

La versión española del *SF-36 Health Survey* (Cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos

J. Alonso, L. Prieto y J. M. Antó

Departamento de Epidemiología y Salud Pública. Instituto Municipal de Investigación Médica (IMIM). Universidad Autónoma de Barcelona.

cuestionarios de salud

FUNDAMENTO: El estudio, que ha sido realizado en el marco del proyecto IQOLA, en el que participan investigadores de 15 países, presenta los resultados iniciales del proceso de adaptación del Cuestionario de la Salud SF-36 para su uso en España.

MÉTODOS: La adaptación se basó en el método de traducción y retrotraducción. Se organizaron reuniones de traductores, investigadores y pacientes para producir versiones sucesivas. Se realizó un estudio en 47 individuos para valorar (a través de una escala analógica visual) cada una de las opciones de respuesta de las preguntas del cuestionario. Finalmente, se administró la versión española del SF-36 a 46 pacientes con enfermedad coronaria estable en dos ocasiones separadas por 2 semanas, para analizar su consistencia interna y su reproducibilidad.

RESULTADOS: La puntuación media de la equivalencia conceptual con la versión original fue alta e independiente de la dificultad de traducción. El orden de las puntuaciones promedio de cada opción de respuesta coincidió con el orden asignado en el cuestionario en todos los casos. El alfa de Cronbach fue superior a 0,7 en todas las dimensiones (rango de 0,71 a 0,94), excepto en la dimensión de relación social (alfa = 0,45). Los coeficientes de correlación intraciase entre las dos administraciones del cuestionario tuvieron un recorrido de 0,58 a 0,99.

CONCLUSIONES: La adaptación del Cuestionario de la Salud SF-36 ha culminado con la obtención de un instrumento aparentemente equivalente al original y con una fiabilidad aceptable. Sin embargo, es necesario evaluar otras características básicas, como son la validez y la sensibilidad a los cambios de salud.

The Spanish version of the SF-36 Health Survey: a measure of clinical outcomes

BACKGROUND: The present study, performed within the Internacional Quality of Life Assessment project (including researchers from 15 countries) presents preliminary results of the process of adaptation of the SF-36 to be used in Spain.

METHODS: The adaptation was based on the translation/back-translation methodology. Meetings of translators, researchers and patients were organized in order to produce successive versions. A study involving 47 individuals was carried out to assess the relative value (through a visual analogue scale) of each response choice of the questionnaire items. Finally, internal consistency and reproductibility of the Spanish version of the SF-36 was assessed by administering the questionnaire to 46 patients with stable coronary heart disease in two different occasions 2 weeks apart.

RESULTS: The average ratings of equivalence of the translated version with the original were high regardless of the difficulty of translation. The rank ordering of mean scores for each responses choice agreed with the ranking assigned in the questionnaire in all cases. Cronbach's Alpha was higher than 0.7 for all dimensions (range: 0.71-0.94) except for Social Functioning scale (alpha = 0.45). Intraclass correlation coefficients between both administrations of the questionnaire ranged from 0.58 to 0.99.

CONCLUSIONS: The adaptation process of the SF-36 has concluded with an instrument apparently equivalent to the original and with an acceptable level of reliability. Nevertheless, other basic characteristics of the adapted questionnaire (i.e. validity and sensitivity to changes) should be also assessed.

Med Clin (Barc) 1995; 104: 771-776

Correspondencia: Dr. J. Alonso.
Instituto Municipal de Investigación Médica (IMIM).
Doctor Aiguader, 80. 08003 Barcelona.

Manuscrito aceptado el 1-3-1995

El conocimiento de la efectividad y la eficiencia de los servicios sanitarios es una necesidad para poder configurar adecuadamente las prioridades sanitarias. De manera creciente, los estudios que evalúan la efectividad de los procedimientos médicos utilizan como variable de resultado la calidad de vida relacionada con la salud o el estado de salud. Resulta innovador que este tipo de medidas, que reflejan la salud desde la perspectiva del paciente y que recogen la información a través de entrevistas o cuestionarios (en vez de instrumentación u observaciones clínicas), haya atraído el interés de los investigadores. Este interés es, probablemente, reflejo de la importancia conceptual y el respaldo metodológico que tienen los instrumentos de medida del estado de salud¹.

Los instrumentos de medida del estado de salud pueden clasificarse en dos grandes tipos: genéricos y específicos. Mientras que los específicos están diseñados para su utilización en pacientes con un tipo de problema de salud concreto (p. ej., asma), los genéricos son independientes del diagnóstico y se han desarrollado para poder ser utilizados en diferentes tipos de pacientes o poblaciones. Entre las ventajas de los instrumentos genéricos destaca la posibilidad de comparar el impacto relativo de diferentes enfermedades sobre el estado de salud, así como la de obtener valores poblacionales de referencia. Algunos instrumentos genéricos, ya disponibles en nuestro medio, como el *Nottingham Health Profile* (NHP)²⁻⁴ y el *Sickness Impact Profile* (SIP)⁵⁻⁷, destacan por su bien establecida validez y por su cada vez más frecuente utilización.

