

# Confiabilidad del cuestionario de calidad de vida en salud SF-36 en Medellín, Colombia\*

Luz Helena Lugo A.<sup>1</sup> • Héctor Iván García G.<sup>2</sup> • Carlos Gómez R.<sup>3</sup>

## Resumen

**Objetivo:** evaluar la consistencia interna y la fiabilidad test-retest e interevaluador del cuestionario de calidad de vida en salud SF-36 en pacientes con dolor músculo-esquelético, depresión mayor, diabetes mellitus y personas sanas. **Métodos:** se aplicó por entrevista y autodiligenciado en 605 personas. Se usaron medidas descriptivas para ítems, escalas y grupos; cálculo de consistencia interna de ítems con sus escalas, fiabilidad de escalas y fiabilidad interobservador e intraobservador. **Resultados:** los datos omitidos fueron de 0% a 1,5%, las correlaciones ítem-escala superaron 0,48, la fiabilidad fue

mayor de 0,70. No hubo diferencias en los puntajes de escalas según tipo de aplicación. La fiabilidad inter-observador fue mayor de 0,80; el test-retest superó el 0,70 y no mostró diferencias significativas en las dos aplicaciones. **Discusión:** el SF-36 es confiable para evaluar calidad de vida en salud después de adaptarse lingüísticamente en adultos colombianos.

## Palabras clave

Calidad de vida, calidad de vida relacionada con la salud, escalas de medición, SF-36, fiabilidad

---

## Reliability of SF-36 quality of life in health questionnaire in Medellín, Colombia

### Summary

**Objective:** to evaluate internal correspondence of questionnaire SF-36, as well as its test-retest and inter-evaluator reliability in patients with muscular-skeletal pain, depression, type II diabetes mellitus and healthy persons. **Methods:** it was completed by interview and individual filling out of the forms in 605 people. Descriptive measures were used for items, scales and groups; internal correspondence between items and their scales, and inter and intra-evaluator reliabilities were calculated. **Results:** omitted data ranged from 0% to 1,5%, item-scale correlations were over 0,48, reliability was higher than

0,70. There were no differences in scales scores according to type of application. Inter-evaluator reliability was higher than 0,80; test-retest was over 0,70, showing no significant differences in the two applications. **Discussion:** SF-36 questionnaire is reliable to evaluate health life quality in Colombian adults after its linguistic adaptation.

### Key words

Life quality, health-related life quality, measure scales, SF-36, reliability

---

\* Este artículo es resultado de la investigación “Adaptación cultural y fiabilidad del Cuestionario de calidad de vida en salud SF-36 en instituciones de salud de Medellín, Colombia”, realizada para optar al título de magister en epidemiología

1 Médica, especialista en medicina física y rehabilitación, magister en epidemiología, profesora de la Facultad de Medicina de la Universidad de Antioquia, integrante del Grupo Académico de Epidemiología Clínica (GRAEPIC)

2 Médico, magister en salud pública, magister en epidemiología, profesor de la Facultad de Medicina de la Universidad de Antioquia, integrante del Grupo Académico de Epidemiología Clínica (GRAEPIC). Cibercorreo: higarca@quimbaya.udea.edu.co

3 Médico, especialista en psiquiatría, magister en epidemiología clínica, profesor de la Facultad de Medicina de la Pontificia Universidad Javeriana, Bogotá, Unidad de Epidemiología Clínica y Bioestadística

## Introducción

En las últimas cuatro décadas ha habido un considerable trabajo teórico sobre la calidad de vida (CV), con significaciones de mayor o menor amplitud según sean las teorías que lo sustentan y el espectro de áreas de la vida de una persona incluidas en su valoración. En un sentido filosófico, CV es lo que hace que una vida sea mejor, que sea “una buena vida”.<sup>1,2</sup> En otras perspectivas, es el bienestar mental y espiritual o la cantidad y calidad de las relaciones interpersonales o el funcionamiento e integridad del cuerpo; en su sentido más restrictivo significa poseer bienes materiales.<sup>3</sup>

La OMS la define en función de la manera en que el individuo percibe el lugar que ocupa en la vida, en el contexto de la cultura y del sistema de valores en que vive, y en relación con sus objetivos, expectativas, normas y preocupaciones.<sup>4</sup> La calidad de vida relacionada con la salud (CVS) es el componente de la CV debido a las condiciones de salud de las personas y referido a partir de las experiencias subjetivas de ellas sobre su salud global.<sup>5</sup> Es un concepto multidimensional que incluye componentes físicos, emocionales y sociales asociados con la enfermedad. Los aspectos de la CV que deben evaluarse son controvertidos, ya que para algunos deben ser solo las condiciones físicas, para otros solo la percepción de los pacientes y para otros, ambas.<sup>6</sup>

La CVS toma cada día más relevancia en dos aspectos: ético, para tomar decisiones de los tratamientos médicos, tanto en el caso del consentimiento informado como en el de intervenciones para mantener la vida, y en el campo de los servicios de salud, para establecer prioridades y diseñar, implementar, seguir y evaluar programas de prevención, atención y rehabilitación, usada como complemento de los indicadores tradicionales de morbilidad, mortalidad o expectativa de vida.<sup>7</sup>

Los instrumentos existentes para medir la CVS se clasifican genéricos y específicos. Los primeros incluyen los puntajes globales, los perfiles de salud y las medidas de la utilidad; se emplean en diferentes tipos de enfermedades, pacientes o poblaciones y permiten comparar el impacto de las enfermedades, hacer análisis de costo-utilidad, obtener valores poblacionales de referencia y evaluar programas para asignar recursos. Los específicos se usan en pacientes o poblaciones para evaluar síntomas, funciones o enfermedades.<sup>8,9</sup>

Uno de los instrumentos genéricos validado en español es el cuestionario SF-36,<sup>10-13</sup> desarrollado en Estados Unidos. Incluye 36 ítems agrupados en 8 escalas: funcionamiento físico (FF), desempeño físico (DF), dolor corporal (DL), desempeño emocional (DE), salud mental (SM), vitalidad (VT), salud general (SG) y funcio-

namiento social (FS) y uno adicional, el *cambio de la salud en el tiempo*.<sup>11,13,14</sup>

El SF-36 en su idioma original mostró confiabilidad y validez de apariencia, de contenido, de criterio (concurrente y predictiva) y de constructo.<sup>14,15</sup> Se usó en más de 40 países en el proyecto International Quality of Life Assessment (IQOLA), está documentado en más de 1.000 publicaciones, su utilidad para estimar la carga de la enfermedad está descrita en más de 130 condiciones y por su brevedad y comprensión se usa ampliamente en el mundo. En la validación realizada en España<sup>10</sup> se obtuvo una versión en español que mostró coincidencia completa con la ordinalidad original esperada, alta equivalencia con los valores originales y reproducibilidad aceptable, menor que la original.<sup>16</sup>

Debido a que en Colombia no hay publicaciones del SF-36, ni disponemos de otro instrumento validado de este tipo, el objetivo de la investigación fue adaptarlo culturalmente y valorar su consistencia interna y su fiabilidad ínterevaluador e intraevaluador en Medellín, Colombia.

## Materiales y métodos

### Diseño y población

Es un estudio descriptivo para adaptar culturalmente y evaluar la fiabilidad del cuestionario SF-36. La población incluyó colombianos de habla hispana, mayores de 18 años, residentes en Medellín y su área metropolitana, que firmaron consentimiento para participar y que estaban en alguna de estas condiciones:

- Dolor músculo-esquelético, que consultaron por primera vez, con evolución de 3 semanas a 3 meses.
- Diabetes mellitus tipo II, con más de 6 meses de evolución, en tratamiento con antidiabéticos orales y/o insulina, sin complicaciones graves (retinopatía, neuropatía o nefropatía).
- Depresión clínica diagnosticada por un psiquiatra según los criterios del DSM IV, con más de tres semanas de tratamiento farmacológico y condición clínica estable.
- Personas sanas, que no hubieran consultado por un trastorno físico ni mental en los últimos 3 meses.

