

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

TESIS DOCTORAL

TRADUCCIÓN, ADAPTACIÓN Y ANÁLISIS DE VALIDEZ Y FIABILIDAD DE LA HOPKINS SYMPTOM CHECKLIST-25 EN ESPAÑOL PARA LA EVALUACIÓN DE LA DEPRESIÓN EN ATENCIÓN PRIMARIA

Doctoranda

María Rodríguez Barragán

Directoras de la tesis doctoral

Dra. María Isabel Fernández San Martín

Dra. Ana Clavería Fontán

Dra. Eva Peguero Rodríguez

Tutor de la tesis doctoral

Dr. Miquel Porta Serra

Programa de Doctorado en Metodología de la Investigación Biomédica y Salud Pública

UAB
Universitat Autònoma
de Barcelona

Barcelona, 2023

Dra. María Isabel Fernández San Martín, técnica de salud de la Unitat Docent Atenció Familiar i Comunitària Barcelona ciutat de la Gerència Territorial d'Atenció Primària de Barcelona del Institut Català de la Salut y de la Unitat de Suport a la Recerca de la Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol),

Dra. Ana Clavería Fontán, técnica de salud del Instituto de Investigación Sanitaria Galicia Sur del Área de Salud de Vigo del Servicio Gallego de Salud. Miembro de la Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS), y

Dra. Eva Peguero Rodríguez, médica de familia y comunitaria en el Equipo de Atención Primaria El Castell del Institut Català de la Salut, profesora asociada de la Facultad de Medicina de la Universidad de Barcelona del Campus Bellvitge, miembro del grupo de investigación en vértigo de la Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol) y del grupo de estudios humanísticos en ciencia y tecnología GEHUCT de la Universidad Autònoma de Barcelona.

CERTIFICAN,

Que la tesis doctoral titulada "Traducción, adaptación y análisis de validez y fiabilidad de la Hopkins Symptom Checklist-25 en español para la evaluación de la depresión en Atención Primaria" presentada por **María Rodríguez Barragán** y codirigida por ellas reúne los méritos para ser presentada y defendida delante del tribunal correspondiente para optar al grado de Doctora.

Y para que así conste, firman este documento en Barcelona, el 19 de septiembre de 2023.

Dra. María Isabel Fernández San Martín Dra. Ana Clavería Fontán Dra. Eva Peguero Rodríguez

AGRADECIMIENTOS

Esta tesis doctoral empezó con una propuesta en un congreso de la semFYC cuando yo era residente de cuarto año. Nunca olvidaré cuando la Dra. Maribel Fernández me preguntó si me interesaría añadirme a un proyecto con el que podría hacer el doctorado. Mi primer agradecimiento es para ella, por aquel ofrecimiento y por creer que yo podría llevarlo a cabo. El segundo es para la Dra. Ana Clavería, a quien en aquel momento no conocía, pero que fiándose del criterio de Maribel aceptó que me incorporase al equipo. Y el tercero es para la Dra. Eva Peguero, en la codirección de esta tesis doctoral no podía faltar una médica de familia, y nadie mejor que ella por su proximidad y practicidad. A mis directoras de tesis doctoral les agradezco el acompañamiento y la dedicación, ha sido un placer empezar a adentrarme en el mundo de la investigación de su mano. Maribel, Ana y Eva son tres mujeres a las que admiro por su inteligencia, generosidad, saber hacer y resolución. Sin duda ellas han sido las mejores personas en las que he podido confiar para ir avanzando y dejarme llevar en este camino de aprendizaje.

Mi agradecimiento al grupo Family Depression and Multimorbidity (FPDM) del European General Practice Research Network (EGPRN), en especial a los doctores Patrice Nabbe y Jean-Yves Le Reste, por acogerme en su grupo e incluirme en su proyecto. Es una suerte formar parte de un grupo de investigación internacional y poder aprender de todo su conocimiento y experiencia.

A la Unitat de Suport a la Recerca (USR) de la Gerencia Territorial de Barcelona, por brindarme un espacio, compañeros y compañeras de los que aprender, y por ofrecerme un inicio de trayectoria de investigación que espero que dure muchos años. En especial mi agradecimiento al director de la USR el Dr. Miguel Ángel Muñoz por su tiempo, sus buenos consejos y por confiar en mí. También a Sandra Sáez por facilitar siempre las cosas y hacer más agradables mis días de investigación. Estaré siempre agradecida al Dr. Juantxo Mendive, por permitirme y animarme a compaginar la investigación con la asistencia desde que terminé la residencia de Medicina de Familia y Comunitaria en el Centro de Atención Primaria de La Mina, en el que él era el director en ese momento.

A título personal, sin duda el primer gran agradecimiento es a mis padres, Nieves y Andrés, por enseñarme desde pequeña el valor de la perseverancia y el placer por el estudio. Si he sido capaz de escribir una tesis doctoral es gracias a ellos. Por su ejemplo de vida y su generosidad gratuita. Por su apoyo y acompañamiento en esta tesis doctoral y en todos los pasos de mi vida.

A mi marido y compañero de vida Marc, sin el que todo esto habría sido imposible. Le agradezco todo su apoyo incondicional, su paciencia infinita, así como sus palabras reconfortantes al final del día. Le agradezco hacer equipo siempre, hacerlo todo fácil sin quejas ni reproches, con su sonrisa inapagable dibujada siempre en su cara.

A mi hija Laura, por ser una todoterreno con tan solo un año, agradezco de corazón su capacidad de adaptarse a todo y la tranquilidad de saber que ella está bien mientras yo estoy redactando este documento. Su manera de ser alegre, sociable y cariñosa han salvado este verano en el que hemos pasado varios días sin vernos. A partir de ahora vendrán muchos momentos juntas.

A mi familia, especialmente a mis hermanos Tomás y Marta y a mi cuñada Amaya. Los incondicionales, los únicos capaces de hacerme reír en los días de mayor estrés. La estética bonita de esta tesis doctoral es gracias a ellos. A la tía, por todo. También a mis suegros, Yolanda y Jordi, y a mi cuñada Marta. Por su apoyo emocional y logístico y por su comprensión.

A todos mis compañeros de camino que me inspiran y hacen que ame la Atención Primaria. A mis compañeros y compañeras de trabajo, en especial a Albert Brau y Rosa Fernández, amigos y confidentes, compañeros de meriendas y reflexiones, el lugar al que siempre puedo acudir. También a Isabel Bobé, Adrià Giralt, Marta Castelló y la Dra. Núria Soldevila, compartir el día a día con ellos me hace ir más feliz al trabajo. A Nani Vall-Llosera y Blanca de Gispert, ha sido una suerte pasar tantas horas sentada en el ordenador de al lado, un regalo inesperado que me llevo de mis horas en la UR. A Jaume Martín y Antonia Raya, con los que he compartido tantas horas amenas de investigación, buenos ratos y conversaciones interesantes. A Judit Pertíñez, por tantos cafés, buenos consejos y reflexiones conjuntas en los días de investigación. Al Dr. Xavier Mundet, por su apoyo, proximidad y confianza. A todas mis coerres, por hacerme avanzar en la reconstrucción de mis propias ideas, porque ellas hacen que quiera ser mejor médica de familia constantemente, porque son soporte y fuente de cariño y fuerza siempre.

A mis amigos y amigas, tengo la suerte de tener tantos que sería imposible nombrarlos a todos. Saber que están ahí y poder contar con ellos es la suerte de mi vida.

Finalmente, un agradecimiento sincero a los investigadores del proyecto EIRA, por aceptar incluir el presente estudio dentro de su proyecto. A los médicos y médicas de familia que colaboraron en este estudio de forma generosa. Y por supuesto a los y las pacientes que participaron respondiendo a la HSCL-25 y a la multitud de preguntas que se les realizaron durante la recogida de datos. Ellos son el motivo de nuestro trabajo e investigación.

Financiación y becas

Este proyecto de tesis doctoral ha obtenido la siguientes becas y financiación para su realización:

Beca	Organismo que la concede
Beca Xb para la Realización de Tesis Doctorales para especialistas de reciente titulación en Medicina y Enfermería Familiar y Comunitaria, 5a convocatoria	Gerencia Territorial de Barcelona a través de la Unidad Docente Multiprofesional Barcelona Ciudad, Instituto Catalán de la Salud (ICS).
19a Beca para la Capacitación en Investigación y Realización del Doctorado en Atención Primaria	Instituto Universitario de Investigación en Atención Primaria Jordi Gol i Gurina (IDIAP Jordi Gol) y el Instituto Catalán de la Salud (ICS).
Ayuda para la realización de tesis doctorales Isabel Fernández 2019	Sociedad Española de Medicina Familiar y Comunitaria (semFYC).
5ª Beca para la Realización de Tesis Doctorales para médicos en periodo de formación o especialistas de reciente titulación en Medicina Familiar y Comunitaria	Cátedra UAB-Novartis juntamente con el Instituto Universitario de Investigación en Atención Primaria Jordi Gol i Gurina (IDIAP Jordi Gol).

La financiación otorgada no ha tenido ningún rol en el diseño del estudio o en la recogida de datos, en el análisis o interpretación de resultados, en la escritura ni en la decisión del proceso de publicación de los manuscritos que conforman esta tesis doctoral.

Índices

I.I Índice de contenido

1	Resumen/Abstract	17
2	Introducción	23
2.1	Investigación en Atención Primaria	23
2.1.1	European General Practice Research Network	24
2.2	La depresión	24
2.2.1	Diagnóstico de la depresión	25
2.2.2	Situación epidemiológica de la depresión.....	29
2.2.3	Evaluación y cribado de la depresión	32
2.3	Hopkins Symptom Checklist	38
2.3.1	Hopkins Symptom Checklist-25	38
2.3.2	Hopkins Symptom Checklist-10 y Hopkins Symptom Checklist-5	41
2.4	Entrevista Diagnóstica Internacional Compuesta	42
2.5	Psicometría y uso de escalas en el ámbito de la salud.....	44
2.5.1	Datos referidos por los pacientes (Patient Reported Outcomes, PRO)	45
2.5.2	Traducción y adaptación de escalas	47
2.5.3	Taxonomía	52
2.5.4	Otros aspectos de los estudios de validez y fiabilidad de un PROM	63
3	Justificación	69
4	Hipótesis y objetivos	73
4.1	Hipótesis.....	73
4.1.1	Objetivo general	73
4.1.2	Objetivos específicos	73
5	Métodos	77
5.1	FASE 1: Traducción y adaptación transcultural de la HSCL-25 al español.....	77
5.1.1	Diseño y ámbito.....	77
5.1.2	Participantes y muestra.....	77
5.1.3	Recogida de datos	78
5.1.4	Análisis estadístico.....	79
5.2	FASE 2: Análisis de la validez y fiabilidad de la HSCL-25 en español.....	80
5.2.1	Diseño	80
5.2.2	Periodo de realización	80
5.2.3	Selección de la población	80
5.2.4	Recogida de datos	82
5.2.5	Análisis estadístico.....	84
5.3	Consideraciones éticas en la Fase 1 y la Fase 2.....	87

6	Resultados.....	91
6.1	Artículo 1.....	93
6.2	Artículo 2.....	103
6.3	Artículo 3.....	121
7	Discusión.....	139
7.1	Resumen de los resultados principales.....	139
7.1.1	Artículo 1.....	139
7.1.2	Artículo 2.....	140
7.1.3	Artículo 3.....	141
7.2	Discusión por objetivos y comparación con otros artículos.....	142
7.3	Fortalezas.....	160
7.4	Limitaciones.....	161
7.5	Implicaciones de los resultados y líneas futuras.....	162
8	Conclusiones.....	171
9	Bibliografía.....	175
10	Anexos.....	197

I.II Índice de Anexos

ANEXO 1	Grupo FPDM del EGPRN	198
ANEXO 2	Ítems en las diferentes versiones de la Hopkins Symptom Checklist (HSCL)	199
ANEXO 3	Taxonomía COSMIN en inglés.....	200
ANEXO 4	Artículo original.....	201
ANEXO 5	Versión original, traducción directa y retrotraducción de la HSCL-25	215
ANEXO 6	Versión HSCL-25 en español utilizada en la Fase 2.....	220
ANEXO 7	Composite International Diagnostic Interview (CIDI)	221
ANEXO 8	Formación a los entrevistadores de la CIDI	227
ANEXO 9	PHQ-9.....	229

I.III Índice de tablas

Tabla 1.	Criterios diagnósticos del trastorno depresivo mayor según el DSM-5	26
Tabla 2.	Clasificación de los episodios depresivos según la Clasificación Internacional de Enfermedades – 10ª revisión (CIE-10)	28
Tabla 3.	Ítems y dimensión correspondiente de la Hopkins Symptom Checklist-25 (HSCL-25)	39
Tabla 4.	Valores de sensibilidad, especificidad, valores predictivo positivo y negativo de la HSCL-25 comparada con una entrevista psiquiátrica en estudios europeos realizados en Atención Primaria	41
Tabla 5.	Parámetros estadísticos de los distintos niveles de medición para el <i>gold standard</i> y el instrumento de medición objeto de estudio	54
Tabla 6.	Tipos de parámetros más frecuentes en el análisis de la fiabilidad	60

I.IV Índice de figuras

Figura 1.	Representación gráfica de un modelo reflectivo (a) y un modelo formativo (b).	46
Figura 2.	Fases del proceso de traducción y adaptación de un PROM.....	51
Figura 3.	Curvas ROC y su correspondiente AUC.....	55
Figura 4.	Representación gráfica de la estructura de una escala con 1 factor (a) y con 2 factores correlacionados (b).	56
Figura 5.	Representación gráfica de la validez y la fiabilidad	59

I.V Palabras clave

Atención Primaria de Salud; Médicos de Familia; Depresión; Trastorno Depresivo; Encuestas y Cuestionarios; Psicometría; Medición de Resultados Informados por el Paciente.

Primary Health Care; Family Physicians; Depression; Depressive Disorder; Surveys and Questionnaires; Psychometrics; Patient Reported Outcome Measures.

I.VI Abreviaturas

ADAMHA	Alcohol, Drug Abuse, and Mental Health Administration (USA organization)
AP	Atención Primaria
APA	American Psychiatric Association
AUC	Area Under Curve
AVAD	Años de Vida Ajustados por Discapacidad
BDI	Beck Depression Inventory
CAP	Centro de Atención Primaria
CD	Coefficient of Determination
CES-D	Center for Epidemiologic Studies-Depression
CESD-R	Center for Epidemiologic Studies Depression Scale-Revised
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CIDI	Composite International Diagnostic Interview
CIE-10	Clasificación Internacional de Enfermedades
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
CS	Centro de Salud
CSMA	Centros de Salud Mental de Adultos
DASS	Depression, Anxiety and Stress Scales
DSM	Diagnostic and Statistical Manual of Mental Disorders
EFA	Exploratory Factor Analysis
EGPRN	European General Practice Research Network
EIRA	Evidence in Research and Action network
ESEMeD	Estudio Epidemiológico sobre trastornos Mentales en España
FPDM	Family Depression and Multimorbidity (Grupo de EGPRN)
GDS	Geriatric Depression Scale
HADS	Hospital Anxiety Depression Scale
HAM-D	ver HSRD
HSCL	Hopkins Symptom Checklist
HSRD	Hamilton Rating Scale for Depression
ICC	Intraclass Correlation Coefficient
ICS	Institut Català de la Salut
ID	Índice de Discriminación
IDIAP	Instituto Universitario de Investigación en Atención Primaria
IRT	Item Response Theory
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MADRS	Montgomery-Asberg Depression Rating Scale
MF	Médicos y médicas de familia
MFyC	Medicina de Familia y Comunitaria
MLM	Maximum Likelihood Mean Adjusted
NA	No Aplicable
NICE	National Institute for health and Care Excellence (UK organization)
OMS	Organización Mundial de la Salud
PHQ	Patient Health Questionnaire
PRO	Patient Reported Outcomes
PROM	Patient Reported Outcome Measures
PSC	Physical Symptom Checklist
rediAPP	Red de Investigación en Actividades Preventivas y de Promoción de la Salud

RICAPPS	Red de Investigación en Cronicidad, Atención Primaria y Prevención y Promoción de la Salud
RMSEA	Root Mean Square Error of Approximation
ROC	Receiver Operating Curve
SCAN	Schedules for Clinical Assessment in Neuropsychiatry
SCL	Symptom Checklist
semFYC	Sociedad Española de Medicina Familiar y Comunitaria
SRMR	Standardized Root Mean Square Residual
STARD	Standards for Reporting Diagnostic accuracy studies
TLI	Tucker-Lewis Fit Index
UAB	Universitat Autònoma de Barcelona
UCLA	University of California-Los Angeles
USPSTF	United States Preventive Services Task Force
WONCA	World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians

1

RESUMEN/ABSTRACT

1 Resumen/Abstract

RESUMEN

La depresión es una patología frecuente entre la población adulta con un impacto negativo en la vida de quienes la padecen. Su prevalencia ha aumentado en los últimos años y en España constituye la quinta causa de años de vida ajustados por discapacidad. La mayor parte de las personas con depresión consultan a su médico/a de familia y reciben seguimiento y tratamiento en Atención Primaria (AP). Su diagnóstico es clínico y se basa en los criterios del DSM-5, sin embargo, existen múltiples herramientas de evaluación de la depresión que han demostrado validez y fiabilidad para el cribado de la depresión y para la evaluación de sus síntomas. La Hopkins Symptom Checklist-25 (HSCL-25) es una herramienta con buenas propiedades psicométricas y fue seleccionada por el grupo *Family Practice Depression and Multimorbidity* del *European General Practice Research Network* por su sencillez, por ser autoadministrada por el propio paciente, y por la facilidad de interpretación de sus puntuaciones para su uso en AP con fines clínicos y de investigación. El objetivo de esta tesis doctoral es realizar la traducción y adaptación transcultural de la HSCL-25 al español y analizar la validez y la fiabilidad de su puntuación en población adulta del ámbito de la AP. También se analizan la validez y la fiabilidad de las versiones breves HSCL-10 y HSCL-5 como objetivo secundario. Este estudio se enmarca en un proyecto europeo del grupo mencionado en el que participaron diferentes países realizando simultáneamente la traducción a nueve idiomas con el mismo objetivo. Se realizó la traducción directa, consenso con metodología Delphi en el que participaron médicos/as de familia, la retrotraducción, la clasificación de equivalencias, la armonización europea y la entrevista cognitiva a pacientes reales de AP. Se obtuvo una versión definitiva de la HSCL-25 en español. Posteriormente se diseñó un estudio transversal multicéntrico con la participación de 6 Centros de Salud para analizar la validez y la fiabilidad. Se incluyeron pacientes participantes en el proyecto EIRA de entre 45 y 75 años con dos o más hábitos no saludables, quienes respondieron también al *Patient Health Questionnaire* (PHQ9) y a la *Composite International Diagnostic Interview* (CIDI), considerada *gold standard* del diagnóstico de la depresión en este estudio. Se administró la HSCL-25 a 790 pacientes con una tasa de respuesta del 97,2%. No se encontraron patrones de no respuesta. El ítem más consistente en todos los análisis y versiones fue el 17 “Se siente triste”. El análisis factorial confirmatorio mostró mejor ajuste para el modelo de dos factores correlacionados (ansiedad y depresión) que para un solo factor. El área bajo la curva (AUC) fue de 0,89 (0,86-0,93) en la HSCL-25; 0,88 (0,84–0,92) en la HSCL-10 y 0,85 (0,81–0,89)

en la HSCL-5. Para el punto de corte de 1,75, la sensibilidad fue 88,1% (77,1-95,1%) y la especificidad 76,7% (73,3-79,8%) en la HSCL-25. Para el punto de corte de 1,85, la sensibilidad fue 79,7% (67,7-88,0%) y la especificidad 83% (80,0-85,7%) en la HSCL-10. Para el punto de corte de 2,00 la sensibilidad fue 78,0% (65,9-86,6%) y la especificidad 72,8% (69,3-76,0%) en la HSCL-5. El alfa de Cronbach fue 0,92 en la HSCL-25; 0,84 en la HSCL-10 y 0,77 en la HSCL-5. La fiabilidad test-retest analizada con 94 pacientes obtuvo un coeficiente de correlación intraclassa de 0,92 (0,87-0,95). La HSCL-25 en español tiene un alto porcentaje de respuestas, su puntuación presenta validez y fiabilidad adecuadas y es bien aceptada por los pacientes de AP. La HSCL-10 y la HSCL-5 también son adecuadas y pueden ser más factibles por su brevedad sin renunciar en exceso a la validez y la fiabilidad.

ABSTRACT

Depression is a common pathology among the adult population, with a negative impact on the lives of those who suffer from it. Its prevalence has increased in the past years and in Spain it is the fifth cause of disability-adjusted life years (DALYs). Most people with depression consult their family doctor and receive follow-up and treatment in Primary Care (PC). Its diagnosis is clinical and is based on the DSM-5 criteria. However, there are multiple depression assessment tools that have shown their validity and reliability as screening for depression and for the assessment of its symptoms. The Hopkins Symptom Checklist-25 (HSCL-25) is a tool with good psychometric properties and was selected by the Family Practice Depression and Multimorbidity group of the European General Practice Research Network because of its simplicity, being self-administered by the patient, and ease of score interpretation for use in PC for clinical and research purposes. The objective of this thesis is to translate and cross-culturally adapt the HSCL-25 into Spanish and analyze the validity and reliability of its score in the adult population in the PC scope. The validity and reliability of the short versions HSCL-10 and HSCL-5 are also analyzed as a secondary objective. This study is part of a European project of the aforementioned group in which different countries participated doing the translations simultaneously into nine languages with the same objective. Forward translation, subsequent consensus of a group of family doctors with Delphi methodology, back translation, equivalence classification, European harmonization and cognitive interview with real PC patients were carried out. A final version of the HSCL-25 was obtained in Spanish. Subsequently, a multicenter cross-sectional study was designed with the participation of 6 Health Centers to analyze validity and reliability. Patients participating in the EIRA project were included, aged between 45 and 75, with two or more unhealthy habits, who also answered the Patient Health Questionnaire (PHQ9) and the Composite International Diagnostic Interview (CIDI), taken as the gold standard for depression diagnosis in this study. HSCL-25 was administered to 790 patients with a response rate of 97.2%. No non-response patterns were found. The most consistent item in all analysis and versions was item 17 "Feeling blue". The confirmatory factor analysis showed better fit for the model of two correlated factors (anxiety and depression) than for the single factor model. The area under the curve (AUC) was 0.89 (0.86-0.93) in HSCL-25; 0.88 (0.84-0.92) in HSCL-10 and 0.85 (0.81-0.89) in HSCL-5. For the cut-off point of 1.75, sensitivity was 88.1% (77.1-95.1%) and specificity 76.7% (73.3-79.8%) in HSCL-25. For the cut-off point of 1.85, sensitivity was 79.7% (67.7-88.0%) and specificity 83% (80.0-85.7%) in HSCL-10. For the cut-off point of 2.00, the sensitivity was 78.0% (65.9-86.6%) and the specificity was 72.8% (69.3-76.0%) in HSCL-5. Cronbach's alpha was 0.92 in HSCL-25; 0.84 in HSCL-10 and 0.77 in HSCL-5. The test-retest reliability analyzed with 94

patients resulted in an intraclass correlation coefficient of 0.92 (0.87-0.95). The Spanish version of HSCL-25 has a high response rate, its score presents adequate validity and reliability, and is well accepted by PC patients. HSCL-10 and HSCL-5 are also suitable and may be more feasible due to their shortness without sacrificing too much validity and reliability.

