

ORIGINAL

Adaptación cultural al español y validación de la escala GSFQ (Gastrointestinal Short Form Questionnaire)

Miguel A. Ruiz Díaz^{a,*}, José Manuel Suárez Parga^b, Antonio Pardo Merino^a, Mercedes García Vargas^c y Virginia Pascual Renedo^c

^aFacultad de Psicología, Universidad Autónoma de Madrid, España

^bHospital Universitario La Paz, Madrid, España

^cDepartamento de Investigación de Resultados en Salud, Unidad Médica, Pfizer España, Madrid, España

Recibido el 15 de julio de 2008; aceptado el 9 de septiembre de 2008

PALABRAS CLAVE
ERGE;
Enfermedad por
reflujo
gastroesofágico;
CVRS;
Adaptación cultural;
Propiedades
psicométricas

Resumen

Objetivo: describir el proceso seguido para la adaptación cultural y la validación al español del cuestionario GSFQ (Gastrointestinal Short Form Questionnaire) de interferencia de los síntomas asociados con la enfermedad de reflujo gastroesofágico e informar de sus propiedades psicométricas.

Material y métodos: el proceso de adaptación fue supervisado por un panel de 5 expertos. Tras la traducción, revisión y retrotraducción del instrumento se obtuvo un cuestionario en español que fue administrado a dos muestras de pacientes: una muestra piloto de 5 pacientes para valorar la comprensibilidad del instrumento y una muestra de 4.000 pacientes para comprobar su validez estructural (análisis factorial y fiabilidad) y la validez de constructo y discriminante.

Resultados: el cuestionario presenta una única dimensión que se ajusta a la estructura original. La fiabilidad es buena ($\alpha = 0,83$) con buena correlación entre ítems pares e impares ($r = 0,69$). La puntuación total correlaciona con las puntuaciones genéricas de CVRS medidas por el EQ-5D ($r = 0,499$) y la escala VAS ($r = -0,481$). La escala discrimina entre los niveles de gravedad (Savary-Miller; $p < 0,008$), aunque no para el nivel más elevado respecto a los inmediatamente anteriores. El cuestionario es capaz de detectar diferencias entre diversos grupos de enfermedades concomitantes y antecedentes. La sensibilidad respecto al criterio clínico de ERGE es del 60,5% y la especificidad del 68,3%. Se ofrecen los baremos normativos de corrección.

Conclusiones: los resultados obtenidos son satisfactorios, demuestran buenas propiedades psicométricas y ponen a disposición de los profesionales un nuevo instrumento para valorar la interferencia de la sintomatología de ERGE teniendo en cuenta el punto de vista del paciente.

© 2008 Elsevier España, S.L. Todos los derechos reservados.

*Autor para correspondencia.

Correo electrónico: miguel.ruiz@uam.es (M.A. Ruiz Díaz).

KEYWORDS

GERD;
Gastroesophageal
reflux disease;
HRQoL;
Cultural adaptation;
Psychometric
properties

Cultural adaptation to Spanish and validation of the Gastrointestinal Short Form Questionnaire

Abstract

Objective: To describe the process followed for the cultural and psychometric adaptation (validation) to Spanish of the Gastrointestinal Short Form Questionnaire (GSFQ), used to measure the interference of symptoms of gastroesophageal reflux disease GERD and to report the psychometric properties of this instrument.

Material and methods: The adaptation process was supervised by a five-member expert panel. After forward and backward translations in duplicate, a Spanish version was obtained, which was administered to two samples; a five-patient pilot sample to check comprehension and face validity, and a 4,000-patient sample to check structural validity (factor analysis and reliability), construct validity, and discriminative validity.

Results: The questionnaire showed a unique dimension that matched that of the original questionnaire. Reliability was high ($\alpha = 0.83$), and the correlation between even-odd items was good ($r = 0.69$). The overall score correlated with generic health-related quality of life measures evaluated by the EQ-5D tariff ($r = 0.499$) and VAS ($r = -0.481$). The scale discriminated between GERD severity levels ($p < 0.008$) as measured by the Savary-Miller scale, except for the most severe level with respect to the levels immediately below. The questionnaire was able to detect differences between diverse concomitant diseases and antecedents. Sensitivity with respect to the GERD clinician criterion was 60.5% and specificity was 68.3%. Normative comparison scaling values are reported.

Conclusions: The results show acceptable psychometric properties. A new instrument to assess the interference of GERD symptoms is thus available to health professionals. This instrument takes the patient's perspective into account.

© 2008 Elsevier España, S.L. All rights reserved.

La enfermedad por reflujo gastroesofágico (ERGE) es el término más reciente acuñado para definir cualquiera de los síntomas o lesiones histológicas que resultan tras episodios de reflujo gastroesofágico de larga evolución¹⁻⁴. Esta definición abarca un espectro clínico que se sitúa desde el reflujo ocasional con apenas síntomas, pasando por el reflujo clínicamente importante sin esofagitis, a la esofagitis complicada con estenosis, erosiones, úlcera, hemorragia, esófago de Barret, carcinoma, además de un cuadro clínico atípico de dolor torácico que se asemeja al dolor por isquemia coronaria³. Toda ERGE tiene un carácter crónico con una tendencia natural a la recidiva, lo que obliga a la utilización de tratamientos de mantenimiento, especialmente en los pacientes con lesiones esofágicas más graves, por lo que es fundamental la utilización de antisecretores, entre los cuales los inhibidores de la bomba de protones (IBP) han demostrado la máxima eficacia y, por tanto, constituyen hoy en día los fármacos de elección en presencia de lesiones de la mucosa esofágica demostradas endoscópicamente^{5,6}.

No toda ERGE tiene la misma gravedad ni causa el mismo impacto sobre la calidad de vida del paciente, y distinguir entre grados requiere contar con herramientas validadas, sencillas y útiles. Los instrumentos de medida de la calidad de vida específicos de los pacientes con ERGE permiten comprender mejor la repercusión de la enfermedad, estratificar a los pacientes según diferentes grados de intensidad sintomática y registrar los cambios conseguidos con el tratamiento. Una correcta evaluación y seguimiento de los pacientes con ERGE requiere la utilización de cuestionarios que permitan poner de manifiesto de forma

sencilla, fiable, reproducible y sensible los síntomas de la enfermedad y su repercusión en la percepción de la calidad de vida del paciente.

El Gastrointestinal Symptom Rating Scale (GSRS)⁷ es un cuestionario de 15 ítems que se responde mediante una escala tipo Likert de 7 puntos. El GSRS permite discriminar síntomas digestivos y ha demostrado ser de utilidad para discriminar síntomas de esofagitis, se correlaciona bien con la calidad de vida relacionada con la salud (CVRS) medida con el SF-36 y el Índice de Bienestar Psicológico⁸, es sensible a los cambios provocados por el tratamiento⁷ y es válido tanto si hay lesiones endoscópicas como en la ERGE sin esofagitis⁷. Sin embargo el GSRS carece de especificidad en la evaluación de síntomas de ERGE y su repercusión, y no permite hacer un seguimiento de los cambios evolutivos con el tratamiento. El Gastroesophageal Reflux Questionnaire (GERQ) es un cuestionario de 76 ítems que cuantifica la intensidad de diferentes síntomas de ERGE (pirosis, regurgitación, dolor, disfagia, etc.) de forma fiable y reproducible, aunque no se ha analizado su validez para determinar la respuesta al tratamiento del reflujo^{9,10}.

Recientemente se han desarrollado otros cuestionarios, como el Symptom Score de Allen et al¹¹ o el GERD Symptom Frequency Questionnaire (GSFQ) de Pare et al¹², que gradúan de forma objetiva los síntomas específicamente para la ERGE.

El GSFQ¹² es un cuestionario específico desarrollado para valorar la frecuencia de los síntomas de ERGE (3 ítems) y su influencia en la CVRS (3 ítems). Posee adecuadas características psicométricas, es sensible a los cambios inducidos por el tratamiento y se correlaciona significativamente con

la CVRS medida mediante cuestionarios genéricos. Está constituido por 6 ítems, 4 de los cuales se contestan mediante una escala ordinal. Los primeros 4 ítems exploran la frecuencia de síntomas de ERGE en los últimos 7 días. Las posibles respuestas son: todo el tiempo, la mayor parte del tiempo, a veces, pocas veces y nunca, que se puntuán de 4 a cero. Una de estas preguntas hace referencia a las limitaciones de la ERGE sobre la comida. Las dos últimas preguntas determinan el número de días de la semana anterior en los que los síntomas de ERGE interfirieron en las actividades de la vida diaria y en el sueño el paciente.