A todos los sujetos elegibles se les aplicó la escala de Zung.<sup>17</sup> En los grupos de sanos, diabetes y dolor músculo-esquelético solo se incluyeron las personas clasificadas con depresión leve o ausente, no los moderados ni los severos. En el grupo de depresión se aceptó cualquier resultado en el test de Zung.

Se captaron consecutivamente en los registros médicos de consulta ambulatoria: dolor músculo-esquelético en la Unidad de Medicina Física y Rehabilitación de la Clínica Las Américas; diabéticos en el programa de

control del Centro de Atención Ambulatoria del Instituto de Seguros Sociales; y los pacientes con depresión en el Hospital Mental de Antioquia, las clínicas psiquiátricas INSAM y SAMEIN y miembros de la Asociación Colombiana de Deprimidos. Los sanos fueron acompañantes de la consulta externa de esos hospitales y en las Universidades de Antioquia y Pontificia Bolivariana.

La fiabilidad del cuestionario, que incluyó la consistencia interna de los ítems (correlación entre ellos y su escala correspondiente) y la confiabilidad entre las escalas (alfa de Cronbach) se hizo con 605 personas, mediante dos formas de aplicación (entrevista y autodiligiado). La fiabilidad interevaluador, cuando se aplicó por parte de dos entrevistadores diferentes en dos días consecutivos, se hizo con 135 de las 341 personas iniciales escogidas al azar. La fiabilidad test-retest se evaluó de 10 a 14 días después de la primera evaluación, a 105 personas diferentes (51 por entrevista y 54 autodiligiado) seleccionadas al azar.

Las personas que se incluyeron en el análisis de la fiabilidad interevaluador debían cumplir el criterio de regresar al día siguiente o que permitieran una visita domiciliaria y las de la fiabilidad test-retest debían tener disponibilidad para la segunda encuesta. La recolección de los datos la hicieron tres profesionales, una médica y dos psicólogos, previamente capacitadas.

Para definir el tamaño de muestra necesario para valorar la consistencia interna de los ítems y la confiabilidad entre escalas no se encontraron fórmulas estadísticas, por lo que se tomó el mayor número muestral entre los obtenidos para la fiabilidad interevaluador y test-retest, que se estimaron con base en el coeficiente de correlación intraclase (CCI), usando las tablas de Walter<sup>18</sup> y con los estimadores mínimos aceptados por la teoría psicométrica y los reportados para el cuestionario SF-36.<sup>10, 15, 19, 20</sup>

Los criterios para la muestra de la fiabilidad interevaluador, para 2 mediciones fueron: hipótesis nula  $\rho_0 = 0,70$ , hipótesis alterna  $\rho_1 > 0,80$ ; error  $\alpha = 0,05$ , error  $\beta = 0,20$ , lo que dio una muestra de 117 para cada grupo de estudio. Se completó hasta 605 personas para ajustar las pérdidas. Los criterios de la fiabilidad test-retest para dos mediciones fueron: hipótesis nula  $\rho_0 = 0,80$ , hipótesis alterna  $\rho_1 > 0,90$ , error  $\alpha = 0,05$ , error  $\beta = 0,20$ , lo cual dio un número de 45 personas por grupo. En todos los casos se buscó mantener el mínimo de sujetos necesarios para las dos formas de aplicación: autodiligiada (en presencia de una encuestadora) y entrevista.

### El cuestionario de calidad de vida SF-36

Se usó la versión colombiana del SF-36 suministrada por sus autores, que corresponde a la versión en inglés,

traducida y retraducida por un grupo de expertos colombianos, con la supervisión de aquellos. La tabla 1 ilustra los 35 ítems según la dimensión de la calidad de vida que evalúan y el que evalúa el cambio de salud en el tiempo y el anexo, el cuestionario con sus opciones de respuesta.

### Prueba piloto

El SF-36 se probó con 100 pacientes (por entrevista 53 y autodiligiadas, 47): depresión (21), diabetes (19), dolor músculo-esquelético (20) y acompañantes (40), seleccionados al azar, diferentes de los que se incluyeron en el estudio. La recolección fue en las mismas instituciones mencionadas por dos de los investigadores. Se calculó la frecuencia de respuesta de cada ítem, el tiempo de aplicación y la comprensión del cuestionario. Se identificaron las preguntas que requerían aclaración o que tenían dificultades y los motivos de ello. Se hizo un análisis independiente para cada tipo de aplicación, para identificar diferencias en la comprensión de las preguntas, tiempo y frecuencia de respuesta de ítems y de datos omitidos. Luego de esta prueba se hicieron las adaptaciones semánticas necesarias que no alteraran la estructura de la escala.

### Análisis de la información

Los análisis estadísticos se hicieron para toda la muestra y para 16 subgrupos: edad (18-34, 35-64,  $\geq 65$ ), sexo, años de estudio ( $\leq 4$ , 5, 6-10, 11,  $\geq 12$ ), diagnóstico (4) y modo de aplicación del cuestionario (2). Se usaron las medidas estadísticas: tendencia central y dispersión y porcentajes de respuestas piso y techo, tanto para ítems como para escalas. Con la escala de Zung se calculó para establecer el índice de depresión: ausencia (25-49), leve o mínima (50-59), moderada a marcada (60-69) y severa a extrema (70-100). La integridad de datos en los ítems se determinó por los datos omitidos, los que se imputaron con puntajes según lo establecido por los autores,<sup>15</sup> siempre que en una escala se hubiera respondido al menos 50% de los demás ítems; se calcularon las frecuencias relativas de las respuestas de cada ítem y se describieron con promedios y desviación estándar. Para la consistencia interna se calcularon los coeficientes de correlación de Pearson entre los ítems y sus escalas hipotetizadas, es decir, ajustadas por sobrestimación del ítem dentro de su escala, tanto para toda la población como para subgrupos.<sup>19</sup>

El nivel de escalas se valoró con estadísticas descriptivas de tendencia central, dispersión, variabilidad, rango y frecuencias relativas del piso y techo. Se calcularon frecuencias relativas de datos faltantes por ítem en su escala. Los puntajes por escala se transformaron