2

INTRODUCCIÓN

2 Introducción

2.1 Investigación en Atención Primaria

La Atención Primaria (AP) es el escenario idóneo para el estudio^{1,2} de aspectos relacionados con la oportunidad de observar la enfermedad desde sus estadios más precoces³, especialmente en aquellas más prevalentes. La AP constituye un área de conocimiento propio⁴ que requiere proyectos de investigación en este ámbito por diversos motivos⁵:

1. La atención clínica y los cuidados preventivos dispensados en AP deben basarse en la evidencia obtenida de la investigación⁶.
2. La mayor parte de la atención a la salud de la población se proporciona desde la AP⁷.
3. La evidencia para el manejo de las patologías crónicas y la mayoría de los cuidados en salud debe apoyarse en la investigación clínica de base poblacional realizada en AP y no únicamente por investigaciones llevadas a cabo desde hospitales o por laboratorios⁸.
4. La ausencia de investigación en AP conduce al sobrediagnóstico, a tratamientos inapropiados y a retrasos diagnósticos⁹.

Los médicos y médicas de familia (MF) dedican gran parte de su atención a las patologías de mayor prevalencia y en estadios precoces de enfermedad¹⁰, son expertos en el manejo de la complejidad, la cronicidad y la incertidumbre; combinan el arte de la anamnesis, la exploración física y el conocimiento profundo de las patologías con las habilidades comunicativas que ayudan a comprender la vivencia individual de la enfermedad y a cuidar la relación médico/a-paciente, tan importante en este ámbito. Durante su jornada laboral, los MF se ayudan de diferentes herramientas y guías basadas en la evidencia científica para realizar diagnósticos, prescribir o desprescribir tratamientos farmacológicos, aconsejar estudios y actuaciones en cada paciente, derivar a otros especialistas, etc. Por este motivo, la investigación en el ámbito de la AP es esencial, ya que sirve de apoyo a la toma de decisiones⁴ y contribuye a la mejora de los servicios y de la salud de la población a la que atiende.

Se han creado diferentes redes, sociedades y fundaciones con la finalidad de promover y gestionar la innovación, la formación, la docencia y la investigación clínica, epidemiológica y en servicios de la salud en el ámbito de la AP. En España destaca especialmente la Red de Investigación en Actividades Preventivas y de Promoción de la Salud¹¹ (redIAPP), que fue fundada en el año 2003 y ha desarrollado diversos proyectos colaborativos y multicéntricos.

Sigue haciéndolo en la actualidad reorganizada como Red de Investigación en Cronicidad, Atención Primaria y Prevención y Promoción de la Salud (RICAPPS)¹².

Como MF, es importante dedicar tiempo, ilusión y empeño a la investigación en AP, ya que hemos elegido una profesión comprometida con la sociedad y con las personas y a ellas debemos nuestro esfuerzo¹³.

2.1.1 European General Practice Research Network

European General Practice Research Network (EGPRN)¹⁴ es una organización europea de MF y otros profesionales de la salud relacionados con la investigación en AP y la Medicina de Familia y Comunitaria (MFyC)¹⁴ creada en el año 1971, que pertenece a la World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians (WONCA)¹⁵. Su objetivo es proporcionar un espacio para compartir y desarrollar líneas de investigación en este ámbito, así como promover estudios internacionales e intercambios de experiencias. Pretende desarrollar una base científica válida para la AP. Sus congresos y jornadas ofrecen la oportunidad de conocer profesionales de otros países y la posibilidad de realizar colaboraciones en investigación.

Dentro de la organización hay diferentes grupos de trabajo entre los que se encuentra el grupo Family Depression and Multimorbidity (FPDM) creado en el año 2011, compuesto por profesionales de distintos países que realizan proyectos colaborativos sobre los conceptos de multimorbilidad y depresión¹⁶. El estudio que se presenta en esta tesis doctoral se enmarca en el proyecto de investigación sobre depresión que desarrolla el citado grupo. Así pues, se trata de un proyecto internacional en el que han participado investigadores de varias nacionalidades europeas, como se detallará más adelante. Los investigadores participantes se pueden consultar en el ANEXO 1.

2.2 La depresión

Los trastornos psiquiátricos son comunes en la práctica médica y pueden presentarse como un trastorno primario o como una comorbilidad asociada a otras patologías¹⁷. Los trastornos del estado de ánimo se caracterizan por una alteración en la regulación del estado de ánimo, el comportamiento y el afecto y se subdividen en 1) trastornos depresivos, 2) trastornos bipolares (depresión con episodios maníacos o hipomaniacos) y 3) depresión en asociación con

enfermedades médicas o abuso de alcohol y toxicomanía¹⁷. Los trastornos depresivos incluyen los siguientes¹⁸:

- Trastorno de desregulación disruptiva del estado de ánimo.
- Trastorno de depresión mayor (incluye el episodio depresivo mayor y los episodios depresivos recurrentes).
- Trastorno depresivo persistente (distimia).
- Trastorno disfórico premenstrual.
- Trastorno depresivo inducido por una sustancia/medicamento.
- Trastorno depresivo debido a otra afección médica.
- Otro trastorno depresivo especificado.
- Otro trastorno depresivo no especificado.

Entre la variedad de síntomas del trastorno depresivo mayor destacan el estado de ánimo deprimido, la pérdida de placer (anhedonia) o el desinterés por actividades durante largos periodos de tiempo¹⁹, como se detallará en el siguiente apartado.

2.2.1 Diagnóstico de la depresión

La sospecha clínica de depresión aparece con la propia manifestación del paciente o por la observación de su expresividad no verbal por parte del facultativo²⁰. En este sentido, la posición del MF es privilegiada, ya que el conocimiento profundo del paciente permite detectar estos signos incluso cuando los individuos acuden a la consulta como acompañantes o cuando consultan por otros motivos. El diagnóstico de la depresión es fundamentalmente clínico y se realiza a partir de la anamnesis basándose en unos criterios diagnósticos definidos. La entrevista clínica es clave en el diagnóstico de la depresión^{21,22} por lo que las habilidades comunicativas del profesional, especialmente la escucha activa²³, son esenciales.

Los criterios diagnósticos que más se utilizan, tanto en clínica como en investigación, son los de la clasificación de la Asociación Americana de Psiquiatría, conocida por sus siglas en inglés como APA (*American Psychiatric Association*) y los de la Clasificación Internacional de Enfermedades, actualmente según su décima revisión (CIE-10)²⁴. La Sociedad Americana de Psiquiatría publicó en el año 2013 la quinta edición del “Manual diagnóstico y estadístico de los trastornos mentales” (en inglés *Diagnostic and Statistical Manual of Mental Disorders*), abreviado como DSM-5¹⁸. En esta última edición se han reordenado algunos conceptos respecto a la versión previa del DSM-IV-TR^{25,26}. El DSM-5 es el manual de referencia para los diagnósticos del campo

de la salud mental ya que define los criterios diagnósticos para cada una de las patologías. La versión actual incluye la armonización con la CIE-10.

2.2.1.1 Criterios diagnósticos del trastorno de depresión mayor

El trastorno depresivo mayor se define como la presencia de tristeza o anhedonia durante más de dos semanas con afectación del funcionamiento de la persona y la suma de otros síntomas hasta sumar un mínimo de 5, según se describe en la Tabla 1.

Tabla 1. Criterios diagnósticos del trastorno depresivo mayor según el DSM-5

A. Cinco (o más) de los síntomas siguientes han estado presentes durante el mismo período de dos semanas y representan un cambio del funcionamiento previo; al menos uno de los síntomas es (1) estado de ánimo deprimido o (2) pérdida de interés o de placer.	
	1. Estado de ánimo deprimido la mayor parte del día, casi todos los días, según se desprende de la información subjetiva (p.ej., se siente triste, vacío, sin esperanza) o de la observación por parte de otras personas (p. ej., se le ve lloroso).
	2. Disminución importante del interés o el placer por todas o casi todas las actividades la mayor parte del día, casi todos los días (como se desprende de la información subjetiva o de la observación).
	3. Pérdida importante de peso sin hacer dieta o aumento de peso (p. ej., modificación de más de un 5% del peso corporal en un mes) o disminución o aumento del apetito casi todos los días.
	4. Insomnio o hipersomnia casi todos los días.
	5. Agitación o retraso psicomotor casi todos los días (observable por parte de otros, no simplemente la sensación subjetiva de inquietud o de enlentecimiento).
	6. Fatiga o pérdida de energía casi todos los días.
	7. Sentimiento de inutilidad o culpabilidad excesiva o inapropiada (que puede ser delirante) casi todos los días (no simplemente el autorreproche o culpa por estar enfermo).
	8. Disminución de la capacidad para pensar o concentrarse, o para tomar decisiones, casi todos los días (a partir de la información subjetiva o de la observación por parte de otras personas).
	9. Pensamientos de muerte recurrentes (no sólo miedo a morir), ideas suicidas recurrentes sin un plan determinado, intento de suicidio o un plan específico para llevarlo a cabo.
B. Los síntomas causan malestar clínicamente significativo o deterioro en lo social, laboral u otras áreas importantes del funcionamiento.	
C. El episodio no se puede atribuir a los efectos fisiológicos de una sustancia o de otra afección médica.	
D. El episodio de depresión mayor no se explica mejor por un trastorno esquizoafectivo, esquizofrenia, un trastorno esquizofreniforme, un trastorno delirante, u otro trastorno especificado o no especificado del espectro de la esquizofrenia y otros trastornos psicóticos.	
E. Nunca ha habido un episodio maníaco o hipomaníaco.	

Tomado de: "Trastorno de depresión mayor. Criterios diagnósticos." por CIBERSAM, 2014, *Manual Diagnóstico y Estadístico de los trastornos mentales. DSM-5*, p. 160-161, Copyright © 2014 Asociación Americana de Psiquiatría (APA).

En la depresión mayor, el núcleo ideativo incluye la pérdida de autoestima y sensación de culpa inapropiada, así como ideas de suicidio²⁰, y puede ser leve, moderada o grave, con o sin síntomas psicóticos. Se puede realizar un diagnóstico basado en un único episodio, aunque en la mayoría de los casos el trastorno suele ser recurrente¹⁸. La duración media de los episodios de depresión mayor es de 16 semanas²⁷.

En los episodios depresivos el rasgo principal es el ánimo depresivo o la pérdida de interés o del placer en casi todas las actividades (criterio A)^{18,19}; la capacidad para disfrutar, interesarse y concentrarse está reducida²⁸. El sueño suele estar alterado en forma de insomnio o hipersomnia²⁹ y el apetito suele estar disminuido²⁸, aunque también puede haber aumento de la ingesta y comportar una ganancia de peso¹⁸. Los pacientes deprimidos suelen referir pérdida de energía, cansancio y fatiga; asimismo, la autoestima y la confianza en uno mismo suelen verse mermadas y son frecuentes las ideas de culpa, inferioridad o inutilidad^{28,29}. Son frecuentes los pensamientos de muerte, la ideación suicida o los intentos autolíticos; esto incluye desde pensamientos pasivos como la creencia de que los demás estarían mejor si el paciente estuviera muerto hasta pensamientos recurrentes de cometer suicidio¹⁸. Todos estos síntomas pueden acompañarse de los denominados síntomas "somáticos", como son: despertar precoz varias horas antes de la hora habitual, empeoramiento matutino del humor depresivo, enlentecimiento psicomotor importante, agitación, pérdida de apetito, pérdida de peso y disminución de la libido²⁸.

En función del número y severidad de los síntomas, un episodio depresivo puede especificarse como leve (malestar manejable y poco deterioro funcional), moderado (número de síntomas e intensidad entre "leve y grave") o grave (gran número de síntomas, con malestar no manejable y disfunción sociolaboral)²⁸.

Frecuentemente se percibe el trastorno depresivo asociado a ansiedad con diversas combinaciones sintomáticas en sus manifestaciones^{21,22}. La ansiedad en este contexto se define como la presencia de dos o más de los síntomas siguientes durante la mayoría de los días de un episodio de depresión mayor¹⁸:

1. Se siente nervioso/a o tenso/a.
2. Se siente inhabitualmente inquieto/a.
3. Dificultad para concentrarse debido a las preocupaciones.
4. Miedo a que le pueda suceder algo terrible.
5. El individuo siente que podría perder el control de sí mismo/a.

El diagnóstico diferencial de la depresión incluye el duelo, la demencia, la enfermedad de Parkinson, los periodos de tristeza, el hipotiroidismo, el síndrome tóxico, un proceso cerebral orgánico, fármacos y drogodependencias, así como otros trastornos psiquiátricos como el trastorno adaptativo, la psicosis, el trastorno obsesivo-compulsivo y especialmente el trastorno bipolar^{20,30}. En este último caso, el diagnóstico diferencial puede resultar complicado, dado que la fase de depresión del trastorno bipolar se puede presentar como un episodio depresivo³¹.

2.2.1.2 Clasificación de los trastornos depresivos según la CIE-10

Los trastornos mentales quedan recogidos en la CIE-10²⁴ con el código que empieza por la letra F seguidos del número correspondiente. Los episodios depresivos se clasifican en el capítulo V “Trastornos mentales y del comportamiento”, apartado 3 “Trastornos del humor (afectivos)”²⁸. Reciben el código alfanumérico F32 y engloban los diagnósticos expuestos en la Tabla 2. Puede utilizarse un quinto carácter para especificar la presencia o ausencia del síndrome somático.

Tabla 2. Clasificación de los episodios depresivos según la Clasificación Internacional de Enfermedades – 10ª revisión (CIE-10)

F.32	Episodios depresivos
F32.0	Episodio depresivo leve
F32.1	Episodio depresivo moderado
F32.2	Episodio depresivo grave sin síntomas psicóticos
F32.3	Episodio depresivo grave con síntomas psicóticos
F32.4	Episodio depresivo en remisión parcial
F32.5	Episodio depresivo en remisión completa
F32.8	Otros episodios depresivos
F32.9	Episodio depresivo, no especificado

Fuente: adaptado de “Clasificación Internacional de Enfermedades - 10.ª Revisión. Modificación Clínica” por *Ministerio de Sanidad, Servicios Sociales e Igualdad*. 2018.

Esta clasificación incluye episodios aislados de:

- Reacción depresiva.
- Depresión psicógena.
- Depresión reactiva.

Excluye:

- Trastorno de adaptación (F43.2).
- Trastorno depresivo recurrente (F33.-).
- Cuando se asocia a trastorno de la conducta en F91.- (F92.0).

Cada uno de los diagnósticos incluidos dentro de la categoría “F32 Episodios depresivos” tiene sus propios criterios diagnósticos. El subdiagnóstico “F32.8 Otros episodios depresivos” incluye la depresión atípica y episodios aislados de depresión enmascarada sin especificación. El apartado “F32.9 Episodio depresivo sin especificación” agrupa el resto de episodios como la depresión sin especificación y el trastorno depresivo sin especificación²⁸.

El trastorno depresivo recurrente (F33) se puede considerar cuando existe un intervalo mínimo de dos meses consecutivos entre los episodios, durante el cual no se cumplen los criterios para un episodio de depresión mayor¹⁸. En la versión DSM-5 el trastorno depresivo persistente (distimia) (F34.1) agrupa el trastorno de depresión mayor crónico y el trastorno distímico del DSM-IV.

2.2.2 Situación epidemiológica de la depresión

La depresión y la ansiedad son las patologías más frecuentes de la esfera mental y la prevalencia de ambas ha aumentado en los últimos 30 años, siendo mayor entre las mujeres³². A escala mundial, aproximadamente 280 millones de personas sufren trastorno depresivo mayor o depresión³³. Según la Organización Mundial de la Salud (OMS), la prevalencia mundial de depresión se sitúa alrededor del 5-6%^{33,34}, aunque hay una gran variabilidad entre los estudios en función de los países y los grupos de edad³⁵. Las diferencias en las prevalencias y en los factores de riesgo asociados reportadas en los diferentes estudios e informes pueden deberse a las múltiples codificaciones utilizadas por los profesionales sanitarios en los sistemas de información³⁶, así como al uso de diferentes herramientas utilizadas en el proceso del diagnóstico de la depresión.

Estudios realizados con población de AP en Europa reportan prevalencias más elevadas que varían entre el 9,2% y el 20,2%^{29,37-40}, siendo mayor en mujeres mayores de 50 años, asociada a condiciones de viudedad, soledad y aislamiento social⁴¹.

En el informe anual del Sistema Nacional de Salud (SNS) de España de 2020-2021⁷ la prevalencia registrada de trastornos mentales fue de 286,7 casos por cada 1000 habitantes en la población española, más elevada en mujeres que en hombres (313,3 frente a 258,8). El 5,4% de la población mayor de 15 años declaró padecer un cuadro depresivo, el doble en mujeres (7,1%) que en hombres (3,5%). Tanto su prevalencia como su severidad aumentan con la edad y más del 10% de las personas mayores de 75 años sufren algún cuadro depresivo. Según el estudio ESEMeD sobre prevalencia de los trastornos mentales y factores asociados en España, el episodio depresivo mayor es el trastorno mental más frecuente, con una prevalencia a lo largo

de la vida del 10,5%⁴². En los datos que se obtienen de la Encuesta Nacional de Salud de 2017⁴³, la prevalencia de depresión en España es de 10,2% en el grupo de personas mayores de 45 años.

Se han descrito diferentes factores de riesgo para el desarrollo de depresión, como son el género femenino, los episodios de depresión previos, niveles elevados de estrés e historia de episodio depresivo mayor en familiares de primer grado¹⁸. También tienen más probabilidades de sufrir depresión las personas que han vivido situaciones personales de trauma, como abusos, pérdidas graves u otros eventos estresantes³³ y las personas que viven en zonas urbanas frente a las que viven en zonas rurales⁴⁴.

La depresión se asocia frecuentemente a otras enfermedades orgánicas, especialmente con trastornos respiratorios como la enfermedad pulmonar obstructiva crónica⁴⁵ o el dolor crónico⁴⁶. Se trata de una relación recíproca, ya que los problemas físicos pueden exacerbar la depresión y a su vez la depresión puede afectar negativamente a la evolución de la patología física^{21,22}.

Por otro lado, la mala salud mental se asocia a un empleo precario, a peor nivel socioeconómico, así como a situaciones estresantes como procesos de desahucio⁴¹. Dentro del papel atribuido a los determinantes sociales como factor de riesgo de depresión se ha destacado el bajo nivel de recursos económicos y las circunstancias laborales^{21,22,44}. Las personas en desempleo y en situación de baja laboral presentan episodios depresivos con mayor frecuencia⁴⁷; la depresión es 2,5 veces más frecuente entre quienes se encuentran en situación de desempleo (7,9%) que en quienes trabajan (3,1%), y alcanza el 30% entre las personas incapacitadas para trabajar⁴³. El riesgo de depresión sigue un gradiente socioeconómico de forma inversamente proporcional⁴⁸, especialmente cuando el trastorno es crónico⁴⁹. El estado civil que implica no vivir en pareja (soltero/a, divorciado/a o viudo/a) así como tener un nivel de estudios básico o inferior también se relacionan con una mayor probabilidad de desarrollar depresión^{21,22,43}.

Las mujeres presentan una probabilidad cercana al doble de padecer depresión respecto a los hombres^{27,32,33,41,50}. También presentan de forma más frecuente ansiedad y trastornos de la conducta alimentaria³², mientras que en los hombres son más frecuentes los trastornos por abuso de sustancias⁴⁸. En todo el mundo, más del 10% de las embarazadas y de las mujeres en situación de postparto experimentan depresión⁵¹. Varios estudios han demostrado que la falta de empoderamiento en las mujeres interactúa con otras adversidades como la pobreza, la violencia de género, el acoso sexual y la inseguridad alimentaria para aumentar la prevalencia de trastornos mentales en las mujeres^{48,52}.

La depresión tiene un gran impacto en la salud y en la vida de quienes la padecen, puede ser incapacitante y tener un efecto negativo en el funcionamiento y bienestar de las personas⁴¹. A nivel mundial, los trastornos mentales han pasado de ser la decimotercera causa principal de años de vida ajustados por discapacidad (AVAD) en 1990 a la séptima en 2019³². Los trastornos depresivos se sitúan entre las 10 primeras posiciones de AVAD en todos los grupos de edad según los últimos datos de 2019 recogidos en *The Global Burden of Disease*⁵³ y representan la mayor proporción de AVAD de todos los trastornos mentales con un 37,3%, seguidos de los trastornos de ansiedad (22,9%) y la esquizofrenia (12,2%)³². De las 25 principales causas de años vividos con discapacidad (AVD) en 2019, los trastornos depresivos ocuparon el segundo lugar, los trastornos de ansiedad ocuparon el octavo y la esquizofrenia ocupó el puesto 20. Dentro de los trastornos mentales, los trastornos depresivos ocuparon el primer lugar en todos los grupos de edad, excepto en la categoría de 0 a 14 años, para la cual el trastorno de conducta fue la principal causa de AVAD³².