Sin embargo, la mayoría de los *Patient Outcome Research Teams* (PORTs) estadounidenses, que constituyen una de las iniciativas de investigación más importantes para la evaluación de la efectividad de los servicios sanitarios⁸, utilizan otro instrumento genérico, el *SF-36 Health Survey* (SF-36)⁹⁻¹¹, que, aunque menos difundido que los anteriormente mencionados, está empezando a ser utilizado en todo el mundo.

El SF-36 fue desarrollado para su uso en el Estudio de los Resultados Médicos

(Medical Outcomes Study[MOS])^{9,12} a partir de una extensa batería de cuestionarios que incluía cuarenta conceptos relacionados con la salud. Para crear el nuevo cuestionario de una manera eficiente, se seleccionó el mínimo número de conceptos necesarios para mantener la validez y las características operativas del cuestionario inicial. El formato final es un instrumento genérico que contiene 36 ítems que cubren ocho dimensiones del estado de salud y proporcionan un perfil del mismo. Sus ítems detectan tanto estados positivos como negativos de la salud. Para cada dimensión, los ítems se codifican, agregan y transforman en una escala que tiene un recorrido desde 0 (el peor estado de salud para esa dimensión) hasta 100 (el mejor estado de salud)¹². El cuestionario no ha sido diseñado para generar un índice global.

Su contenido, que incluye tanto salud física como mental, su robustez psicométrica y su relativa simplicidad son factores que pueden facilitar su utilización en la investigación. Dichos factores, unidos a la existencia de un proyecto internacional de adaptación del cuestionario original en quince países (*International Quality of Life Assessment -IQOLA- Project*)^{13,14} están convirtiendo al *SF-36 Health Survey* en uno de los instrumentos genéricos de medida del estado de salud con mayor potencial de uso internacional en la evaluación de los resultados clínicos.

Este artículo presenta los resultados iniciales del proceso de adaptación del *SF36 Health Survey* para su uso en España.

Sujetos y métodos

Una descripción de las dimensiones del *SF-36 Health Survey* así como ejemplos de algunos de los ítems aparecen en las **tablas 1 y 2**, respectivamente.

Metodología de adaptación. La traducción al castellano del cuestionario siguió un protocolo común a todos los países que participan en el proyecto IQOLA^{13,14}. Se basó en el método de traducción y retrotraducción¹⁵ por profesionales y la realización de estudios piloto con pacientes. Este proceso se representa gráficamente en la **figura 1** y se describe brevemente a continuación.

El cuestionario original fue traducido al castellano por dos personas bilingües (de origen español y viviendo en los Estados Unidos, con experiencia clínica) que, además, puntuaron la dificultad para encontrar una expresión española conceptualmente equivalente a la original. Ambas traducciones fueron discutidas simultáneamente entre los traductores y uno de los investigadores (JA) hasta alcanzar un consenso. Otras dos personas bilingües evaluaron la equivalencia conceptual, la claridad y la naturalidad de cada una de las frases y de las opciones de respuesta de esta primera versión. Finalmente, en una nueva reunión entre el investigador principal y los traductores iniciales, se llegó a una primera versión adaptada (versión 1.0) (**fig. 1**).

Para evaluar la equivalencia de esta versión, se realizaron tres actividades diferentes. En primer lugar, la versión 1.0 fue independientemente traducida de nuevo al inglés (retrotraducción) por dos personas bilingües (lingüistas de origen norteamericano que vivían en España). Sus dos retrotraducciones resultantes fueron comparadas con la versión original por un equipo de expertos del centro estadounidense que dirige y coordina todo el proyecto IQOLA, quienes

Fig. 1 .Descripción del proceso de adaptación del proyecto IQOLA.

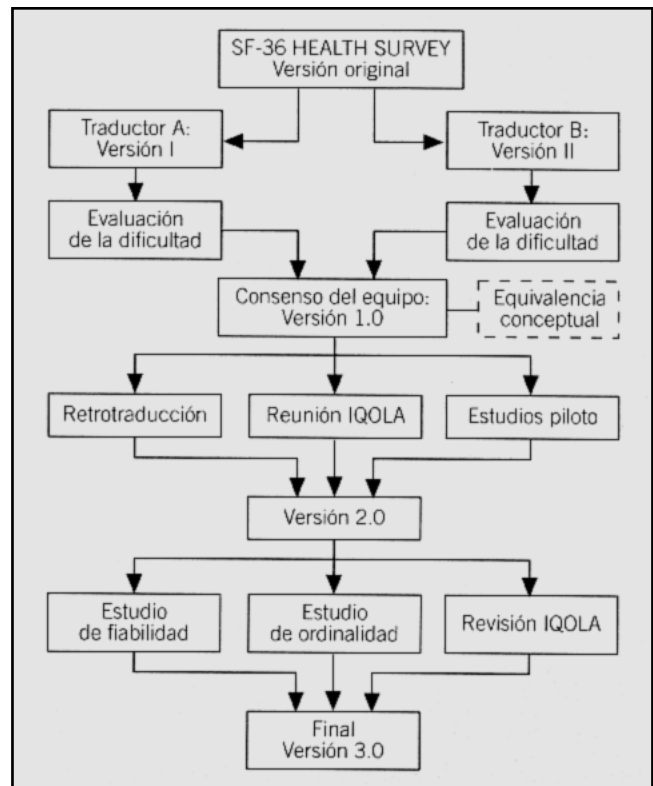


TABLA 1
Descripción del SF-36 Health Survey

Dimensión	N.º de ítems
Función física	10
Limitaciones del rol: problemas físicos	4
Dolor	2
Percepción de la salud general	5
Vitalidad	4
Función social	2
Limitaciones del rol: problemas emocionales	3
Salud mental	5
Cambio de la salud en el tiempo*	1

* Este ítem no está incluido en las ocho dimensiones.