**Tabla 1.** Contenido del cuestionario de calidad de vida SF-36 por ítems

Escala de salud	Contenido resumido del ítem	Media (DE)	Distribución porcentual de ítems					
			1	2	3	4	5	6
<b>Función física (FF)</b>	Actividades intensas tales como correr, levantar objetos pesados (FF1)	1,91 (0,87)	40,8	26,0	32,7	--	--	--
	Actividades moderadas tales como mover una mesa, trapear, lavar (FF2)	2,36 (0,82)	21,5	20,7	57,7	--	--	--
	Levantar o llevar bolsas de compras (FF3)	2,55 (0,73)	14,0	16,0	69,8	--	--	--
	Subir varios pisos por una escalera (FF4)	2,40 (0,80)	19,0	21,7	59,0	--	--	--
	Subir un piso por una escalera (FF5)	2,74 (0,59)	5,5	12,7	81,0	--	--	--
	Agacharse, arrodillarse o ponerse de cuclillas (FF6)	2,40 (0,82)	20,8	17,4	61,5	--	--	--
	Caminar más de un kilómetro (FF7)	2,59 (0,73)	13,4	13,9	72,4	--	--	--
	Caminar medio kilómetro (FF8)	2,71 (0,65)	8,6	10,1	80,7	--	--	--
	Caminar cien metros (FF9)	2,86 (0,47)	3,5	6,0	90,1	--	--	--
	Bañarse o vestirse (FF10)	2,89 (0,41)	2,6	5,8	91,4	--	--	--
<b>Desempeño físico (DF)</b>	Menor tiempo dedicado al trabajo u otras actividades (DF1)	1,59 (0,50)	41,0	58,8	--	--	--	--
	Hacer menos actividades de las que hubiera querido hacer (DF2)	1,51 (0,50)	48,9	51,1	--	--	--	--
	Limitación en el tipo de trabajo o actividades (DF3)	1,54 (0,50)	46,4	53,6	--	--	--	--
	Dificultades para hacer el trabajo u otras actividades (DF4)	1,52 (0,50)	47,8	52,2	--	--	--	--
<b>Dolor físico (DL)</b>	Intensidad del dolor físico (DL1)	4,07 (1,67)	6,4	18,0	21,7	10,1	16,9	26,8
	Interferencia del dolor físico para el trabajo diario (DL2)	4,08 (1,57)	6,1	13,4	15,7	21,2	17,9	25,6
<b>Salud general (SG)</b>	Su salud es: excelente, muy buena, buena, regular, mala (SG1)	3,24 (1,17)	4,8	32,1	29,1	24,3	9,8	--
	Me parece que me enfermo más fácil que otras personas (SG2)	3,89 (1,45)	10,7	8,1	11,9	14,9	53,4	--
	Estoy tan sano como cualquiera (SG3)	3,56 (1,47)	12,6	14,7	9,9	25,0	36,9	--
	Creo que mi salud va a empeorar (SG4)	3,92 (1,41)	7,9	8,9	15,5	11,6	54,7	--
	Mi salud es excelente (SG5)	3,41 (1,47)	14,9	16,2	8,4	30,1	29,6	--
<b>Vitalidad (VT)</b>	Sentirse lleno de vitalidad (VT1)	4,35 (1,59)	5,8	7,9	18,5	9,6	26,1	31,4
	Tener mucha energía (VT2)	4,15 (1,62)	6,8	10,1	20,2	10,2	25,6	26,4
	Sentirse agotado (VT3)	4,16 (1,50)	6,1	8,9	11,1	32,4	16,0	24,8
	Sentirse cansado (VT4)	3,88 (1,40)	6,3	10,4	15,0	36,9	16,5	14,2
<b>Función social (FS)</b>	Frecuencia de problemas de salud que interfieren con actividades sociales (FS1)	3,83 (1,34)	8,4	10,2	18,0	16,5	46,8	--
	Tiempo que problemas de salud interfieren con actividades sociales (FS2)	3,55 (1,34)	9,8	12,6	26,3	15,7	35,7	--
<b>Desempeño emocional (DE)</b>	Menor tiempo para actividades por problemas emocionales (DE1)	1,57 (0,50)	42,6	57,4	--	--	--	--
	Hacer menos actividades por problemas emocionales (DE2)	1,52 (0,50)	47,6	52,4	--	--	--	--
	Menor cuidado en actividades por problemas emocionales (DE3)	1,68 (0,47)	31,7	68,3	--	--	--	--
<b>Salud mental (SM)</b>	Ha estado muy nervioso (SM1)	4,01 (1,66)	10,4	11,2	9,8	25,6	18,0	24,3
	Sentirse con el ánimo decaído (SM2)	4,57 (1,63)	7,4	6,8	6,6	21,5	13,4	43,8
	Sentirse tranquilo y sereno (SM3)	4,13 (1,65)	7,8	9,8	19,5	9,1	27,1	26,0
	Sentirse desanimado y triste (SM4)	4,17 (1,52)	6,8	8,8	9,4	31,6	18,3	24,5
	Sentirse feliz (SM5)	4,20 (1,62)	7,4	8,4	18,7	13,1	23,3	28,4
<b>Cambio de salud</b>	Calificación del estado de salud comparado con el de hace 1 año (CT)	3,07 (1,09)	8,4	16,5	45,3	17,4	12,1	--

en una puntuación de 0 a 100 mediante la siguiente fórmula:<sup>15</sup>

$$\text{Escala transformada} = [(\text{puntaje obtenido} - \text{puntaje más bajo posible}) / \text{rango posible}] \times 100$$

La consistencia interna de la escala se determinó con el coeficiente alfa de Cronbach.<sup>13, 19</sup> Para la fiabilidad interevaluador se usó el CCI. El valor de comparación del coeficiente fue de 0,70 recomendado por los autores y por la teoría psicométrica.<sup>19</sup> Para evaluar la fiabilidad test-retest se calculó el CCI para cada escala.

## Resultados

### Descripción de la primera evaluación

El SF-36 se aplicó por primera vez a 605 personas, 341 (56,4%) fueron por entrevista y 264 (43,6%) autodiligienciados (con presencia de un encuestador), pertenecientes a los siguientes grupos: depresión 144 (23,8%), diabetes 172 (28,4%), dolor músculo-esquelético 150 (24,8%) y sanos 139 (23,0%). El promedio de edad fue 45,7 años (DE: 16,6) y el 66% eran mujeres. En la tabla 2 se describen las demás características de estudio.

### Nivel de los ítems

La frecuencia de datos sin respuesta fue baja: máximo en SG4 con 1,5%. La imputación de los datos faltantes se hizo hasta en el 0,7% para SG1 y 0,6% para FF5. Todos los niveles de los ítems tuvieron alguna respuesta. En general, las distribuciones fueron asimétricas, con porcentajes más altos hacia el mejor estado de salud en los ítems. En las preguntas de la escala funcionamiento físico que implican una menor demanda física, como subir un piso (FF5), caminar cien metros (FF9) o bañarse (FF10), se observó una mayor asimetría hacia la mejor condición de salud. El FF1 que implica la mayor demanda física fue el único que tuvo una desviación a la izquierda. En las escalas desempeño físico y desempeño emocional, que tienen dos opciones de respuesta, la distribución fue similar para cada una, excepto en DE3 que fue más del doble en la mejor condición de salud. En la pregunta cambio del estado de salud, la respuesta del centro (más o menos igual ahora que hace un año) tuvo el mayor número de respuestas (tabla 1).

**Tabla 2.** Características de la muestra y de los grupos según tipo de diagnóstico

	<i>Depresión</i>	<i>Diabetes</i>	<i>Dolor músculo-esquelético</i>	<i>Sanos</i>	<i>Total</i>
	N.º (%)	N.º (%)	N.º (%)	N.º (%)	N.º (%)
<b>Total</b>	144 (23,8)	172 (28,4)	150 (24,9)	139 (23,1)	605
<b>Edad (años)</b>					
Media (DE)	41,9 (14,0)	61,3 (12,2)	43,5 (14,9)	32,9 (9,3)	45,7 (16,6)
18–34	45 (31,5)	7 (4,1)	41 (27,3)	87 (62,6)	180 (29,9)
35–64	89 (62,2)	89 (52,0)	92 (61,3)	51 (36,7)	321 (53,2)
≥ 65	9 (6,3)	75 (43,9)	17 (11,3)	1 (0,7)	102 (16,9)
Sin dato	1	1	0	0	2
<b>Mujeres (%)</b>	109 (75,7)	103 (59,9)	78 (52,0)	109 (78,4)	399 (66,0)
<b>Escolaridad (años)</b>					
Media (DE)	10,5 (4,5)	6,7 (4,3)	12,3 (4,4)	13,6 (4,0)	10,6 (5,1)
≤ 4	12 (8,5)	48 (29,3)	3 (2,1)	2 (1,5)	65 (11,1)
5	16 (11,3)	39 (23,8)	11 (7,5)	5 (3,6)	71 (12,1)
6–10	31 (22,0)	38 (23,2)	23 (15,8)	10 (7,3)	102 (17,3)
11	29 (20,6)	19 (11,6)	40 (27,4)	26 (19,0)	114 (19,4)
≥ 12	53 (37,6)	20 (12,2)	69 (47,3)	94 (68,6)	236 (40,1)
Sin dato	3	8	4	2	17
<b>Depresión según escala de Zung</b>					
Ausencia	26 (19,0)	148 (86,0)	131 (87,3)	122 (87,8)	427 (71,4)
Leve	21 (15,3)	24 (14,0)	19 (12,7)	17 (12,2)	81 (13,6)
Moderada	27 (19,7)	0 (0,0)	0 (0,0)	0 (0,0)	27 (4,5)
Severa	63 (46,0)	0 (0,0)	0 (0,0)	0 (0,0)	63 (10,5)
Sin dato	7	0	0	0	7