Sobre estas premisas de utilidad, se presenta mediante este estudio la validación al castellano de este cuestionario: GSFQ (Gastrointestinal Short Form Questionnaire).

Objetivo

El objetivo de este estudio es realizar la adaptación cultural al castellano y la validación clínimétrica del cuestionario GSFQ, para su uso en el entorno social y sanitario de habla española (España). También se realizará un estudio sobre la validez discriminante del instrumento para evaluar la capacidad de clasificación de este cuestionario y valorar la viabilidad de su utilización como herramienta de cribado.

Metodología

El presente estudio es de tipo transversal, no intervencionista, observacional epidemiológico y multicéntrico. La investigación se desarrolló según las siguientes fases: traducción, revisión, pilotaje, construcción de la versión adaptada, retrotraducción y estudios psicométricos. El proceso de adaptación cultural fue supervisado por un panel de expertos compuesto por un gastroenterólogo clínico, un experto en investigación clínica, un experto en investigación de resultados en salud y dos metodólogos.

En primer lugar el cuestionario fue traducido por dos traductores independientes. Ambas traducciones fueron conciliadas en una versión integradora por el panel de expertos, prestando especial atención a la propiedad clínica de los términos utilizados y el uso habitual de modismos en la práctica clínica. La versión candidata fue administrada a una muestra de 5 pacientes para valorar el grado de comprensión y los posibles problemas de interpretación, presentación o formato de respuesta. Los comentarios recogidos en la fase piloto fueron integrados en la versión definitiva del cuestionario, la cual fue utilizada en la fase de evaluación psicométrica y retrotraducida al inglés.

En la fase de evaluación psicométrica el cuestionario fue administrado a 4.000 pacientes, junto con el diagnóstico clínico de la gravedad de la ERGE, según la escala de Savary-Miller¹³, se recogieron en campo cerrado datos demográficos y de tratamiento para la ERGE previo, así como los siguientes factores de riesgo asociados a la enfermedad: hábito tabáquico, consumo habitual de bebidas carbonáticas y alcohólicas, y dieta alimenticia dominante (por posible consumo de picantes). También se registraron los datos sobre comorbilidades asociadas y el tratamiento concomitante en campo abierto, y se pasó el cuestionario de CVRS EuroQol-5D¹⁴ (EQ-5D), al que acompaña una escala visual analógica VAS.

En primer lugar se llevó a cabo un análisis de los ítems. Se estudió el número de respuestas dejadas en blanco (valores perdidos) para cada uno de los ítems como indicador de la pertinencia de cada uno de ellos (validez aparente). También se estudió la distribución de las respuestas a lo largo de las posibles categorías en busca de la posible presencia de asimetría y de efecto suelo o efecto techo. Se valoró la diferencia entre las medias y las varianzas de los ítems y se comprobó si se puede considerar que todas las medias son iguales mediante la prueba T^2 de Hotelling. La fiabilidad de la escala se estimó valorando la consistencia interna mediante el alfa de Cronbach¹⁵ y la estabilidad de las respuestas mediante el índice de correlación intraclasé (ICC)¹⁶ y el método de dos mitades, valorando la correlación entre mitades y estimando la fiabilidad total a partir de las dos partes mediante el coeficiente de Spearman-Brown¹⁷.

La validez estructural y la dimensionalidad del escalamiento se comprobaron mediante un análisis factorial exploratorio. El método de extracción utilizado fue el de componentes principales y se utilizaron dos métodos de rotación, uno ortogonal (Varimax) y otro oblicuo (Oblimin). Como regla de decisión para determinar el número de dimensiones se utilizaron el número de autovalores >1 (regla de Kaiser-Gutman^{18,19}), la prueba de sedimentación de Cattell²⁰ y el porcentaje de varianza explicada por la solución.

La validez concurrente se valoró comprobando la correlación con un cuestionario de CVRS genérico, el EQ-5D, tanto para su puntuación tarifa global²¹ como para la escala VAS que lo acompaña. Los autores del instrumento original definen que el concepto valorado por el GSFQ es la frecuencia de los síntomas de ERGE y su repercusión sobre la calidad de vida y, por ello, cabe esperar que presente cierta correlación con otros instrumentos de CVRS.

Para valorar la validez de constructo, se llevó a cabo la comparación en las puntuaciones promedio del GSFQ entre grupos de pacientes con distintos niveles de gravedad. Si el cuestionario es válido, cabe esperar que los pacientes con una mayor gravedad de su enfermedad obtengan puntuaciones sensiblemente mayores. Las diferencias entre los grupos son también una estimación de la diferencia clínicamente relevante²²⁻²⁴. La comparación de las medias de los distintos grupos de gravedad definidos según la escala de Savary-Miller se llevó a cabo utilizando un análisis de la varianza de un factor de grupos completamente al azar, y las comparaciones múltiples entre grupos se llevaron a cabo utilizando la prueba de Tukey.

También se compararon las puntuaciones promedio entre los pacientes con distintos factores de riesgo (tabaquismo, consumo de bebidas alcohólicas, consumo de bebidas carbonáticas y tipo de alimentación). Las comparaciones se realizaron tanto con la puntuación total en el GSFQ como con las puntuaciones tarifa y escala VAS del EQ-5D.

Las comorbilidades registradas por los clínicos fueron agrupadas siguiendo un procedimiento de reducción semántica, de manera que las enfermedades se agruparon en categorías clínicas cada vez más amplias hasta llegar a un número razonablemente reducido de enfermedades de etiología similar asociadas entre sí. Posteriormente, se compararon los grupos de antecedentes respecto a las puntuaciones en el GSFQ y las puntuaciones tarifa y VAS del EQ-5D. La frecuencia de las comorbilidades puede

considerarse una estimación de la prevalencia de éstas en este tipo de pacientes.

Las enfermedades concomitantes recogidas en campo abierto por los clínicos responsables fueron muy heterogéneas y se clasificaron en tres grandes grupos generales: antecedentes de ERGE, comorbilidades y otras enfermedades concomitantes. Se llevaron a cabo comparaciones entre los grupos de pacientes con algún tipo de antecedente y los pacientes de los que no se notificó ninguna enfermedad. Adicionalmente, se compararon los grupos de pacientes con comorbilidades específicas y aquellos de los que no se notificó ninguna comorbilidad o antecedente. Debe tenerse en cuenta que un paciente podía mostrar más de una comorbilidad.

La corrección del GSFQ se realiza mediante la suma de los puntos obtenidos en las preguntas, lo que permite obtener una puntuación lo que aumenta con la cantidad de molestias experimentadas por el paciente. Para obtener la puntuación, primero se debe recodificar las 4 primeras preguntas, asignando el valor 0 cuando el paciente no percibe nunca la molestia y el valor 4 cuando el paciente percibe la molestia todo el tiempo. La suma de las 4 primeras preguntas obtiene una puntuación de 0 a 16. Las 2 últimas preguntas se suman directamente a las anteriores de manera que, en cada pregunta, se asigna el valor 0 cuando no se ha experimentado molestias ningún día y un valor de 7 cuando se han experimentado molestias los 7 días de la semana. La puntuación total resultante oscila entre un mínimo de 0 y un máximo de 30 puntos. Para obtener una puntuación más fácil de interpretar, se divide por 0,3 para obtener una puntuación entre 0 (mínimas molestias) y 100 (máximas molestias).

Para valorar la capacidad predictiva del GSFQ a la hora de discriminar entre pacientes con y sin ERGE, se crearon dos grupos de pacientes, según los criterios de segmentación utilizados para seleccionarlos (ver más adelante), asignando a un grupo a los pacientes con un nivel 0 o I en la escala de Savary-Miller y en otro los pacientes con niveles II, III y IV. Utilizando este criterio, se estimó el punto de corte óptimo en la escala para discriminar entre pacientes con ERGE leve y paciente con ERGE grave. El punto de corte fue designado utilizando la curva característica de operación del receptor (COR) mediante el programa SPSS²⁵.

Todos los cálculos estadísticos se realizaron con el programa SPSS para Windows v.15.0, excepto el gráfico de malla que fue realizado con el programa Clementine v.10.0.