TABLA 2
Ejemplos de ítems de la versión española del Cuestionario de Salud SF-36[®]

Ítem	3. Las siguientes preguntas se refieren a actividades o cosas que usted podría hacer en un día normal. Su salud actual, ¿le limita para hacer estas actividades o cosas? Si es así, ¿cuánto?			9. Las preguntas que siguen se refieren a cómo se ha sentido y cómo le han ido las cosas durante las 4 últimas semanas. En cada pregunta responda lo que se parezca más a cómo se ha sentido usted. Durante las 4 últimas semanas ¿cuánto tiempo...					
	Sí, me limita mucho	Sí, me limita un poco	No, no me limita nada	Siempre	Casi siempre	Muchas veces	Algunas veces	Sólo alguna vez	Nunca
c. Coger o llevar la bolsa de la compra	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
d. Subir varios pisos por la escalera	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
f. Agacharse o arrodillarse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
a. ¿Se sintió lleno de vitalidad?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
b. ¿Estuvo muy nervioso?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
c. ¿Se sintió tan bajo de moral que nada podía animarle?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

[®] Correspondientes a la versión 3.0.

[®] 1985 Medical Outcomes Trust. Todos los derechos reservados. El cuestionario está a la disposición de los interesados, tras petición.

señalaron algunos ítems o palabras que no parecían tener una completa equivalencia conceptual con el original.

En segundo lugar, se realizó una reunión con los autores de todas las versiones ya existentes en diversos países (Suecia, Reino Unido, Italia, Francia, Canadá, Australia, Alemania y Holanda) durante la cual se trató de armonizar el contenido del cuestionario en los casos en los que existían diferentes expresiones alternativas del mismo concepto (p. ej., sustituir la distancia de una milla por la de un kilómetro, etc.).

En tercer lugar, se realizaron estudios piloto con diversos grupos de pacientes crónicos para valorar la comprensibilidad y factibilidad de administración del cuestionario. A partir de la versión castellana se obtuvo, mediante la utilización del mismo método con cuatro personas bilingües, la versión catalana del cuestionario. Finalmente, antes de producir la versión definitiva del

cuestionario se llevaron a cabo dos estudios empíricos. El primero consistió en la calibración de las opciones de respuesta del cuestionario, con el fin de comparar su ordinalidad o gradación con la de la versión original americana. El segundo fue la evaluación de la fiabilidad del cuestionario.

Comprobación de la ordinalidad de las opciones de respuesta. Las opciones de respuesta para cada uno de los ítems del SF-36 Health Survey están organizadas en categorías, descripciones adjetivales¹⁶ (para cada ítem, de tres a siete opciones de respuesta dentro de cada categoría -ver la columna 2 de la tabla 4-) que permiten expresar la respuesta a cada ítem particular dentro de una escala de respuestas discretas.

Existen 6 categorías o grupos diferentes de opciones de respuesta a los ítems del SF-36 Health Survey. El estudio pretendió establecer el orden de las diferentes opciones de respuesta posibles dentro de cada

categoría y compararlo con el del cuestionario original. El recorrido de cada categoría está limitado por las dos opciones que señalan los extremos de la misma (en negrita en la tabla 4). La tarea consistió en solicitar a los participantes en el estudio que situaran sobre una línea de 10 cm (una escala analógica visual) la posición relativa de cada una de las opciones de respuesta respecto a los extremos de su propia categoría. Por ejemplo, la categoría correspondiente a uno de los ítems sobre la Salud General (cuarta en la tabla 4) consta de las siguientes opciones de respuesta: *Excelente, Muy buena, Buena, Regular y Mala*. Para estimar la posición relativa de cada una de ellas, a cada participante se le mostraba una línea horizontal de 10 cm con las palabras «mala» en el extremo izquierdo (0 cm) y «excelente» en el extremo derecho de la misma (10 cm). Para cada una de las opciones de respuesta restantes, la tarea del encuestado consistió en marcar, con una cruz sobre la línea, el lugar donde creía que se situaría. Así por

ejemplo, cabría esperar que el encuestado situase la opción «muy buena» cerca del extremo derecho de la línea, más cerca de la expresión «excelente» que de la expresión «mala». Ello significaría que la opción de respuesta «muy buena» tiene un valor alto, comparado con los dos extremos de su categoría.