El promedio y la variabilidad de las respuestas de los ítems dentro de cada escala tuvieron valores similares entre sí, excepto en el ítem FF1 (actividades intensas) de la escala FF. La desviación estándar de los ítems con 5 opciones de respuesta (escalas salud general y función social) fue cercana a 1,0. Los coeficientes de variación de los ítems estuvieron entre 30 a 90%.

Todas las correlaciones de Pearson del ítem con su escala hipotetizada fueron superiores a 0,48. Estas correlaciones dentro de la escala fluctuaron poco, con diferencias entre el valor más alto y más bajo de máximo 0,23 para FF (tabla 3). Las medianas de las correlaciones ítem-escala hipotetizadas para cada escala fluctuaron entre 0,57 en SG y 0,82 en dolor.

Las correlaciones de cada ítem con las demás escalas (validez discriminante) están en la tabla 3 y tuvieron valores más bajos que las correlaciones ítem-escala hipotetizada en su escala, excepto en vitalidad y FS. En VT el no éxito lo tuvo el ítem 2 con una correlación más alta con salud mental y con FS.

### Nivel de las escalas

El máximo porcentaje de datos incompletos por escala antes de hacer imputaciones fue 1,1% para SG. Las escalas con más datos omitidos fueron SG, SM y VT. Después de hacer las asignaciones de respuestas a los datos omitidos, las proporciones de personas que tuvieron todos los datos completos variaron entre 98,7 y 100%. El porcentaje de personas que respondieron la opción mínima (efecto piso) fue notorio en DF (32,2%)

y DE (27,4%), en las demás no fue importante. El efecto techo (proporción de respuestas para la mayor opción) fue alto en las escalas DE (47,6%), DF (42,1%), FS (30,1%), DL (25,7%) y FF (24,0%). Las escalas bipolares (que varían desde una condición negativa hasta una condición positiva) SG, VT y SM tuvieron un efecto piso-techo insignificante.

Las medias de los puntajes transformados para las escalas variaron desde 54,0 (desviación estándar: 44,3) en DF y 59,3 (DE: 43,2) en DE hasta 77,5 (DE: 24,1) en FF. Las desviaciones estándar fluctuaron entre 24,1 y 44,3 (de un rango posible de 0 a 100 puntos) y los coeficientes de variación lo hicieron entre 31,1 y 82,1%. En la tabla 3 se observa el coeficiente alfa de Cronbach de las escalas, cuyo valor mínimo fue de 0,80 para FS. Los valores más altos fueron para DF, DL y SM.

### Análisis por grupos

Las correlaciones ítem-escala en los grupos de edad, sexo y escolaridad fueron mayores de 0,40. La consistencia ítem-escala para los diagnósticos fue mayor de 0,40. Los coeficientes de Cronbach superaron en todos los grupos el estándar de 0,70 (tabla 4). El 56,4% (341) de las evaluaciones se hicieron por entrevista, con un promedio de edad de 47,4 años y predominio de mujeres (64%); 38,3% tenían estudios universitarios. El grupo a quienes la encuesta se hizo autodiligenciada tenía una edad media de 43,5 años, 68,2% eran mujeres y 42,6% tenían más de 11 años de escolaridad. Los puntajes promedio de las escalas fueron similares según las

**Tabla 3.** Confiabilidad ítem-escala del SF-36\*

Escala	Rango de correlaciones de ítems		Prueba ítems-escalas		Confiabilidad (alfa de Cronbach)
	Consistencia interna <sup>†</sup>	Validez discriminante <sup>‡</sup>	N.º éxitos / total <sup>¶</sup>	% éxito escala	
Función física	0,48–0,71	0,14–0,47	80/80	100	0,89
Desempeño físico	0,76–0,82	0,19–0,53	32/32	100	0,91
Dolor físico	0,82	0,21–0,54	16/16	100	0,90
Salud general	0,55–0,71	0,21–0,59	40/40	100	0,82
Vitalidad	0,62–0,68	0,23–0,67	31/32	97	0,83
Función social	0,67	0,25–0,67	15/16	94	0,80
Desempeño emocional	0,64–0,80	0,25–0,62	24/24	100	0,86
Salud mental	0,75–0,81	0,29–0,70	40/40	100	0,91

\*  $n = 605$

<sup>†</sup> Rango de los coeficientes de correlación ítem-escala hipotetizada con corrección por sobreestimación

<sup>‡</sup> Rango de los coeficientes de correlación entre los ítems y otras escalas

<sup>¶</sup> Número de correlaciones ítem-escala hipotetizadas mayores de 0,40/número total de correlaciones

dos formas de aplicación. Ambos grupos tuvieron un promedio alto en las escalas DF y luego en la DE. El rango de respuestas para casi todas las escalas en ambos grupos estuvo entre 0 y 100. El comportamiento de los efectos piso y techo fue similar a los obtenidos en la muestra total y no hubo diferencias significativas entre ellos. La confiabilidad ítem-escala fue similar en los dos modos de aplicación y estuvo por encima de 0,50, excepto en FF de los entrevistados (tabla 4). Los coeficientes de Cronbach fueron similares en las dos formas y estuvieron por encima de 0,80 menos en FS de los autodiligenciados (0,77).

**Fiabilidad interobservador e intraobservador**

La fiabilidad de las escalas entre dos observadores distintos se ilustra en la figura 1a. Los valores de los CCI de todas las escalas fueron superiores a 0,80. En la figura 1b se muestran los resultados de los CCI para dos observaciones del mismo evaluador, que estuvieron por encima de 0,70. Solo los límites inferiores del intervalo de confianza de los CCI de las escalas DL y cambio de salud fueron menores de ese valor.

**Calidad de vida de los grupos**

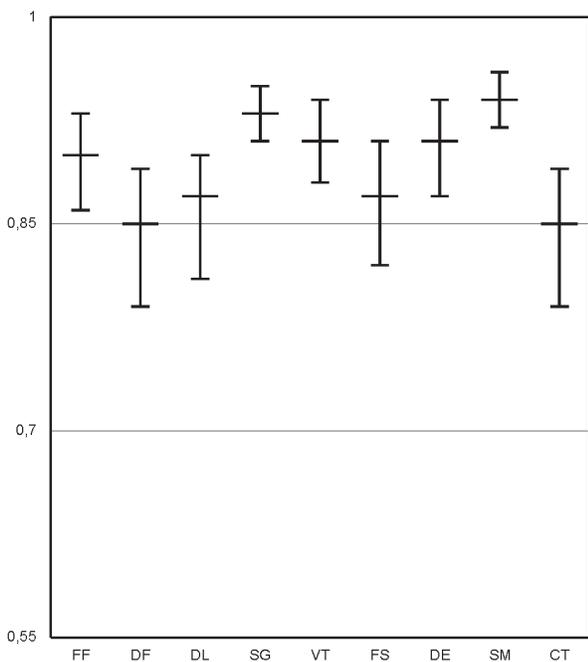
Los puntajes promedio de las escalas transformadas de los grupos de diagnóstico están en la figura 2. Todos los puntajes de los pacientes deprimidos, diabéticos y con dolor músculo-esquelético fueron menores a los sanos.