En España, actualmente la depresión se sitúa como quinta causa de AVAD por detrás de la cardiopatía isquémica, la enfermedad de Alzheimer y otras demencias, la Diabetes Mellitus y el cáncer de pulmón, bronquios y/o tráquea⁵⁴. Al realizar este análisis por género, la depresión se sitúa como segunda causa de AVAD en las mujeres sin tener en cuenta rangos de edad, por detrás de la enfermedad de Alzheimer y otras demencias, y como décima causa en los hombres.

Los trastornos depresivos tienen un gran impacto en la calidad de vida, aumentan la morbilidad y la mortalidad^{34,55}. La depresión está asociada a un incremento de la mortalidad global⁴¹, especialmente asociada a enfermedades cardiovasculares como el infarto y al suicidio^{21,22,56}. Cada año se suicidan más de 700000 personas en el mundo³³, Europa es la región de la OMS con la mayor incidencia de suicidio⁵⁷. El suicidio se sitúa entre las 10 principales causas de pérdida de años de vida estandarizados por edad en toda Europa, siendo las tasas masculinas superiores a las de las mujeres en todos los países del mundo y para todos los grupos de edad, excepto en el grupo de 15 a 19 años⁵⁸. En el año 2020 en España se registraron 3941 defunciones por suicidio (2930 en hombres y 1011 en mujeres)⁷. Es bien conocida la asociación entre depresión y suicidio⁵⁷; la ideación autolítica está presente en el 29% de pacientes con depresión mayor⁵⁹ y hasta un 4% de ellos cometen el suicidio²⁰. Los pacientes con depresión presentan un riesgo de suicidio 4 veces mayor que la población general^{21,22}, que puede aumentar hasta ser 20 veces mayor en el caso de la depresión grave⁶⁰. El principal factor de riesgo de suicidio descrito son los antecedentes de intentos autolíticos o las amenazas de que el sujeto se suicidará, otros factores de riesgo son ser varón, no vivir acompañado y los sentimientos de desesperanza¹⁸.

Desde hace años se reconoce que existe estigma asociado a las enfermedades de la esfera mental⁶¹, especialmente en pacientes con ansiedad y depresión⁶². Hasta el 27% de la población muestra actitudes estigmatizadoras hacia la depresión y el 46,9% las perciben en los demás (estigma percibido)⁶¹. El estigma internalizado o autoestigma se refiere al grado en que las personas con un trastorno depresivo internalizan las creencias negativas (estigmatizantes) de la sociedad, hecho que puede provocar vergüenza, menor autoestima y sensación de fracaso personal⁶³. En este sentido, los pacientes que sufren depresión se encuentran con el problema añadido del estigma, que suele ser permanente y refractario a la intervención, empeorando su calidad de vida⁶⁴.

El estigma provoca la creencia errónea de que el paciente con depresión debe recuperarse sin ayuda dado que su situación se debe a una debilidad personal⁶⁵ y puede conducir a posponer el contacto con los profesionales y a un retraso terapéutico. Los datos demuestran que solo la mitad de las personas que padecen trastornos mentales desean buscar ayuda⁴⁸. Estudios globales realizados por el consorcio de la Encuesta Mundial de Salud Mental⁶⁶⁻⁶⁸ muestran que solamente el 41% de las personas con ansiedad, el 57% de las personas con depresión y el 39% de las personas con trastornos por consumo de sustancias consideran que tienen un problema de salud mental. El estigma asociado a los trastornos mentales provoca efectos negativos y supone una barrera al tratamiento efectivo de la depresión^{33,69}.

2.2.3 Evaluación y cribado de la depresión

2.2.3.1 Evaluación de la depresión en Atención Primaria

La AP constituye la puerta de entrada al sistema sanitario para la mayoría de la población de España⁷. La atención a la salud mental de las personas, al igual que la de la mayoría de las patologías más prevalentes, forma parte del día a día de las consultas de AP. Los problemas relacionados con la salud mental suponen al menos una de cada cinco consultas que reciben los MF⁷⁰ y las principales patologías de esta esfera que presentan los pacientes adultos atendidos en AP en España y Europa son la ansiedad y la depresión^{71,72}, a menudo asociadas a factores sociales, personales, profesionales y de salud⁴¹. Los pacientes deprimidos que acuden a la consulta de AP suelen encontrarse en alguna de estas tres situaciones: acontecimientos adversos de la vida (familiares, sociales, laborales, etc.)⁶⁵, enfermedad crónica, especialmente si causa dolor, o un trastorno primario del estado de ánimo²⁰.

La mayor parte de los pacientes con depresión son diagnosticados y tratados exclusivamente en el ámbito de la AP⁷². Se estima que el 61% de la asistencia médica a la salud mental en casos de

depresión y ansiedad se realiza únicamente por parte de los MF en centros de AP⁷³. Hasta el 79% de los tratamientos antidepresivos se prescriben por especialistas diferentes a los psiquiatras y los MF se encuentran entre los más habituales⁷⁴. Un estudio muestra que el 64% de los pacientes que han realizado una tentativa de suicidio han visitado un profesional de la salud en las 4 semanas previas, la mayoría de ellos se visitaron en centros de AP⁷⁵.

La evaluación de la depresión no se debe limitar al recuento de síntomas, sino que debe incluir la duración e intensidad de los síntomas, así como una evaluación psicosocial y la evaluación del riesgo de suicidio^{21,22}, considerando intentos previos, abuso de sustancias, síntomas específicos como la ideación autolítica o la desesperanza y los factores de riesgo^{76,77}.

Los MF tienen las mejores condiciones para conocer al paciente y a su entorno, por lo que tienen una gran importancia en el reconocimiento y el manejo de los síntomas de depresión⁶⁹ y son profesionales competentes para realizar este diagnóstico⁷⁸. El diagnóstico de los trastornos mentales es complejo³⁴, siendo los trastornos depresivos una de las patologías mentales mejor detectadas por los MF⁷⁹. La variabilidad en el diagnóstico de depresión entre facultativos es amplia debido a factores personales, sociales y culturales, que pueden llevar a sobrevalorar o infravalorar el malestar psicológico de los pacientes⁸⁰. Los MF son más propensos a diagnosticar depresión a medida que aumenta la gravedad de los síntomas^{72,81}.

Son varios los estudios y guías de práctica clínica que reflejan la preocupación por el infradiagnóstico e infratratamiento de la depresión en muchos países^{21,22,34,82-84}, asociado al género masculino, la ancianidad y el bajo nivel económico y académico⁸⁵. Un metaanálisis refleja que los MF identifican correctamente la depresión en el 47% de los casos⁸⁶. En España se ha estimado que los MF diagnostican el 78% de los pacientes que acuden con depresión a la consulta de AP. El porcentaje de pacientes no diagnosticado suele corresponderse con aquellos pacientes que presentan síntomas leves⁷² y que por lo tanto pueden beneficiarse menos del diagnóstico⁸⁷ y probablemente no requieren tratamiento farmacológico⁸⁸. El estudio ESEMeD pone de manifiesto que casi el 60% de pacientes con depresión no acuden a consulta de AP ni de atención especializada⁴⁷. El retraso diagnóstico está asociado a peor pronóstico⁸³.

Por otro lado, no se debe restar importancia al sobrediagnóstico que conduce a tratamientos innecesarios⁹ y que también ocurre en el caso de la depresión^{86,87}. En España se estima que se diagnostica de depresión a un 26,5% de pacientes que no la padecen, especialmente en aquellos que han presentado episodios de depresión previos⁸⁹. Es importante una adecuada evaluación que lleve a un diagnóstico de certeza, hecho que puede requerir algunas visitas sucesivas. En

este sentido, la longitudinalidad y el conocimiento profundo de los pacientes es una gran ventaja de la que gozan los MF.

2.2.3.2 Cribado de la depresión

Se ha propuesto el cribado de depresión por parte de algunos autores y asociaciones para asegurar la identificación de casos que pueden pasar desapercibidos. Los grupos de trabajo del servicio preventivo de los Estados Unidos (*United States Preventive Services Task Force, USPSTF*) y la Academia Americana de Médicos de Familia (*American Academy of Family Physicians*) recomiendan el cribado universal en todos los adultos en AP siempre y cuando vaya seguido de una confirmación diagnóstica posterior, así como de un tratamiento y visitas de seguimiento⁹⁰⁻⁹². En la reciente declaración de recomendaciones de la USPSTF publicada en junio de 2023, recomienda (grado de recomendación B) por su beneficio moderado el cribado de la depresión en la población adulta, incluidas las mujeres embarazadas y las puérperas, así como los adultos mayores de 65 años⁹². Ambos grupos coinciden en que no existe evidencia sobre posibles efectos nocivos de hacerlo. No hay consenso sobre la periodicidad adecuada del cribado⁹².

La guía NICE, actualizada en 2022, recomienda estar alerta ante la posible existencia de depresión particularmente en los casos de episodios depresivos previos o enfermedad física asociada a deterioro funcional⁹³. Algunos estudios han demostrado que el cribado ambulatorio y en urgencias están al mismo nivel de coste-efectividad que las intervenciones preventivas para otras enfermedades⁶⁹ y son más coste-efectivas que el cribado del virus de la inmunodeficiencia humana (VIH)⁹⁴, por ejemplo.

A pesar de las recomendaciones de algunas asociaciones, existe controversia sobre la indicación del cribado^{21,22}, y sobre si este debe ser universal o selectivo (solo en aquellos pacientes con un riesgo específico)²³. En el documento multidisciplinar “Abordaje compartido de la depresión”²³ en el que han participado diferentes sociedades científicas entre ellas la Sociedad Española de Medicina de Familia y Comunitaria (semFYC) citan que, en cualquier caso, la AP es la indicada para detectar a estos pacientes⁶⁹ tanto de forma oportunista como de forma activa utilizando herramientas de cribado sencillas y breves que ayuden a orientar el diagnóstico²³, por su facilidad de acceso y atributos que posee.

Si el cribado da positivo para un posible caso de depresión tras la administración de una escala, el diagnóstico debe confirmarse mediante la entrevista clínica utilizando los criterios del DSM-5⁹¹.

2.2.3.3 Herramientas de evaluación de la depresión

Los cuestionarios de evaluación de la depresión pueden ser útiles en la ayuda al diagnóstico o al cribado, como instrumentos de valoración de la gravedad y también para el seguimiento clínico y de la respuesta al tratamiento^{21,22}. Tienen como objetivo la evaluación sintomática del paciente en un momento determinado a partir de las respuestas del paciente a cada ítem y la obtención de una puntuación final para la que se establecen unos puntos de corte. Estos instrumentos no conducen inmediatamente a un diagnóstico, pero pueden identificar con precisión los posibles casos de depresión o pacientes en riesgo⁹⁵.

Las escalas más frecuentes que han demostrado adecuadas propiedades psicométricas y están adaptadas al español se listan continuación.

Patient Health Questionnaire (PHQ)

La escala de depresión *Patient Health Questionnaire* (PHQ-9)⁹⁶ consta de 9 ítems que evalúan la presencia de síntomas depresivos en las últimas dos semanas y ha demostrado utilidad para el cribado de la depresión y para la monitorización de la respuesta al tratamiento²³. La puntuación total puede variar entre 0 y 27. Se considera depresión leve entre 10 y 14, moderada entre 15 y 19 y grave si la puntuación es mayor o igual a 20 puntos^{21,22,96}.

La PHQ-9 cuenta con traducción al español⁹⁷ y a múltiples idiomas, es la escala de depresión que se ha evaluado más veces en los estudios de análisis de propiedades psicométricas⁹⁸. Existe la versión de dos ítems (PHQ-2)⁹⁹ que también muestra elevada sensibilidad aunque menor especificidad que la PHQ-9 para detectar casos de depresión mayor⁹².

Inventario de Depresión de Beck (en inglés *Beck Depression Inventory, BDI*)

El Inventario de Depresión de Beck (BDI)¹⁰⁰ es uno de los más utilizados. Su versión más reciente es la BDI-II¹⁰¹ y todas sus versiones han sido traducidas al castellano y validadas en España con diferentes grupos poblacionales¹⁰¹⁻¹⁰³. Cuenta con propiedades psicométricas adecuadas y ha demostrado un buen rendimiento diagnóstico¹⁰⁴.

Se trata de un instrumento autoadministrado de 21 ítems que valoran de 0 a 3 puntos el estado del paciente en función de la gravedad durante las dos semanas previas. La puntuación total puede oscilar entre 0 y 63 y se ha propuesto como punto de corte 14 o más puntos para el diagnóstico de depresión^{104,105}.

Escala de Hamilton para la depresión (*Hamilton Rating Scale for Depression, HRSD o HAM-D*)

La escala de Hamilton para la depresión (HRSD)¹⁰⁶ es heteroadministrada y su versión original consta de 21 ítems, aunque existen versiones de 6, 17 y 24 ítems. La HRSD cuenta con traducción al español¹⁰⁷ y las propiedades psicométricas de las versiones de 6, 17 y 21 ítems han sido analizadas en la población de España¹⁰⁸. No todos los ítems constan del mismo número de categorías de respuesta y se acepta un punto de corte de 8 o superior para considerar depresión^{21,22}.

Se utiliza para evaluar la gravedad, realizar el seguimiento de los pacientes con depresión y para monitorizar la respuesta al tratamiento, que se define como una disminución $\geq 50\%$ de la puntuación inicial y la remisión cuando se alcanza una puntuación total ≤ 7 puntos²¹⁻²³.

Escala de depresión de Montgomery (*Montgomery-Asberg Depression Rating Scale, MADRS*)

La escala de Montgomery (MADRS)¹⁰⁹ se cumplimenta a través de una entrevista realizada por un clínico y contiene 10 ítems que evalúan la gravedad de la depresión y el perfil de síntomas²³, aunque también existe una versión autoadministrada que ha demostrado una buena correlación con la versión original¹¹⁰. Cada ítem se puntúa del 0 al 6 en función de los síntomas presentados durante la semana previa o los tres días anteriores, por lo que la puntuación total puede oscilar entre 0 y 60. Se ha realizado la traducción, adaptación y análisis de las propiedades psicométricas en español¹¹¹. Se propone el punto de corte ≥ 7 para considerar depresión²¹⁻²³. Igual que en la HDRS, la respuesta al tratamiento se considera si hay una disminución \geq del 50% de la puntuación inicial^{21,22}.

Escala de depresión de Zung (*Zung Self-Rating Depression Scale*)

La escala de depresión de Zung¹¹² consta de 20 ítems que cuantifican la frecuencia de los síntomas de depresión en cuatro categorías de respuesta que van de 1 a 4 puntos, por lo que la puntuación total puede tomar valores entre 20 y 80. El punto de corte propuesto es 50 puntos^{21,22}. Cuenta con una versión traducida al español¹¹³ y ha demostrado adecuadas propiedades psicométricas en pacientes de AP de España para detectar casos de pacientes con depresión¹¹⁴.

Escala Center for Epidemiologic Studies-Depression (CES-D)

La escala *Center for Epidemiologic Studies-Depression* (CES-D)¹¹⁵ contiene 20 ítems que evalúan la frecuencia de los síntomas de depresión durante la última semana en una escala de Likert de 4 categorías que puntúa de 0 a 3, por lo que la puntuación máxima puede ser de 60 puntos. El punto de corte establecido para la depresión es igual o superior a 16^{21,22}. Las propiedades

psicométricas de la versión en español son adecuadas tanto en población con patología psiquiátrica¹¹⁶ como sin ella¹¹⁷.

Escala de Depresión Geriátrica de Yesavage (*Geriatric Depression Scale, GDS*)

La Escala de Depresión Geriátrica de Yesavage (GDS)¹¹⁸ es una de las más utilizadas para el cribado de la depresión en población anciana. Consiste en una escala de 15 ítems dicotómicos que se responden con “sí” o “no” en relación con los síntomas presentados durante la semana previa. Es autoadministrada y el punto de corte recomendado para considerar depresión es 6 o más respuestas afirmativas²³.

Existe una versión de 5 ítems y otra de 30 ítems. Tanto la versión de 15 ítems¹¹⁹ como la de 30 ítems¹²⁰ han sido traducidas al español y han mostrado buena validez y fiabilidad, así como elevada sensibilidad y especificidad para el cribado de depresión en población de 65 años o más en AP en España. La versión de 5 ítems también ha demostrado ser válida para el cribado en esta población a pesar de ser la más breve^{23,121}.

Escala de depresión, ansiedad y estrés (*Depression, Anxiety and Stress Scales, DASS-21*)

La *Depression, Anxiety and Stress Scales* (DASS-21)¹²² contiene tres escalas que evalúan la presencia de síntomas de depresión, ansiedad y estrés. Cada escala tiene siete ítems que se valoran desde 0 puntos (“nunca”) a 3 puntos (“casi siempre/siempre”). Cada dimensión se estratifica en normal, leve, moderada, severa o extremadamente severa. Según las instrucciones de puntuación del manual del cuestionario, los respectivos puntos de corte para la escala de depresión son: < 5, 5-6, 7-10, 11-13, > 13; para la escala de ansiedad: < 4, 4, 5-7,8-9, > 9; y para la escala de estrés: < 8, 8-9, 10-12, 13-16, > 16. Existe la versión adaptada al español¹²³ que ha demostrado adecuadas propiedades psicométricas¹²⁴.

Escala de Ansiedad y Depresión Hospitalaria (*Hospital Anxiety Depression Scale, HADS*)

La Escala de Ansiedad y Depresión Hospitalaria (HADS)¹²⁵ es un instrumento autoadministrado diseñado para la evaluación de síntomas de ansiedad y depresión en pacientes ambulatorios del hospital. Es una escala de 14 ítems con dos subescalas (ansiedad y depresión) de 7 ítems cada una con puntuaciones de 0 a 3¹²⁶. Los puntos de corte originales recomendados son > 8 para casos posibles y > 10 para casos probables en ambas subescalas¹²⁵.

La HADS cuenta con traducción al español¹²⁷ y se ha utilizado para el diagnóstico y para la evaluación de la gravedad tanto en población general como en pacientes con enfermedad física¹²⁶.

En el proyecto de esta tesis doctoral se ha utilizado la Hopkins Symptom Checklist (HSCL) y la Entrevista Diagnóstica Internacional Compuesta (CIDI), por lo que se le dedica un apartado concreto a continuación a cada una de ellas. El uso de la HSCL en este estudio proviene de un proyecto europeo en el que se seleccionó dicha herramienta^{128,129}, tal y como se detalla más adelante en la justificación.

2.3 Hopkins Symptom Checklist

La Hopkins Symptom Checklist, abreviada como HSCL, es un cuestionario de síntomas autoadministrado que fue diseñado por Derogatis et al. en la Universidad John Hopkins de Baltimore (Estados Unidos de América) en el año 1974¹³⁰ a partir de la *Discomfort Scale* de Parloff (1954) y las aportaciones de otros investigadores.

La escala original Symptom Checklist (SCL) está escrita en inglés y se compone de 58 ítems (SCL-58) que recogen los síntomas observados en pacientes psiquiátricos no ingresados. Los mismos autores desarrollaron otro listado más extenso de 90 síntomas (SCL-90) en el año 1976¹³¹, que posteriormente fue revisada y modificada y pasó a llamarse SCL-90-R. La SCL-90-R tiene su versión en español que data de 1991¹³². Tanto la versión de 58 como la de 90 ítems se basan en un modelo reflectivo y son multidimensionales. La SCL está orientada a la medición de psicopatología tanto en pacientes psiquiátricos hospitalarios como en el resto de los pacientes ambulatorios.

Los 58 síntomas preguntados en los ítems de la escala original se agrupan en 5 dimensiones: somatización, obsesivo-compulsividad, sensibilidad interpersonal, depresión y ansiedad¹³⁰. La SCL-90 contiene 9 dimensiones, que están compuestas por las cinco de la SCL-58 más las siguientes: hostilidad, fobia ansiosa, ideación paranoide y psicoticismo¹³¹. Ambas se responden con una escala de Likert de 0 a 4 puntos (5 categorías). Existen múltiples versiones de la SCL que van desde la más breve de 5 ítems hasta la más larga de 90 ítems¹³³.

2.3.1 Hopkins Symptom Checklist-25

La HSCL-25 es la versión de 25 ítems^{130,134} derivada de la SCL-90-R e incluye ítems de las dimensiones originales de ansiedad, depresión y síntomas somáticos¹³⁵. De los 25 ítems, 10 ítems hacen referencia a síntomas de ansiedad y 15 ítems hacen referencia a síntomas de depresión, conformando las dos dimensiones o subescalas de ansiedad y depresión, respectivamente, tal y como se observa en la Tabla 3.

Tabla 3. Ítems y dimensión correspondiente de la Hopkins Symptom Checklist-25 (HSCL-25)

Ítem en la versión original (inglés)	Dimensión
1. Being scared for no reason	Ansiedad
2. Feeling fearful	
3. Faintness	
4. Nervousness	
5. Heart racing	
6. Trembling	
7. Feeling tense	
8. Headache	
9. Feeling panic	
10. Feeling restless	
11. Feeling low in energy	Depresión
12. Blaming oneself	
13. Crying easily	
14. Losing sexual interest	
15. Feeling lonely	
16. Feeling hopeless	
17. Feeling blue	
18. Thinking of ending one's life	
19. Feeling trapped	
20. Worrying too much	
21. Feeling no interest	
22. Feeling that everything is an effort	
23. Worthless feeling	
24. Poor appetite	
25. Sleep disturbance	

Fuente: elaboración propia.