Determinación del tamaño muestral

La determinación del tamaño se realizó a partir de varios criterios acumulativos. En primer lugar, se tuvo en cuenta el número de preguntas del cuestionario y la representatividad de la solución factorial obtenible. Se recomienda un mínimo de 4-10 sujetos por pregunta^{26,27} y un mínimo de 3 preguntas por dimensión^{28,29}. Los autores originales asumen que el cuestionario es unidimensional, si bien es posible obtener dos puntuaciones relacionadas entre sí, una de frecuencia de los síntomas y otra de interferencia con la calidad de vida. Según estos valores, se infiere que es recomendable un mínimo de 60 pacientes (el máximo de los valores anteriormente mencionados). Sin embargo, otros autores

recomiendan que se compruebe la estructura factorial con un mínimo de 150-300 sujetos³⁰.

Por otra parte, con la finalidad de obtener baremos poblacionales de corrección, es necesario ampliar la muestra para asegurar la representatividad de la muestra normativa de comparación. Se estima conveniente asegurar la representatividad de la muestra respecto al sexo, respecto a dos grupos de edad (<45 y ≥ 45 años) y respecto a dos grupos de gravedad según la escala de Savary-Miller (0-I y \geq II). El cruce de los estratos da lugar a 8 tipos de pacientes, por lo que el clínico responsable debe seleccionar a un paciente de cada uno de los tipos.

Los pacientes fueron seleccionados por orden secuencial de demanda de tratamiento y cubriendo cada cuota de forma exclusiva. La escala se administró en una sola visita.

Los clínicos fueron captados de forma aleatoria y proporcional a la extensión y demanda de servicio de las comunidades autónomas del territorio español. El estudio fue propuesto a 510 gastroenterólogos, que debían incluir 8 pacientes, según su capacidad de reclutamiento en la práctica clínica habitual, lo que debía permitir asegurar un mínimo próximo a los 4.000 pacientes.

Sujetos y criterios de inclusión

La población en estudio son los pacientes con diagnóstico de ERGE que demandan atención especializada debido a problemas gastrointestinales asociados a una ERGE reciente o que hayan sido tratados con anterioridad sobre el padecimiento de ERGE, pero que no se encuentren en tratamiento, al menos, en el último mes. Se contemplaron los siguientes criterios de inclusión: ser mayor de 18 años, capaz de hablar y entender la lengua española y otorgar el consentimiento oral para participar en el estudio. Se consideraron los siguientes criterios de exclusión: encontrarse con una limitación física (minusvalía) que, a juicio de los investigadores participantes, no permitiera ser sometido a la administración de la escala, disponer de un nivel cultural que le incapacite para entender o responder a las preguntas de la escala, seguir tratamiento con fármacos, que a juicio del investigador participante, puedan interferir con su capacidad para entender o responder a las preguntas de la escala, sufrir un nivel de deterioro cognitivo que impida la comprensión de las preguntas, estar en tratamiento actual o con IBP durante el último mes o estar embarazada.

Todos los sujetos participantes en el estudio fueron informados del propósito de su participación y se les aseguró su derecho a renunciar a participar. Todos ellos firmaron el consentimiento informado en presencia de su responsable clínico. En todo momento se preservó la confidencialidad y la identidad de los sujetos durante el tratamiento de los datos.

Resultados

La muestra piloto estuvo formada por 4 mujeres y un hombre de edades comprendidas entre los 32 y los 86 años (media \pm desviación estándar [DE], $57,6 \pm 21,2$). El tiempo de administración osciló entre 3 y 10 min, con un tiempo promedio de 5 ± 2 min. Todos los pacientes respondieron a la

totalidad de las preguntas y sólo necesitó aclaraciones una mujer de 86 años que manifestó que resultaba fácil confundir la pregunta 1 con la 4.

En la muestra de valoración de las propiedades psicométricas se reclutaron 4.000 pacientes, si bien sólo 3.997 contaban con información relevante. El 50,5% fueron hombres, con una edad promedio de $49,4 \pm 14,4$ años, y edades comprendidas entre los 18 y los 91 años. El 9,7% no disponía de estudios, el 29,7% de estudios primarios, el 22,7% de estudios secundarios, el 16,3% de formación profesional y el 21,6% de estudios superiores. La estatura media fue de $167,7 \pm 8,9$ cm (rango, 136–224), el peso promedio de $73,49 \pm 12,5$ kg (rango, 34–150) y el índice de masa corporal promedio de $26,1 \pm 3,6$ (rango, 14,7–51,9).

El 41,8% de los pacientes se encontraba en tratamiento con IBP, el 14,4% en tratamiento con anti-H2 y el 22,7% con antiácidos.

El 41,7% manifestó ser fumador, el 47,9% consumía bebidas carbónicas, el 42,5% consumía bebidas alcohólicas y sólo un 0,3% manifestó seguir una dieta culinaria distinta de la española.

Análisis de los ítems

La tasa de respuestas perdidas en el cuestionario fue baja. Tan sólo un total de 88 (2%) sujetos no contestó a ninguna de las preguntas del cuestionario (o no lo entregó). Estos sujetos fueron descartados de análisis posteriores. De los 3.909 sujetos que entregaron el cuestionario, el 97,5% respondió a todas las preguntas, un 2% dejó en blanco una pregunta y los restantes dejaron en blanco entre 2 y 4 preguntas. La pregunta con mayor tasa de valores perdidos fue la referente a la sensación de ardor (pregunta 4) con un 0,4% de respuestas perdidas.

Las 4 preguntas tipo Likert (1 a 4) no presentaron efecto suelo ni efecto techo. En la **tabla 1** se muestra el porcentaje de respuestas acumulado en cada una de las categorías de respuesta, así como el porcentaje de respuestas en blanco.

Recordemos que las preguntas 5 y 6 constan de dos partes. En primer lugar, un filtro para saber si ha habido alteraciones de las actividades y, en segundo lugar, una valoración del grado de alteración. En lo que respecta a la alteración de las actividades diarias, el 60% de los pacientes manifestó no haber sufrido ninguna alteración, mientras que este valor se redujo al 46,9% en el caso de la ausencia de alteraciones durante el sueño.

Siguiendo el método de corrección propuesto por los autores originales, se asigna el valor 0 en el número de días afectados a los sujetos que manifiestan no haber sentido ninguna alteración (durante el día o durante la noche). Un total de 129 sujetos (3,2%) no contestó al número de días afectados. De ellos, 37 (28,7%) seleccionaron haber sufrido alguna alteración diaria pero no indicaron el número de días afectados, mientras que 92 (71,3%) no contestaron ni a la pregunta sobre el número de días afectados ni sobre el hecho de haber sufrido o no molestias diurnas. Por otra parte, 137 sujetos (3,4%) no contestaron al número de noches afectadas. De ellos, 46 (33,6%) seleccionaron haber sufrido alguna alteración nocturna pero no indicaron el número de noches afectadas, mientras que 91 (66,4%) no contestaron ni a la pregunta sobre el número de noches afectadas ni sobre el hecho de haber sufrido o no molestias nocturnas.

El promedio de días afectados por la sintomatología fue de $1,23 \pm 1,89$ (rango, 0–7) y el promedio de noches afectadas fue de $1,50 \pm 1,87$ (rango, 0–7). Las distribuciones de ambos ítems resultaron ser asimétricas positivas, con una desviación significativa de la normal ($p < 0,05$). En el caso del ítem 5, el 61,9% de los sujetos se encuentran en la categoría 0 (asimetría = 1,48; $et = 0,039$) y en ítem 6 el 47,4% (asimetría = 1,25; $et = 0,039$) (**tabla 2**).

La puntuación total promedio fue de $34,17 \pm 19,685$, con una puntuación mínima de 0 puntos y una puntuación máxima de 100 puntos, recorriendo la totalidad de la métrica del cuestionario. La mediana obtenida fue de 30 puntos. La distribución de la puntuación total presenta un ligero sesgo positivo (índice de asimetría = 0,60;

Tabla 1 Frecuencia y porcentaje de respuestas por cada categoría de respuesta de las 4 primeras preguntas del GSFQ

	Todo el tiempo	La mayor parte del tiempo	A veces	Pocas veces	Nunca	NC
Molestias en el abdomen						
n	114	1.180	1.706	679	229	1
%	2,9%	30,2%	43,6%	17,4%	5,9%	0%
Molestias tras el esternón						
n	79	872	1.662	866	423	7
%	2%	22,3%	42,5%	22,2%	10,8%	0,2%
Limitación de la comida normal						
n	155	796	1.493	912	548	5
%	4%	20,4%	38,2%	23,3%	14%	0,1%
Sensación de ardor ascendente						
n	89	975	1.597	810	423	15
%	2,3%	24,9%	40,9%	20,7%	10,8%	0,4%

NC: no contesta.