En los deprimidos, los puntajes más bajos fueron en la escala salud mental (DE, SM, FS, VT); en los diabéticos, la escala de menor puntaje fue DF (60,0); entre los sujetos con dolor músculo-esquelético, la escala DF fue muy baja (37,2) y en los sanos, todos los puntajes estuvieron por encima de 70. La calificación que las personas hacen sobre su estado de salud presente comparado con el de hace un año tuvo puntajes bajos (menores de 62) en todos los grupos.

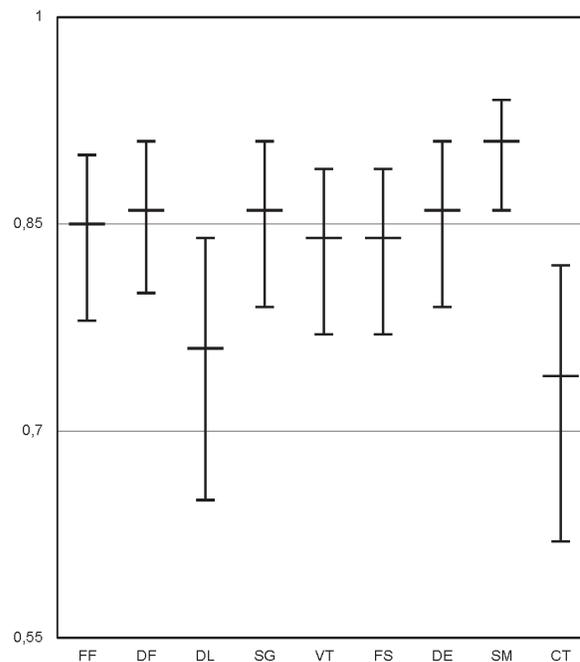
**Discusión**

Esta investigación se hizo para suplir la falta de un instrumento genérico de evaluación de la cvs disponible para los trabajadores de salud en Colombia. Se seleccionó el cuestionario SF-36 por usarse ampliamente en enfermedades crónicas,<sup>21-26</sup> degenerativas<sup>27,28</sup> y agudas;<sup>29,30</sup> en diferentes grupos de edad y sexo,<sup>31,32</sup> en salud pública y en medición de consecuencias de las enfermedades.<sup>33,34</sup> No se validó sino que se adaptó culturalmente y se evaluó su fiabilidad porque hay estudios que la hacen en cuanto al contenido,<sup>11</sup> constructo,<sup>12</sup> validez convergente, divergente y predictiva<sup>11,12</sup> y al cambio en el tiempo,<sup>35</sup> y porque hay una versión española con equivalencia lingüística entre los ítems traducidos y los originales y con coeficientes de Cronbach entre 0,7 y 0,94.<sup>10</sup>

Los datos omitidos encontrados (máximo de 1,5% en el ítem SG4) fueron menores que en IQOLA y en países escandinavos que hicieron autoaplicación del ins-



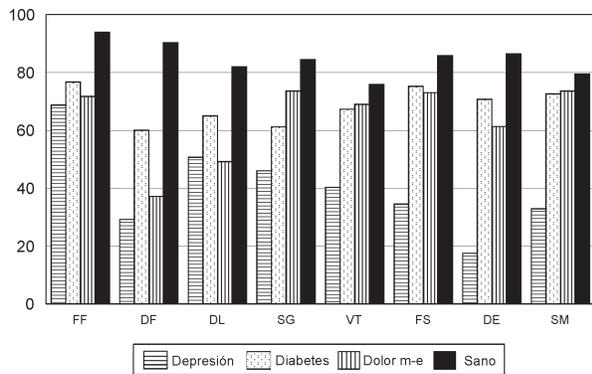
**Figura 1a.** Coeficiente de correlación intraclassa de la fiabilidad interevaluador



**Figura 1b.** Coeficiente de correlación intraclassa de la fiabilidad intraevaluador

**Tabla 4.** Confiabilidad ítem-escala y coeficientes de confiabilidad alfa de Cronbach del SF-36 por grupos

	Rango de correlaciones de ítems-escala										Coeficientes de confiabilidad									
	FF	DF	DL	SG	VT	FS	DE	SM	FF	DF	DL	SG	VT	FS	DE	SM				
<b>Total</b>	0,48-0,71	0,76-0,82	0,82	0,55-0,71	0,62-0,68	0,67	0,64-0,80	0,75-0,81	0,89	0,91	0,90	0,82	0,83	0,80	0,86	0,91				
<b>Edad</b>																				
18-34	0,48-0,71	0,82-0,86	0,83	0,57-0,75	0,51-0,72	0,71	0,62-0,80	0,77-0,86	0,88	0,93	0,90	0,86	0,82	0,83	0,86	0,92				
35-64	0,42-0,69	0,75-0,81	0,83	0,55-0,72	0,64-0,72	0,68	0,68-0,81	0,76-0,82	0,87	0,90	0,91	0,82	0,85	0,81	0,87	0,92				
≥ 65	0,42-0,80	0,67-0,83	0,78	0,37-0,62	0,44-0,61	0,55	0,60-0,81	0,64-0,72	0,90	0,89	0,87	0,73	0,72	0,71	0,85	0,87				
<b>Sexo</b>																				
Mujer	0,44-0,70	0,75-0,84	0,83	0,57-0,69	0,44-0,64	0,69	0,66-0,84	0,72-0,80	0,88	0,92	0,90	0,82	0,83	0,82	0,88	0,91				
Hombre	0,57-0,77	0,77-0,82	0,82	0,45-0,77	0,54-0,64	0,62	0,59-0,73	0,76-0,83	0,90	0,91	0,90	0,81	0,79	0,77	0,82	0,92				
<b>Escolaridad</b>																				
≤ 4	0,39-0,73	0,68-0,81	0,83	0,45-0,61	0,51-0,70	0,68	0,63-0,85	0,67-0,81	0,87	0,88	0,91	0,76	0,81	0,81	0,88	0,91				
5	0,52-0,80	0,72-0,85	0,80	0,28-0,67	0,57-0,68	0,63	0,66-0,83	0,71-0,81	0,89	0,92	0,89	0,72	0,81	0,77	0,87	0,91				
6-10	0,47-0,75	0,73-0,83	0,79	0,56-0,71	0,51-0,69	0,60	0,67-0,76	0,72-0,80	0,87	0,90	0,89	0,83	0,79	0,75	0,85	0,91				
11	0,30-0,65	0,71-0,85	0,83	0,51-0,76	0,48-0,61	0,57	0,50-0,72	0,68-0,82	0,83	0,89	0,90	0,82	0,76	0,73	0,80	0,91				
≥ 12	0,52-0,73	0,81-0,87	0,83	0,57-0,75	0,68-0,77	0,74	0,65-0,85	0,74-0,83	0,89	0,93	0,90	0,84	0,87	0,85	0,88	0,92				
<b>Lugar de residencia</b>																				
Medellín	0,48-0,72	0,76-0,81	0,80	0,54-0,73	0,59-0,67	0,65	0,63-0,79	0,72-0,80	0,88	0,91	0,89	0,82	0,81	0,79	0,86	0,90				
Resto Área Metropolitana	0,53-0,74	0,76-0,94	0,91	0,52-69	0,67-76	0,71	0,67-0,82	0,81-0,86	0,90	0,94	0,95	0,82	0,87	0,83	0,88	0,94				
<b>Diagnóstico</b>																				
Depresión	0,54-0,71	0,73-0,87	0,85	0,54-0,62	0,57-0,69	0,61	0,41-0,69	0,72-0,80	0,90	0,91	0,92	0,79	0,81	0,75	0,72	0,90				
Diabetes	0,45-0,76	0,74-0,80	0,81	0,38-0,66	0,46-0,55	0,48	0,57-0,77	0,57-0,71	0,87	0,88	0,89	0,74	0,71	0,64	0,83	0,84				
Dolor musculoesquelético	0,45-0,72	0,69-0,74	0,69	0,40-0,62	0,41-0,54	0,47	0,59-0,71	0,54-0,65	0,86	0,87	0,81	0,72	0,70	0,64	0,81	0,80				
Sano	0,14-0,54	0,64-0,72	0,80	0,36-0,61	0,55-0,66	0,50	0,51-0,72	0,50-0,63	0,67	0,83	0,89	0,73	0,68	0,65	0,79	0,78				
<b>Aplicación del cuestionario</b>																				
Autodilucidado	0,51-0,77	0,76-0,79	0,80	0,52-0,75	0,65-0,68	0,62	0,66-0,81	0,77-0,81	0,90	0,90	0,89	0,84	0,84	0,77	0,88	0,91				
Entrevista	0,46-0,70	0,75-0,86	0,84	0,53-0,70	0,58-0,69	0,70	0,62-0,79	0,73-0,82	0,87	0,92	0,92	0,81	0,82	0,82	0,85	0,91				



