rejas J, Ruiz M, Pardo A. El error estándar de medida: ¿una alternativa a la diferencia mínimamente importante para evaluar cambios en las medidas de salud autopercebida. An Med Interna (Madrid) 2007; 24: 415-420.*

## INTRODUCCIÓN

La utilización de cuestionarios de salud y la investigación en el área de la calidad de vida relacionada con la salud (CVRS) se remonta varias décadas atrás. Sin embargo, no ha sido hasta hace relativamente pocos años cuando se ha despertado un inusitado interés por el uso y aplicación en la práctica médica de este tipo de resultados en salud (1,2).

La interpretación de la puntuación obtenida por un paciente en los instrumentos de CVRS se puede realizar comparando

## STANDARD ERROR OF MEASUREMENT: AN ALTERNATIVE TO MINIMALLY IMPORTANT DIFFERENCE TO ASSESS CHANGES IN PATIENT-REPORTED-HEALTH-OUTCOMES?

## ABSTRACT

**Objective:** To assess the concordance level between the standard Error of Measurement (EEM) and the minimal important difference (MID) criteria when evaluating the magnitude of change in self-perceived health status due to a clinical intervention.

**Method:** Two scales of patient self-reported health are used [a measure of benign prostatic hyperplasia impact (BPH-PIM) and a life satisfaction scale (LISAT-8)] in two independent studies carried out in order to adapt each scale to Iberian Spanish population. These studies included 129 and 537 males with benign prostatic hyperplasia (BPH) and erectile dysfunction (ED), respectively. After computing EEM and MID values, patients were classified in three groups attending to each criterion: improvement, without change, and worsening. Cohen's kappa agreement index and Kendall's tau-b lineal correlation coefficient were used to assess the level of concordance between criteria.

**Results:** The concordance level between EEM and MID criteria was found to be high for both scales: kappa agreement index attained values of 0.81 and 0.76, while tau-b correlation coefficient attained values of 0.87 and 0.79.

**Conclusion:** The agreement level found allows to state that the EEM criterion is an adequate alternative to the MID criterion when assessing the magnitude of change produced in patient reported health.

**KEY WORDS:** Minimal important difference. Standard error of measurement. Patient self-reported output. Clinical relevance. Magnitude of change.

la puntuación observada en el cuestionario con normas poblacionales que permiten situar al paciente en un determinado percentil (3,4). De forma equivalente, también es posible tipificar las puntuaciones observadas para valorar la distancia (expresada en unidades de desviación típica) existente entre el estado actual del paciente y la media de la población (5). Estas estrategias se basan en la distribución de las puntuaciones.

En el caso de las mediciones repetidas en el tiempo interesa valorar, además, si el instrumento de medida es capaz de detectar un cambio en el estado de salud del paciente y la

*Trabajo aceptado:* 14 de mayo de 2007

*Correspondencia:* Javier Rejas Gutiérrez. Departamento de Investigación de Resultados en Salud. Unidad Médica Pfizer España. Parque Empresarial de La Moraleja. Avda. Europa, 20B. 28108 Alcobendas, Madrid. e-mail: javier.rejas@pfizer.com

magnitud del mismo (lo que habitualmente se conoce como *tamaño del efecto*) (6,7). La estimación del tamaño del efecto también se basa en la distribución de las puntuaciones.

Otra forma de interpretar las puntuaciones de una escala consiste en utilizar estados de referencia (estados ancla) para calcular lo que se ha dado en llamar *diferencia mínimamente importante* (DMI) o diferencia clínicamente relevante, o cambio mínimo con relevancia clínica (8-10). Es decir, el cambio mínimo que un paciente es capaz de reconocer en su estado de salud (9). Mientras que el tamaño del efecto se calcula a partir de las puntuaciones observadas en la escala, sin implicar al paciente más allá de la mera cumplimentación de la misma (es una estrategia puramente estadística), la DMI requiere de la implicación directa del paciente y del uso de una valoración global de cambio que permita calcular estados ancla. Esta estrategia para medir la magnitud del cambio incorpora la percepción que tiene el paciente del cambio que ha experimentado su estado de salud durante el tiempo de observación o exposición a la intervención médica, lo que aumenta la complejidad de su medición (11,12).