La HSCL-25 es una escala autoadministrada en la que el paciente debe puntuar cada ítem siguiendo una escala de Likert con relación a cómo se ha encontrado durante la semana previa. Cada ítem puede puntuarse del 1 al 4 con la siguiente relación de puntuaciones: 1 = “En absoluto”; 2 = “Un poco”; 3 = “Bastante” y 4 = “Mucho”. Para obtener la puntuación total, se deben sumar las puntuaciones obtenidas en cada ítem y dividirla por el número total de ítems respondidos. También se puede obtener la puntuación de las subescalas de ansiedad y depresión con el mismo método. La puntuación total, por tanto, puede oscilar entre 1 y 4 y se suele presentar con dos decimales. Se contemplan dos puntos de corte diferentes, se considera caso probable de depresión si la puntuación media de la escala es $\geq 1,55$ y caso psiquiátrico que requiere tratamiento si la puntuación media es $\geq 1,75$ ¹³⁶⁻¹³⁹.

La HSCL-25 se considera una escala con buena ergonomía por ser autoadministrada y fácil de cumplimentar, por la posibilidad de ser introducida durante la consulta del profesional de una forma cómoda y por la sencillez de su interpretación¹²⁸. Requiere tan solo entre 5 y 10 minutos para responder a los ítems¹⁴⁰.

Ha sido traducida a muchos idiomas y su uso se extiende a diferentes continentes¹⁴¹ y poblaciones como personas refugiadas^{142,143}, ancianos^{144,145}, adolescentes^{146,147}, población con otras patologías asociadas como VIH^{148,149}, etc. Previamente a la realización de este estudio no se ha encontrado en la bibliografía la versión española de la HSCL-25 y no se han analizado sus propiedades psicométricas en AP de nuestro país.

La HSCL-25 ha demostrado adecuada validez y fiabilidad en pacientes de AP al compararla con el diagnóstico de depresión realizado por profesionales médicos, con una concordancia del 86,7%¹³⁴. Los valores de fiabilidad reportados en la bibliografía están medidos principalmente con el coeficiente alfa de Cronbach y son superiores a 0,70^{137,140,148}, lo que demuestra que las puntuaciones de la escala son fiables.

Por lo que se refiere a la estructura interna de la HSCL-25, se han propuesto diferentes modelos¹³⁷. La mayoría de los estudios abogan por un único factor^{140,147} o dos factores correlacionados (depresión y ansiedad)^{137,150}. También se han propuesto modelos de tres factores (depresión, ansiedad y síntomas somáticos) correlacionados en algunos contextos donde las quejas somáticas tienen un peso considerable¹⁴⁸ y de cinco factores (añadiendo los de malestar general y ansiedad autonómica o fisiológica)¹⁵⁰.

La HSCL-25 ha demostrado valores de buena utilidad diagnóstica para identificar síntomas de ansiedad y depresión en pacientes de AP¹³⁵. Diferentes estudios europeos reportan valores adecuados de sensibilidad y especificidad de la HSCL-25 para el diagnóstico de la depresión en pacientes de AP al comparar sus puntuaciones con una entrevista psiquiátrica tal y como se puede observar en la Tabla 4. También ha demostrado una elevada sensibilidad y especificidad en AP al ser comparada con la escala MADRS basada en los criterios DSM-IV vigentes en el momento del estudio^{129,145}.

Tabla 4. Valores de sensibilidad, especificidad, valores predictivo positivo y negativo de la HSCL-25 comparada con una entrevista psiquiátrica en estudios europeos realizados en Atención Primaria

Autor principal (año)	País	Sensibilidad	Especificidad	VPP	VPN	Entrevista psiquiátrica (criterios diagnósticos)
Nettelbladt ¹³⁸ (1993)	Suecia	76	73	58	86	PSE9 (CIE-8)
Veijola ¹³⁶ (2003)	Finlandia	40	94	26	97	SCID (DSM-III)
Lundin ¹⁵¹ (2015)	Suecia	67	78	30	95	SCAN (DSM-III)
Nabbe ¹⁴⁰ (2019)	Francia	59	91	70	87	PSE9 (CIE-8)

Fuente: elaboración propia. VPP: valor predictivo positivo; VPN: valor predictivo negativo; PSE9: 9ª edición del Examen del estado presente (en inglés *Present State Exam*); ICD-8: 8ª edición de la Clasificación Internacional de las Enfermedades; SCID: Entrevista Clínica Estructurada de los trastornos del DSM-III (en inglés *Structured Clinical Interview for DSM-III Disorders*); DSM-III: 3ª edición del Manual diagnóstico y estadístico de los trastornos mentales (en inglés *Diagnostic and Statistical Manual of Mental Disorders*); SCAN: de sus siglas en inglés *Schedules for Clinical Assessment in Neuropsychiatry*.

De la Tabla 4 destacan sobre todo los valores elevados de especificidad y valor predictivo negativo. Otros estudios realizados en encuestas poblacionales muestran valores de sensibilidad cercanos al 40%^{139,144}, por lo que en este contexto la HSCL-25 se considera adecuada como primer cribado¹³⁹.

2.3.2 Hopkins Symptom Checklist-10 y Hopkins Symptom Checklist-5

Existen otras versiones de la misma escala HSCL con un número reducido de ítems¹⁵² que resultan interesantes por su brevedad y menor tiempo de cumplimentación. Estos aspectos son relevantes ya que pueden ahorrar tiempo en contextos complejos como el de la AP.

La HSCL-10 y la HSCL-5 contienen 10 y 5 ítems, respectivamente. Para su desarrollo se seleccionaron los ítems teniendo en cuenta su fuerte correlación con la puntuación media total de la HSCL-25^{133,153}. Los ítems de estas dos versiones están incluidos en la versión HSCL-25 (ANEXO 2). La puntuación total se calcula de la misma manera, por lo que se obtiene una puntuación entre 1 y 4. Existe muy poca bibliografía acerca del punto de corte adecuado para estas versiones, Strand et al. recomiendan 1,85 para el para la HSCL-10 y 2,00 para la HSCL-5 para considerar depresión¹⁵⁴.

Ambas versiones cortas, especialmente la HSCL-10, se han utilizado en contextos como encuestas poblacionales^{154,155} y en diferentes poblaciones como adolescentes¹⁵⁶⁻¹⁵⁸,

trabajadores de la industria¹⁵⁹, pacientes con trastorno por consumo de alcohol¹⁶⁰, refugiados¹⁶¹, etc. Recientemente también se ha utilizado para medir síntomas de depresión y ansiedad debidos a la pandemia por COVID-19¹⁶²⁻¹⁶⁴. Sus propiedades psicométricas recomiendan estas versiones para su uso clínico como cribado y evaluación de síntomas así como para fines de investigación^{155,158,165}.

Las puntuaciones de la HSCL-10 y de la HSCL-5 han mostrado una fiabilidad y validez elevadas^{154,155}, sin embargo, las propiedades psicométricas son mejores en la HSCL-10^{158,165,166}. Tanto la HSCL-10 como la HSCL-5 han mostrado una fuerte correlación de su puntuación total con respecto a la puntuación total que se obtiene en la HSCL-25 y una alta fiabilidad (coeficiente alfa de Cronbach > 0,80)¹⁵⁴.

Solamente dos estudios han utilizado la HSCL-10 con población de AP, ambos están realizados por los mismos autores y se han llevado a cabo en Dinamarca y Noruega con la misma población de 294 adolescentes de entre 14 y 16 años de etnias minoritarias en estos países. En uno de los estudios comparan la HSCL-10 con otra versión de la HSCL de 6 ítems y otros dos cuestionarios breves¹⁵⁶, y en el otro la utilizan para cribado de depresión en esta población¹⁶⁵. Ambos estudios obtienen buenos resultados. No se han encontrado estudios realizados con la población adulta de AP en su forma más amplia para ninguna de estas dos versiones breves.

2.4 Entrevista Diagnóstica Internacional Compuesta

La Entrevista Diagnóstica Internacional Compuesta¹⁶⁷, conocida por sus siglas en inglés como CIDI (*Composite International Diagnostic Interview*), es una entrevista diagnóstica estandarizada que se diseñó a petición de la OMS y la *Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA)* de los Estados Unidos. Ambas organizaciones se unieron en 1979 con el fin de evaluar el alcoholismo, el abuso de drogas y el diagnóstico y clasificación de los trastornos mentales. En 1982 se creó un grupo para desarrollar entrevistas diagnósticas basadas en las definiciones y criterios de la CIE-10 y el DSM-III entonces vigente, entre otros¹⁶⁷. En esta etapa se desarrollaron dos instrumentos, uno diseñado especialmente para estudios clínicos, el *Schedules for Clinical Assessment in Neuropsychiatry (SCAN)*¹⁶⁸ y otro pensado principalmente para encuestas en población general, la CIDI.

La CIDI fue diseñada para servir a estudios transculturales de tipo epidemiológico y comparativo que se centran en estimar las tasas de prevalencia de trastornos específicos entre países y regiones, describir la comorbilidad y evaluar la universalidad de los factores de riesgo y las

consecuencias sociales de los trastornos mentales¹⁶⁹. Se utiliza principalmente en estudios epidemiológicos, aunque puede servir para propósitos clínicos y de investigación.

El idioma original de la entrevista CIDI es el inglés, pero ha sido adaptada y validada en múltiples idiomas, entre ellos el español¹⁷⁰, utilizando un proceso de adaptación protocolizado de la OMS común y riguroso¹⁷¹.

Se trata de una entrevista diagnóstica completamente estandarizada que proporciona diagnósticos actuales (un mes y 12 meses antes de la entrevista) y a lo largo de la vida según criterios aceptados, como la CIE-10 y el DSM^{167,172}. Debe ser administrada por entrevistadores entrenados específicamente a partir del material de formación existente para aprender a administrar la CIDI¹⁶⁷. Los entrevistadores no requieren formación clínica previa¹⁷⁰. En el documento de la entrevista se detallan unas instrucciones diseñadas para que los entrevistadores consigan respuestas concisas y completas de los entrevistados. También hay materiales disponibles para los supervisores de los estudios con el fin de controlar la calidad de la recogida de datos¹⁶⁷.

La CIDI original ha ido evolucionando tras las revisiones realizadas. La versión CIDI 3.0¹⁶⁷ es la más reciente e incluye un módulo de cribado y 40 secciones centradas en diagnósticos (22 secciones), funcionamiento (cuatro secciones), tratamiento (dos secciones), factores de riesgo (cuatro secciones), correlatos sociodemográficos (siete secciones) y factores metodológicos (dos secciones). A los encuestados se les hacen preguntas sobre síntomas relacionados con trastornos mentales, que son leídas por los entrevistadores exactamente tal y como están redactadas y siguiendo las instrucciones rigurosamente. Si se manifiestan suficientes síntomas en un patrón que sugiere que podría haber un diagnóstico, se pregunta a los encuestados sobre el inicio y la frecuencia del grupo concreto de síntomas que han manifestado¹⁷².

La CIDI ha demostrado ser una entrevista válida y fiable en diferentes culturas del mundo^{173,174}. Un metaanálisis reciente que analiza la capacidad de clasificar correctamente a individuos con depresión sugiere que la CIDI es mejor que otras entrevistas semiestructuradas en individuos con síntomas depresivos subclínicos o leves, sin embargo su correcta clasificación no aumenta a medida que aumenta la intensidad de los síntomas¹⁷⁵.

La CIDI ha sido utilizada en estudios llevados a cabo con población española^{47,50}. Asimismo, es frecuente el uso de la entrevista CIDI como medida *gold standard* para el análisis de la validez de criterio y la precisión diagnóstica de escalas y cuestionarios de evaluación de síntomas depresivos en estudios de diferentes países¹⁷⁶⁻¹⁷⁹. Varios estudios han utilizado la CIDI en estudios de análisis de validez de diferentes versiones de la HSCL^{139,146,156,165}.

2.5 Psicometría y uso de escalas en el ámbito de la salud

El término “psicometría” hace referencia a la disciplina que tradicionalmente se ha encargado de proporcionar los criterios de calidad para el desarrollo de medidas estandarizadas de las características psicológicas y/o valores subjetivos¹⁸⁰. La psicometría tiene sus raíces en el campo de la investigación psicológica, aunque esta metodología se aplica también a otros aspectos médicos y de la salud¹⁸¹. Las ciencias sociales y de la psicología utilizan para su desarrollo aspectos que son poco objetivables, difíciles de definir o incluso algo inciertos. En este sentido se habla de características latentes, aquellas que están o se intuyen, pero son difíciles de objetivar y por lo tanto de medir. La conceptualización es el proceso mental mediante el cual se definen constructos (conceptos) difusos e imprecisos y sus componentes concretos y precisos. En psicometría, la palabra “constructo” se refiere a cualquier concepto que se desea medir¹⁸⁰. Los test o escalas son cuestionarios preparados con el fin de medir uno o más constructos¹⁸⁰.

En el ámbito de la salud, es necesario medir y cuantificar los constructos tanto con objetivos clínicos como de investigación. El uso de escalas para obtener datos es una herramienta habitual para los profesionales clínicos e investigadores. En la prevención se utilizan tests de cribado, mientras que en la clínica se usan para recabar datos que permitan establecer un diagnóstico, conocer el estado de los pacientes o para objetivos de tratamiento¹⁸². Cabe destacar el uso de escalas para cuantificar conceptos subjetivos como la calidad de vida^{183,184}, el dolor^{185,186}, síntomas de la esfera mental^{96,101,187,188} o aspectos relacionados con la valoración social y los cuidados¹⁸⁹⁻¹⁹¹ de los pacientes, entre muchos otros.

COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) es una iniciativa multidisciplinar llevada a cabo por investigadores internacionales del ámbito de la epidemiología, psicometría, medicina, investigación cualitativa y asistencia sanitaria, expertos en el desarrollo y la evaluación de instrumentos de medición de resultados¹⁹².

Según sus guías, una vez identificado el constructo que se pretende medir en una población determinada, se debe comprobar si existe un instrumento adecuado para medirlo. Si existe, puede utilizarse directamente siguiendo sus instrucciones de uso. Puede ocurrir que exista un cuestionario adecuado validado para otra población distinta a la población de estudio; como por ejemplo idiomas y/o culturas diferentes o personas de otros grupos de edad (niños, adolescentes, ancianos, etc.). En tal caso, se debe valorar la opción de adaptar el cuestionario a las características de la población de estudio¹⁸². En el caso de que no exista ningún cuestionario que mida dicho constructo, se deberá desarrollar uno nuevo. Valorar una adaptación antes que una elaboración *de novo* es obligado, no se justifica la elaboración indiscriminada de tests¹⁹³. Sin

embargo, también la traducción, adaptación y validación de escalas es un proceso costoso que requiere tiempo y planificación, así como el seguimiento de una metodología rigurosa para asegurar una medida válida y fiable del constructo en cuestión en la población diana¹⁹⁴.

2.5.1 Datos referidos por los pacientes (Patient Reported Outcomes, PRO)

Existen diferentes formas de medir aspectos subjetivos en los pacientes. La entrevista clínica destaca como método cualitativo para obtener información del paciente, aunque está sujeto a las habilidades comunicativas del profesional. Otra forma de medir los síntomas subjetivos de los pacientes de forma cuantitativa es a través de la utilización de escalas.

Los datos referidos por pacientes, más conocidos por sus siglas en inglés PRO (*Patient-Reported Outcomes*), se definen como la medida de cualquier aspecto del estado de salud que proviene directamente del paciente, sin interpretación de las respuestas del paciente por parte del clínico u otra persona¹⁹⁵. Los PROs suelen medirse a partir de las puntuaciones obtenidas en cuestionarios autoadministrados por los propios pacientes, conocidos por sus siglas en inglés como PROMs (*Patient-Reported Outcome Measures*)¹⁹⁶ y pueden contener un solo ítem o varios (multi-ítem).

En los instrumentos de medida multi-ítem es necesario conocer el marco conceptual que relaciona los ítems con el constructo que se mide. Este aspecto es relevante porque determina el tipo de análisis que se realizará para la evaluación del instrumento¹⁸¹. En este sentido, existen dos modelos básicos:

- Modelo reflectivo: el constructo se manifiesta en los ítems, que están correlacionados entre sí y con la puntuación total.
- Modelo formativo: el constructo es el resultado de los ítems, que no están necesariamente correlacionados entre sí.

Se ilustra un ejemplo de ambos modelos en la Figura 1: en el modelo reflectivo (a) se observa como el constructo se manifiesta por sí mismo en los ítems; mientras que en el modelo formativo (b) el constructo es el resultado de los ítems o, dicho de otra manera, los ítems son la causa del constructo. La forma más sencilla de distinguir un modelo del otro es preguntarse si los ítems cambiarían cuando el constructo cambiase¹⁸¹. Si la respuesta es sí, se trata de un modelo reflectivo, mientras que, si la respuesta es no, se trataría de un modelo formativo. En el caso del constructo de la depresión medida a partir de varios ítems, se esperaría encontrar un cambio en

la puntuación de los ítems que preguntan por sus síntomas si esta aparece o desaparece (modelo reflectivo).

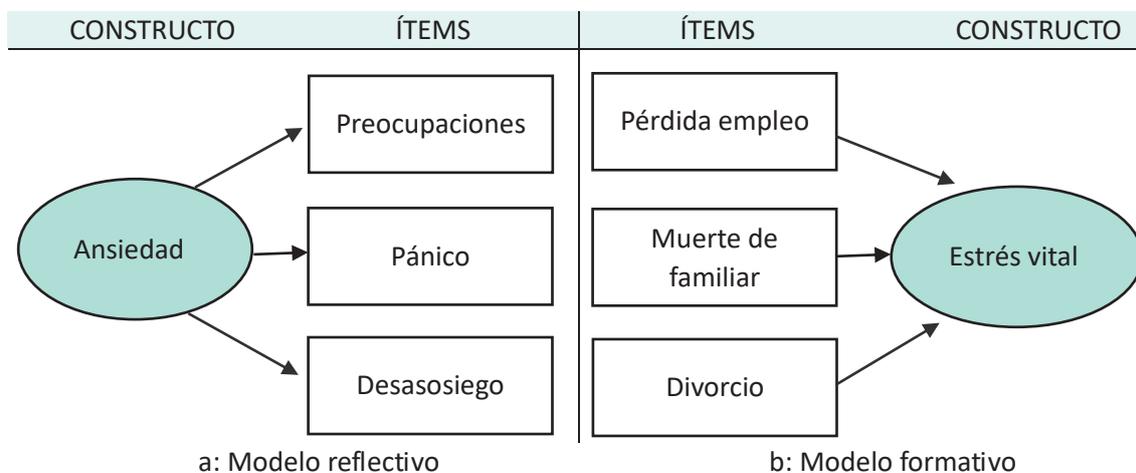


Figura 1. Representación gráfica de un modelo reflectivo (a) y un modelo formativo (b).

Traducido de: "Graphical representation of a reflective model (a) and a formative model (b)" por de Vet H, Terwee CB, Mokkink L, Knol DL; 2018, *Measurement in Medicine*, p. 14, Copyright © Cambridge University Press".

Por lo que se refiere a la estructura de las escalas multi-ítem, estas pueden ser unidimensionales o multidimensionales en función de si los ítems se correlacionan con un solo factor o con varios factores dentro de la misma escala. En la literatura se habla de factor o de dimensión lo que puede dar lugar a confusión ya que se refieren al mismo concepto: el análisis de la estructura de una escala puede señalar uno o más factores en función de cómo los ítems representan uno o más constructos, los factores dan lugar a las dimensiones de las escalas, que son las agrupaciones de ítems que miden cada constructo dentro de una misma escala. Las escalas con más de un factor miden más de un constructo. En psicometría, los constructos inobservables (como por ejemplo la extroversión, la habilidad espacial, la autoeficacia u otras habilidades) que forman parte una escala se denominan variables latentes¹⁹⁷, pueden estar correlacionadas entre sí o ser independientes. Puede haber tantas dimensiones como número de ítems contenga la escala. Este aspecto se explicará con más detalle en el apartado de validez de constructo.

En cuanto al tipo de instrumento de medida, existen diferentes tipos de escalas, en el ámbito de la salud destacan las escalas sumativas tipo Likert, en las que todos los ítems son relevantes para medir el constructo¹⁸⁰. Cada ítem tiene un rango variable de respuestas posible ordenado de forma ascendente y comúnmente tratado como una variable continua¹⁹⁸. La posición del individuo se valora a partir de la suma de la puntuación total. Este tipo de escala es útil para medir conceptos con manifestaciones variadas y poco específicas¹⁸⁰, como sería el caso de la depresión.

2.5.2 Traducción y adaptación de escalas

Incluso si no interviene la escritura, una escala es siempre lingüística y cultural¹⁸². Una escala requiere una adaptación cultural si se pretende utilizar con población de un país diferente para el que fue diseñada, aunque el idioma en ambos países sea el mismo. Si el idioma también es distinto requiere además una traducción previa¹⁹⁹. Este proceso debe seguir una metodología específica que asegure la equivalencia conceptual, la lingüística y la normativa. La equivalencia conceptual hace referencia a que los conceptos implicados en la medida, como es el caso de “la depresión” en el presente estudio, existen en las poblaciones diana respectivas. La equivalencia lingüística se refiere a que las estructuras semánticas seleccionadas cumplen la misma función y poseen el mismo significado²⁰⁰, es decir, que se utilizan palabras y frases con significado similar en ambas poblaciones. Este concepto trasciende la simple traducción, ya que la forma de expresar verbalmente un concepto no es la misma en todos los países, incluso aunque utilicen una lengua común como por ejemplo es el caso de los países de habla hispana. Finalmente, la equivalencia normativa tiene en cuenta la forma de expresar las ideas y hablar de determinados temas en culturas distintas¹⁸². En este sentido, sería prácticamente imposible administrar una escala de forma no anónima sobre “la depresión” si este fuera un tema tabú en una población concreta.

Existen diversos textos que sugieren estándares para realizar el proceso de adaptación transcultural²⁰⁰⁻²⁰³, la mayoría de ellos proponen una adaptación secuencial en diferentes fases tal y como se muestra en la Figura 2 (al final de este apartado). En cada fase se obtiene una versión que se va consensuando y modificando en las fases posteriores.