Tabla 2 Frecuencia y porcentaje de respuestas por cada categoría de respuesta de las 2 últimas preguntas del GSFQ

Número de días afectados	Días afectados		Noches afectadas	
	Recuento	%	Recuento	%
0	2.396	61,3%	1.831	46,8%
1	203	5,2	435	11,1
2	404	10,3	619	15,8
3	324	8,3	428	10,9
4	225	5,8	216	5,5
5	152	3,9	139	3,6
6	40	1	59	1,5
7	124	3,2	133	3,4
NC	41	1	49	1,3

NC: no contesta.

$\text{et} = 0,040$) y no se puede considerar normal (Kolmogorov-Smirnov = 0,094; $gl = 3811$; $p < 0,001$).

Fiabilidad

La fiabilidad de la escala adaptada es buena, con un coeficiente de consistencia interna alfa de Cronbach de 0,825 sin estandarizar y 0,864 estandarizado.

Una vez invertida la métrica de los 4 primeros ítems (para que una mayor puntuación indique una mayor gravedad), todos los ítems correlacionan de forma positiva con una correlación promedio de 0,515, oscilando entre una correlación mínima de 0,423 y una correlación máxima de 0,662.

Preservando la métrica original de los ítems (0 a 4 puntos para los ítems 1 a 4 y 0 a 7 para los ítems 5 y 6), la media de los ítems individuales osciló entre 1,23 puntos para el ítem 5 y 2,06 puntos para el ítem 1, con un valor promedio de las medias obtenidas de 1,708. La variabilidad de los ítems individuales osciló entre una varianza mínima de 0,821 para el ítem 1 y una varianza máxima de 3,57 para el ítem 5, con un valor promedio de 1,82. A partir de estos datos no se puede asumir que las medias de todos los ítems sean similares (T^2 de Hotelling = 1156,7; $F(5,3806) = 231$; $p < 0,001$).

El ICC arroja un valor de 0,825, con un intervalo de confianza (IC) del 95% comprendido entre 0,816 y 0,833. La correlación entre formas (ítems pares respecto a impares) fue de 0,689 y la fiabilidad estimada por el método de Spearman-Brown fue de 0,816.

Validez estructural y escalamiento

Sólo se obtuvo un autovalor mayor que la unidad ($\lambda_1 = 3,45$), correspondiente a la primera dimensión, la cual explicó el 57,5% de la variabilidad disponible. Este resultado sugiere que la escala es unidimensional. Sin embargo, el segundo autovalor se encuentra próximo a 1 ($\lambda_2 = 0,78$) y entre ambas dimensiones explicarían el 70,5% de la varianza disponible. La prueba de sedimentación

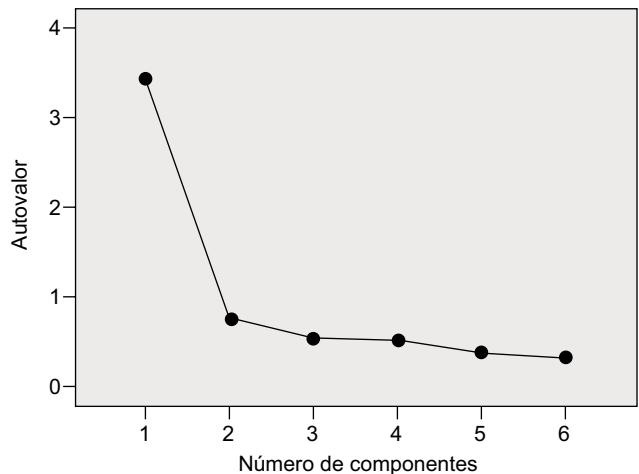


Figura 1 Prueba de sedimentación.

Tabla 3 Matrices de configuración (saturaciones) de las soluciones unifactorial y bifactorial (rotación Oblimin)

	Dimensión 1	Solución bifactorial	
		Dimensión 1	Dimensión 2
Molestias en el abdomen	0,814	0,876	-0,038
Molestias tras el esternón	0,823	0,894	-0,049
Limitación de la comida normal	0,751	0,693	0,110
Sensación de ardor ascendente	0,819	0,835	0,020
Limitación de las actividades diarias	0,683	0,107	0,758
Afectado el sueño	0,640	-0,063	0,915

sugiere también la presencia de dos autovalores. Por ello, se estudiaron las soluciones de 1 y 2 factores (fig. 1).

La solución de una sola dimensión permite la saturación elevada de todos los ítems en ella, con saturaciones por encima de 0,64, en valor absoluto (tabla 3). Los ítems correspondientes a la frecuencia de los síntomas saturan en menor medida que los ítems correspondientes a la ubicación e intensidad de los síntomas. Las comunidades resultantes de esta solución se encuentran entre 0,410 (ítem 6) y 0,677 (ítem 3), lo que indica que ésa es la proporción de varianza común de los ítems.

Dado que la regla K1 puede llegar a infraestimar el número de factores^{31,32}, se comprobó también una solución bidimensional. La solución bidimensional dio lugar a dos factores que separan los ítems de intensidad y ubicación de los ítems de frecuencia, aumentando la saturación de cada ítem en su dimensión correspondiente. Sin embargo la correlación entre las dos dimensiones es alta ($r_{xy} = 0,56$), lo

que permite asumir que en realidad se trata de una única dimensión, decisión que se ve también avalada por la alta fiabilidad de la escala (el estadístico α de Cronbach asume unidimensionalidad).

Validez concurrente

La correlación de la puntuación total del GSFQ con la puntuación global del EQ-5D fue moderada ($r_{xy} = 0,499$) y similar con la puntuación obtenida en la VAS de dicho instrumento ($r_{xy} = -0,481$), significativa en ambos casos ($p < 0,001$). Por otra parte, la correlación entre la puntuación del EQ-5D y la escala VAS fue sensiblemente superior ($r_{xy} = -0,674$; $p < 0,001$).

Validez de constructo

Comparando las puntuaciones promedio obtenidas en el GSFQ según el nivel de gravedad diagnosticado de ERGE, se aprecian diferencias significativas entre los grupos ($F[4, 3035] = 84,921$; $p < 0,005$). Los pacientes con grado 0 y 1 de gravedad de la ERGE obtuvieron puntuaciones promedio significativamente diferentes entre sí ($d = 3,03$; $et = 0,91$; $p = 0,008$) y más bajas que los restantes grupos de pacientes (grados 2, 3 y 4). Los grupos con nivel de gravedad 2 y 3 también difieren significativamente entre sí ($d = 7,14$; $et = 0,84$; $p < 0,001$). Sin embargo, el grupo de máximo grado de gravedad (grado 4) muestra una gran heterogeneidad y no difiere de los pacientes de grado 2 ($d = 1,52$; $et = 2,15$; $p = 0,954$) y grado 3 ($d = 5,61$; $et = 2,39$; $p = 0,135$). A partir de estos resultados, parece razonable establecer como diferencia clínicamente relevante una puntuación superior a 7,1, correspondiente a la diferencia entre los grupos de nivel de gravedad 2 y 3, dado que una diferencia menor no resulta significativa en la comparación del grupo de mayor gravedad (tabla 4 y fig. 2).

En la tabla 5 se muestran los estadísticos descriptivos (media, mínimo, máximo y DE) obtenidas por la muestra del estudio en las puntuaciones del GSFQ, puntuación tarifa del EQ-5D y escala VAS del EQ-5D, como referencia para la interpretación de las diferencias entre los grupos que presentan distintos antecedentes (tabla 6). El 41,7% de los pacientes presentó hábito tabáquico, el 47,9% consumía bebidas carbónicas, el 42,5% bebidas alcohólicas y un 4,5% seguía hábitos culinarios distintos de los nacionales. El GSFQ sólo detectó diferencias significativas entre el grupo con

hábito tabáquico y los pacientes no fumadores ($d = 3,67$; $et = 0,647$; $p < 0,001$). El grupo con hábito tabáquico también mostró diferencias significativas en las puntuaciones de utilidad ($d = 0,035$; $et = 0,006$; $p < 0,001$) y en la escala VAS del EQ-5D ($d = 1,81$; $et = 0,595$; $p = 0,002$). Por otra parte, también se detectaron diferencias significativas en las puntuaciones promedio de utilidad y el estado general de salud entre los grupo que consumen bebidas carbónicas y los que consumen alcohol respecto al resto de pacientes (tabla 5).