En los últimos años, el grupo de Wyrwich y cols en primer lugar (13-15) y otros autores posteriormente (16-18), han venido utilizando, como estimador directo de la DMI, un conocido indicador estadístico basado en la distribución de las puntuaciones: el *error estándar de medida* (EEM) (3,5,8). El cálculo del EEM no requiere de la utilización de valores de referencia o estados ancla convencionales, lo que permite valorar el cambio relevante en un instrumento de medida con criterios estadísticos, teniendo además en cuenta la precisión del instrumento de medida (aspecto éste obviado con las estimaciones de cambio basadas en estados ancla) (13). Estos autores han encontrado una elevada concordancia entre los criterios EEM y DMI (14-16): la mayoría de los pacientes cuyas puntuaciones en CVRS experimentan un cambio superior a la DMI, muestran también un cambio superior a un EEM. Si el grado de concordancia fuera alto, el criterio EEM podría utilizarse para determinar si el cambio experimentado en una escala de CVRS es o no clínicamente relevante (15,16). Nosotros hemos demostrado previamente que el grado de concordancia es alto en medidas de salud autopercibida que miden la presencia de síntomas de ansiedad, depresión, alteración del sueño o discapacidad en pacientes con dolor neuropático, pero que no incluían la medición de otras facetas de la salud auto-percibida como la CVRS o la satisfacción con la vida (20). El presente trabajo pretende valorar si la utilización del criterio EEM como estimador del criterio DMI se mantiene en dos escalas distintas de salud autopercibida que miden CVRS y satisfacción con la vida en varones con hiperplasia prostática y disfunción eréctil.

## MATERIAL Y MÉTODOS

### DESCRIPCIÓN DE LAS ESCALAS UTILIZADAS

Cuestionario de satisfacción con la vida de Fugl-Meyer de 8 ítems (LISAT-8) (21): cuestionario desarrollado para evaluar la satisfacción con la vida de población adulta sometida a programas de rehabilitación y que, posteriormente, ha sido validado en varones con disfunción eréctil (DE). Es un cuestionario autoadministrado de 8 ítems tipo Likert con 6 categorías de respuesta (de “muy insatisfactoria” a “muy satisfacto-

ria”) que miden la satisfacción con 8 facetas diferentes de la vida del paciente: vida en general, vida sexual, relación de pareja, vida familiar, relación con amigos y conocidos, ocio, situación laboral y situación económica. Ofrece una puntuación total sumando las puntuaciones de todos los ítems (8 puntos indican máxima insatisfacción con la vida y 48 puntos máxima satisfacción). Esta puntuación se puede expresar, también, en una métrica estandarizada con puntuación mínima de 0 y máxima de 100.

Medida de impacto en el paciente de la hiperplasia benigna de próstata (BPH-PIM) (22): instrumento específico, auto-aplicado con supervisión, que explora CVRS en pacientes con hiperplasia benigna de próstata (HBP). Consta de dos partes diferenciadas: una escala de 15 preguntas tipo Likert con 5 categorías de respuesta (de “mucho” a “nada en absoluto”; mayor puntuación indica menor impacto) y un cuestionario con 16 actividades de la vida cotidiana de las que el paciente debe elegir 5 y evaluar, en una escala tipo Likert con las categorías de respuesta ya mencionadas, el grado en que la HBP le limita cada actividad. Las puntuaciones de la primera parte oscilan entre 15 y 75 puntos; las de la segunda, entre 5 y 25. Las puntuaciones de ambas partes se suman en una puntuación global con rango 20-100 (mayor y menor impacto, respectivamente, de la HBP en la CVRS del paciente). Para facilitar la interpretación, la puntuación global se transforma en una puntuación derivada con rango 0-100 (0 = mayor impacto de la HBP y, por tanto, peor calidad de vida; 100 = menor impacto de la HBP, lo que equivale a mejor calidad de vida). Se pueden obtener puntuaciones derivadas para cada parte por separado.