**Figura 2.** Perfil de la calidad de vida en grupos de diagnóstico

trumento fue de 11,5%,<sup>36</sup> y fueron parecidos a los de países que hicieron entrevistas o administración telefónica:<sup>37, 38</sup> Italia (0,6% en VT1), España (2,5% en SF1), Reino Unido (1,9% en FF1) y Estados Unidos (1,8% en FF1 y DE3).

Las escalas con más datos omitidos fueron SG, SM y VT, que tienen posibilidades de respuesta múltiple tipo Likert. De acuerdo con la prueba piloto, este tipo de respuestas fueron las que presentaron mayor grado de dificultad. Nuestros hallazgos son superiores al IQOLA y cercanos a los de Italia, en donde la mitad del estudio fue por entrevista. Después de asignar respuestas a los datos omitidos según las normas del SF-36, las proporciones de personas que tuvieron todos los datos completos fueron mayores a 98,7%, también superiores a IQOLA, excepto Italia y Japón. Lo anterior muestra una buena integridad de los datos y permite concluir que este instrumento fue comprendido en la población y que es posible con él obtener información adecuada en diferentes grupos.

La distribución de frecuencias de respuestas de cada uno de los ítems fue similar a los reportados por el MOS.<sup>11, 36</sup> En el ítem de percepción de salud de la escala salud general, es notorio el bajo porcentaje de respuesta de la última opción, en ambos estudios.

Los puntajes promedio de los ítems de este estudio fueron similares a MOS e IQOLA, aunque en función social todos fueron menores a los primeros. La desviación estándar de los ítems con cinco opciones de respuesta (escalas SG y FS) fue cercana a 1,0, tal como lo recomienda Levy.<sup>39</sup> La magnitud y amplitud de los coeficientes de variación, que variaron desde 30 a 90% (datos no mostrados), indican la capacidad de las preguntas de medir todo el espectro posible de condiciones de salud de las personas y es una de las propiedades de la escala en otros estudios.

Todas las correlaciones ítem-escala hipotetizadas fueron superiores a 0,48, que es mayor a 0,40 aceptado en psicometría para definir que el ítem mide el concepto que pretende medir dentro de la escala.<sup>13, 14, 19, 40</sup> Los puntajes más bajos se obtuvieron en FF10, que es el único ítem que mide limitaciones severas, y en el FF1 (limitaciones para hacer actividades intensas), lo que es consistente con hallazgos en que las correlaciones más bajas se obtienen en los extremos del concepto. Las correlaciones de cada ítem con las demás escalas tuvieron valores más bajos que las correlaciones ítem-escala hipotetizada en su escala. Esto es necesario para garantizar la fiabilidad del instrumento, ya que indica que el ítem no está midiendo lo que las otras escalas pretenden medir (validez discriminante).<sup>19</sup> El estándar de 0,40 en las correlaciones ítem-escala se alcanzó en 98,5% (414/420), mientras que en el proyecto MOS fue de 97% (814/840). La confiabilidad ítem-escala fue similar en los dos modos de aplicación y estuvo por encima de 0,50.

Con respecto a las escalas, las limitaciones en las comparaciones de nuestros hallazgos con los de otros países son mayores, debido a que en estos los muestreos fueron poblacionales. Al comparar con el MOS (que tiene una muestra no poblacional), los promedios fueron similares, excepto en la FS (80,5 frente a 67,4), SM (71,4 frente a 64,9) y en VT (53,5 frente a 63,2), donde las diferencias son mayores de 10 puntos. La magnitud y amplitud de los valores de las desviaciones estándar (24,1-44,3) para un rango posible de 0 a 100 puntos, y de los coeficientes de variación (31 a 82%) reflejan la diversidad de la CV de las personas del estudio. El rango posible de calificaciones de cada una de las escalas (0 a 100 puntos) se observó en todas ellas. Todos los coeficientes de asimetría (valores no mostrados) fueron negativos, lo que indica una distribución con mayor proporción de personas en los estados de salud positivos notorios con una asimetría derecha evidente. La curtosis fue negativa en todas las escalas, excepto para la FF.

El comportamiento de los efectos piso y techo para las escalas DE y DF fue notorio y similar a otros estudios,<sup>11, 12</sup> explicado porque solo tienen dos opciones de respuesta.

El SF-36 mostró su fiabilidad en los 16 grupos (mayor de 0,70). El coeficiente de Cronbach de las escalas fue mayor que 0,80 (el mínimo aceptado internacionalmente es 0,70).<sup>19</sup> Los valores mayores de 0,90 fueron para DF, DL y SM, similares al MOS y al IQOLA. Los coeficientes alfa de Cronbach fueron también similares entre las dos formas de aplicación y en ambos estuvieron por encima de 0,70, lo que coincide con el MOS que evaluó las diferencias entre la aplicación de la encues-

ta por teléfono y por correo o el estudio de Australia, que comparó una aplicación autodiligiada con la administrada por teléfono.<sup>38, 39</sup> Esto permite considerar la aplicación futura por cualquiera de las dos formas. Como en este estudio, en más de 25 que incluyen comparaciones de grupos la confiabilidad siempre excedió el mínimo de 0,70; la mayoría superó 0,80.

Los valores del CCI superiores de 0,80 indican que los resultados no se modificaron por haber sido aplicado por dos observadores con 24 horas de diferencia, por lo que se puede recomendar su aplicación por parte de diferentes observadores. No se encontraron estudios similares en bibliografía consultada.

El instrumento es estable en el tiempo, ya que la fiabilidad intraobservador fue alta (por encima de 0,70). Solo los límites inferiores del intervalo de confianza de los CCI de las escalas DL y cambio de salud fueron menores de ese valor, resultados diferentes de España,<sup>10</sup> donde obtuvieron tres escalas con valores menores de 0,70: DL, FS y DE.