Para llevar a cabo esta tarea, se facilitaron instrucciones estandarizadas y se presentó en todos los casos un ejemplo previo que describía la tarea a realizar. Las opciones de respuesta se agruparon según su categoría y se incluyeron en orden aleatorio dentro de cada una de ellas. Cada opción de respuesta a calibrar aparecía por separado en una hoja. A su vez, se recogió información sobre el estado de salud, la edad, el sexo y el nivel de estudios completados de los sujetos participantes.

En aquellos casos en los que se había obtenido más de una versión o traducción para una opción de respuesta, se evaluaron todas, con el fin de identificar la que tenía el valor escalar más parecido a la del cuestionario original. De esta manera se compararon dos traducciones diferentes de las opciones de respuesta «a little of the time» («casi nunca» y «sólo alguna vez») y «moderately» («regular» y «moderadamente»). Asimismo, todas las opciones de respuesta de la categoría «nada» a mucho, se incluyeron dos veces en el cuestionario para medir su reproducibilidad test-retest.

Para el estudio se seleccionaron, mediante un muestreo por cuotas de sexo y grupos de edad (20-49 y más de 50 años), 50 personas (pacientes y acompañantes) procedentes de las consultas externas del Hospital Universitari del Mar de Barcelona. El cuestionario autoadministrado que contenía las escalas analógicas visuales fue completado por 47 individuos.

Evaluación de la reproducibilidad del cuestionario. Se administró el cuestionario a 50 pacientes con enfermedad coronaria estable que se habían reclutado de manera consecutiva durante una visita de control a las consultas externas de cardiología de un hospital

TABLA 3

Dificultad y equivalencia medias de la traducción de los ítems de las dimensiones de la versión española del SF-36 Health Survey

Dimensión	Dificultad ^a (rango)	Equivalencia ^b (rango)
Función física	24,2 (5-45)	96,4 (85-100)
Limitaciones del rol: problemas físicos	21,3 (18-27)	97,0 (90-100)
Dolor	37,5 (32-43)	90,0 (80-100)
Percepción de la salud general	17,8 (7-27)	95,4 (85-100)
Vitalidad	21,3 (10-28)	99,5 (98-100)
Función social	22,5 (17-28)	97,5 (95-100)
Limitaciones del rol: problemas emocionales	20,0 (15-27)	97,5 (95-100)
Salud mental	18,9 (12-27)	97,5 (90-100)
Cambio de la salud en el tiempo	18,3	90,0

^a Dificultad para traducir (0 = ninguna dificultad · 100 = mucha dificultad).

^b Equivalencia conceptual de la traducción (0 = mala · 100 = muy mala)

TABLA 4

Resultado del estudio de ordinalidad de las opciones de respuesta de los ítems del SF-36 y comparación con la puntuación en el cuestionario y con un estudio similar realizado en los Estados Unidos

Escala SF-36	Opciones de respuesta en inglés	Opciones de respuesta en castellano	Puntuación en el cuestionario	Muestra española revisada (n = 47) Hospital del Mar			Muestra EE.UU. revisada (n = 117)		
				Media	Mediana	DE	Media	Mediana	DE
Salud General	Definitely false	Totalmente falsa	1						
	Mostly false	Bastante falsa	2	1,83	1,68	0,73	1,92	1,76	0,62
	Don't know	No lo sé	3	2,26	2,32	0,70	2,70	2,70	0,57
	Mostly true	Bastante cierta	4	3,85	4,00	0,64	0,64	3,82	0,58
	Definitely true	Totalmente cierta	5						
Vitalidad	None of the time	Nunca	1						
	A little of the time	Casi nunca/sólo alguna vez	2	1,79/2,64	1,55/2,40	0,65/1,06	2,50	2,30	0,79
	Some of the time	Algunas veces	3	3,29	3,50	0,92	2,98	3,00	0,77
	A good bit of the time	Muchas veces	4	4,52	4,75	0,67	4,25	4,40	0,78
	Most of the time	Casi siempre	5	5,03	5,25	0,68	4,48	4,60	0,77
	All of the time	Siempre	6						
Función social	Not at all	Nada	1						
	Slightly (Social)/ A little bit (Pain)	Un poco	2	2,06	1,88	0,75	1,96/2,06	1,90/1,96	0,48/0,61
	Moderately	Regular/Moderadamente	3	2,60/2,79	2,72/2,84	0,44/0,66	2,84	2,92	0,53
	Quite a bit	Bastante	4	4,02	4,04	0,43	3,58	3,72	0,63
	Extremely	Mucho	5						
Salud General	Poor	Mala	1						
	Fair	Regular	2	2,64	2,56	0,65	2,54	2,56	0,56
	Good	Buena	3,4	3,72	3,80	0,57	3,29	3,12	0,65
	Very good	Muy buena	4,4	4,12	4,20	0,48	3,87	3,98	0,63
	Excellent	Excelente	5						
Función física	No, not limited at all	Ninguna dificultad^a	1						
	Yes, limited a little	Poca dificultad ^b	2	1,56	1,46	0,36	1,95	1,98	0,38
	Yes, limited a lot	Mucha dificultad^c	3						
Dolor	None	Ninguno	6						
	Very mild	Muy poco	5,4	5,22	5,30	0,63	4,67	4,85	0,78
	Mild	Un poco	4,2	4,27	4,65	1,02	4,44	4,55	0,72
	Moderate	Bastante ^d	3,1	2,38	2,20	0,81	3,69	3,58	0,65
	Severe	Mucho	2,2	2,02	1,75	0,94	2,27	2,10	0,73
	Very severe	Muchísimo	1						