Como instrumentos de anclaje para el cálculo de la diferencia mínimamente importante se ha utilizado el “Índice internacional de función eréctil” (IIEF) y la “Escala de valoración global del cambio de Junniper” (ver más adelante).

### MUESTRA DE PACIENTES DEL ESTUDIO

Se han utilizado las muestras de pacientes incluidas en los estudios de adaptación a la población española de las dos escalas utilizadas en este trabajo (21,22). En la adaptación de la escala LISAT-8 se utilizó una submuestra de 537 pacientes seleccionados aleatoriamente por ordenador de una muestra de 2.816 pacientes con DE incluidos en un estudio previo (estudio IDEAP: abierto, de corte naturalístico, prospectivo y multicéntrico, desarrollado en centros de atención primaria) en el que se evaluaba la efectividad del tratamiento con dosis flexibles de sildenafil durante 10 semanas en pacientes con DE y patología orgánica concomitante (23). La muestra estaba estratificada según la intensidad de la disfunción eréctil medida por el cuestionario *Índice internacional de función eréctil* (IIEF) (24).

En la validación del BPH-PIM (22) se utilizó una muestra de 129 varones con HBP diagnosticada por tacto rectal y/o ecografía, sintomática, con una puntuación media basal en el IPSS (Cuestionario internacional de síntomas prostáticos) igual o mayor que 7 puntos. Los pacientes eran varones ambulatorios, de más de 40 años de edad, con prostatismo de intensidad leve a severo medido por la escala IPSS, capaces de cumplimentar por sí mismos los cuestionarios IPSS y BPH-PIM, y susceptibles de ser tratados con un alfa-1-bloqueante (Doxazosina GITS). Los criterios de exclusión fueron: insufi-

ciencia hepática grave, antecedentes de obstrucción gastrointestinal, obstrucción esofágica o cualquier grado de disminución del diámetro de la luz del tubo digestivo, reacción farmacológica grave documentada o sospechada clínicamente o idiosincrasia a los bloqueadores alfa-adrenérgicos o a las quinazolininas en general, antecedentes de hipotensión ortostática, síncope o presencia de presión arterial sistólica y/o diastólica inferior a 95 y/o 60 mmHg, y tratamiento concomitante con otros bloqueadores alfa.

#### DIFERENCIA MÍNIMAMENTE IMPORTANTE (DMI) Y MAGNITUD DEL CAMBIO

La magnitud del cambio experimentado en las escalas incluidas en este estudio se ha estimado mediante el tamaño del efecto (TE), la respuesta media estandarizada (RME) y la diferencia mínimamente importante o diferencia mínima con relevancia clínica (DMI).

El tamaño del efecto se ha calculado, siguiendo a Kazis y cols, (6) dividiendo la diferencia entre las respuestas medias observadas antes y después de la intervención farmacológica (4 semanas con doxazosina y 12 semanas con sildenafil, respectivamente) por la desviación típica observada en el momento basal. La respuesta media estandarizada se ha calculado dividiendo el cambio medio entre las visitas inicial y final por la desviación típica del cambio (3,5,8).

Para el cuestionario LISAT-8, la DMI se ha calculado utilizando como criterio de referencia o anclaje el cambio observado en la respuesta de los pacientes en la dimensión *función eréctil* de la escala IIEF (24). Este cambio se obtiene a partir de la variación experimentada en esa dimensión entre la respuesta basal y el postratamiento. Se han considerado cuatro posibles niveles de cambio: sin cambio (hasta 5 puntos), cambio pequeño (de 6 a 10 puntos), cambio moderado (de 11 a 15 puntos) y cambio grande (más de 15 puntos). Se han seleccionado estos rangos de puntuación porque son los que, en la dimensión *función eréctil* de la escala IIEF, indican que se ha producido un cambio de nivel en la de intensidad de la DE. Así, una variación de 5 puntos o menos puntos indica que la intensidad de la disfunción permanece en el mismo nivel o que ha cambiado un nivel; una variación de 6 a 10 puntos indica un cambio de dos niveles; una variación de 11 a 15 puntos indica un cambio de tres niveles; y una variación superior a 15 puntos indica un cambio de cuatro niveles (25). Esta estrategia concuerda con la seguida por Jaeschke y cols, (8) y Juniper y cols (9) para definir el concepto de diferencia mínimamente importante, y ha sido posteriormente ratificada por otros autores (10). La DMI se calcula como la diferencia entre la puntuación media de los sujetos que no experimentan cambio en su función eréctil según la escala LISAT-8 y la puntuación media de los sujetos que cambian un nivel.