No se detectaron diferencias significativas entre los pacientes procedentes de culturas con distintos hábitos alimentarios en ninguna de las medidas. El 91,8% de los sujetos eran de procedencia europea, distribuyéndose los restantes entre las culturas gastronómicas centroamericanas (3,3%), norteafricana (0,7%) y otras procedencias (0,3%) (árabe, oriental, libanesa, rusa y sudamericana). En el 3,9% se desconoce la cultura de origen del sujeto.

En el 56,3% de los casos los clínicos responsables no han recogido ningún antecedente ni ninguna enfermedad concomitante en el CRD. Este grupo de pacientes es el que se utiliza como grupo comparación en las restantes enfermedades.

Tras la agrupación de enfermedades concomitantes, se consideraron posibles antecedentes de ERGE las siguientes: anorexia, hábitos alimentarios inadecuados, hernia de hiato, estrés, dispepsia y obesidad. En general,

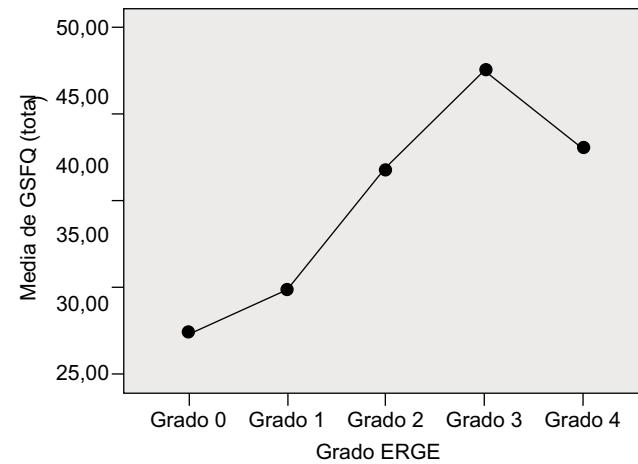


Figura 2 Puntuación total promedio del GSFQ por grado de gravedad de la enfermedad por reflujo gastroesofágico (ERGE).

Tabla 4 Estadísticos descriptivos para la puntuación total del GSFQ por grupo de gravedad (Savary-Miller)

	n	Media	DE	Error típico	IC del 95%		Mínimo	Máximo
					Límite inferior	Límite superior		
Grado 0	710	27,9484	18,32854	0,68786	26,5979	29,2988	0	100
Grado 1	881	30,9799	17,68226	0,59573	29,8107	32,1492	0	100
Grado 2	972	39,7188	18,25337	0,58548	38,5698	40,8677	0	100
Grado 3	319	46,8548	21,47023	1,20210	44,4897	49,2198	0	100
Grado 4	158	41,2447	26,02053	2,07008	37,1559	45,3335	0	100
Total	3040	35,2654	19,96787	0,36216	34,5553	35,9754	0	100

DE: desviación estándar; IC: intervalo de confianza para la media.

Tabla 5 Frecuencia de los antecedentes y significación de las diferencias respecto al grupo que no registra ninguna enfermedad en el GSFQ, utilidad del EQ-5D y escala VAS del EQ-5D

	Muestra (%)	Frecuencia	d GSFQ*	GSFQ	Utilidad	EQ 5D VAS
Tabaquismo	41,7	1665	3,67	<0,001	<0,001	0,002
Consumo de bebidas carbónicas	47,9	1916	-0,64	0,316	<0,001	<0,001
Consumo de bebidas alcohólicas	42,5	1697	0,41	0,524	<0,001	<0,001
Otros hábitos alimentarios	4,5	171	1,34	0,281	0,228	0,167

*Diferencia entre el grupo patológico y los pacientes que no refieren dicha enfermedad (antecedente-sin antecedente).

Tabla 6 Estadísticos descriptivos para las puntuaciones del GSFQ, utilidad del EQ-5D y escala VAS del EQ-5D

	n	Mínimo	Máximo	Media	DE
GSFQ (total)	3.811	0	100	34,1678	19,68513
Puntuación Tarifa (EQoL-VAS)	3.840	-0,08	1	0,8616	0,17907
Estado de salud actual (VAS)	3.744	0	100	68,63	17,993
n válido (según lista)	3.548				

DE: desviación estándar.

Tabla 7 Frecuencia de los antecedentes registrados y significación de las diferencias respecto al grupo que no registra ninguna patología en el GSFQ, utilidad del EQ-5D y escala VAS del EQ-5D

	Muestra (%)	Frecuencia	d GSFQ ^a	GSFQ	Utilidad	EQ 5D VAS
Anorexia	0,1	3	-14,11	0,187	0,006	0,010
Dispepsia	0,1	2	-10,22	0,819	0,861	0,840
Estrés	0,3	12	-8,28	0,123	0,032	0,290
Hábitos alimentarios inadecuados	0,2	8	-21,89	0,004	0,205	0,049
Hernia de hiato	0,7	29	-6,18	0,201	0,002	0,001
Obesidad	2,2	86	-4,50	0,040	0,001	<0,001
Ninguno ^b	56,3	2250	6,25	<0,001	<0,001	<0,001

^aDiferencia entre el grupo patológico y los pacientes que no refieren enfermedad (ninguno-enfermedad).

^bDiferencia entre el grupo que no refiere enfermedad y todos los que refieren alguna enfermedad (enfermedad-ninguno).

los antecedentes fueron poco frecuentes; sólo destacaba la obesidad mórbida, registrada en el 2,2% de los pacientes. El siguiente antecedente más frecuente fue la hernia de hiato, presente en el 0,7% de los pacientes (29 casos). El GSFQ mostró ser sensible y detectar la diferencia entre pacientes obesos y pacientes sin ninguna enfermedad ($d = -4,48$; $et = 2,19$; $p = 0,041$), y también mostró detectar diferencias significativas para el grupo de pacientes con hábitos alimentarios inadecuados ($d = -21,87$; $et = 7,58$; $p = 0,004$). En ambos casos la puntuación fue mayor para el grupo que presentaba el antecedente. En el caso de la hernia de hiato, la diferencia de puntuaciones entre grupos no alcanzó la significación ($d = -6,18$; $et = 4,71$; $p = 0,201$).

Sin embargo, sí se apreciaron diferencias significativas en la utilidad promedio del EQ-5D entre los pacientes sin enfermedad y los grupos de pacientes con anorexia ($d = 0,22$; $et = 0,08$; $p = 0,006$), estrés ($d = 0,087$; $et = 0,040$; $p = 0,033$), hernia de hiato ($d = 0,14$;

$et = 0,026$; $p = 0,002$) y obesidad ($d = 0,067$; $et = 0,016$; $p = 0,001$). Por su parte, la escala VAS del EQ-5D fue capaz de discriminar entre las siguientes enfermedades: anorexia ($d = 25,26$; $et = 9,75$; $p = 0,01$), hábitos alimentarios inadecuados ($d = 13,59$; $et = 8,46$; $p = 0,05$), hernia de hiato ($d = 11,16$; $et = 3,20$; $p = 0,001$) y obesidad ($d = 9,76$; $et = 1,98$; $p < 0,001$) (tabla 7).

De entre las enfermedades concomitantes recogidas por los investigadores clínicos responsables, se consideran comorbilidades asociadas a la ERGE las siguientes: ansiedad-depresión, artritis, artrosis, cirugía gástrica, diabetes mellitus, dolores musculares y osteoarticulares, enfermedades del hígado y vías biliares, esofagitis, gastritis crónica, hipertensión arterial (HTA), obstrucción esofágica, otros trastornos digestivos, síndrome de intestino irritable, trastornos respiratorios y úlcera (gástrica, esofágica o duodenal). Las comorbilidades más frecuentes son: HTA (15,8%), diabetes mellitus (5,9%), artrosis (4,7%), ansiedad-depresión (4,3%), trastornos respiratorios (3,3%) y dolores

musculares y osteoarticulares (2,2%). Los resultados para las comorbilidades de esofagitis y gastritis crónica deben considerarse con cautela por el reducido número de pacientes que las presentan.