2.5.2.1 Traducción directa

Se recomienda realizar un mínimo de dos traducciones independientes realizadas por traductores bilingües y a poder ser con experiencia en ambas culturas. Hay consenso sobre que dichos traductores deben tener como lengua materna el idioma de la población diana^{200,201}. Se recomienda que uno de ellos sea conocedor de terminología sanitaria y el otro conocedor de matices lingüísticos como son las expresiones coloquiales en el idioma final¹⁹⁴. Incorporar traductores cualificados es clave para obtener traducciones de calidad.

Este primer paso debe estar enfocado a comprobar que existe equivalencia conceptual y lingüística con la versión original¹⁹³ y debe ser revisado por un panel de expertos conjuntamente con los investigadores que promueven la adaptación de la escala o por personas nativas de la población diana²⁰¹. En caso de conformar un panel de expertos que evalúe las traducciones obtenidas, se sugiere que esté integrado por miembros con amplio conocimiento en el

constructo que se pretende medir y que no tengan conocimiento sobre la finalidad de obtener la versión adaptada para evitar sesgos¹⁹³. Para llegar a acuerdos se utilizan métodos de consenso formal, siendo el más utilizado el método Delphi.

Metodología Delphi

La técnica Delphi consiste en un proceso estructurado de consenso a partir de la opinión colectiva de un panel de expertos²⁰⁴. Este método sirve para estructurar la discusión y el debate necesarios para llegar a un consenso mediante una serie de encuestas iterativas a los expertos sobre una cuestión concreta. Es especialmente útil cuando el tema que se debate no puede ser estudiado de forma empírica²⁰⁵, pero también cuando no existe consenso o hay una falta de conocimiento sobre un tema concreto²⁰⁶. De esta forma se consigue modular la comunicación entre individuos de una forma efectiva y anónima.

Los miembros que participan en el proceso anónimo de votación reciben el nombre de panelistas²⁰⁴, su adecuada selección es el elemento más crucial para el buen funcionamiento del método²⁰⁷. El panel de expertos debe seleccionarse teniendo en cuenta la experiencia, aunque este es un aspecto difícil de medir de forma cuantitativa. Por ello, se deben definir los criterios de selección del panel en función del objetivo de estudio. Se necesitará un grupo homogéneo para evaluar problemas concretos, mientras que quizás conviene un grupo más heterogéneo si se pretenden evaluar aspectos más amplios²⁰⁴.

La técnica Delphi se desarrolla a partir de las “rondas Delphi”, que son series estructuradas de preguntas dirigidas de forma anónima e individual a los panelistas. Los principios básicos incluyen el feedback individual en relación con un tema o cuestionario; la evaluación de las opiniones del conjunto del grupo; la oportunidad para que los individuos revisen sus puntos de vista al conocer la opinión del resto de expertos y el anonimato de las respuestas individuales²⁰⁸. En cada ronda se evalúa de forma cuantitativa el acuerdo entre los panelistas, y se propone una segunda ronda para aquellas cuestiones en las que no se ha alcanzado el umbral de acuerdo definido. En esta nueva ronda se pueden reformular las preguntas o hacer modificaciones en función del feedback obtenido²⁰⁸. En el caso de los cuestionarios, los ítems se pueden reescribir en la segunda ronda en función de las respuestas de la primera ronda. Puede realizarse una tercera ronda si es necesario.

No es obligatorio el uso de metodología Delphi o de otros métodos de consenso formales en el paso de la traducción directa de una escala, sin embargo, se considera muy pertinente que la traducción sea valorada por expertos del entorno en el que se va a utilizar.

2.5.2.2 Traducción inversa o retrotraducción

Se debe realizar al menos una traducción inversa o retrotraducción del idioma final al idioma original para conciliar las posibles discrepancias^{182,200}, aunque algunos autores sugieren realizar también dos retrotraducciones¹⁹⁴. Este proceso debe llevarse a cabo por otro traductor bilingüe cuya lengua nativa sea la de la versión original²⁰¹ y que no haya estado implicado en la traducción previa. Se aconseja que la retrotraducción se haga de forma literal en caso de que el constructo que se mide sea de aspectos prácticos, mientras que se recomienda que se haga teniendo en cuenta la equivalencia lingüística (mismo significado) cuando se miden constructos subjetivos o relacionados con la calidad de vida²⁰¹.

Tras este paso, el equipo investigador debe comparar minuciosamente la versión original con la versión retrotraducida para comprobar la equivalencia conceptual¹⁹³, este paso es clave a pesar de que frecuentemente se le ha dado poca importancia²⁰¹. Se debe prestar atención a la retrotraducción, ya que es posible que en una mala traducción el grado de equivalencia entre la versión original y la retrotraducida sea muy alta. La razón de esta singularidad se debe a que habitualmente las malas traducciones se apoyan en traducciones literales en lugar de en una esmerada adaptación de significados²⁰³. En caso de detectar discrepancias conceptuales, se deberían llevar a cabo revisiones de las versiones traducidas y de la retrotraducida.

2.5.2.3 Entrevista cognitiva o estudio piloto

Los estándares coinciden en la importancia de realizar una prueba piloto con un grupo reducido de personas pertenecientes a la población diana, sin embargo, no existe consenso sobre el número de participantes necesario. Este paso consiste en administrar la versión obtenida en la fase previa a personas que serán sus destinatarios finales. La International Society for Pharmacoeconomics and Outcomes Research (ISPOR)²⁰¹ recomienda que la versión traducida sea probada por 5-8 personas para analizar la equivalencia cognitiva. Los encuestados deben tener el idioma final como lengua materna y representar a la población diana de forma adecuada por lo que se refiere a las características sociodemográficas²⁰¹.

Los datos obtenidos de esta pequeña muestra permiten analizar, revisar y modificar algunos aspectos muy relevantes en una fase previa al análisis de las propiedades psicométricas. Este paso permite asegurar que los ítems e instrucciones son claros y correctamente comprendidos e interpretados por las personas que deben responderlos¹⁹⁴; asimismo, permite conocer el tiempo necesario para la administración de la escala²⁰³. En caso de detectarse dificultades, estas pueden ser subsanadas antes de administrarla a una muestra mayor de participantes.

2.5.2.4 Armonización internacional

La armonización entre las distintas traducciones es esencial para garantizar la validez intertraducción²⁰¹. En el caso de proyectos internacionales colaborativos en los que la adaptación se realice de manera simultánea a varios idiomas de los países participantes, puede ser necesario realizar una armonización internacional, consistente en un encuentro de los investigadores para consensuar y debatir las versiones finales en cada idioma¹⁸². Con esta reunión se pretende mantener una metodología común que permita la comparación de las versiones de la escala en los distintos idiomas.

2.5.2.5 Revisión de la redacción

Se debe realizar una revisión ortográfica y sintáctica antes de obtener la versión final¹⁸², que será la escala lingüísticamente validada. En esta revisión conviene verificar la puntuación, las estructuras gramaticales y comprobar si existen referencias al género u otros aspectos²⁰³.

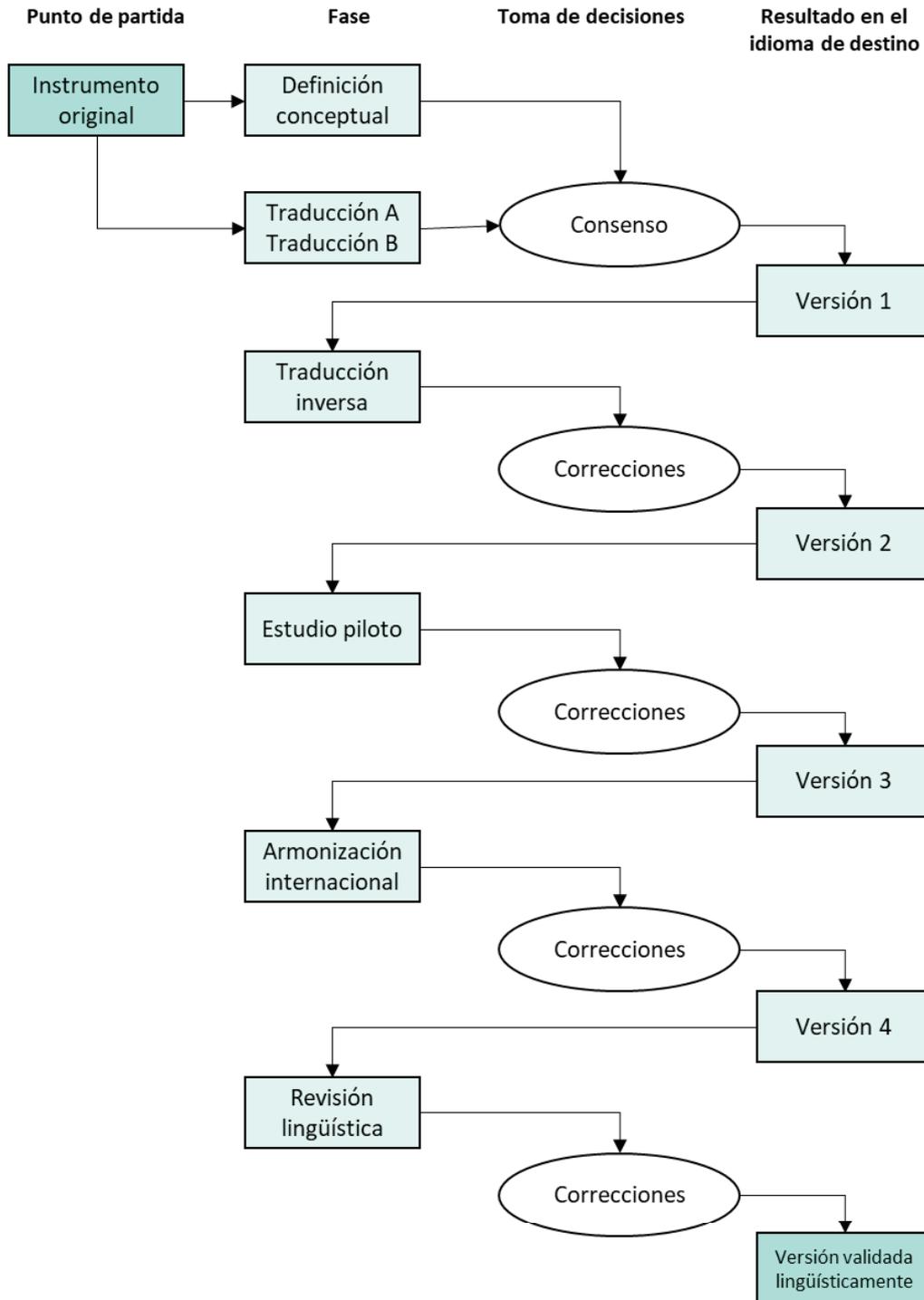


Figura 2. Fases del proceso de traducción y adaptación de un PROM

Tomado de: "Fases del proceso de validación lingüística de un cuestionario" por Doval E, Viladrich C, 2019, *Desarrollo y adaptación de cuestionarios en el ámbito de la Salud*, p. 77, Copyright © Eduardo Doval Diéguez, Carme Viladrich Segués.

Una vez obtenida la versión final en el idioma de la población diana, se debe realizar un estudio para comprobar la validez y la fiabilidad de las puntuaciones que se obtienen en dicha población respecto al constructo que se desea medir. Para ello, se debe diseñar un estudio que permita analizar dichas propiedades psicométricas.

2.5.3 Taxonomía

Uno de los principales objetivos de la iniciativa COSMIN fue establecer unas guías para el consenso en la terminología y definiciones utilizadas para referirse a las propiedades psicométricas²⁰⁹. Se pueden consultar las definiciones del consenso COSMIN de cada término en inglés en el ANEXO 3.

2.5.3.1 Validez

El término validez, en inglés *validity*, es el grado en el que un PROM mide el constructo o constructos que se propone medir²⁰⁹; se refiere a los argumentos disponibles para convencer de que realmente puede utilizarse para ese fin¹⁸⁰.

La definición clásica de validez es aquella según la que “la validez consiste en asegurarse de que la escala mide lo que dice medir”. La American Psychological Association (APA) abandonó esta definición a partir de los estándares publicados en 1974 por la sensación que transmite de que se está validando un test y no sus usos¹⁸⁰. En esta versión ya se hace explícito que lo que se valida son las puntuaciones que se obtienen con el test y no el test en sí mismo. La idea de que no se valida un instrumento de medida, sino el uso que se hace de él fue ratificada por Nunally en 1978²¹⁰. Por este motivo, no se considera correcto decir que un instrumento es válido, sino que proporciona puntuaciones válidas para el contexto concreto en el que se está utilizando²¹¹. Este es el motivo por el cual se deben realizar pruebas de validez en diferentes poblaciones para un mismo instrumento. La validación es un proceso continuo en el que se va acumulando evidencia de forma progresiva¹⁸¹ a medida que se evalúa dicha validez en diferentes grupos de población, ya sea por grupos de edad o por características sociodemográficas concretas.

En términos generales, existen tres tipos de validez^{181,200,209}: la validez de contenido, la validez de criterio y la validez de constructo.

Validez de contenido (en inglés *content validity*)

La validez de contenido se define como el grado en que el contenido de un PROM refleja adecuadamente el constructo que se pretende medir²⁰⁹.

La validez de contenido incluye la validez aparente (en inglés *face validity*), COSMIN la define como el grado en que un instrumento de medida parece reflejar adecuadamente el constructo que debe medirse²⁰⁹. El primer paso del análisis de la validez de contenido consiste en realizar una mirada general a todos los ítems para obtener una primera impresión de la escala. No existe un método concreto sobre cómo debe ser evaluada la validez aparente, ya que se trata de un

análisis subjetivo¹⁸¹. Sin embargo, la falta de validez aparente es un motivo de peso como para no utilizar un cuestionario, esto podría ocurrir si al echar un vistazo a un cuestionario se obtiene la impresión de que la población diana no querría contestar a los ítems por su falta de adecuación (por ejemplo: pregunta por actividades que la población no suele realizar, pregunta por un tema incómodo para dicha población...). Los ítems de un PROM deben ser relevantes y significativos, no ambiguos, fáciles de responder, libres de juicio y no deben causar malestar en los usuarios destinatarios²¹².

Una vez superada la validez aparente, se debe analizar el contenido con más detalle. Este paso es especialmente importante en el desarrollo de nuevos instrumentos de medida, más que en instrumentos adaptados o traducidos de otros ya existentes. En este sentido, se analizan la relevancia y la representatividad de los ítems en relación con el constructo que se pretende medir. La relevancia hace referencia a que todos los ítems deben relacionarse con alguno de los objetivos de medición, mientras que la representatividad es el concepto inverso de que todos los objetivos de medición deben estar representados por uno o más ítems¹⁸⁰. La validez de contenido se basa en el juicio, por lo que no se utilizan pruebas estadísticas para comprobarla¹⁸¹, para ello es muy importante que el constructo a medir esté claramente definido. Se puede evaluar a partir de la información proporcionada por los diseñadores de la escala sobre su desarrollo. Para comprobar la validez de contenido también se pueden utilizar valoraciones por personas expertas que evalúen la adecuación de los ítems y/o por personas pertenecientes a la población diana, para comprobar que aceptan el cuestionario y reconocen su finalidad¹⁸⁰.

Validez de criterio (en inglés *criterion validity*)

La validez de criterio se define como el grado en que las puntuaciones de un PROM son un reflejo adecuado de un estándar-oro (en inglés *gold standard*)²⁰⁹. Este hecho implica que la validez de criterio solamente puede analizarse cuando existe un *gold standard* para dicha medida¹⁸¹.

La validez de criterio incluye la validez concurrente (en inglés *concurrent validity*) y la validez predictiva (en inglés *predictive validity*). La validez concurrente se analiza en los instrumentos que se utilizan para propósitos evaluativos y diagnósticos y mide la puntuación del PROM y del *gold standard* en el mismo momento, mientras que la validez predictiva se analiza en instrumentos que se utilizan para aplicar predicciones y mide si la puntuación del PROM puede predecir la puntuación del *gold standard* en el futuro¹⁸¹. Estas medidas se deben tomar siempre de forma independiente y el resultado de la prueba cuya puntuación se pretende validar no puede estar influido por la medida de la prueba de referencia. En el caso de las pruebas diagnósticas, se utiliza la validez de criterio de tipo concurrente para convencer a los usuarios

de que una prueba es adecuada al compararla con la medida *gold standard*, que suele ser menos práctica o más compleja. Se recomienda establecer a priori el grado de acuerdo entre ambas medidas que se considerará aceptable¹⁸¹. En función del tipo de variables, el parámetro estadístico que se debe utilizar para el análisis es diferente, tal y como se puede observar en la Tabla 5.

Tabla 5. Parámetros estadísticos de los distintos niveles de medición para el *gold standard* y el instrumento de medición objeto de estudio

Nivel de medida		Mismas unidades	Parámetro estadístico
Estándar	Instrumento de medida		
Dicotómico	Dicotómico	Sí	Sensibilidad y especificidad
	Ordinal	NA	ROC
	Continuo	NA	ROC
Ordinal	Ordinal	Sí	Kappa ponderada
		No	r de Spearman u otras medidas de asociación
	Continuo	NA	ROC/r Spearman
Continuo	Continuo	Sí	ICC o acuerdo de Bland y Altman
		No	r de Spearman o Pearson

r: coeficiente de correlación; ROC: curva ROC (en inglés *receiver operating curve*) para un *gold standard* ordinal se puede utilizar un conjunto de curvas ROC, dicotomizando el instrumento por los diversos puntos de corte; ICC: coeficiente de correlación intraclass (en inglés *intraclass correlation coefficient*); NA, no aplicable. Traducido de: "Overview of statistical parameters for various levels of measurement for the gold standard and measurement instrument under study" por de Vet H, Terwee CB, Mokkink L, Knol DL; 2018, *Measurement in Medicine*, p. 163, Copyright © Cambridge University Press".

Para evaluar la validez de criterio, las directrices COSMIN establecen lo siguiente: la correlación es el método estadístico preferido si tanto las puntuaciones del PROM como las del instrumento de referencia son continuas; si las puntuaciones del PROM son continuas y las puntuaciones del *gold standard* son dicotómicas, el área bajo la curva (en inglés *Area Under Curve, AUC*) de la característica operativa del receptor (ROC) es el método preferido; mientras que si ambas puntuaciones son dicotómicas, la sensibilidad y la especificidad son los métodos preferidos²¹³. El AUC está se utiliza como medida del poder discriminatorio de una prueba diagnóstica²¹⁴, a modo de ejemplo, un AUC de 0,8 refleja una probabilidad del 80% de clasificación correcta del paciente entre sano o enfermo. Un valor AUC de 0,5 equivale a la misma precisión diagnóstica que tirar una moneda al aire, ya que acierta en el 50% de las ocasiones en clasificar adecuadamente a los individuos. Se consideran buenos los valores de $AUC \geq 0,8$ ²¹⁵.

La representación gráfica de la curva ROC de la Figura 3 muestra diferentes curvas con sus correspondientes AUC. La curva ROC refleja una relación entre la sensibilidad y la especificidad de la prueba diagnóstica que se evalúa. Se utiliza también para establecer los puntos de corte óptimos que maximizan el conjunto de la sensibilidad y la especificidad a partir del cálculo del

Índice de Youden (J), que se define como “Sensibilidad + Especificidad - 1”. Gráficamente corresponde al punto de la curva ROC más cercano al ángulo superior-izquierdo, es decir, más cercano al punto del gráfico cuya sensibilidad y especificidad son del 100%²¹⁶.

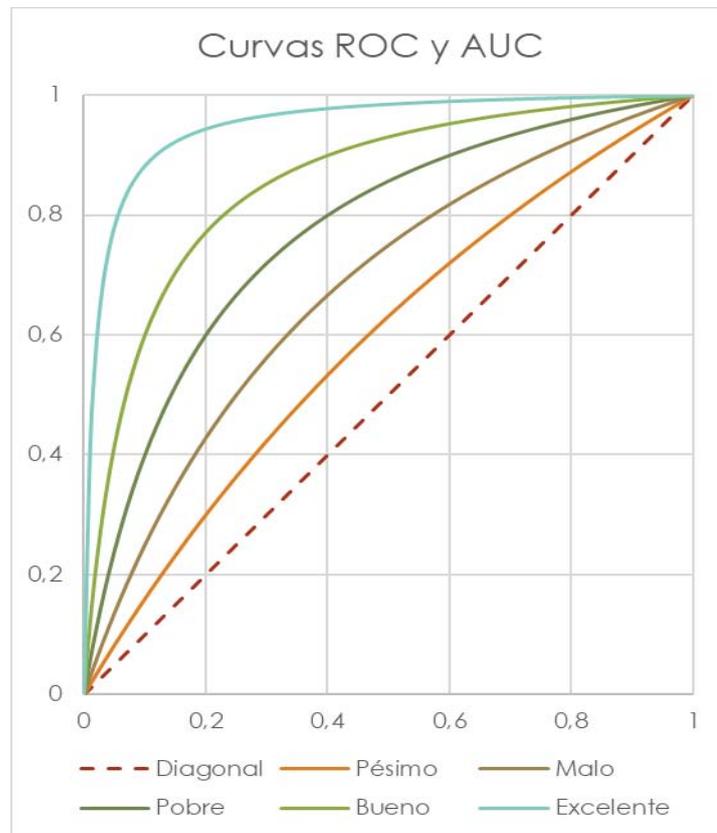


Figura 3. Curvas ROC y su correspondiente AUC

Fuente: elaboración propia. ROC: curva ROC (en inglés *Receiver Operating Curve*), AUC: área bajo la curva (en inglés *Area Under Curve*)

Validez de constructo (en inglés *construct validity*)

La validez de constructo se define como el grado en que las puntuaciones de un PROM son coherentes con las hipótesis (por ejemplo, con respecto a las relaciones internas, las relaciones con las puntuaciones de otros instrumentos o las diferencias entre grupos relevantes) basadas en el supuesto de que el PROM mide de forma válida el constructo que debe medirse²⁰⁹. La validez de constructo es menos potente que la validez de criterio¹⁸¹, pero es fundamental en los casos en los que no exista un *gold standard* disponible.

Hay tres aspectos en la validez de constructo: la validez estructural (en inglés *structural validity*), la comprobación de hipótesis (en inglés *hypotheses testing*) y la validez transcultural (en inglés *cross-cultural validity*).