Para calcular la DMI con el cuestionario BPH-PIM, los estados de referencia o anclaje se han fijado a partir de la escala de valoración global de Juniper y cols. (9). Se ha considerado la categoría “más o menos lo mismo” representa la ausencia de cambio; las categorías “casi lo mismo”, “un poco mejor (peor)” o “algo mejor (peor)” representan un cambio pequeño; las categorías “moderadamente mejor (peor)” o “bastante mejor (peor)” representan un cambio moderado; y las categorías “mucho mejor (peor)” o “muchísimo mejor (peor)” representan un cambio grande. La DMI en el cuestionario BPH-

PIM se ha calculado como la diferencia entre la puntuación media de los sujetos que no experimentan cambio según la escala de referencia y la puntuación media de los sujetos que experimentan un cambio pequeño.

Una vez calculada la DMI, se han utilizado las puntuaciones de ambas escalas para clasificar a los pacientes en tres categorías (cada paciente ha sido clasificado utilizando la escala correspondiente a su estudio): *mejoría* (si la puntuación aumenta más de una DMI), *sin cambio* (si la puntuación cambia entre más una y menos una DMI), *empeoramiento* (si la puntuación disminuye una DMI o más).

#### ERROR ESTÁNDAR DE MEDIDA (EEM)

El error estándar de medida (EEM) se ha obtenido multiplicando la desviación típica de la correspondiente escala en el momento basal por la raíz cuadrada de uno menos el coeficiente de fiabilidad de esa escala (3). La fiabilidad de las escalas se ha estimado mediante el coeficiente  $\alpha$  de Cronbach (3). Al valor obtenido se le ha llamado EEM (1). Tomando como referencia este valor y la diferencia (d) entre las puntuaciones basal y final de cada escala, los pacientes se han clasificado en tres grupos: a) *mejoría*: pacientes con  $d > EEM$  (1); b) *sin cambio*: pacientes con d comprendida entre  $\pm EEM$  (1); y c) *empeoramiento*: pacientes con  $d < -EEM$  (1). Esta misma clasificación se ha llevado a cabo utilizando los criterios EEM (1,96) y EEM (2,58). Para evaluar la concordancia entre los diferentes criterios, las clasificaciones resultantes de aplicar los 3 criterios EEM se han comparado con la clasificación resultante de aplicar el criterio DMI.

#### ANÁLISIS ESTADÍSTICO

En todas las variables de interés se han calculado estadísticos de tendencia central y de dispersión, y se ha valorado el ajuste a la normalidad mediante la prueba de Kolmogorov-Smirnoff. Tras clasificar a los pacientes en las categorías *mejoría*, *sin cambio* y *empeoramiento* con los criterios DMI, EEM(1), EEM(1,96) y EEM(2,58), se ha utilizado el coeficiente de acuerdo kappa y el coeficiente de correlación tau-b de Kendall para valorar la concordancia entre el criterio DMI y los criterios EEM.

También se ha calculado el tamaño del efecto, la respuesta media estandarizada y el coeficiente de fiabilidad alpha de Cronbach. Los valores DMI y EEM se han calculado con la escala total y con cada ítem. Todos los análisis se han realizado con el paquete estadístico SPSS (versión 13.0).