Se detectaron diferencias significativas, en las puntuaciones promedio del GSFQ, entre los siguientes grupos de comorbilidades y los pacientes para los que no se registró ninguna enfermedad concomitante: ansiedad-depresión ($d = -12,07$; $et = 1,79$; $p < 0,001$), artritis ($d = -11,15$; $et = 4,38$; $p = 0,011$), artrosis ($d = -8,89$; $et = 1,44$; $p < 0,001$), diabetes mellitus ($d = -8,70$; $et = 1,32$; $p < 0,001$), dolores musculares y osteoarticulares ($d = -7,50$; $et = 2,13$; $p < 0,001$), HTA ($d = -6,67$; $et = 0,87$; $p < 0,001$), síndrome de intestino irritable ($d = -10,69$; $et = 3,74$; $p = 0,036$), trastornos respiratorios ($d = -7,93$; $et = 1,73$; $p < 0,001$) y úlcera gástrica ($d = -21,89$; $et = 6,56$; $p = 0,001$) (tabla 8).

Por su parte, se detectan diferencias significativas en la puntuación promedio de utilidad del EQ-5D para las siguientes comorbilidades: ansiedad-depresión, artritis, artrosis, diabetes mellitus, dolores musculares y osteoarticulares, HTA, otros trastornos digestivos, síndrome de intestino irritable, trastornos respiratorios y úlcera gástrica (tabla 8). Adicionalmente se aprecian diferencias significativas en las puntuaciones promedio de la escala VAS de CVRS del EQ-5D en la úlcera de esófago.

También se valoraron las enfermedades concomitantes comunicadas por los investigadores clínicos responsables y que no pudieron ser clasificadas como antecedentes o comorbilidades. Las enfermedades concomitantes restantes más frecuentes fueron: dislipemias (4,1%), hipercolesterolemia (3,2%), enfermedades vasculares (2,3%) y trastornos hormonales (1,3%).

Entre estos grupos de pacientes se detectaron diferencias significativas en el GSFQ en las siguientes enfermedades: alteraciones del ritmo cardíaco ($d = -6,43$; $et = 3,13$; $p < 0,001$), alteraciones del metabolismo ($d = -10,39$; $et = 3,35$; $p = 0,002$), carcinoma colorrectal ($d = -13,89$; $et = 5,88$; $p = 0,018$), cirugías ($d = -7,96$; $et = 3,53$; $p = 0,024$), dislipemia ($d = -3,84$; $et = 1,52$; $p = 0,012$), enfermedades vasculares ($d = -4,32$; $et = 2,06$; $p = 0,036$), fracturas y otros traumatismos ($d = -19,39$; $et = 9,29$; $p = 0,037$), hipercolesterolemia ($d = -8,06$; $et = 1,75$; $p < 0,001$), trastornos del metabolismo óseo ($d = -8,40$; $et = 2,86$; $p = 0,003$), trastornos hematopoyéticos ($d = -10,41$; $et = 4,39$; $p = 0,018$) y trastornos hormonales ($d = -9,67$; $et = 2,81$; $p = 0,008$) (tabla 9).

Por su parte, se apreciaron diferencias significativas en las puntuaciones promedio de utilidad obtenidas a partir del EQ-5D en las siguientes enfermedades: alteraciones del ritmo cardíaco, alteraciones del metabolismo, alteraciones neurológicas, cirugías, dislipemia, enfermedades autoinmunes, enfermedades del corazón, enfermedades vasculares, hipercolesterolemia, otras enfermedades neoplásicas, trastornos del metabolismo óseo, trastornos del riñón y vías urinarias, trastornos hematopoyéticos y trastornos psiquiátricos (tabla 9). También se apreciaron diferencias significativas en el valor promedio de la escala VAS de calidad de vida en las siguientes enfermedades: carcinoma colorrectal, enfermedades infecciosas y fracturas y otros traumatismos.

Es importante resaltar que algunas de las afecciones son de baja prevalencia y, por ello, la representatividad de los grupos de enfermedad correspondientes debe considerarse con cautela. Tampoco debemos olvidar la existencia de varias comorbilidades de maneras conjunta, por lo que algunos de los grupos pueden ser significativos por la

Tabla 8 Frecuencia de las comorbilidades registradas y significación de las diferencias respecto al grupo que no registra ninguna enfermedad en el GSFQ, utilidad del EQ-5D y escala VAS del EQ-5D

	Muestra (%)	Frecuencia	d GSFQ ^a	GSFQ	Utilidad	EQ 5D VAS
Ansiedad/depresión	4,3	173	-12,07	<0,001	<0,001	<0,001
Artritis	0,5	19	-11,15	0,011	0,007	0,012
Artrosis	4,7	189	-8,89	<0,001	<0,001	<0,001
Cirugía gástrica	0,1	2	-18,56	0,157	0,917	0,681
Diabetes mellitus	5,9	236	-8,70	<0,001	<0,001	<0,001
Dolores musculares y osteoarticulares	2,2	86	-7,50	<0,001	<0,001	0,010
Enfermedades del hígado y vías biliares	1,6	64	-3,78	0,208	0,002	0,001
Esofagitis	0,0	1	11,44	0,537	0,614	0,653
Gastritis crónica	0,1	2	1,47	0,912	0,031	0,840
Hipertensión arterial	15,8	631	-6,67	<0,001	<0,001	<0,001
Obstrucción esofágica	0,7	26	2,00	0,599	0,587	0,163
Otros trastornos digestivos	0,3	10	-3,89	0,508	0,047	0,005
Síndrome de intestino irritable	0,7	26	-10,69	0,036	0,002	0,033
Trastornos respiratorios	3,3	132	-7,93	<0,001	<0,001	<0,001
Úlcera duodenal	0,5	19	-4,39	0,345	0,572	0,590
Úlcera gástrica	0,2	8	-21,89	0,001	<0,001	<0,001
Úlcera de esófago	0,1	2	-6,87	0,600	0,539	0,012
Ninguno ^b	56,3	2250	6,25	<0,001	<0,001	<0,001

^aDiferencia entre el grupo patológico y los pacientes que no refieren enfermedad (ninguno-enfermedad).

^bDiferencia entre el grupo que no refiere enfermedad y todos los que refieren alguna (enfermedad-ninguno).

Tabla 9 Frecuencia de otras enfermedades registradas y significación de las diferencias respecto al grupo que no registra ninguna en el GSFQ, utilidad del EQ-5D y escala VAS del EQ-5D

	Muestra (%)	Frecuencia	d GSFQ ^a	GSFQ	Utilidad	EQ 5D VAS
Accidente cerebrovascular	0,3	13	-2,72	0,612	0,003	<0,001
Alergia	0,7	27	-7,79	0,129	0,161	0,379
Alteraciones del ritmo cardíaco	1,0	40	-6,43	0,040	<0,001	<0,001
Alteraciones del metabolismo	0,8	33	-10,39	0,002	0,001	0,017
Alteraciones neurológicas	1,0	39	-3,60	0,241	0,001	<0,001
Carcinoma colorrectal	0,3	10	-13,89	0,018	0,974	0,015
Cefaleas/migraña	0,6	23	-5,53	0,164	0,874	0,677
Cirugías	0,7	28	-7,96	0,024	0,023	0,043
Corticoides	0,1	3	8,11	0,449	0,331	0,552
Dislipemia	4,1	165	-3,84	0,012	<0,001	<0,001
Embarazo	0,0	1	-18,56	0,316	0,913	0,302
Enfermedad autoinmune	0,4	14	-4,20	0,415	0,016	0,086
Enfermedades del corazón	0,3	10	-11,15	0,072	0,020	<0,001
Enfermedades infecciosas	0,5	20	-0,72	0,862	0,181	0,024
Enfermedades vasculares	2,3	91	-4,32	0,036	<0,001	<0,001
Fracturas y otros traumatismos	0,1	4	-19,39	0,037	0,239	<0,001
Hipercolesterolemia	3,2	126	-8,06	<0,001	<0,001	<0,001
Otras enfermedades	0,5	18	-2,26	0,606	0,213	0,340
Otras enfermedades neoplásicas	2,0	79	-3,94	0,065	<0,001	<0,001
Trastornos de piel y anejos	0,1	5	6,78	0,414	0,651	0,544
Trastornos del metabolismo óseo	1,2	46	-8,40	0,003	<0,001	<0,001
Trastornos del riñón y vías urinarias	0,7	29	-5,10	0,156	0,006	0,001
Trastornos hematopoyéticos	0,5	20	-10,41	0,018	0,009	0,094
Trastornos hormonales	1,3	51	-9,67	0,008	0,001	0,001
Trastornos psiquiátricos	0,4	14	-9,58	0,208	0,021	<0,001
Ninguno ^b	56,3	2250	6,25	<0,001	<0,001	<0,001

^aDiferencia entre el grupo patológico y los pacientes que no refieren enfermedad (ninguno-enfermedad).