Las opciones de respuesta subrayadas no se utilizan en el instrumento final, pero fueron incluidas en el estudio. Formatos finales: ^a «No, no me limita nada»; ^b «Sí, me limita un poco»; ^c «Sí, me limita mucho»; ^d «Moderado»

TABLA 5
Puntuaciones de la primera administración de la versión española del Cuestionario de Salud SF-36 (n = 46)

	Función física	Rol físico	Dolor	Salud general	Vitalidad	Función social	Rol emocional	Salud mental
Media	69,9	80,1	74,5	56,1	66,8	90,5	89,6	77,8
Mediana	75	100	80	51	70	100	100	80
Desviación estándar	20,0	35,2	25,6	22,2	23,7	16,3	29,1	16
Tanto por ciento con puntuación mínima (100) (ceiling)	4,4	70,5	33,3	4,3	10,9	64,3	86,7	8,7
Tanto por ciento con puntuación mínima (0) (floor)	0	11,4	2,2	0	0	0	8,9	0

terciario (Hospital General Universitari de la Vall d'Hebron, Barcelona). El cuestionario fue administrado por una enfermera adiestrada, al acabar la visita y tras ser comprobada la estabilidad clínica del paciente por parte del cardiólogo.

Transcurridas dos semanas, los pacientes completaron de nuevo personalmente el cuestionario y, al acabar, respondieron a un cuestionario que incluía preguntas sobre cualquier cambio (mejora o empeoramiento) de su salud en general y de su enfermedad cardíaca durante las 2 semanas anteriores. En total, se obtuvo información completa de 46 pacientes (92%).

Para la evaluación de la consistencia interna, se calculó el coeficiente alfa de Cronbach¹⁷ para la primera administración del cuestionario. Este coeficiente se utiliza habitualmente para evaluar el grado en que los ítems de una misma escala están evaluando un concepto común a todos ellos. El cálculo del coeficiente alfa se basa en la correlación media de cada ítem de la escala con el total de la misma, y en el número de ítems que contiene. El valor del coeficiente tiene un recorrido entre 0 y 1 y se recomiendan los valores superiores a 0,5¹⁸ o 0,7¹⁹ como criterio mínimo para asegurar una apropiada consistencia interna de la escala cuando se utiliza para la comparación de grupos¹⁹.

La reproducibilidad del cuestionario se estableció mediante la comparación de las puntuaciones medias obtenidas en las dos administraciones del cuestionario (test de t de Student) y el cálculo de los coeficientes de correlación intraclase²⁰. Sólo tres pacientes manifestaron haber sufrido algún pequeño cambio en su estado de salud transcurridas las dos semanas desde la primera administración del cuestionario; a pesar de ello, se incluyeron en los análisis que se presentan, ya que éstos son idénticos a los que se obtuvieron tras su exclusión.

Resultados

Algunas etapas del proceso de traducción del proyecto IQOLA fueron de naturaleza fundamentalmente cualitativa, mientras que otras son susceptibles de análisis cuantitativo. A pesar de la influencia sobre el formato final del cuestionario de las etapas cualitativas, la descripción de los resultados se centra en los aspectos cuantitativos.

La variabilidad de las puntuaciones asignadas a la dificultad de traducción de ítems fue mayor que las asignadas a su equivalencia con la versión original: mientras que la equivalencia entre un ítem traducido y el original no se situó por debajo del 90 por ciento en ningún caso, las puntuaciones de dificultad variaron entre el 5 y el 45% (tabla 3). En promedio, las dimensiones que cubren el concepto de salud en general presentaron una dificultad de traducción menor que el resto de dimensiones mientras que la dimensión del dolor resultó significativamente más difícil de

TABLA 6
Consistencia interna y reproducibilidad test-retest del Cuestionario de Salud SF-36 (n = 46)

Dimensión	Consistencia interna del test (alfa de Cronbach)	Reproducibilidad test-retest	
		Diferencia media	Coefficiente de correlación intraclase
Función física	0,84	-3,7 ^a	0,83
Rol físico	0,90	0	0,99
Dolor	0,88	-4	0,65
Salud general	0,71	1,4	0,85
Vitalidad	0,74	-3,1	0,70
Función social	0,45	-2,2	0,65
Rol emocional	0,94	2,2	0,58
Salud mental	0,77	-3	0,76

traducir que las demás.

Ordinalidad de las opciones de respuesta. El orden de las puntuaciones promedio de cada opción de respuesta o descriptor coincidió, en todos los casos, con el orden asignado en el cuestionario (columna cuarta de la tabla 4). Los valores fueron, en general, muy similares a los valores obtenidos en un estudio similar realizado en los Estados Unidos (observaciones no publicadas, columnas 8-10 de la tabla 4), aunque algunos de ellos presentaron diferencias estadísticamente significativas. Estos resultados fueron invariables independientemente de la medida de tendencia central utilizada (p. ej., medias o medianas).