La validez estructural se define como el grado en que las puntuaciones de un PROM reflejan adecuadamente la dimensionalidad del constructo que se pretende medir²⁰⁹. La dimensionalidad hace referencia al número de factores o variables latentes que contiene el PROM. La puntuación de cada individuo en cada ítem está generada por una variable no observada (latente) que explica la variabilidad de las puntuaciones en el ítem. La variable nunca explicará de forma totalmente satisfactoria la variabilidad de las respuestas del ítem, a esta parte no explicada se le denomina error de medida (e)²¹⁷. En la Figura 4 se muestra un ejemplo de la estructura de un modelo multi-ítem con un solo factor (a) y con dos factores correlacionados (b). En este ejemplo se representa una escala de 7 ítems con dos propuestas de modelos estructurales diferentes. Cada ítem se correlaciona con un constructo o variable latente; se espera encontrar coeficientes de correlación elevados entre cada ítem y el constructo, y de tamaño moderado entre los ítems¹⁸⁰. Si la correlación entre los ítems fuera demasiado elevada, ello indicaría que el contenido de los ítems es redundante. En el modelo de un solo factor (a), los 7 ítems comparten entre sí la relación con una estructura latente, en este caso configurada por un único factor que explica en buena parte las correlaciones entre los ítems. En el modelo (b) existe una estructura factorial con 2 factores correlacionados que se relacionan entre sí (tal y como indica la flecha bidireccional). Los tres primeros ítems se correlacionan con un constructo (factor o variable latente) mientras que los cuatro últimos con otro diferente, aunque ambos están correlacionados por lo tanto tienen que ver el uno con el otro.

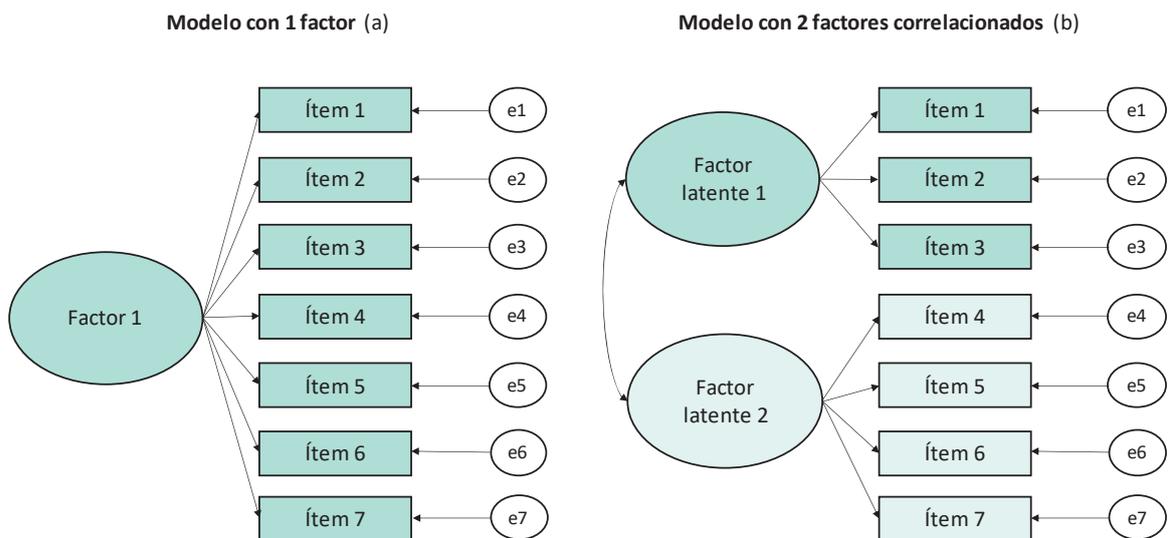


Figura 4. Representación gráfica de la estructura de una escala con 1 factor (a) y con 2 factores correlacionados (b).

Fuente: elaboración propia. e: error de medida.

La validez estructural se analiza mediante el análisis factorial: si no se conoce el tipo y número de dimensiones del PROM, se aconseja realizar un análisis factorial exploratorio (en inglés *exploratory factor analysis, EFA*), mientras que si existe evidencia o estudios previos sobre dichas dimensiones, se aconseja realizar un análisis factorial confirmatorio (en inglés *confirmatory factor analysis, CFA*)^{180,181}. El propósito del CFA es comprobar si los resultados del análisis factorial son coherentes con la base conceptual o el marco especificado del PROM, por lo tanto, se lleva a cabo cuando se conoce claramente la dimensionalidad de un concepto o constructo. Cuando se analiza la validez de una versión traducida, se aconseja realizar el CFA para comprobar la hipótesis de que la estructura interna es igual en ambas poblaciones¹⁸⁰. El CFA asume que una variable observada específica se ve necesariamente afectada sólo por un factor relacionado (variable latente), pero no por otros, basándose en una sólida base teórica o en investigaciones previas²¹⁸. El CFA se analiza mediante la carga factorial (o peso) de cada ítem en el factor esperado. Se considera baja una carga factorial inferior a 0,3²¹⁹.

La comprobación de hipótesis para el análisis de la validez de constructo incluye la comparación de los resultados del PROM con la puntuación de otras medidas del mismo constructo o contrarias; o las diferencias de puntuación del PROM en subgrupos de pacientes¹⁸¹. Se deben formular hipótesis con antelación a los resultados sobre la relación esperada de las puntuaciones del PROM con otros instrumentos de medida^{200,213}. Se esperaría encontrar correlaciones positivas entre las puntuaciones de las mediciones del mismo constructo (validez convergente, en inglés *convergent validity*), como por ejemplo, el uso de dos escalas para medir síntomas de depresión; mientras que se esperaría una ausencia de correlación o correlación débil en el caso de utilizar medidas no relacionadas con el constructo (validez discriminante, en inglés *discriminant validity*), por ejemplo en el caso de comparar la puntuación obtenida en una escala de depresión frente a otra escala para medir dolor (otro constructo diferente). Las correlaciones de cada ítem con su propia escala también son una medida de validez convergente¹⁸⁰. Otro método de comprobación de hipótesis consiste en la comparación por subgrupos de pacientes, lo que también se conoce como validez discriminatoria o de grupos conocidos (en inglés *discriminative or known-group validity*). Este método consiste en comparar las puntuaciones obtenidas en dos grupos en los que se espera obtener una puntuación diferente según la literatura disponible^{181,213}. Por ejemplo, en el caso de la depresión, se espera encontrar una puntuación más elevada entre las mujeres, frente a la puntuación en los hombres, ya que se les conoce una prevalencia mayor^{32,41}.

Finalmente, según la taxonomía COSMIN, la validez de constructo incluye la validez transcultural, que se define como el grado en que el funcionamiento de los ítems de un PROM

traducido o adaptado culturalmente refleja adecuadamente el funcionamiento de los ítems de la versión original²⁰⁹. Este aspecto es especialmente relevante en nuestro país, ya que la mayoría de PROMs disponibles son adaptaciones de cuestionarios originalmente desarrollados en otros países²⁰⁰. Algunos ítems de un cuestionario pueden ser irrelevantes para algunas culturas concretas. Este proceso se inicia con el proceso adecuado de traducción y adaptación transcultural previamente detallado en el apartado 2.5.2. Una vez se dispone de la versión definitiva en el idioma y cultura concreta, se puede analizar el funcionamiento de los ítems mediante la evaluación de la invariancia de la medición (en inglés *assessment of measurement invariance*) o el funcionamiento diferencial de los ítems (en inglés *differential item functioning, DIF*)¹⁸¹. Existen diferentes métodos para evaluarlos²²⁰, entre los más frecuentes se encuentran el análisis factorial y la teoría de respuesta a los ítems (en inglés *item response theory, IRT*). En este sentido, el análisis factorial puede mostrar que alguno de los ítems no pesan en el factor esperado, lo que implicaría que estos ítems tienen un significado diferente en la población destinataria respecto a la original y podría llevar a una estructura factorial diferente en la nueva población¹⁸¹. Tal y como se ha comentado previamente, para este propósito se utiliza el CFA.

2.5.3.2 Fiabilidad

La fiabilidad, en inglés *reliability*, se define como el grado en que la medición está libre de errores de medición²⁰⁹. COSMIN propone una definición extendida en la que se refiere a la fiabilidad como la medida en que las puntuaciones de los pacientes que no han cambiado son las mismas en mediciones repetidas en distintas condiciones, por ejemplo, utilizando conjuntos diferentes de ítems del mismo PROM (consistencia interna); a lo largo del tiempo (test-retest); por diferentes personas en la misma ocasión (inter-evaluador); o por las mismas personas en diferentes ocasiones (intra-evaluador)²⁰⁹. Al igual que ocurre con el término de la validez, se hace referencia a la fiabilidad como propiedad de las puntuaciones, y no de los PROMs^{180,181}.

Los errores sistemáticos son recurrentes y se repiten en el mismo sentido y cantidad, por lo que afectan a la validez de los PROMs, mientras que los errores aleatorios son impredecibles y afectan al azar en cada evaluación realizada, por lo que afectan a su fiabilidad¹⁸⁰. La Figura 5 muestra una representación simple y gráfica de las posibles combinaciones de validez y fiabilidad. En la casilla (a) se observa una puntuación válida y fiable, en la (b) y en la (c) se observan puntuaciones no válidas y no fiables, en la casilla (d) se observa una puntuación no válida pero fiable.



Figura 5. Representación gráfica de la validez y la fiabilidad

Fuente: elaboración propia. Cada cuadro representa un instrumento de medida con una combinación diferente de validez y fiabilidad. La diana representa la puntuación verdadera del paciente. Los puntos representan las puntuaciones obtenidas por el mismo paciente cada vez que responde a los ítems del instrumento.

La puntuación observada de una medida se puede representar mediante la siguiente fórmula¹⁸¹:

$$Y = \eta + \varepsilon$$

donde Y es la puntuación observada, η es la puntuación verdadera del paciente y ε es el error inherente a cualquier proceso de medición. Por lo tanto, para conocer la puntuación verdadera hay que conocer la cantidad de error, y ese es el objetivo de estudio de la fiabilidad¹⁸⁰.

Spearman demostró que el coeficiente de fiabilidad de una medida puede calcularse a partir de la correlación entre dos repeticiones de dicha medida²²¹. Según la teoría clásica de los tests²²¹, el coeficiente de fiabilidad de las puntuaciones obtenidas en un grupo de personas mediante un cuestionario, es el coeficiente de correlación entre ese conjunto de puntuaciones y otro conjunto de puntuaciones obtenidas al aplicar, de forma independiente, el mismo instrumento o uno equivalente al mismo grupo de personas. Puede entenderse como el coeficiente de correlación entre dos mediciones y como la proporción de variancia verdadera contenida en la variancia observada. La taxonomía COSMIN para la propiedad de medida de la fiabilidad indica que ésta es la proporción de la variancia total en la medición (σ_y^2) que se debe a diferencias "verdaderas" entre pacientes (σ_p^2)²⁰⁹. "Verdadero" en este contexto se refiere a la puntuación media que se obtendría si las mediciones se realizaran un número infinito de veces, por lo que este coeficiente hace referencia a la consistencia de la puntuación, y no a su validez²²². Según la definición COSMIN mencionada, el coeficiente de fiabilidad se puede expresar mediante:

$$\text{Coeficiente de fiabilidad} = \frac{\sigma_p^2}{\sigma_y^2} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{error}^2}$$

Como se observa en esta fórmula, los parámetros de fiabilidad son una proporción y pueden obtener valores entre 0 (ausencia total de fiabilidad) y 1 (fiabilidad perfecta)^{180,181}.

Según la taxonomía COSMIN, el término fiabilidad de un PROM incluye la consistencia interna, definida como el grado de interrelación entre los ítems; la medida de la fiabilidad en sí misma,

definida previamente como la proporción de la variancia total en la medición que se debe a diferencias verdaderas entre pacientes; y la medida del error, que es el error sistemático y aleatorio de la puntuación de un paciente que no se atribuye a verdaderos cambios en el constructo a medir²⁰⁹.

La evaluación de la fiabilidad admite diferentes diseños, entre los cuales se encuentran¹⁸⁰:

- Fiabilidad de medidas equivalentes: uso de dos instrumentos equivalentes en el mismo grupo de individuos.
- Fiabilidad de consistencia interna: aplicación de un único instrumento a una muestra de personas en un diseño de dos mitades, dividiendo el grupo en dos y calculando el coeficiente de cada mitad.
- Fiabilidad test-retest: uso de un único instrumento de medida en un único grupo en más de una ocasión para valorar la estabilidad de la puntuación en el tiempo. En algunos contextos se denomina coeficiente de reproducibilidad o de estabilidad de las puntuaciones.

El análisis de la fiabilidad utiliza diferentes parámetros en función del diseño de obtención de datos, la métrica de las variables y el tipo de acuerdo deseado¹⁸⁰, como se detalla en la Tabla 6.

Tabla 6. Tipos de parámetros más frecuentes en el análisis de la fiabilidad

Tipo de acuerdo	Métrica de las variables	Medidas equivalentes y test-retest	Dos mitades	Consistencia interna
Consistencia	Cuantitativa	Pearson ICC	Spearman-Brown Gutmann-Flanagan	Alfa de Cronbach Alfa estandarizada Theta de Armor Omega de McDonald ICC
	Ordinal	Spearman W de Kendall		v. ordinal de alfa v. ordinal de theta v. ordinal de omega
	Nominal	Phi C de Pearson		KR-20 v. ordinal de alfa v. ordinal de theta v. ordinal de omega
Acuerdo	Cuantitativa	Bland-Altman ICC	ICC	ICC
	Ordinal	Kappa ponderada		
	Nominal	Kappa		

ICC: coeficiente de correlación intraclase (en inglés *intraclass correlation coefficient*); KR-20: coeficiente de Kuder y Richarson; v.: versión. Tomado de: "Tipos de índices más frecuentes en el análisis de la fiabilidad en función del diseño de obtención de datos, la métrica de las variables y el tipo de acuerdo deseado" por Viladrich C, Doval E, Penelo E, 2019, *Medición: Fiabilidad y Validez*, p. 196, Copyright © Carme Viladrich Segué, Eduardo Doval Diéguez, Eva Penelo Werner.

Fiabilidad test-retest

Para evaluar la fiabilidad o el error de medición de un PROM se requieren mediciones repetidas en pacientes estables²⁰⁵. La fiabilidad test-retest se evalúa mediante la administración repetida del PROM a una misma población que se haya mantenido estable en relación con el constructo objeto de medición²⁰⁰. Por lo tanto, es necesario un diseño adecuado del estudio para poder llevar a cabo esta medida, ya que se requieren unas condiciones especiales tales como: que los pacientes se mantengan estables, respetar un intervalo de tiempo adecuado entre ambas administraciones del PROM, que las condiciones de medida sean similares en ambas ocasiones, y que las mediciones y las puntuaciones sean independientes²⁰⁵. COSMIN no especifica cuál es el intervalo de tiempo adecuado entre el test y el retest, pero indica que debe ser lo suficientemente largo para evitar la repetición de respuestas por memorización y lo suficientemente corto para garantizar que los pacientes permanezcan estables²¹³. Algunos autores recomiendan que este intervalo sea de alrededor de 2 semanas²⁰⁰, pero pueden existir razones que requieran otro criterio²²³.

Para el cálculo de la fiabilidad test-retest se emplean medidas de concordancia, habitualmente el coeficiente de correlación intraclass (en inglés *Intraclass correlation coefficient, ICC*)²⁰⁰, que es un coeficiente de consistencia y acuerdo para una puntuación cuantitativa¹⁸⁰. El consenso COSMIN recomienda el uso del ICC para medidas cuantitativas y el índice Kappa (o Kappa ponderado) en variables ordinales²⁰⁵. Existen diferentes fórmulas para calcular el ICC, pero todas ellas consisten en una proporción de variancias¹⁸¹: la variancia en la población (variancia interindividual) dividida por la variancia total, que es la variancia interindividual más la variancia intraindividual (error de medición), expresada como una relación entre 0 y 1²²³. Existen diferentes fórmulas para el cálculo del ICC en función de si se obtiene a partir de una o varias evaluaciones, uno o dos factores y de si se mide acuerdo o consistencia¹⁸⁰. El valor del ICC se aproxima a 1 cuando la variancia de error es despreciable comparada con la variancia del paciente, y se aproxima a 0 cuando la variancia de error es extremadamente grande comparada con la variancia del paciente, y este valor se obtiene en muestras muy homogéneas (se obtendría ICC = 0 si todos los pacientes obtuvieran la misma puntuación en el PROM)¹⁸¹. Se considera adecuado un valor de ICC $\geq 0,70$ ^{200,224}, pero se considera mejor si son superiores a 0,80 e incluso 0,90¹⁸¹.

Consistencia interna

La consistencia interna es el grado en que todos los ítems de un PROM miden el mismo constructo, en otras palabras, es una medida del grado en que los ítems están correlacionados

entre sí (son homogéneos)^{209,223}. La consistencia interna solamente es relevante en PROMs basados en un modelo reflectivo²¹³ y que pretenden medir un único concepto subyacente (constructo) utilizando múltiples ítems. En cambio, para los cuestionarios en los que los ítems evalúan diferentes aspectos de un fenómeno clínico complejo, los ítems no tienen por qué estar correlacionados, y por lo tanto la consistencia interna no es relevante²²³.

Para el cálculo de la consistencia interna se ha empleado habitualmente el coeficiente alfa de Cronbach (en inglés *Cronbach's Alpha coefficient*)²⁰⁰, que es el coeficiente de elección para variables continuas²¹³ porque es el promedio de todos los coeficientes de dos mitades¹⁸⁰. Para su cálculo no se necesitan mediciones con otros instrumentos equivalentes, ni mediciones repetidas del mismo PROM, ni diferentes observadores disponible¹⁸¹. El coeficiente se calcula de la siguiente manera:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

donde k corresponde al número de ítems de la escala, σ_i^2 a la variancia del ítem i , σ_x^2 a la variancia de la puntuación total de la escala. De esta fórmula se deduce la siguiente, que facilita la interpretación del coeficiente alfa de Cronbach¹⁸⁰.

$$\alpha = \frac{k^2 \bar{\sigma}_{ij}}{\sigma_x^2}; \text{ siendo: } \bar{\sigma}_{ij} = \frac{\sum_{i<j} \sigma_{ij}}{k(k-1)/2}$$

donde $\bar{\sigma}_{ij}$ es el promedio de las covariancias entre los ítems. Para su interpretación, hay que tener en cuenta que el promedio de covariancias entre los ítems i , es una estimación de la variancia común entre ellos. Es decir, si los ítems no tuvieran información común entre ellos, sus covariancias valdrían 0 ($\bar{\sigma}_{ij} = 0$) y el alfa de Cronbach tendría el valor 0, indicando que los ítems no son consistentes¹⁸⁰. Se debe tener en consideración que el valor alfa de Cronbach aumenta al aumentar el número de ítems del PROM¹⁸¹.

Se han propuesto valores de alfa de Cronbach de 0,70 como estándares mínimos apropiados para realizar comparaciones entre grupos²²³. Para comparaciones individuales que puedan conllevar decisiones diagnósticas el estándar mínimo se fija en 0,90 e incluso 0,95 para evitar que el intervalo de confianza de puntuaciones sea demasiado amplio^{200,210}.

2.5.4 Otros aspectos de los estudios de validez y fiabilidad de un PROM

Tal y como afirman los estándares APA²²⁵, la validación de un PROM es un proceso continuo por el cual se acumulan pruebas sólidas que dan soporte a la interpretación de las puntuaciones y a demostrar que se trata de una medida adecuada para el uso concreto que se le quiere dar.

Los estándares COSMIN¹⁹² son la referencia para el diseño y la evaluación de estudios de validación de PROMs. Entre ellos destaca el listado de comprobación del diseño de estudios para instrumentos de medición de resultados comunicados por los pacientes (PROMs)²¹³. Este documento establece los estándares de calidad para la evaluación de las diferentes propiedades psicométricas de los PROMs y se ha utilizado como referencia para el desarrollo de esta tesis doctoral. Además, el documento da unas recomendaciones generales para el diseño de estudios de propiedades psicométricas entre las que se encuentra la clara definición del objetivo de estudio y del constructo que se pretende medir con el PROM que se evalúa, así como la descripción de la población diana que será usuaria del PROM en el futuro. Se establece que se deben proporcionar de forma clara los criterios de selección de los pacientes y si estos son representativos de la población diana.

Cabe destacar que dicho documento proporciona un listado minucioso de los aspectos a tener en cuenta durante el análisis de la validez y la fiabilidad de un PROM, la mayoría de ellos comentados en los apartados previos. En términos prácticos, la validez aparente y de contenido deberían evaluarse siempre en estudios de validación de cualquier PROM¹⁸¹ como requisito para demostrar su adecuación al propósito específico del instrumento en una población concreta²⁰⁰. Si se dispone de un *gold standard* se debe analizar la validez de criterio, mientras que si no se dispone de él, se debe analizar la validez de constructo. En el caso de la validez de criterio, se suele hacer una única prueba de comparación¹⁸¹, mientras que en el caso de la validez de constructo se pueden realizar diferentes pruebas: validación estructural en el caso de multidimensionalidad y otras pruebas que confirman las hipótesis más relevantes formuladas. Cada prueba que se realiza añade evidencia a la validez de un PROM en un contexto específico¹⁸¹. En cuanto a la fiabilidad, existe un apartado concreto que establece los criterios de calidad de la consistencia interna y otro para el resto de las medidas de error y fiabilidad.

Para poder considerar como “muy buena” la calidad del estudio se necesita un mínimo de muestra para cada tipo de análisis. Para el análisis de la validez estructural se considera necesaria una muestra de 7 veces el número de ítems de la escala y un número superior a 100 participantes²¹³.

2.5.4.1 Prueba de campo (en inglés, *field-testing*)

En el caso de realizar una adaptación de un PROM, se debe dar información comparativa de las características psicométricas de los ítems con el objetivo de facilitar el análisis de la equivalencia estadística entre las versiones¹⁸². Después de la prueba piloto con una muestra reducida de participantes, se requiere de una muestra de mayor tamaño para poder realizar el análisis cuantitativo¹⁸¹ de las propiedades psicométricas.