^bDiferencia entre el grupo que no refiere enfermedad y todos los que refieren alguna (enfermedad-ninguno).

inclusión de pacientes con enfermedades más incapacitantes respecto a las molestias gástricas.

En la figura 3 se muestra mediante enlaces las enfermedades concomitantes que se han comunicado en más de 20 pacientes. El grosor de los enlaces representa la frecuencia de la comorbilidad. Dada la alta prevalencia de HTA, no es de extrañar que los enlaces con esa afección sean los de mayor frecuencia. Las comorbilidades más frecuente son HTA y diabetes (97 pacientes), HTA y dislipemia (66), HTA y artrosis (61), HTA e hipercolesterolemia (49), HTA y trastornos respiratorios (31), HTA y enfermedades vasculares (28), HTA y obesidad (28), HTA y ansiedad/depresión (26), artrosis y diabetes (22), HTA y otras enfermedades neoplásicas (22) y dislipemia junto con diabetes (21).

La diferencia mínima significativa detectada entre grupos es de 3,84 puntos (en valor absoluto) y la diferencia mayor detectada es de 21,89 puntos, lo que puede dar una indicación del valor clínicamente relevante.

Validez predictiva

El área bajo la curva COR para las puntuaciones de la escala GSFQ respecto a la clasificación de los pacientes según el nivel de gravedad de ERGE, valorada por el clínico mediante la escala de Savary-Miller (niveles 0–I frente a II–IV) fue de

0,677 ($et = 0,01$; $p < 0,001$), generando un IC del 95% comprendido entre los valores 0,658 y 0,696. Estos valores permiten rechazar la hipótesis nula de un área bajo la curva igual a 0,5 (ausencia de discriminación). El valor de la escala que permite una mejor discriminación entre grupos en el punto de corte 35. Clasificando como no pacientes de ERGE a los que obtienen puntuaciones menores o iguales que este valor de corte y como pacientes con ERGE a los que obtengan una puntuación mayor, se alcanza una sensibilidad del 60,5% y una especificidad del 68,3%. La capacidad predictiva negativa para este punto de corte es del 89,9% y la capacidad predictiva positiva es del 24,1%.

Baremos

La puntuación total en el GSFQ sólo admite 31 puntuaciones distintas, por lo que es posible informar de los porcentajes acumulados de sujetos para cada una de las puntuaciones observadas. La tabla 10 muestra el porcentaje de sujetos que obtienen una determinada puntuación o menor. Por ejemplo, un paciente con una puntuación de 70 deja por debajo al 95% de los sujetos de la muestra normativa, de manera que un 70% de la población tiene una sintomatología de ERGE más leve (que interfiere menos en su calidad de

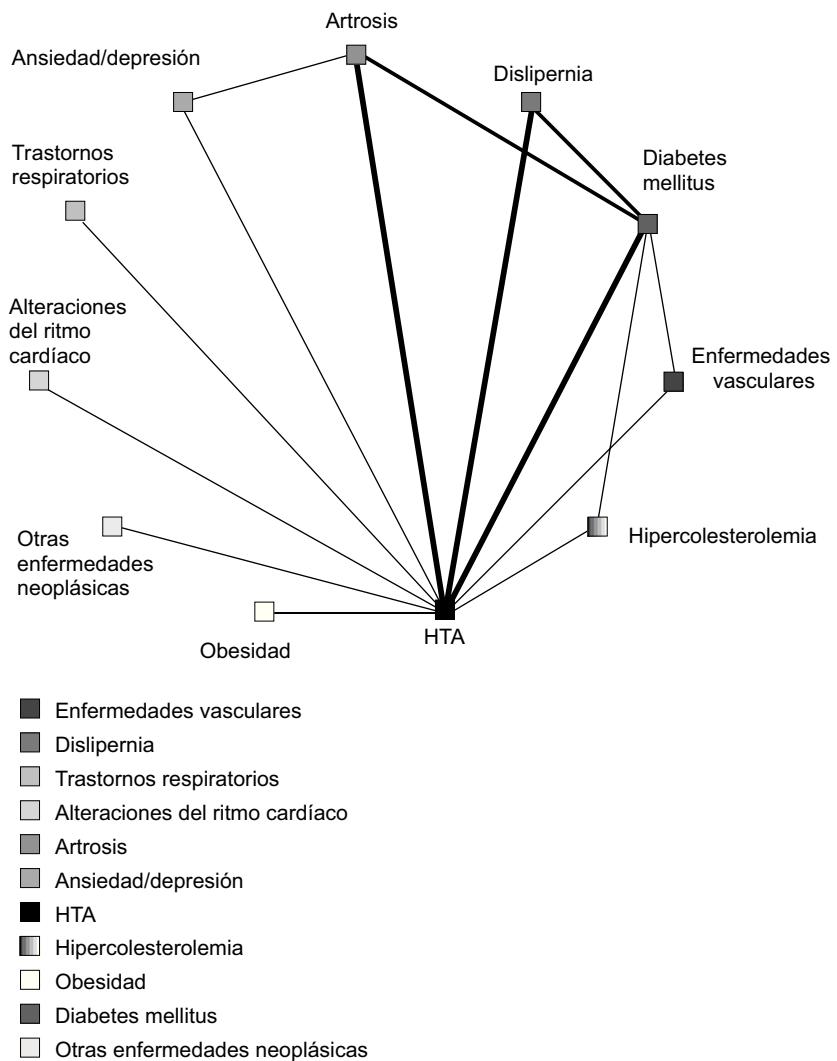


Figura 3 Gráfico de malla de las comorbilidades. La intensidad de las líneas representa la frecuencia conjunta de la asociación. HTA: hipertensión arterial.

vida) y un 30% de la población tiene una sintomatología que interfiere más con su calidad de vida.

Discusión

La muestra obtenida excede ampliamente el tamaño requerido para la validación cultural de un instrumento y, dado su gran tamaño y método de muestreo, se puede considerar una muestra representativa de la población nacional de pacientes que padecen ERGE. De hecho, dado su gran tamaño, pensamos que los resultados obtenidos respecto a la prevalencia de antecedentes y comorbilidades pueden considerarse una estimación de la prevalencia de dichas enfermedades.

El cuestionario resultó fácil de responder de manera generalizada y el número de respuestas en blanco fue muy reducido. Mientras que la distribución de las respuestas en las 4 preguntas medidas mediante escalas tipo Likert es adecuada y no muestran un sesgo apreciable, las dos últimas preguntas referentes a la frecuencia de los síntomas presentan un efecto suelo, acumulando más del 40% de las

respuestas en la categoría de ausencia de días afectados. Este hecho hace que la puntuación total también muestre un ligero sesgo positivo.

Dado que los 4 primeros ítems se encuentran en una escala de 0-4 puntos y los 2 últimos en una escala 0-7, no se puede asumir que todas las medias de los ítems sean iguales, y los últimos ítems influyen más en la puntuación final.

La fiabilidad obtenida es buena y supera el valor 0,8, tanto con la estimación mediante el alfa de Cronbach como con la estimación de test de longitud doble a partir de las dos mitades de Spearman-Brown³³. La estimación de estabilidad de las medidas es aceptable, ya que hay correlación entre ambas mitades. El IC del 95% para el ICC también se mantiene por encima del valor 0,8 en su totalidad.

A la vista de los resultados del AFE, el cuestionario puede considerarse unidimensional y no se estima pertinente la utilización de dos dimensiones que tan sólo darían lugar a una discriminación de la frecuencia de los síntomas, que además mostraría una alta correlación con la dimensión de influencia de la sintomatología sobre la CVRS.