#### RESULTADOS

En el estudio de validación del cuestionario LISAT-8 se analizaron un total de 537 pacientes con DE [edad media (desviación típica) = 58,2 (9,1) años, y 3,2 (2,9) años de evolución de la DE], de los que 323 (60,2%) sufrían DE severa [6 a 10 puntos en el Índice Internacional de Función Eréctil (IIEF)]; 144 (26,8%) moderada [11 a 16 puntos], y 70 (13,0%) leve [17 a 24 puntos]. En el estudio de adaptación del cuestionario BPH-PIM se incluyeron 129 varones con una edad media de 66,2 (8,32) años. De éstos, 116 completaron el estudio y 115

complimentaron los cuestionarios en el momento basal y en el postratamiento. Todos los pacientes presentaban HBP, con una evolución media de 3,6 (3,1) años, diagnosticada por tacto rectal y/o ecografía, sintomática, con una puntuación media basal en el cuestionario de síntomas prostáticos IPSS de 17,11 (6,17) puntos.

Las dos escalas analizadas mostraron una elevada fiabilidad (coeficientes alpha de Cronbach de 0,879 y 0,843 (Tabla I). Y las estimaciones obtenidas para la magnitud del efecto son coherentes cuando se expresan corregidas por el número de ítems de la escala (Tabla I).

TABLA I

VALORES OBTENIDOS PARA LA DIFERENCIA MÍNIMAMENTE IMPORTANTE (DMI), EL ERROR ESTÁNDAR DE MEDIDA (EEM) Y OTROS ESTIMADORES DE LA MAGNITUD DEL EFECTO EN LAS ESCALAS BPH-PIM Y LISAT-8

Escala	Magnitud del cambio				TE	RME	Coeficiente de fiabilidad alpha de Cronbach
	DMI		EEM				
	Escala total	Por ítem	Escala total	Por ítem			
BPH-PIM	2,5	0,13	4,1	0,21	0,62	0,71	0,879
LISAT-8	8,1	1,01	5,8	0,73	0,67	1,09	0,843

TE: tamaño del efecto; RME: respuesta media estandarizada.

La tabla II muestra los valores del índice de acuerdo kappa y del coeficiente de correlación tau-b obtenidos al cruzar la clasificación efectuada con el criterio DMI y la efectuada con los criterios EEM(1), EEM(1,96) y EEM(2,58). Según el criterio DMI, un 73,1% de los pacientes con HBP muestran mejoría en la escala BPH-PIM tras 4 semanas de tratamiento con doxazosina, mientras que un 19,4% no modifica su nivel de calidad de vida y un 7,4% empeora. Utilizando el criterio EEM (1), los valores correspondientes fueron, respectivamen-

TABLA II

PORCENTAJE DE CASOS CLASIFICADOS EN CADA CATEGORÍA CON LOS CRITERIOS DMI (DIFERENCIA MÍNIMAMENTE IMPORTANTE) Y EEM (ERROR ESTÁNDAR DE MEDIDA), Y CONCORDANCIA ENTRE AMBOS CRITERIOS EN LAS DOS ESCALAS INCLUIDAS EN EL ESTUDIO: BPH-PIM (N = 129) Y LISAT-8 (N = 537)

	Categorías			Kappa de Cohen*	Tau-b de Kendall*
	Mejoría	Sin cambio	Empeoram.		
<b>BPH-PIM</b>					
DMI	73,1	19,4	7,4		
<b>EEM</b>					
EEM(1)	67,6	27,8	4,6	0,814	0,873
EEM(1,96)	54,6	42,6	2,8	0,551	0,672
EEM(2,58)	36,1	63,0	0,9	0,290	0,458
<b>LISAT-8</b>					
DMI	54,9	43,9	1,1		
<b>EEM</b>					
EEM(1)	66,5	32,0	1,5	0,759	0,788
EEM(1,96)	44,9	54,4	0,7	0,797	0,819
EEM(2,58)	26,3	73,4	0,4	0,448	0,544

\*:  $p < 0,0001$  en todos los casos.

te, 67,6, 27,8 y 4,6%. El grado de concordancia entre estos dos criterios de clasificación es alto: el índice de acuerdo kappa toma un valor de 0,814 y el coeficiente de correlación tau-b un valor de 0,873. Los porcentajes de sujetos clasificados en cada categoría con los criterios EEM(1,96) y EEM(2,58) son muy diferentes de los porcentajes obtenidos con el criterio DMI. Y los valores del índice de acuerdo kappa y del coeficiente de correlación tau-b, aunque estadísticamente significativos, son moderados o bajos.