Existen otras pruebas diferentes de las hasta ahora mencionadas que suelen llevarse a cabo durante el desarrollo de nuevos PROMs, pero que también deberían realizarse en la adaptación de PROMs a un nuevo idioma y/o cultura. Idealmente este proceso debería completarse antes de aplicarlo a la práctica clínica o a la investigación, sin embargo, debido a la falta de tiempo y recursos, es habitual que se evalúe durante su uso real¹⁸¹.

Análisis de las respuestas

El análisis de los valores perdidos (respuestas en blanco) y los patrones de no respuesta pueden poner de manifiesto algunos problemas. Una respuesta en blanco puede significar que el paciente no entiende el ítem, que no es aplicable a su contexto, que su respuesta no encaja en ninguna de las opciones disponibles, que no saben o que no quieren responder, etc¹⁸¹. Un patrón de no respuesta que muestre que varios participantes dejan en blanco los últimos ítems del PROM podría tener que ver con el hartazgo si hay un número excesivo de ítems o si estos son redundantes. No hay consenso sobre qué número de valores perdidos se considera aceptable; algunos autores de las guías COSMIN señalan que menos del 3% es aceptable, mientras que más del 15% es inaceptable¹⁸¹.

Se debe dar una descripción detallada de cómo se han tratado los valores perdidos²¹³. Existen diferentes métodos de imputación de valores perdidos que se deben sopesar en función de los datos faltantes, el patrón de no respuestas y el tipo de análisis estadístico²²⁶. Una de las formas más frecuentes y simples es la de atribuir al dato faltante la media de las respuestas de la misma persona en el resto de los ítems, lo que sería el valor más representativo de esa persona para ese ítem, dado que todos los ítems miden de forma homogénea el mismo constructo¹⁸⁰.

Otro aspecto relevante en el análisis de las respuestas de los ítems es la distribución de las puntuaciones en las diferentes categorías de respuesta. Se deben inspeccionar las respuestas para averiguar si en alguno de los ítems se acumulan las respuestas en una misma categoría. La manera de analizarlos es mediante tablas de frecuencia, donde se observa la media de puntuación de cada ítem y donde la variancia de cada ítem se expresa en la desviación estándar de cada uno de ellos.

La distribución de las respuestas en las diferentes categorías de respuesta de cada ítem también da información sobre la capacidad de discriminación de los ítems. Los ítems en los que la mayoría de la población obtiene la misma puntuación son poco capaces de discriminar entre pacientes (sanos y enfermos), por lo que contienen menos información¹⁸¹. El “efecto suelo” es el fenómeno que se produce al agruparse un porcentaje elevado de las respuestas de un ítem en la categoría de respuestas inferior. El “efecto techo” se refiere al mismo fenómeno, pero con la agrupación de respuestas en la categoría alta. Ambas situaciones restan capacidad discriminativa al ítem²²⁷.

Análisis relacionados con la consistencia interna

Otro aspecto interesante del análisis de los ítems es la correlación entre sí mismos y con la puntuación total; este aspecto tiene que ver con la consistencia interna. Las correlaciones entre-ítems y la correlación ítem-total (correlación entre el ítem y la puntuación total de la escala) indican si el ítem es consistente como parte de una escala. Las correlaciones entre-ítem de una misma dimensión o de una escala unidimensional deberían oscilar entre 0,2 y 0,5¹⁸¹. Correlaciones superiores a 0,7 indican redundancia en los ítems, por lo que en caso de estar desarrollando un nuevo PROM, se aconsejaría eliminar uno de los dos ítems¹⁸¹.

Otro parámetro relacionado con la consistencia interna que se suele calcular durante el desarrollo o la adaptación de un PROM es el valor alfa de Cronbach sin el ítem. Consiste en calcular el valor que tendría el alfa de Cronbach de la escala en caso de eliminar ese ítem. Esto sirve para averiguar qué ítems restan consistencia a la escala en un contexto y población concreta.

3

JUSTIFICACIÓN

3 Justificación

El EGPRN trazó una agenda específicamente diseñada para la investigación metodológica y de instrumentos, que incluía el desarrollo de la epidemiología de la AP basada en la salud centrada en el paciente. El EGPRN mostró interés por la búsqueda de una herramienta validada y factible para la evaluación de la depresión en AP, con el fin de utilizarla para proyectos colaborativos en toda Europa¹²⁹. En el año 2010 se creó un equipo internacional (ANEXO 1), formado por miembros del EGPRN y liderado por MF franceses, mediante la cooperación y la voluntad de participar en este estudio¹⁶.

El grupo FPDM del EGPRN se planteó seleccionar una herramienta para diagnosticar la depresión del paciente adulto por los MF y que permitiera su comparación en el ámbito de la AP europea. El estudio se diseñó y se ha llevado a cabo en cuatro fases:

1. En el año 2011 se realizó una revisión sistemática para identificar las escalas validadas para el diagnóstico de la depresión que hubiesen sido comparadas con una entrevista psiquiátrica basada en la clasificación DSM y que contaran con la participación de profesionales de AP. Esta revisión sistemática se actualizó en 2015 y se publicó en 2017¹²⁹. Tras ella se seleccionaron 7 escalas diferentes: las versiones de la Geriatric Depression Scale de 5 ítems (GDS 5), 15 ítems (GDS 15), y 30 ítems (GDS 30), Hospital Anxiety Depression Scale (HADS), Center for Epidemiologic Studies Depression Scale-Revised (CESD-R), Physical Symptom Checklist de 51 ítems (PSC-51), y Hopkins Symptoms Checklist de 25 ítems (HSCL-25).
2. En segundo lugar, siguiendo la metodología RAND/UCLA Appropriateness Method User's manual²²⁸, la HSCL-25 fue elegida por consenso por su eficiencia y fiabilidad, así como por su facilidad de uso en AP y la sencillez de su comprensión por parte de los pacientes. Esta fase se realizó en el año 2013 y los resultados se publicaron en 2018¹²⁸.
3. En la tercera fase, se realizó la traducción y adaptación transcultural de esta herramienta al español y a otros 8 idiomas de los equipos integrantes del grupo FPDM, incluyendo también el catalán y el gallego. Los investigadores principales del del proyecto europeo, el Dr. Patrice Nabbe y el Dr. Jean Yves le Reste, obtuvieron los permisos para realizar la traducción de la versión original. Durante el desarrollo de esta tesis doctoral se ha publicado un artículo donde se detalla el proceso de traducción y adaptación conjunta con armonización transcultural en los 9 idiomas participantes²²⁹. La doctoranda es

coautora de este artículo que se adjunta en el ANEXO 4, aunque no forma parte del compendio de esta tesis.

4. Finalmente, se analizaron la validez y la fiabilidad de la escala con la traducción obtenida en población de AP, comparándola con una entrevista psiquiátrica como referencia.

El proyecto de esta tesis doctoral se enmarca en la tercera y cuarta fases del estudio internacional con la versión española de la escala. Esta última fase se ha completado hasta la fecha únicamente con la versión francesa¹⁴⁰. Otros países participantes están en proceso de completar la última fase en el momento de elaborarse este manuscrito.

En el marco de este proyecto se llevó a cabo la traducción, adaptación y análisis de la validez y la fiabilidad de la HSCL-25 en español en la población de AP de España. Se ha completado el estudio analizando dos versiones breves de la misma escala: la HSCL-10 y la HSCL-5.

Durante el desarrollo del proyecto se ha potenciado la colaboración de redes y grupos de investigación, tanto en España como en el ámbito europeo. La justificación de este diseño es que, una vez finalizada la validación de las escalas con una metodología común en todos los idiomas de los países participantes en este proyecto, se dispondrá de una herramienta que permitirá obtener datos comparables procedentes de poblaciones dispares y con diferencias culturales.

4

HIPÓTESIS Y OBJETIVOS

4 Hipótesis y objetivos

4.1 Hipótesis

La Hopkins Symptom Checklist-25 (HSCL-25) en español ofrece una puntuación válida y fiable para el cribado de la depresión en las consultas de Atención Primaria de España.

4.1.1 Objetivo general

Evaluar la validez y fiabilidad de la HSCL-25 en español para el cribado de la depresión en pacientes de Atención Primaria en España.

4.1.2 Objetivos específicos

1. Describir el proceso de traducción y adaptación transcultural de la HSCL-25 para su uso en Atención Primaria en España.
2. Analizar el proceso de respuesta de los pacientes de Atención Primaria a la HSCL-25 en español.
3. Analizar la validez de constructo de la HSCL-25 en español.
4. Analizar la validez de criterio de la HSCL-25 en español.
5. Analizar la fiabilidad de la HSCL-25 en español.
6. Evaluar la validez y fiabilidad de las escalas breves Hopkins Symptom Checklist-10 (HSCL-10) y Hopkins Symptom Checklist-5 (HSCL-5) como alternativas a la HSCL-25.

5

MÉTODOS

5 Métodos

Se han publicado tres artículos originales en revistas indexadas en el proceso de dar respuesta a los objetivos de esta tesis doctoral.

Metodológicamente hablando, la tesis doctoral consta de dos fases distintas:

1. **FASE 1: Traducción y adaptación transcultural de la HSCL-25 al español.** Se realiza un análisis cuantitativo y cualitativo para realizar el proceso de traducción y adaptación transcultural de la escala a partir de su versión original en inglés. En ella participan MF y un grupo reducido de pacientes.
2. **FASE 2: Análisis de la validez y fiabilidad de la HSCL-25 en español.** Se realiza un análisis cuantitativo a partir de la administración de la versión española de la escala obtenida previamente a pacientes reales de AP de diferentes CS de España, con el objetivo de analizar las propiedades psicométricas.

5.1 FASE 1: Traducción y adaptación transcultural de la HSCL-25 al español

5.1.1 Diseño y ámbito

Traducción directa, consenso con metodología Delphi, retrotraducción, clasificación de equivalencias, armonización europea, entrevista cognitiva a pacientes reales del ámbito de la AP.

Este diseño fue común y simultáneo con el del resto de los países miembros del grupo FPDM del EGPRN²²⁹, que lo realizaron en sus idiomas correspondientes.

5.1.2 Participantes y muestra

Se realizó una selección a conveniencia de MF de diferentes equipos de AP de Cataluña y Galicia que hubieran colaborado en estudios de Salud Mental para participar en el estudio piloto con metodología Delphi. Se solicitó evaluación a 73 MF, los profesionales debían tener nivel medio de lengua inglesa, actividad clínica e investigadora en AP, y acreditación de publicaciones en lengua inglesa o actividad como tutores de la especialidad de MFyC.

Se realizó una selección a conveniencia de pacientes captados en consulta por sus MF en días consecutivos que quisieran participar en la entrevista cognitiva. Se entrevistó a 10 pacientes mayores de 50 años que debían tener el español como lengua materna.

5.1.3 Recogida de datos

5.1.3.1 Variables

HSCL-25:

- Se entregó a los MF la versión original en inglés y la traducción directa al español (disponibles en el ANEXO 5) de la HSCL-25. En dicho documento constaban los 25 y ítems y las normas de administración, que también fueron traducidas.

De los MF participantes en el estudio Delphi:

- Género: Hombre/mujer.
- Años de experiencia como médico/a: menos de 10, entre 10 y 20, más de 20.
- Años de experiencia como investigador/a: menos de 10, entre 10 y 20, más de 20.
- Años de experiencia como tutor/a de residentes de MFyC: menos de 10, entre 10 y 20, más de 20.
- Grado de acuerdo sobre la adecuación de la traducción al español de la HSCL-25 respecto a la escala original: puntuando entre 1 (totalmente en desacuerdo) y 9 (totalmente de acuerdo) cada ítem y las instrucciones. Las puntuaciones inferiores o iguales a 3 indicaron que la traducción no se consideró apropiada; entre 4 y 6 se estimaron dudosas; y entre 7 y 9, apropiadas.

De los pacientes participantes en las entrevistas cognitivas:

- Género: Hombre/mujer.
- Edad (años).

5.1.3.2 Fases de la recogida de datos

Este proceso se llevó a cabo en el año 2013 y se realizó siguiendo estos pasos:

Traducción directa: La escala fue traducida al español por una graduada en Traducción e Interpretación.

Metodología Delphi: Mediante encuesta electrónica se solicitó a los MF que emitiesen su grado de acuerdo sobre la adecuación de la traducción respecto a la escala original para cada ítem y para las instrucciones. Se definió como traducción consensuada aquella a la que un porcentaje mínimo del 70% de MF le otorgaba una puntuación igual o superior a 7. De no alcanzar este valor umbral, se planteaba una segunda traducción y una nueva ronda Delphi hasta obtener consenso.

Retrotraducción: Realizada por una filóloga para el español que no había participado en la fase inicial ni conocía la escala original.

Clasificación de equivalencias: Se clasificaron las equivalencias de las traducciones como semánticas, idiomáticas, experienciales y conceptuales. Existe equivalencia semántica cuando hay correspondencia de significado entre el cuestionario en la lengua de partida y su versión correspondiente en la lengua final. Por equivalencias idiomáticas se entienden aquellas expresiones idiomáticas y coloquiales, que muchas veces entrañan dificultades a la hora de traducir, puesto que no cuentan con un equivalente fácil en la lengua materna. Las equivalencias experienciales tienen que ver con aspectos socioculturales y pueden diferir considerablemente con respecto a la cultura de la lengua del cuestionario original. Finalmente, en las equivalencias conceptuales se deberá atender a si las palabras del cuestionario original denotan el mismo concepto en ambas lenguas. Este proceso fue realizado por otra filóloga que no había trabajado en las fases previas.

Armonización europea: Se armonizaron las traducciones en todas las lenguas participantes del grupo FPDM del EGPRN y se compararon las retrotraducciones con la versión original. En aquellos ítems en los que había discrepancia entre las diferentes versiones, se valoró si estas eran culturales o semánticas.

Entrevista cognitiva: Se entrevistó a pacientes reales de AP pertenecientes a la población diana de la HSCL-25. Se les preguntó con relación a cada ítem: “¿Tiene alguna dificultad para entender esta pregunta?”, “Si tiene alguna dificultad, ¿cómo lo diría usted?”. Los pacientes debían responder sí o no a la primera pregunta. En caso de presentar alguna dificultad, podían escribir una redacción alternativa del ítem en cuestión.

5.1.4 Análisis estadístico

Para evaluar el primer objetivo específico sobre el proceso de traducción, se realizó una descripción detallada del proceso de traducción, retrotraducción y armonización transcultural.

Asimismo, se calculó la mediana para el grado de adecuación de la traducción directa, así como el grado de consenso alcanzado para cada ítem y las instrucciones en cada ronda Delphi.

En cuanto a la entrevista cognitiva, se recogieron las respuestas de los pacientes sobre las dificultades en la comprensión de los ítems.

5.2 FASE 2: Análisis de la validez y fiabilidad de la HSCL-25 en español

5.2.1 Diseño

Estudio transversal multicéntrico para el análisis de la validez y la fiabilidad de la HSCL-25 en español.

5.2.2 Periodo de realización

El periodo de reclutamiento y entrevista fue de 6 meses durante el año 2017.

5.2.3 Selección de la población

5.2.3.1 Ámbito de realización

Pacientes de AP atendidos en los 6 siguientes Centros de Salud (CS) participantes en el estudio EIRA²³⁰: Almozara, en Aragón; Safra Alcudi, en las Islas Baleares; y los siguientes 4 en Galicia: Matamá, Teis, Pintor Colmeiro-Nicolás Peña y Val Miñor.

Estos CS y sus profesionales cumplían con los requisitos de selección del proyecto EIRA²³⁰, que son los siguientes:

- **Centros de Salud:** 1) tener acceso a Internet; 2) tener la posibilidad de llevar a cabo actividades comunitarias; 3) no estar situados en zonas con una gran diversidad cultural y lingüística o en zonas turísticas; y 4) contar con un equipo directivo muy comprometido y activo.
- **Profesionales:** personal de enfermería y MF que aceptaron voluntariamente participar, que compartían la atención al mismo paciente y tenían previsto continuar en la plaza los próximos 30 meses.

5.2.3.2 Criterios de inclusión

Los propios del estudio EIRA²³⁰. Se incluyeron pacientes de 45 a 75 años con dos o más conductas o riesgos para la salud entre las siguientes:

1. Consumo de tabaco: consumo de 1 o más cigarrillos diarios durante el mes previo a la visita inicial.
2. Baja adherencia a la dieta mediterránea: se utilizaron dos preguntas validadas sobre el consumo diario de frutas y verduras: *“Habitualmente, ¿cuántas raciones de verdura*

comes al día?” y “Habitualmente, ¿cuántas piezas de fruta comes al día?”. Según este cuestionario, se considera un consumo insuficiente si el consumo de frutas y verduras diario es igual o inferior a 4²³¹.

3. Insuficiente actividad física: medida con el *Brief Physical Activity Assessment Tool*, que es un cuestionario breve de dos preguntas que miden la frecuencia y la duración de la actividad física moderada y vigorosa en una semana habitual de un individuo. Se considera actividad insuficiente una puntuación de entre 0 y 3 sobre 8^{232,233}.

5.2.3.3 Criterios de exclusión

Se excluyeron las personas con enfermedades graves avanzadas, deterioro cognitivo, dependencia, enfermedad mental severa, incluidas en un programa de atención domiciliaria y con cáncer en tratamiento activo o en fase terminal. También se excluyeron las personas que preveían ausentarse de su domicilio habitual durante el periodo de estudio.

5.2.3.4 Procedimiento de reclutamiento

Selección sistemática de personas que acudieron a la consulta por cualquier motivo y cumplían los criterios de selección, considerando cuotas de edad y sexo.

5.2.3.5 Tamaño de la muestra

Se siguieron las guías COSMIN (Consensus-based Standards for the selection of health Measurement Instruments) para el cálculo del tamaño muestral. En ellas se establece que para el análisis de la validez estructural son necesarios 7 cuestionarios por cada ítem que tiene la escala, con un mínimo de 100 pacientes que respondan al cuestionario. Dado que la escala consta de 25 ítems, se necesitan 175 pacientes como mínimo.

Para analizar la curva ROC de la puntuación de la HSCL-25 frente a la entrevista psiquiátrica CIDI y la correspondiente área bajo la curva (AUC), se utilizó el aplicativo BIOSOFT a través del siguiente enlace: <http://www.biosoft.hacettepe.edu.tr/easyROC/> y se calculó el tamaño de la muestra utilizando los siguientes parámetros:

1. El AUC para identificar casos de depresión descrita por Nettelbladt fue 0,742¹³⁸.
2. Se seleccionó un error tipo I de 5% y una potencia del estudio de 95%.

Bajo estos supuestos, se necesitan 87 casos y 174 controles. (AUC test nuevo: 0,80. AUC test estándar: 0,74. Razón caso/control 2). Teniendo en cuenta que la prevalencia estimada de depresión en la consulta de AP es del 16.3%³⁸, para obtener 87 casos se precisa que 533 pacientes respondan a la HSCL-25.

Para el cálculo de la muestra necesaria para el análisis de la fiabilidad test-retest, se tuvieron en cuenta las mismas consideraciones y la posibilidad de un 20% de pérdidas.

5.2.4 Recogida de datos

5.2.4.1 Variables

Se seleccionaron aquellas variables necesarias de las incluidas en el proyecto EIRA²³⁰.

Variables sociodemográficas:

- Sexo (hombre/mujer).
- Edad (años).
- País de nacimiento.
- Estado civil: Soltero/a; Casado/a o vive en pareja; Separado/a o divorciado/a; viudo/a.
- Actividad laboral: Estudiante; Está trabajando por cuenta ajena; Está trabajando por cuenta propia; Trabaja pero tiene una baja laboral de más de 3 meses; Paro con subsidio; Paro sin subsidio; Amo/a de casa; Incapacidad permanente; Jubilado/a.
- Ocupación actual.
- Nivel de estudios: Titulado Superior, universidad o similares; Escuela secundaria; Escuela primaria; Sin Estudios.

Variables relacionadas con estilos de vida:

- Tabaco: fumador sí/no.
- Frecuencia de actividad física:
 1. Frecuencia de actividad física moderada: 5 o más veces por semana; 3-4 veces por semana; 1-2 veces por semana; Nunca.
 2. Frecuencia de actividad física intensa: 3 o más veces por semana; 1-2 veces por semana; Nunca.
- Alimentación saludable: número de raciones de verdura consumida diariamente, número de piezas de fruta consumida diariamente.

Variables relacionadas con escalas y entrevistas:

Hopkins Symptom Checklist-25 (HSCL-25)

La HSCL-25 es la versión de 25 ítems^{130,134} derivada de la SCL-90-R y está formada por dos dimensiones: 10 ítems que hacen referencia a síntomas de ansiedad y 15 ítems que hacen

referencia a síntomas de depresión. Se utilizó la versión española²³⁴ obtenida de la Fase 1 y que se puede consultar en el ANEXO 6.

Se trata de una escala autoadministrada en la que el paciente debe puntuar cada ítem siguiendo una escala de Likert en relación con cómo se ha encontrado durante la semana previa. Cada ítem puede puntuarse del 1 al 4 con la siguiente relación de puntuaciones: 1 = “En absoluto”; 2 = “Un poco”; 3 = “Bastante” y 4 = “Mucho”. Para obtener la puntuación total, se suman las puntuaciones obtenidas en cada ítem y se dividen por el número total de ítems respondidos. Se puede obtener la puntuación de las subescalas de ansiedad y depresión de la misma manera. La puntuación total oscila entre 1 y 4. Para este estudio se utilizaron los puntos de corte propuestos en la bibliografía: caso probable de depresión si la puntuación media de la escala es $\geq 1,55$ y caso psiquiátrico que requiere tratamiento si la puntuación media es $\geq 1,75$ ^{136–139}.

Hay dos versiones más breves de esta escala que contienen 10 y 5 ítems y conforman la HSCL-10 y la HSCL-5 respectivamente. Los ítems de ambas escalas están incluidos en la versión de la HSCL-25 y pueden consultarse en el (ANEXO 2). Para considerar depresión se utilizaron los puntos de corte propuestos de 1,85 para el para la HSCL-10 y 2,00 para la HSCL-5¹⁵⁴.