Los puntajes promedio de las escalas según los grupos de edad tuvieron una tendencia decreciente según ella; fueron más altos para las personas más jóvenes, entre los 18 y 34 años, y menores en los mayores de 64 años, excepto en la escala de SM. Todos los puntajes

de CV de los hombres fueron mayores que los de las mujeres. Según la escolaridad, las puntuaciones de las escalas de quienes habían cursado estudios universitarios fueron notoriamente mayores que las de los otros grupos. No existió una tendencia creciente a mejores puntuaciones de CV según una mayor escolaridad y llama la atención que los promedios de los del grupo que no completó el bachillerato son similares a los del grupo que no terminó la educación primaria (datos no mostrados).

Los resultados obtenidos muestran que la versión colombiana en español del cuestionario SF-36 es un instrumento adaptado y confiable que puede utilizarse en el país para el estudio de la CVS de personas sanas o enfermas, mayores de 18 años de edad, alfabetas con cualquier grado de escolaridad, mediante entrevista o autodiligiado asistido, debido a la dificultad de algunas preguntas y para garantizar un menor número de datos omitidos.

## Financiación

Esta investigación fue financiada por la International Clinical Epidemiology Network (INCLEN) y cofinanciada por la Vicerrectoría de Investigaciones de la Universidad de Antioquia.

## Anexo

### Cuestionario SF-36 sobre su estado de salud (español, Colombia), versión 1.2

**Instrucciones:** las preguntas que siguen se refieren a lo que usted piensa sobre su salud. Sus respuestas permitirán saber cómo se encuentra usted y hasta qué punto es capaz de hacer sus actividades habituales.

Conteste cada pregunta tal como se indica. Si no está seguro/a de cómo responder a una pregunta, por favor conteste lo que le parezca más cierto.

1. En general, ¿diría usted que su salud es: (marque un solo número.)

excelente? ..... 1  
 muy buena?..... 2  
 buena?..... 3  
 regular?..... 4  
 mala? ..... 5

2. ¿Cómo calificaría usted su estado general de salud actual, comparado con el de hace un año?  
 (Marque un solo número.)

Mucho mejor ahora que hace un año..... 1  
 Algo mejor ahora que hace un año..... 2  
 Más o menos igual ahora que hace un año..... 3  
 Algo peor ahora que hace un año ..... 4  
 Mucho peor ahora que hace un año..... 5

3. Las siguientes preguntas se refieren a actividades que usted puede hacer durante un día normal. ¿Su estado de salud actual lo/la limita en estas actividades? Si es así, ¿cuánto?  
(Marque un número en cada línea.)

	Sí, me limita mucho	Sí, me limita poco	No, no me limita para nada
a. Actividades intensas, tales como correr, levantar objetos pesados, participar en deportes agotadores	1	2	3
b. Actividades moderadas, tales como mover una mesa, empujar una aspiradora, trapear, lavar, jugar fútbol, montar bicicleta.	1	2	3
c. Levantar o llevar las bolsas de compras	1	2	3
d. Subir varios pisos por las escaleras	1	2	3
e. Subir un piso por la escalera	1	2	3
f. Agacharse, arrodillarse o ponerse en cuclillas	1	2	3
g. Caminar más de un kilómetro (10 cuadras)	1	2	3
h. Caminar medio kilómetro (5 cuadras)	1	2	3
i. Caminar cien metros (1 cuadra)	1	2	3
j. Bañarse o vestirse	1	2	3

4. Durante las últimas cuatro semanas, ¿ha tenido usted alguno de los siguientes problemas con su trabajo u otras actividades diarias normales a causa de su salud física?  
(Marque un número en cada línea.)

	Sí	No
a. ¿Ha disminuido usted el tiempo que dedicaba al trabajo u otras actividades?	1	2
b. ¿Ha podido hacer menos de lo que usted hubiera querido hacer?	1	2
c. ¿Se ha visto limitado/a en el tipo de trabajo u otras actividades?	1	2
d. ¿Ha tenido dificultades en realizar su trabajo u otras actividades (por ejemplo, le ha costado más esfuerzo)?	1	2

5. Durante las últimas cuatro semanas, ¿ha tenido usted alguno de los siguientes problemas con su trabajo u otras actividades diarias normales a causa de algún problema emocional (como sentirse deprimido/a o ansioso/a)?  
(Marque un número en cada línea.)

	Sí	No
a. ¿Ha disminuido el tiempo que dedicaba al trabajo u otras actividades?	1	2
b. ¿Ha podido hacer menos de lo que usted hubiera querido hacer?	1	2
c. ¿Ha hecho el trabajo u otras actividades con menos cuidado de lo usual?	1	2

6. Durante las últimas cuatro semanas, ¿en qué medida su salud física o sus problemas emocionales han dificultado sus actividades sociales normales con su familia, amigos, vecinos u otras personas?  
(Marque un solo número.)

Nada en absoluto .....	1
Ligeramente.....	2
Moderadamente.....	3
Bastante .....	4
Extremadamente.....	5

7. ¿Cuánto dolor físico ha tenido usted durante las últimas cuatro semanas?  
(Marque un solo número.)

Ninguno.....	1
Muy poco.....	2
Poco .....	3
Moderado.....	4
Mucho.....	5
Muchísimo.....	6

8. Durante las últimas cuatro semanas, ¿cuánto ha dificultado el dolor su trabajo normal (incluyendo tanto el trabajo fuera del hogar como las tareas domésticas)?  
(Marque un solo número.)

- Nada en absoluto ..... 1
- Un poco ..... 2
- Moderadamente ..... 3
- Bastante ..... 4
- Extremadamente ..... 5

9. Las siguientes preguntas se refieren a cómo se siente usted y a cómo le han salido las cosas durante las últimas cuatro semanas. En cada pregunta, por favor elija la respuesta que más se aproxime a la manera como se ha sentido usted.

¿Cuánto tiempo durante las últimas cuatro semanas...  
(Marque un número en cada línea.)

	Siempre	Casi siempre	Muchas veces	Algunas veces	Casi nunca	Nunca
a. se ha sentido lleno/a de vitalidad?	1	2	3	4	5	6
b. ha estado muy nervioso/a?	1	2	3	4	5	6
c. se ha sentido con el ánimo tan decaído/a que nada podría animarlo/a?	1	2	3	4	5	6
d. se ha sentido tranquilo/a y sereno/a?	1	2	3	4	5	6
e. ha tenido mucha energía?	1	2	3	4	5	6
f. se ha sentido desanimado/a y triste?	1	2	3	4	5	6
g. se ha sentido agotado/a?	1	2	3	4	5	6
h. se ha sentido feliz?	1	2	3	4	5	6
i. se ha sentido cansado/a?	1	2	3	4	5	6

10. Durante las últimas cuatro semanas, ¿cuánto tiempo su salud física o sus problemas emocionales han dificultado sus actividades sociales (como visitar amigos, parientes, etc.)?  
(Marque un solo número.)

- Siempre ..... 1
- Casi siempre ..... 2
- Algunas veces ..... 3
- Casi nunca ..... 4
- Nunca ..... 5

11. ¿Cómo le parece cada una de las siguientes afirmaciones?  
(Marque un número en cada línea.)