Los resultados obtenidos con la escala LISAT-8 no son del todo idénticos. Aquí, el criterio DMI muestra un alto grado de concordancia tanto con el criterio EEM(1) como con el criterio EEM(1,96): el índice de acuerdo kappa entre estos criterios es de 0,759 y 0,797, respectivamente, y el coeficiente de correlación tau-b toma valores de 0,788 y 0,819.

## DISCUSIÓN

Los estimadores del tamaño del efecto o del cambio clínicamente relevante obtenidos mediante cuestionarios de salud tras una intervención médica constituyen un referente de indudable importancia en el manejo de los pacientes cualquiera que sea el problema de salud abordado. No sólo permiten valorar el estado de salud actual de los pacientes, sino monitorizar la efectividad de una intervención médica y realizar comparaciones con otros estados de salud o con otras intervenciones terapéuticas. El problema es que existen diferentes procedimientos para determinar el cambio clínicamente relevante y, a pesar de que todos ellos están intentando medir lo mismo, no es posible asegurar que todos ellos ofrezcan los mismos resultados (5-8,11-13,27).

La DMI parece haberse instaurado como el criterio más fiable y válidamente acorde con las percepciones del paciente. Sin embargo, adolece de algunos problemas. El primero de ellos es que se requiere del uso de escalas de referencia para establecer los valores ancla a partir de los cuales poder calcular el valor de la DMI. Estas escalas de referencia son cumplimentadas por el propio paciente y, por tanto, adolecen de subjetividad, interpretabilidad, etc. La utilización de criterios con propiedades estadísticas independientes de la subjetividad del paciente y que no necesiten de mediciones de referencia debería ser considerado como de enorme ayuda para los clínicos, a la vez que simplificaría y estandarizaría completamente la medición del cambio producido por las intervenciones médicas. La estimación de esta magnitud del cambio mediante el error estándar de medida parece apuntar en esa dirección (14-16,20).

La comparación de diferentes criterios para medir la magnitud del cambio en el estado de salud de los pacientes llevada a cabo en este trabajo muestra una elevada concordancia entre los criterios DMI y EEM en dos muestras de pacientes varones con patologías diferentes, y en dos aspectos distintos de la salud del individuo: la calidad de vida y la satisfacción con la vida. El criterio de EEM (1) ha mostrado mejor concordancia con la DMI en la escala que explora CVRS que los otros criterios analizados, EEM (1,96) y EEM (2,58), tal como se hipotetizaba a priori y en consonancia con los resultados previos de la literatura científica (14-16, 20). Sin embargo, en la escala de satisfacción con la vida, tanto el criterio EEM (1) como el criterio EEM (1,96) han mostrado valores similares de concordancia con el criterio DMI. Estos resultados no invalidan la hipótesis inicial de utilizar el EEM (1) como cri-

terio de la magnitud del cambio que miden los instrumentos de salud autopercibida, y viene a confirmar los resultados encontrados por Wyrwich y cols. (14-16,20). Los coeficientes de acuerdo encontrados han sido muy elevados en todos los casos (superiores o próximos a 0,8), lo cual indica que las clasificaciones efectuadas con ambos criterios son muy similares.