En el caso de las tres categorías de opciones de respuesta para las que existía alguna versión alternativa, este ejercicio sirvió para incluir en la versión final aquella cuya puntuación fuera más similar a la original americana. Así, se incluyó «moderado» en vez de «bastante» y la versión «sólo alguna vez» sustituyó a la de «casi nunca» (tabla 4). Sin embargo, como las diferencias entre «regular» y «moderadamente» (versión americana «moderately») no resultaron significativas, se mantuvo la opción consensuada originalmente («regular») (tabla 4).

Las opciones de respuesta de la pregunta 6 (cuyos extremos son «nada» y «mucho») se evaluaron en dos ocasiones. Los valores de ambas puntuaciones fueron muy similares en todos los casos («un poco»: 4,0 y 3,9; «regular»: 3,5 y 3,4; «moderadamente»: 3,2 y 3,2; «bastante»: 2,2 y 2,0).

Consistencia interna y fiabilidad del cuestionario. Los pacientes participantes en este estudio tenían una edad promedio de 61,2 ± 7,6 años, y el 89% eran varones. La respuesta al cuestionario exigió aproximadamente unos 10-15 minutos. El 75% lo rellenó personalmente.

Las puntuaciones medias de las dimensiones del cuestionario en 46 pacientes corronarlos se especifican en la tabla 5. Las puntuaciones más altas (mejor estado de salud) correspondieron a la dimensión de la función social (SF) (90,5) y la emocional (RE) (89,6). La salud general (GH) fue la más baja, con 56,1 puntos seguida de la vitalidad (VT) (66,8) y la función física (PF) (69,9).

También aparece en la tabla 5 el tanto por ciento de pacientes que obtuvo la puntuación máxima de cada dimensión (mejor estado de salud medible o efecto «ceiling») y el tanto por ciento de pacientes que obtuvo la peor puntuación (o efecto «floor»). En la tabla 6 se muestran los coeficientes alfa para cada escala del cuestionario. Los valores fluctuaron entre 0,45 y 0,94, con una media de 0,78 y se situaron siempre por encima del valor 0,7 (valor mínimo recomendado), excepto para la escala de función social (SF), que no pasó del valor 0,45.

Las puntuaciones medias de la administración inicial y la realizada al cabo de 2 semanas fueron muy similares; se encontró una diferencia marginalmente significativa para la dimensión de función física (FP) (-3,7, p = 0,041). Los coeficientes de correlación entre ambas administraciones fluctuaron entre 0,51 para la dimensión de rol físico (RF) y el 0,85 de la salud ge-

neral (GH), mientras que los coeficientes de correlación intraciase, que combinan información del test de la t y de la correlación²⁰, fluctuaron entre 0,58 y 0,99.

Discusión

El presente estudio ha permitido obtener la versión española definitiva del *SF-36 Health Survey*, el Cuestionario de Salud SF-36. Los resultados de la calibración de las opciones de respuesta muestran una coincidencia completa con la ordinalidad esperada y una alta equivalencia con los valores americanos. Además, el cuestionario ha mostrado una alta consistencia interna para la mayoría de las dimensiones que lo componen y una reproducibilidad aceptable, similar aunque algo menor a las descritas para el instrumento original^{11,12,21-23}. Todo ello sugiere que el proceso de adaptación ha culminado con la obtención de un instrumento aparentemente equivalente al original que puede utilizarse en la investigación clínica y epidemiológica. Finalmente, el hecho de que el método utilizado fuera estrictamente comparable al seguido por investigadores de otros catorce países asegura la adecuación de la versión española del cuestionario para la realización de estudios internacionales. Sin embargo, para poder establecer su completa equivalencia y para poder recomendar definitivamente su uso en España, será necesario establecer la validez (discriminante y convergente) y la sensibilidad a los cambios de salud de esta versión y compararlas con las del cuestionario original, así como reproducir los estudios de fiabilidad a una mayor escala. El proceso de adaptación se ha basado en el método de traducción-retrotraducción¹⁵, que es el más comúnmente aceptado para asegurar la equivalencia conceptual al adaptar un cuestionario desarrollado en una cultura diferente²⁴⁻²⁷. La equivalencia conceptual es un requisito importante para poder interpretar adecuadamente las evaluaciones realizadas con un instrumento de medida del estado de salud. Además, resulta indispensable para poder replicar estudios y, sobre todo, para realizar estudios internacionales^{24,27}. El presente diseño ha incluido una mayor formalización y estandarización del método propuesto por Brislin y utilizado en estudios anteriores similares, ya que se ha cuantificado la equivalencia y la dificultad para identificar expresiones equivalentes. El alto grado de equivalencia encontrado, incluso para ítems de moderada a alta dificultad de traducción, sugiere que los diversos mecanismos de consenso han conseguido una versión suficientemente equivalente. Sin embargo, las consecuencias prácticas de las puntuaciones de equivalencia han sido escasas, ya que no se utilizaron en ningún caso para la modificación de los ítems consensuados. Esta escasa

operatividad contrasta con la del estudio de calibración de las opciones de respuesta.