Tabla 10 Frecuencia y centiles observados por puntuación total del GSFQ (escala 0–100)

Puntuación	Frecuencia	Centil (porcentaje acumulado)
0	91	2,4
3,33	79	4,5
6,67	95	7
10	143	10,7
13,33	218	16,4
16,67	219	22,2
20	253	28,8
23,33	280	36,2
26,67	321	44,6
30	234	50,7
33,33	241	57
36,67	217	62,7
40	201	68
43,33	199	73,2
46,67	174	77,8
50	146	81,6
53,33	114	84,6
56,67	104	87,4
60	92	89,8
63,33	75	91,7
66,67	70	93,6
70	53	95
73,33	55	96,4
76,67	29	97,2
80	34	98,1
83,33	25	98,7
86,67	29	99,5
90	6	99,6
93,33	7	99,8
96,67	1	99,8
100	6	100

La puntuación global en el GSFQ correlaciona de manera sensible con las medidas concurrentes de CVRS, pero dichas correlaciones son menores que las de las medidas genéricas de CVRS entre sí, lo que invita a pensar que el instrumento realmente mide aspectos específicos de la CVRS no recogidos con precisión en los instrumentos genéricos.

En lo que respecta a la validez de constructo, el cuestionario demuestra ser capaz de discriminar entre los grupos de pacientes con distinto nivel de gravedad, si bien el grupo de mayor gravedad se comporta de forma anómala, mostrando una gran heterogeneidad, y con valores en el cuestionario más bajos. Esto puede deberse a la efectividad de los tratamientos impuestos para las condiciones más graves de ERGE. El cuestionario también es capaz de detectar diferencias para los pacientes que presentan hábito tabáquico, pero no detecta diferencias respecto a los que manifiestan consumir alcohol o bebidas carbónicas. Un análisis más segmentado de las conductas alimenticias inadecuadas sí mostró la influencia del consumo de alcohol y bebidas carbónicas (resultados no mostrados aquí).

Los clínicos participantes en el estudio informaron con frecuencia de la existencia de enfermedades concomitantes

y, tras la agrupación conceptual de éstas, el GSFQ demostró ser capaz de detectar diferencias entre los pacientes con antecedentes de obesidad y hábitos alimentarios inadecuados, pero no respecto a la anorexia, hernia de hiato, estrés o dispepsia. El EQ-5D demostró ser más sensible a la presencia de estas enfermedades. El GSFQ ha mostrado ser sensible a la presencia de las comorbilidades siguientes: ansiedad-depresión (posiblemente por la mayor sensibilidad al malestar de estos pacientes), distintas enfermedades habitualmente tratadas con AINE (artritis, artrosis, dolores musculares y osteoarticulares), diabetes, HTA, síndrome de intestino irritable, trastornos respiratorios y úlcera gástrica. El instrumento también fue sensible a la presencia de otras enfermedades concomitantes de interés.

La escala GSFQ muestra tener capacidad para discriminar entre pacientes con ERGE leve y ERGE grave, con una sensibilidad y una especificidad aceptables. Sin embargo, la capacidad predictiva positiva es reducida. Este valor es mejorable si se pretende discriminar entre los pacientes con nivel 0 y los restantes (resultados no mostrados aquí), pero en detrimento de la capacidad predictiva negativa. Para estimar con mayor precisión las propiedades discriminativas de la escala, sería conveniente ampliar estos resultados con la aplicación de la escala a pacientes que no hayan demandado tratamiento específico por problemas de estómago.

Conclusión

La versión del cuestionario GSFQ adaptada al castellano presenta buenas propiedades métricas, se comporta de manera unidimensional y permite discriminar entre pacientes con distintos niveles de gravedad de ERGE diagnosticada, y también es sensible a la presencia de diversas enfermedades concomitantes.

Bibliografía

1. Winkelstein A. Peptic esophagitis: a new clinical entity. *JAMA*. 1935;104:906–9.
2. Allison PR. Peptic ulcer of the esophagus. *J Thorac Surg*. 1946;15:308–17.
3. Ing AJ, Ng MC, Breslin ABX. The pathogenesis of chronic persistent cough associated with gastroesophageal reflux. *Am J Respir Crit Care Med*. 1994;149:160–7.
4. Marzo M, Alonso P, Bonfill X, Fernández M, Fernández J, Martínez G, et al. Guía de práctica clínica sobre el manejo del paciente con enfermedad por reflujo gastroesofágico (ERGE). *Gastroenterol Hepatol*. 2002;25:85–110.
5. Hardin SM, Rchter JE, Guzzo MR, Schan CA, Alexander RW, Bradley LA. Asthma and gastroesophageal reflux: acid suppressive therapy improves asthma outcome. *Am J Med*. 1996; 100:395–405.
6. Kahrilas PJ. Gastroesophageal reflux disease. *JAMA*. 1996; 276:983–8.
7. Svedlund J, Sjödin I, Dotevall G. GSRS-A clinical rating scale for gastrointestinal symptoms in patients with irritable bowel syndrome and peptic ulcer disease. *Dig Dis Sci*. 1988;33:129–34.
8. Revicki DA, Wood M, Wiklund I. Reliability and validity of the gastrointestinal symptom rating scale in patients with gastroesophageal reflux disease. *Qual Life Res*. 1998;7:75–83.
9. Talley NJ, Fullerton S, Junghard O, Wiklund I. Quality of life in patients with endoscopy-negative heartburn: reliability and

- sensitivity of disease-specific instruments. *Am J Gastroenterol.* 2001;96:1998–2004.
10. Locke GR, Talley NJ, Weaver AL, Zinsmeister AR. A new questionnaire for gastroesophageal reflux disease. *Mayo Clin Proc.* 1994;69:539–47.
 11. Allen CJ, Parameswaran K, Belda J, Anvari M. Reproducibility, validity and responsiveness of a disease-specific symptom questionnaire for gastroesophageal reflux disease. *Dis Esophagus.* 2000;13:265–70.
 12. Pare P, Meyer F, Armstrong D, Pyzyk M, Pericak D, Goeree R. Validation of the GSFQ, a self-administered symptom frequency questionnaire for patients with gastroesophageal reflux disease. *Can J Gastroenterol.* 2003;17:307–12.
 13. Savary M, Miller G. The esophagus. Handbook and atlas of endoscopy. Geneva: Verlag Gassman; 1978. p. 135–42.
 14. Badia X, Roset M, Montserrat S, Herdman M, Segura A. La versión española del EuroQoL: descripción y aplicaciones. *Med Clin (Barc).* 1999;112(Suppl 1):79–86.
 15. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
 16. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psych Meth.* 1996;1:30–46.
 17. Kristof W. The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. *Psychometrika.* 1963;28:221–38.
 18. Kaiser HF. The application of electronic computers to factor analysis. *Educational and Psychological Measurement.* 1960;20: 141–51.
 19. Kaiser HF. An index of factorial simplicity. *Psychometrika.* 1974;39:31–6.
 20. Cattell RB. The Scree test for the number of factors. *Multivariate Behav Res.* 1996;1:245–76.
 21. Badia X, Roset M, Herdman M, et al. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making.* 2001;21:7–16.
 22. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials.* 1989;10:407–15.
 23. 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.
 24. Hays RD, Woolley JM. The concept of clinically meaningful difference in Health-related Quality of Life Research. *Pharmacoconomics.* 2000;18:419–23.
 25. Pardo A, Ruiz MA. SPSS 11. Guía para el análisis de datos. Madrid: McGraw-Hill; 2002.
 26. Cattell RB. Factor analysis. An introduction and manual for the psychologist and social scientist. New York: Harper and Row; 1952.
 27. Everitt BS. Multivariate analysis: the need for data, and other problems. *Br J Psychiatr.* 1975;126:237–40.
 28. Anderson TW, Rubin H. Statistical inference in factor analysis. En: Neyman J, editor. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley: University of California Press; 1956.p. 111–50.
 29. Wiley DE. The identification problem for structural equation models with unmeasured variables. En: Goldberger AS, Duncan OD, editors. *Structural equation models in the social sciences.* New York: Seminar; 1973. p. 69–83.
 30. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. *Psychological Methods.* 1999;4:84–99.
 31. Ruiz MA, San Martín R. Una simulación sobre el comportamiento de la regla K1 en la estimación del número de factores. *Psicothema.* 1992;4:543–50.
 32. Zwick WR, Velicer WF. A comparison of five rules for determining the number of components to retain. *Psych Bull.* 1986;3: 432–42.
 33. Nunnally JC, Bernstein IH. *Psychometric theory.* 3rd ed. New York: McGraw-Hill; 1994.