El criterio EEM tiene ciertas ventajas sobre la DMI. En primer lugar no depende de escalas adicionales de referencia para calcular los estados ancla necesarios para detectar el nivel de cambio clínico. Las propias respuestas de la escala permiten establecer los límites que determinan si la magnitud del cambio es o no relevante. Esto simplifica la valoración del paciente y descarga al clínico y al paciente de trabajo adicional. Las unidades en las que se expresa el criterio EEM son, por otra parte, las mismas que la escala original, lo que facilita su interpretabilidad. Otra ventaja del criterio EEM es la de ser independiente de la variabilidad de las puntuaciones observadas en la muestra de pacientes. De hecho, el EEM se considera más una propiedad del instrumento de medida que una característica de la muestra de sujetos.

Sin embargo, para calcular el EEM, se necesita estimar la fiabilidad de la escala en la muestra de pacientes en la que se desea valorar. Esta estimación puede hacerse con el coeficiente alpha de Cronbach (si se desea utilizar la consistencia interna como medida de la fiabilidad) o con el coeficiente de correlación intraclase (si lo que se requiere es la fiabilidad test-retest). No existe acuerdo hasta la fecha sobre cual de esas dos estimaciones de la fiabilidad se debería utilizar, pero, al menos desde un punto de vista práctico, la estimación de la consistencia interna ofrece la ventaja de que sólo requiere efectuar una medición (al menos con escalas cuya fiabilidad test-retest haya sido evaluada previamente) y, además, no le

afecta el riesgo de que se produzca un cambio (no imputable al azar) en el estado de salud de los pacientes entre el test y el retest. Además, el coeficiente alpha de Cronbach parece menos propenso que el coeficiente de correlación intraclase a sobrestimar o infraestimar la fiabilidad de la escala (al menos con tamaños muestrales grandes) (28-30).

Este trabajo presenta algunas *limitaciones* que deben ser tenidas en cuenta. En primer lugar cabe señalar que el procedimiento utilizado para calcular la DMI en la escala de satisfacción con la vida es algo diferente del seguido por Jaeschke y cols (8) y Juniper y cols (9), pues en los pacientes con DE la DMI no se ha calculado a partir de la escala de Juniper y cols, sino a partir del Índice de función eréctil del IIEF. No obstante, a pesar de que las escalas de referencias son distintas, la metodología utilizada es, en esencia, la misma. Por otro lado, en esta evaluación no se ha tenido en cuenta el efecto de regresión a la media para corregir los valores del EEM, como han hecho otros autores (13). La razón para no hacerlo es que la aplicación de esta corrección no parece mostrar resultados diferentes, como han señalado otros autores (31). Además, tampoco existe un consenso generalizado a este respecto y serían necesarias investigaciones adicionales que nos orientaran sobre cual es el mejor procedimiento a seguir (32).

En *conclusión*, y a pesar de las limitaciones apuntadas, la utilización del error estándar de medida (EEM) puede ser un criterio alternativo a la diferencia mínimamente importante (DMI) para evaluar la magnitud del cambio producido en los instrumentos de medida de salud comunicadas por los pacientes, con la ventaja de no necesitar de escalas o valoraciones de referencia externas al propio instrumento de medida, aunque, por supuesto, estas exploraciones siempre serán complementarias al juicio clínico.

## Bibliografía

1. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 2002; 11: 193-205.
2. Ware JE, Keller SD. Interpreting general health measures. In *Quality of life and pharmacoeconomics in clinical trials*. Second Edition, edited by B. Spilker. Lippincott Raven Publishers, Philadelphia, 1996.
3. Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press, 1989.
4. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *J Consult Clin Psychol* 1999; 3: 300-7.
5. Fairclough D. Design and analysis of quality of life studies in clinical trials. Chapman & Hall/CRC, 2002.
6. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989; 27: S178-S89.
7. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992; 30: 917-25.
8. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clin Trials* 1989; 10: 407-15.
9. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994; 47: 81-7.
10. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality of life research. *Pharmacoeconomics* 2000; 18: 419-23.
11. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991; 59: 12-9.
12. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003; 56: 395-7.
13. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999; 37: 469-78.
14. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol* 1999; 52: 861-73.
15. Wyrwich KW, Tierney WM, Wolinsky FD. Using the standard error of measurement to identify important changes on the Asthma Quality of Life Questionnaire. *Qual Life Res* 2002; 11: 1-7.
16. Cella D, Eton DT, Lai J, Peterman AH, Merkel DE. Combining anchor and distribution-based methods to derive minimal clinically important differences on the functional assessment of cancer therapy (FACT) Anemia and Fatigue Scales. *J Pain Symptom Manage* 2002; 24: 547-61.
17. Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol* 2004; 57: 898-910.
18. Shikar R, Harding G, Leahy M, Lennox RD. Minimal important difference (MID) of the Dermatology Life Quality Index (DLQI): Results from patients with chronic idiopathic urticaria. *Health Qual Life Outcomes* 2005; 3: 36.
19. Wyrwich KW. Minimal important difference thresholds and the standard error of measurement: Is there a connection? *J Biopharm Stat* 2004; 14: 97-110.