El estudio de ordinalidad ha puesto en evidencia la completa concordancia en el orden de las opciones de respuesta del Cuestionario de Salud SF-36 y el original. En los casos de opciones alternativas, el estudio ha servido para escoger la opción más próxima al valor teórico o al valor obtenido en la muestra americana. Este método se utiliza poco en la adaptación de cuestionarios; sin embargo, nuestros resultados sugieren que es una tarea factible y relevante. Además, este estudio ha sido replicado en cada uno de los países que participan en el proyecto IQOLA, y en ellos se han obtenido unos resultados muy semejantes tanto en el orden como en los valores escalares²⁸, lo que confirma todavía más su interés.

Sin embargo, el estudio de calibración no pretendía obtener valores ponderados que sustituyeran a los originales para el análisis de la versión española del Cuestionario de Salud SF-36, ya que, excepto para una pregunta sobre la salud en general y una sobre el dolor, el cuestionario se puntúa como una escala adjetival: es decir, cada opción de respuesta tiene una puntuación superior en una unidad a la anterior. Una comparación de las puntuaciones o ponderaciones de cada opción de respuesta del SF-36 sería interesante desde el punto de vista de la investigación transcultural. Sin embargo, requeriría un diseño específico y la existencia de valores de referencia de la escala original.

La consistencia interna de todas las dimensiones del cuestionario, excepto la de función social (SF), es satisfactoria (coeficiente alfa superior a 0,7). Ello sugiere que los ítems de cada una de ellas están midiendo un constructo unitario y que poseen poco error aleatorio (alto grado de fiabilidad) para la comparación de grupos. Asimismo, la mayoría de las escalas muestran también una alta reproducibilidad cuando se administra el cuestionario en dos ocasiones separadas por dos semanas. La baja consistencia interna de la dimensión de la función social, así como la correlación moderada de sus puntuaciones en ambas administraciones, se ha descrito en algunos de los estudios con el cuestionario original¹², aunque los valores reportados se han situado siempre por encima (mayor fiabilidad) que el presentado en este estudio. Esto sugiere, en general, que los ítems de esta dimensión están poco relacionados entre sí y que, probablemente, informan de diversos aspectos independientes de la salud social. Este problema de la dimensión de relación social se ha descrito previamente en otros instrumentos genéricos de salud adaptados en España²⁹. A su vez, el escaso número de ítems de esta dimensión (dos) puede disminuir la fiabilidad, ya que al estandarizar el cálculo del coeficiente alfa

(datos no presentados), éste se sitúa en 0,80. De cualquier forma, en estudios en los que los aspectos sociales deben medirse de manera detallada, es probable que se deba complementar este cuestionario con otras preguntas o escalas específicas de este concepto. Sin duda, este hallazgo merece un estudio específico en poblaciones donde la prevalencia de los problemas de función social sea mayor (p. ej., ancianos).

Como conclusión, cabe destacar la necesidad de tener en cuenta la influencia cultural que existe en cualquier cuestionario de medida del estado de salud y, por tanto, la importancia de utilizar métodos que aseguren la equivalencia conceptual en el proceso de adaptación. El método de la traducción-retrotraducción utilizado en este proyecto ha dado lugar a una versión española del Cuestionario de Salud SF-36 que resulta conceptualmente equivalente a la original y cuya administración es factible y comprensible. La demostración de estas características debe preceder a cualquier estudio posterior sobre las características psicométricas del instrumento.

Respecto a la versión española del Cuestionario de Salud SF-36 (ver pie de página), el estudio muestra que su consistencia interna y su reproducibilidad han resultado aceptables y similares a la del instrumento original. De todas formas, la total equivalencia de la versión española con la original no podrá ser completamente establecida hasta que sean evaluadas otras características psicométricas básicas, como son la validez y la sensibilidad a los cambios de salud y un estudio más detallado de la fiabilidad y de los patrones de respuesta (p. ej., teoría de las respuestas a los ítems).

Agradecimientos

El presente trabajo ha sido realizado bajo contrato con el *New England Medical Center Hospitals*(NEMCH), en el marco del proyecto International Quality of Life Assessment (IQOLA). El proyecto IQOLA está financiado por Glaxo Research Institute, Research Triangle Park, North Carolina (EE.UU.) y Schering-Plough Corporation, Kenilworth, New Jersey (EE.UU.).

Al igual que la versión original americana, la versión española del Cuestionario de Salud SF-36 se distribuye de manera gratuita Gunto con un manual de instrucciones de administración, y de puntuación e interpretación) a través del *Medical Outcomes Trust*, una fundación sin ánimo de lucro cuyo objetivo es la evaluación y difusión de instrumentos de medida de los resultados clínicos. Su dirección es: 20 Park Plaza, Suite 1041, Boston, MA 02116-4313, USA. Teléfono: (617) 426 40 46, Fax: (617) 426-4131. Los interesados pueden también dirigirse a los autores para mayor información.