	Totalmente cierta	Bastante cierta	No sé	Bastante falsa	Totalmente falsa
a. Me parece que me enfermo más fácilmente que otras personas.	1	2	3	4	5
b. Estoy tan sano/a como cualquiera.	1	2	3	4	5
c. Creo que mi salud va a empeorar.	1	2	3	4	5
d. Mi salud es excelente.	1	2	3	4	5

## Referencias

1. Brock D. Medidas de calidad de vida en el cuidado de la salud y la ética médica. En: Nussbaum MC, Sen A, compiladores. La calidad de vida. México, D. F.: The United Nations University, Fondo de Cultura Económica, 1998. p.135-181.
2. Parfit D. Reasons and persons, citado por Dan Brock. Medidas de la calidad de vida en el cuidado de la salud y la ética médica. En: Nussbaum MC, Sen A, compiladores. La calidad de vida. México, D. F.: The United Nations University, Fondo de Cultura Económica, 1998. p. 135-182.
3. Dijkers M. Quality of life after spinal cord injury: a meta-analysis of the effects of disablement components. *Spinal Cord* 1997;35:829-40.
4. Organización Mundial de la Salud, Grupo sobre la Calidad de Vida. ¿Qué calidad de vida? *Foro Mundial de la Salud* 1996;17(4):385-87.
5. O'Connor R. Development of the health effects scales. Working Paper 43, National Centre for Health Program Evaluation. Melbourne: NHMRC; 1995. p. 3-17.
6. Spilker B, Revicki DA. Taxonomy of quality of life. En: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trials. 2.<sup>a</sup> ed. Filadelfia: Lippincott-Raven Publishers; 1996. p. 25-36.
7. O'Connor R. The development of models for allocation ATSC housing and infrastructure funds. For the Aboriginal and Torres Strait Islander Commission. En: O'Connor R. Development of the health effects scales. Working Paper 43. National Centre for Health Program Evaluation. Melbourne: NHMRC; 1995. p. 3-17.
8. Guyatt GH, Feeny DH, Patrick DL. Measuring Health-related Quality of Life. *Ann Int Med* 1993;118(8):622-629.
9. Bergner M, Bobitt RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787.
10. Alonso J, Prieto L, Antó JM. La versión española del SF-36 Health Survey (Cuestionario de Salud SF-36): un instrumento para la medida de los resultados clínicos. *Medicina Clínica* 1995;104(20):771-776.
11. Ware JE Jr, Sherbourne CD. The MOS 36-ítem short form health survey: I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-483.
12. McHorney CA, Ware JE Jr, 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(3):247-263.
13. McHorney CA, Ware JE Jr, Rachel JF, Sherbourne CD. The MOS 36-Item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patients groups. *Med Care* 1994;31(1):40-66.
14. Ware JE Jr. SF-36 Health Survey Update. *Spine* 2000;25(24):3130-3139.
15. Ware JE Jr, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993. p. 1-31.
16. Ware JE Jr, Gandek BL, Keller SD, IQOLA Project Group. Evaluating instruments used cross-nationally: methods from the IQOLA Project. En: Spilker B, editor. Quality of life and pharmacoeconomics in clinical trial. 2.<sup>a</sup> ed. Filadelfia: Lippincott-Raven Publishers; 1996. p. 681-92.
17. Zung W. A self-rating depression scale. *Arch Gen Psych* 1965;12:63-70.
18. Walter SD, Elias M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med* 1998;17:101-110.
19. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: The IQOLA Project Approach. *J Clin Epidemiol* 1998; 51(11):945-952.
20. Nunnally JC, Bernstein IR. Psychometric theory. 3.<sup>a</sup> ed. Nueva York: McGraw-Hill; 1994.
21. Stier DM, Greenfield S, Lubeck DP, Dukes KA, Flanders SC, Henning JM *et al*. Quantifying comorbidity in a disease-specific cohort: adaptation of the total illness burden index to prostate cancer. *Urology* 1999;54(3):424-429.
22. Eiser C, Kopel S, Cool P, Grimer R. The Perceived Illness Experience Scale (PIE): reliability and validity revisited. *Child: Care, Health & Development* 1999;25(3):179-190.
23. Keller SD, Ware JE, Hatoum HT. The SF-36 Arthritis-Specific Health Index (ASHI): II. Tests of validity in four clinical trials. *Med Care* 1999;60(5Suppl):60.
24. Kosinski M, Keller SD, Ware JE Jr, Hatoum HT, Kong SX. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: relative validity of scales in relation to clinical measures of arthritis. *Med Care* 1999;37(5 Suppl):MS3-39.
25. Thumboo J, Fong KY, Ng TP, Leong KH, Feng PH, Boey ML. Initial construct cross-cultural validation of the Short Form 36 for quality of life assessment of systemic lupus erythematosus patients in Singapore. *Annals of the Academy of Medicine* 1997;26(3):282-284.

26. Durham SR, Sun PP, Sutton LN. Surgically treated lumbar disc disease in the pediatric population: an outcome study. *J Neurosurgery* 2000;921 Suppl 1:1-6.
27. Wagner AK, Ehrenberg BL, Tran TA, Bungay KM, Cynn DJ, Rogers WH. Patient-based health status measurement in clinical practice: a study of its impact on epilepsy patients' care. *Quality of Life Research* 1997;6(4):329-341.
28. Hermann BP, Vickrey B, Hays RD, Cramer J, Devinsky O, Meador K *et al.* A comparison of health-related quality of life in patients with epilepsy, diabetes and multiple sclerosis. *Epilepsy Research* 1996;25(2):113-118.
29. Bessette L, Sangha O, Kuntz KM, Keller RB, Lew RA, Fossel AH *et al.* Comparative responsiveness of generic versus disease-specific and weighted versus unweighted health status measures in carpal tunnel syndrome. *Med Care* 1998;36(4):491-502.
30. Krousel-Wood MA. Predicting work status for patients in an occupational medicine setting who report back pain. *Archives Family Medicine* 1994;3(4):349-355.
31. Wolinsky FD, Wan GJ, Tierney WM. Changes in the SF-36 in 12 months in a clinical sample of disadvantaged older adults. *Med Care* 1998;36(11):1589-1598.
32. Kutner NG, Schechtman KB, Ory MG, Baker DI. Older adults' perceptions of their health and functioning in relation to sleep disturbance, falling, and urinary incontinence. FICSIT Group. *J Am Geriatrics Soc* 1994;42(7):757-762.
33. Keller SD, Bayliss MS, Ware JE Jr, Hsu MA, Damiano AM, Goss TF. Comparison of responses to SF-36 Health Survey questions with one-week and four-week recall periods. *Health Serv Res* 1997;32(3):367-384.
34. Lawrence WF, Fryback DG, Martin PA, Klein R, Klein BE. Health status and hypertension: a population-based study. *J Clin Epidemiol* 1996;49(11):1239-1245.
35. Ware JE Jr. The SF-36 Health Survey (SF-36). En: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2.<sup>a</sup> ed. Filadelfia: Lippincott-Raven Publishers; 1996. p. 337-345.
36. Rob W., Fisher S., Perkins J. Adaptation and validation of the SF-36 Health Survey for use in Australia. *J Clin Epidemiol* 1998;51(11):961-967.
37. McHorney CA, Kosinski M, Ware JE Jr. Comparisons of the costs and quality of norms for the SF-36 Health Survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994;32(6):551-567.
38. Perkins JJ, Sanson-Fisher RW. An examination of self and telephone-administered modes of administration for the Australian SF-36. *J Clin Epidemiol* 1998;51(11):969-973.
39. Levy KJ. Some multiple range tests for variances. *Educ Psychol Measmt* 1975;35:599-604.
40. Ware JE Jr., Keller SD. Interpreting general health measures. En: Spilker B, editor. *Quality of life and pharmacoeconomics in clinical trials*. 2.<sup>a</sup> ed. Filadelfia: Lippincott-Raven Publishers; 1996. p. 445-460.