20. Rejas J, Pardo A, Ruiz M. Minimal important difference and standard error of measurement: Further evidences supporting the connection of these two measures of meaningful changes in patient reported outcomes. 2006 International Society for Quality of Life Research meeting abstracts [www.isoqol.org/2006mtgabstracts]. The QLR Journal, A-103, Abstrac#1631.
21. Rejas J, Gil de Miguel A, San Isidro C, Palacios G, Carrasco P. Adaptación cultural y validación psicométrica del Cuestionario de Satisfacción con la Vida de Fugl-Meyer -LISAT 8- en población española masculina con disfunción eréctil. *Pharmacoeconomics Sp Res Art* 2004; 1: 87-99.
22. Llorente C, Ruiz M, Rejas J, Esteban M, Villasante F, Hareendran A, et al. Linguistic adaptation and validation of the Benign Prostatic Hyperplasia-Patient Impact Measure (BPH-PIM) into Iberian Spanish. *Clin Drug Invest* 2006; 26: 103-112.
23. Gil A, Martínez E, Oyagüez I, Palacios G, Rejas J. Erectile dysfunction in primary care setting: results of an observational, no-control-group, prospective study with sildenafil under routine conditions of use. *Int J Impotence Res* 2001; 13: 338-47.
24. Rosen RC, Riley A, Wagner G, Osterloh TH, Kirkpatrick J, Mishra A. The International Index of Erectile Function (IIEF): A multidimensional scale for assessment of erectile dysfunction. *Urology* 1997; 49: 822-30.
25. Rejas J, Gil A, San Isidro C, Palacios G, Carrasco P. Sensibilidad al cambio y Diferencia Mínimamente Importante de la versión española del Cuestionario de Satisfacción con la Vida LISAT 8 en pacientes varones con Disfunción Eréctil. *Med Clin (Barc)* 2005; 124: 165-71.
26. Guyat GH, Osoba D, Wu AW, Wyrwich KW, Norman GR, the Clinical Significance Consensus Meeting Group. Methods to explain the clinical significance of health status measures. *Mayo Clinic Proc* 2002; 77: 371-83.
27. Sloan JA, Cella D, Frost M, Guyatt GH, Sprangers M, Symonds T, et al. Assessing clinical significance in measuring oncology patient quality of life: Introduction to the symposium, content overview, and definition of terms. *Mayo Clinic Proc* 2002; 77: 367-70.
28. McHorney CA, Tarlov AR. Individual patient monitoring in clinical practice: Are available health status surveys adequate? *Qual Life Res* 1995; 4: 293.
29. Nunnally JC, Bernstein IJ. *Teoría Psicométrica*. 3ª Edición. México. McGraw-Hill; 1995.
30. Anastasi A, Urbina S. *Psychological testing (7th edition)*. Upper Saddle River; NJ: Prentice-Hall, 1997.
31. Bauer S, Lambert MJ, Nielsen SL. Clinical significance methods: A comparison of statistical techniques. *J Per Assess* 2004; 82: 60-70.
32. Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui KK. Evaluating the statistical significance of health-related quality-of-life change in individual patients. *Eval Heal Proff* 2005; 28: 160-71.