Financiación adicional proviene de: Procter & Gamble Pharmaceuticals, Solvay Duphar B.V., Laboratoires Synthelabo, France, Parke-Davis, The Upjohn Company, Bayer S.p.A. y once compañías más.

Los autores agradecen a los Drs. Gaietà Permanyer y Jaume Figueras y a las enfermeras Puri Cascant y Rosa M. Laínez su colaboración en el estudio de reproducibilidad.

REFERENCIAS BIBLIOGRÁFICAS

- Bowling A. La medida del estado de salud. Una revisión de las medidas de la calidad de vida. Barcelona: SG, 1994.
- Hunt SM, McKenna SP, McEwen J, Backett E, Williams J, Papp E. A quantitative approach to perceived health status: a validation study. *J Epidemiol Commun Health* 1980; 34: 281-286.
- Alonso J, Antó JM, Moreno C. The Spanish Version of the Nottingham Health Profile: translation and preliminary validity. *Am J Public Health* 1990; 80: 704-708.
- Alonso J, Prieto L, Antó J M. The Spanish Version of the Nottingham Health Profile: a review of adaptation and instrument characteristics. *Qual Life Res* 1994; 3: 385-393.
- Bergner M, Bobbit RA, Pollard WE, Martin DP, Gilson BS. The Sickness impact Profile: validation of a health status measure. *Med Care* 1976; 14: 57-67.
- Badía X, Alonso J. Adaptación de una medida de la disfunción relacionada con la enfermedad: la versión española del Sickness Impact Profile. *Med Clin (Barc)* 1994; 102: 90-95.
- Badía X, Alonso J. Re-scaling the Spanish version of the Sickness impact Profile: an opportunity for assessment of cross-cultural equivalence. *J Clin Epidemiol*. En prensa.
- DeFriese GH. Measuring the effectiveness of medical interventions. *Health Serv Res* 1990; 25: 691-695.
- Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36) I. *Med Care* 1992; 30: 473-483.
- McHorney CA, Ware JE, Raczek AE. The MOS 36-item short form health survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993; 31: 247-263.
- McHorney CA, Ware JE, Lu R, Sherbourne CD. The MOS 36-item short form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. *Med Care* 1994; 32: 40-66.
- Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey. Manual and Interpretation Guide. Boston MA: The Health Institute, New England Medical Center, 1993.
- Aaronson NK, Acquadro J, Alonso J, Apolone G, Bucquet D, Bullinger M et al. International quality of life assessment (IQOLA) project. *Qual Life Res* 1992; 1: 349-351.
- Ware JE, Gandek B, and the IQOLA Project Group. The SF-36 Health Survey: development and use in mental health research and the IQOLA project. *Int J Ment Health* 1994; 23: 49-73.
- Brislin RW. The wording and translation of research instruments. En: Lonner WJ, Berry W, editores. *Field methods in cross-cultural research*. Beverly Hills, CA: Sage publications, 1986; 137-164.
- Streiner DL, Norman GR. *Health measurement scales. A practical guide to their development and use*. Oxford: Oxford University Press, 1989; 25-26.
- Cronbach LJ. Coefficient alpha and the internal structure of a test. *Psychometrika* 1951; 16: 297-334.
- Helmstader GC. *Principles of psychobiological measurement*. Nueva York: Appleton-Century-Crofts, 1964.
- Nunnally JC. *Psychometric theory*, 2.^a ed. Nueva York: McGraw-Hill, 1978.
- Deyo RA, Diehr PD, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials* 1991; 12: 142s-158s.
- Nerenz DR, Repasky DP, Whithouse FW, Kahionen DM. Ongoing assessment of health status in patients with diabetes mellitus. *Med Care* 1992; 30 Supl: MS112-MS124.
- Brazier JE, Harper R, Jones NMB, O' Cathain A, Thomas KJ, Unsherwood T et al. Validating the SF-36 Health Survey Questionnaire: new outcome measure for primary care. *BMJ* 1992; 305: 160-164.
- Jenkinson C, Coulter A, Writh L. The Short Form 36 (SF-36) Health Survey Questionnaire: normative data for adults of working age. *BMJ* 1993; 1.437-1.440.
- Patrick DL, Sittampalm Y, Somerville SM, Carter WB, Bergner M. A cross-cultural comparison of health status values. *Am J Public Health* 1985; 75: 1.402-1.407.
- Guillemin F, Bombardier C, Beaton D. Crosscultural adaptation of health-related quality of life instruments: literature review and proposed guidelines. *J Clin Epidemiol* 1994; 46: 1.417-1.432.
- Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res* 1993; 2: 451-460.
- Hunt SM, Alonso J, Bucquet D, Niero M, Wiklund I, McKenna S. Cross-cultural adaptation of health measures. *Health Policy* 1991; 19: 33-34.
- Ware JE, Keller SD, Gandek B, Brazier JE, Sullivan M and the IQOLA Project Group. Evaluating translations of health questionnaires: methods from the IQOLA Project. *Int J Technol Assessment*. En prensa.
- Badía X, Alonso J, Brosa M, Lock P. Reliability of the Spanish version of the Nottingham Health Profile in patients with stable end-stage renal disease. *Soc Sci Med* 1994; 1: 153-158.