Composite International Diagnostic Interview (CIDI)

La CIDI¹⁶⁹ es una entrevista psiquiátrica estructurada basada en los criterios CIE-10 y DSM que pregunta sobre síntomas psiquiátricos¹⁶⁷. Ha sido ampliamente utilizada en estudios epidemiológicos, clínicos y como medida de referencia para los estudios de validez y precisión diagnóstica de diferentes escalas. Para este estudio se utilizó la versión CIDI 2.1²³⁵, con la versión traducida al español¹⁷¹.

Los participantes respondieron a las preguntas de la sección E de depresión que hacen referencia al episodio depresivo (código CIE-10: F32), sin tener en cuenta el episodio recurrente (código CIE-10: F33), con el objetivo de detectar los pacientes con depresión actual. La sección E de depresión contiene dos preguntas iniciales de sondeo, solo es necesario continuar con el cuestionario CIDI en caso de una respuesta afirmativa a alguna de las dos preguntas. Las preguntas de la CIDI utilizadas en este estudio se pueden consultar en el ANEXO 7. La CIDI fue administrada por entrevistadores entrenados (el entrenamiento que obtuvieron los entrevistadores está detallado en el ANEXO 8).

El diagnóstico obtenido mediante la CIDI se utiliza como *gold standard* del diagnóstico de la depresión en este estudio.

Patient Health Questionnaire-9 (PHQ-9)

La PHQ-9 es una escala de evaluación de la depresión que contiene 9 ítems, cada uno de ellos puede puntuarse de 0 a 3 en función de la frecuencia de la sintomatología del paciente durante las últimas dos semanas de la siguiente forma: 0 = “nunca”, 1 = “algunos días”, 2 = “más de la mitad de los días” y 3 = “casi todos los días”. La puntuación total puede variar entre 0 y 27. El punto de corte establecido para depresión mayor moderada es 15 puntos^{21,22,96}. Se utilizó la versión de la escala validada en español⁹⁷ que está disponible en el ANEXO 9.

5.2.4.2 Recogida de datos

La recogida de datos se realizó mediante citación de pacientes y rellenando de la hoja de recogida de datos propia del estudio EIRA. Las personas seleccionadas fueron entrevistadas por la Unidad Externa del proyecto. Se les informó de la autoadministración de la HSCL-25, que devolvieron en un sobre cerrado, y posteriormente fueron entrevistados con las preguntas de la sección E de la entrevista estructurada CIDI. La entrevista estuvo a cargo de personal especialmente formado y entrenado en esta técnica, que desconocía los resultados obtenidos en la HSCL-25.

El retest de la HSCL-25 se realizó telefónicamente en el plazo de un mes. Para el retest participaron pacientes de los cuatro CS gallegos Matamá, Teis, Pintor Colmeiro-Nicolás Peña y Val Miñor. La muestra se seleccionó a conveniencia en aquellos pacientes que quisieron participar en el retest.

En todo momento se siguieron los criterios de práctica habitual para el tratamiento del paciente en el caso de hallazgos patológicos.

5.2.5 Análisis estadístico

Para evaluar el objetivo principal sobre el análisis de la fiabilidad y la validez de la HSCL-25 en español se llevaron a cabo a varios análisis que se detallan a continuación.

Para evaluar el segundo objetivo específico sobre el análisis del proceso de respuesta, se realizó inicialmente la búsqueda de valores perdidos y patrones de no respuesta. Posteriormente se llevó a cabo la imputación de valores perdidos de las respuestas a los ítems de la escala con la media de respuestas del individuo al resto de los ítems de la escala (valor más representativo de la persona)¹⁸². Se descartaron individuos con más del 50% de los ítems sin respuesta. Se realizó la misma imputación para los valores del retest. Se analizó la distribución de frecuencia de respuesta a los ítems por categorías de respuesta de forma global y por sexo. Se analizó la

capacidad discriminante de los ítems mediante comparación de dos grupos extremos dividiendo la muestra en tercios en función de la puntuación total obtenida. El índice de discriminación (ID) de cada ítem se calculó a partir de la diferencia de medias en cada ítem del tercio de la muestra con una puntuación total de la escala más elevada con el tercio que obtuvo la puntuación total más baja. Dado que las opciones de respuesta de la escala contienen 4 categorías posibles, el ID puede tomar valores entre -3 y +3. Los ID positivos y con valor más elevado son los que mejor discriminan entre población sana y enferma, mientras que valores ID negativos indicarían que la población sana obtiene una puntuación más elevada en dicho ítem (lo que podría significar que se necesita recodificar las categorías de respuesta de ese ítem de forma inversa).

Para evaluar el tercer objetivo específico sobre la validez de constructo se llevó a cabo el análisis de la estructura interna de la HSCL-25. Se condujo el análisis factorial confirmatorio (CFA) en base a la estructura conocida de la versión original inglesa de un único factor y dos factores (ansiedad y depresión) correlacionados. Se calcularon las cargas factoriales para ambos modelos. Se utilizó el método de estimación robusto de MLM (*Maximum Likelihood Mean Adjusted*) para realizar el análisis factorial con valores estandarizados. Para valorar el ajuste del modelo estimado se calculó el índice de ajuste absoluto con Chi cuadrado. Dado que este valor puede estar afectado por el tamaño muestral^{180,181}, se utilizaron complementariamente los siguientes índices: error de aproximación cuadrático (en inglés *Root Mean Square Error of Approximation, RMSEA*), residual estandarizado cuadrático medio (*Standardized Root Mean square Residual, SRMR*), el coeficiente de determinación (en inglés *coefficient of determination, CD*) y los índices comparativos: índice de ajuste comparativo (en inglés *Comparative Fit Index, CFI*) y el índice de Tucker-Lewis (en inglés *Tucker-Lewis Fit Index, TLI*)^{181,236}. Los valores de corte considerados adecuados fueron SRMR < 0,05; RMSEA ≤ 0,08; CFI > 0,90 y TLI > 0,90¹⁸⁰. Se consideró adecuada una carga factorial ≥ 0,3 en el factor esperado²¹⁹.

Se analizó también la validez convergente de la escala PHQ-9 y la HSCL-25 como medidas del mismo constructo “depresión”. Se calculó el coeficiente de correlación de Pearson para dos medidas cuantitativas; se consideró adecuada una correlación positiva superior o igual a 0,7²³⁷. Como medida adicional y convirtiendo ambas variables en dicotómicas a partir de sus puntos de corte conocidos (1,75 para la HSCL-25^{144,145} y de 15 para el PHQ-9⁹⁶), se calculó el índice Kappa ajustado por prevalencia y sesgo.

Para evaluar el cuarto objetivo específico sobre el análisis de la validez de criterio, se calcularon las curvas operativas del receptor (ROC) para la HSCL-25 en comparación con el *gold standard* CIDI. COSMIN establece que en el caso de variables cuantitativas como la puntuación de la HSCL-25, el cálculo AUC es una muy buena medida de la validez de criterio. En caso de utilizar puntos

de corte y convertir la variable HSCL-25 en dicotómica (sano/enfermo) se deben determinar la sensibilidad y especificidad como parámetros para el análisis de la validez de criterio²¹³. La curva ROC presenta de forma gráfica la sensibilidad en función de los falsos positivos (complementario de la especificidad) para distintos puntos de corte de una variable cuantitativa frente a una dicotómica de referencia, por este motivo es útil para elegir el punto de corte más adecuado de una prueba, conocer el rendimiento global de esta y comparar la capacidad discriminativa de dos o más pruebas diagnósticas²¹⁶. El área bajo la curva ROC (en inglés *Area Under Curve, AUC*) está ampliamente reconocida como la medida del poder discriminatorio de una prueba diagnóstica²¹⁴, los valores de AUC se interpretan de la siguiente forma: 0,5-0,6 (pésimo), 0,6-0,7 (malo), 0,7-0,8 (pobre), 0,8-0,9 (bueno), > 0,9 (excelente)²¹⁵. Para el análisis de la validez de criterio se calculó el AUC con un intervalo de confianza del 95% de forma global y por sexo. También se calcularon la sensibilidad, la especificidad, el valor predictivo positivo (VPP) y el valor predictivo negativo (VPN).

Con el fin de seleccionar el mejor punto de corte que conjuntamente obtuviera la mayor sensibilidad y especificidad, se calculó el Índice de Youden, que se define como:

$$\text{Índice de Youden (J)} = \text{Sensibilidad} + \text{Especificidad} - 1$$

su valor indica (J) la validez de un instrumento para un punto de corte específico²³⁸. El Índice de Youden puede tomar valores entre 0 y 1: La separación completa de las distribuciones de los valores de los marcadores para las poblaciones enferma y sana da como resultado $J = 1$, mientras que la superposición completa da como resultado $J = 0$. El índice de Youden tiene una característica atractiva que no está presente en el AUC, ya que proporciona un criterio para elegir el valor umbral o punto de corte óptimo para el que se maximiza la "sensibilidad + especificidad - 1"²³⁹. Se considera un valor adecuado del Índice de Youden de alrededor de 0,6²³⁹.

Para la evaluación del quinto objetivo específico sobre el análisis de la fiabilidad, se analizó la contribución de los ítems a la consistencia interna con indicadores basados en la correlación (homogeneidad), en la covariancia (coeficiente alfa de Cronbach) y en la regresión (coeficiente de determinación). Se calcularon las correlaciones entre los 25 ítems y se consideraron adecuados valores de correlación entre-ítem de entre 0,2 y 0,5¹⁸¹. Asimismo, se calculó la correlación ítem-total con el coeficiente de correlación de Spearman entre el ítem y la puntuación total de la escala calculada a partir de la suma de todos los ítems de la escala menos el ítem evaluado¹⁸². Se consideraron adecuados valores $\geq 0,4$ ²⁴⁰. Se calculó el coeficiente alfa de Cronbach total y para cada una de las dos subescalas de ansiedad y depresión. Se consideró adecuado un valor igual o superior a 0,7²²³. También se calculó el valor alfa de Cronbach sin el

ítem, que corresponde al valor que se obtendría en este parámetro en caso de eliminar dicho ítem. Valores alfa sin el ítem superiores al alfa de Cronbach global indican mala consistencia de dicho ítem, ya que en caso de eliminarlo la fiabilidad global de la escala mejoraría. Para este cálculo puede ser necesario utilizar varios decimales por lo que se recomienda presentar los coeficientes con un mínimo de tres decimales¹⁸². Como medida adicional para el análisis de la consistencia de los ítems se calculó el coeficiente de determinación (R^2) a partir de la regresión múltiple de la puntuación de cada ítem como variable dependiente y la puntuación de los demás ítems como variables independientes. Este valor se puede interpretar como la proporción de la variancia explicada por el resto de los ítems y cuanto más alto es su valor, más homogéneo es el ítem¹⁸².

Como otra medida de fiabilidad diferente, se calculó la fiabilidad test-retest mediante el coeficiente de correlación intraclass (ICC) para el modelo de dos factores con efectos fijos de tipo acuerdo, para el promedio de las dos evaluaciones (basal y retest). El ICC es el parámetro de elección y se considera adecuado un valor igual o superior a 0,7¹⁸¹.

Finalmente, para evaluar el sexto objetivo específico sobre la validez y la fiabilidad de las escalas breves HSCL-10 y HSCL-5, se realizaron los análisis a partir de las respuestas obtenidas a la versión larga de 25 ítems. El análisis fue similar al realizado con la HSCL-25. Para el análisis de la validez de criterio se realizaron los mismos cálculos que se hicieron con la HSCL-25. Para el análisis de la validez de constructo se realizó el CFA de la misma forma que con la HSCL-25 y se añadió además el análisis de comprobación de hipótesis mediante la comparación de grupos conocidos. Se comparó la puntuación obtenida por sexos, dado que se conoce una prevalencia de depresión más elevada en mujeres^{19,32,41}. Se realizó la prueba t de Student para medidas independientes y se consideró satisfactorio un resultado estadísticamente significativo ($p < 0,05$)²⁴¹. La fiabilidad se analizó mediante el coeficiente alfa de Cronbach de manera global y para ambas subescalas ansiedad y depresión. También se calculó el valor alfa de Cronbach sin el ítem. Como valores adecuados para todos los parámetros calculados se utilizaron los mismos que se han mencionado en los cálculos para la HSCL-25.

5.3 Consideraciones éticas en la Fase 1 y la Fase 2

El estudio realizado se llevó a cabo siguiendo las normas de la Declaración de Helsinki y Tokio sobre aspectos éticos para las investigaciones médicas en seres humanos²⁴².

La participación de los MF y pacientes de la primera fase fue voluntaria. Para la primera fase se obtuvo el consentimiento por escrito de los MF participantes. En el caso de los pacientes no se

solicitó por escrito al no cumplimentar la escala ni grabar sus opiniones en ningún formato, ya que únicamente se solicitó su opinión sobre la comprensión e interpretación de los ítems.

Para la segunda fase, el protocolo del estudio fue aprobado por el Comité de Ética de la Fundación Instituto Universitario para la Investigación en Atención Primaria de Salud Jordi Gol i Gurina (número de protocolo P16/025). Se obtuvo el consentimiento informado por escrito de todos los participantes. Los datos se trataron de forma anónima. Los cuestionarios se codificaron con un número de identificación para mantener la confidencialidad de los participantes y solo se destinaron para los objetivos del estudio. Se garantizó la confidencialidad de los datos personales conforme lo que dispone la Ley Orgánica de Protección de Datos de Carácter Personal (15/1999 del 13 de Diciembre, LOPD)²⁴³ vigente en el momento de la aprobación del estudio y es conforme a lo que dispone la ley Orgánica 03/2018, de 5 de diciembre de protección de Datos Personales y garantía de los derechos digitales²⁴⁴.

En la muestra de población participante se encontró un individuo cuya categoría en cuanto a estado civil era única para esa persona. En el tercer artículo se presentó un análisis que relacionaba el estado civil con las puntuaciones obtenidas en las versiones de la HSCL. Este individuo fue incluido en otra de las categorías para proteger su intimidad dada la excepcionalidad de su situación personal.

6

RESULTADOS

6 Resultados

A continuación, se adjuntan los tres artículos originales publicados con los resultados de esta tesis doctoral:

1. Clavería A, Rodríguez-Barragán M, Fernández-San-Martín MI, Nabbe P, Le Reste JY, Miguéns-Blanco I, Cossio-Gil Y. Traducción y adaptación transcultural al español, catalán y gallego de la escala Hopkins Symptom Checklist-25 para la detección de depresión en Atención Primaria [Translation and cross-cultural adaptation into Spanish, Catalan and Galician of the Hopkins Symptom Checklist-25 to identify depression in Primary Care]. *Aten Primaria*. 2020 Oct;52(8):539-547. Spanish. doi: 10.1016/j.aprim.2020.05.017. Epub 2020 Jul 20. PMID: 32703629; PMCID: PMC7505899.
2. Rodríguez-Barragán M, Fernández-San-Martín MI, Clavería-Fontán A, Aldecoa-Landesa S, Casajuana-Closas M, Llobera J, Oliván-Blázquez B, Peguero-Rodríguez E. Validation and Psychometric Properties of the Spanish Version of the Hopkins Symptom Checklist-25 Scale for Depression Detection in Primary Care. *Int J Environ Res Public Health*. 2021 Jul 24;18(15):7843. doi: 10.3390/ijerph18157843. PMID: 34360136; PMCID: PMC8345472.
3. Rodríguez-Barragán M, Fernández-San-Martín MI, Clavería A, Le Reste JY, Nabbe P, Motrico E, Gómez-Gómez I, Peguero-Rodríguez E. Measuring depression in Primary Health Care in Spain: Psychometric properties and diagnostic accuracy of HSCL-5 and HSCL-10. *Front Med (Lausanne)*. 2023 Jan 9;9:1014340. doi: 10.3389/fmed.2022.1014340. PMID: 36698836; PMCID: PMC9869680.



ORIGINAL

Traducción y adaptación transcultural al español, catalán y gallego de la escala Hopkins Symptom Checklist-25 para la detección de depresión en Atención Primaria



Ana Clavería^a, María Rodríguez-Barragán^{b,c,d,*},
María Isabel Fernández-San-Martín^{c,e}, Patrice Nabbe^f, Jean Yves Le Reste^f,
Iria Miguéns-Blanco^g e Yolima Cossio-Gil^h

^a Xerencia Xestión Integrada de Vigo, Servizo Galego de Saúde, Instituto de Investigación Sanitaria Galicia-Sur, Red de Investigación en Actividades Preventivas y de Promoción de la Salud, Vigo, Pontevedra, España

^b Centro de Atención Primaria La Mina, Gerencia Territorial de Atención Primaria de Barcelona, Instituto Catalán de la Salud, Sant Adrià de Besòs, Barcelona, España

^c Fundación Instituto Universitario de Investigación en Atención Primaria de Salud Jordi Gol i Gurina (IDIAP Jordi Gol), Barcelona, España

^d Universidad Autónoma de Barcelona, Bellaterra, Barcelona, España

^e Gerencia Territorial de Atención Primaria de Barcelona, Instituto Catalán de la Salud, Barcelona, España

^f Département de Médecine Générale, ERCR SPURBO, Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale, Brest, Bretagne, Francia

^g Servicio de Urgencias, Hospital General Universitario Gregorio Marañón, Madrid, España

^h Dirección de sistemas de información y soporte a la decisión, Hospital Universitario Vall d'Hebron, Barcelona, España

Recibido el 4 de febrero de 2020; aceptado el 8 de mayo de 2020

Disponible en Internet el 20 de julio de 2020

PALABRAS CLAVE

Atención primaria de salud;
Depresión;
Trastorno depresivo;
Encuestas y cuestionarios;
Técnica Delphi

Resumen

Objetivo: Describir el proceso de traducción y adaptación transcultural de la escala Hopkins Symptom Checklist-25 (HSCL-25) al español, catalán y gallego.

Diseño: Traducción, adaptación transcultural y análisis de la comprensibilidad mediante entrevistas cognitivas.

Emplazamiento: Unidades de Investigación de Atención Primaria de Barcelona y Vigo.

Participantes: Médicos de familia y pacientes de Atención Primaria.

Mediciones principales: Siguiendo las guías de la International Society for Pharmacoeconomics and Outcomes Research (ISPOR), se realizaron: 1) traducción directa; 2) estudio piloto basado en metodología Delphi con médicos de familia; 3) retrotraducción; 4) análisis de equivalencias; 5) análisis de comprensibilidad de las versiones obtenidas en español, catalán y gallego mediante entrevista cognitiva en una muestra de pacientes, y 6) armonización transcultural.

* Autor para correspondencia.

Correo electrónico: marodriguez.bcn.ics@gencat.cat (M. Rodríguez-Barragán).

KEYWORDS

Primary health care;
Depression;
Depressive disorder;
Surveys and
questionnaires;
Delphi technique

Resultados: En el estudio Delphi participaron 73 médicos de familia. El consenso se estableció en la primera ronda para la traducción española y catalana, y en la segunda ronda para la gallega. Las retrotraducciones fueron similares en los 3 idiomas. Todas las versiones fueron equivalentes entre ellas y respecto a la versión original inglesa. En la entrevista cognitiva participaron 10 pacientes por cada idioma, sin que se modificara la redacción de los ítems.

Conclusiones: Las traducciones de la escala HSCL-25 en español, catalán y gallego son equivalentes semántica y conceptualmente a la versión original. Las traducciones son comprensibles y bien aceptadas por los pacientes.

© 2020 Los Autores. Publicado por Elsevier España, S.L.U. Este es un artículo Open Access bajo la licencia CC BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Translation and cross-cultural adaptation into Spanish, Catalan and Galician of the Hopkins Symptom Checklist-25 to identify depression in Primary Care

Abstract

Aim: To describe the translation and cross-cultural adaptation process of the Hopkins Symptom Checklist-25 (HSCL-25) scale into Spanish, Catalan and Galician.

Design: Translation, cross-cultural adaption and comprehensibility analysis through cognitive debriefing.

Location: Research Units of Primary Care in Barcelona and Vigo.

Participants: Family doctors and Primary Care patients.

Main measurements: Following the guidelines of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR): 1) Direct translation. 2) Pilot study based on Delphi methodology with family doctors. 3) Back-translation. 4) Equivalence analysis. 5) Comprehension analysis of versions obtained in Spanish, Catalan and Galician through cognitive debriefing in a sample of patients. 6) Transcultural harmonization.

Results: 73 family doctors participated in the Delphi study. The consensus was established in the first round for the Spanish and Catalan translations, and in the second round for the Galician. The back-translations were similar in all 3 languages. All versions were equivalent between them and compared to the original English version. In the cognitive interview, 10 patients participated for each language, without modifying the writing of the items.

Conclusions: The translations of the HSCL-25 scale in Spanish, Catalan and Galician are semantically and conceptually equivalent to the original version. Translations are understandable and well accepted by patients.

© 2020 The Authors. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introducción

La depresión es un problema de salud frecuente en la población general. Según el estudio ESEMed, su prevalencia en la población española a lo largo de la vida es del 10,6%¹; en el año 2017, su prevalencia fue del 10,2% en mayores de 45 años². Estudios realizados con pacientes de Atención Primaria (AP) sitúan esta prevalencia entre un 9,6³ y un 20,2%^{4,5}, siendo más elevada en mujeres mayores de 50 años, asociada a condiciones de viudez, soledad y aislamiento social⁶.

El diagnóstico de trastornos mentales desde AP es complejo, siendo la depresión una de las enfermedades mentales mejor detectadas por los médicos de familia (MF)⁷. La variabilidad en su diagnóstico entre facultativos es amplia debido a factores personales, sociales y culturales, que pueden llevar a sobrevalorar o infravalorar el malestar psicológico de los pacientes⁸. A pesar de que el Diagnostic and Statistical Manual of Mental Disorders (DSM-5)⁹ es el referente para su diagnóstico clínico, raramente se utiliza en AP y

el debate sobre qué cuestionario/s utilizar para ayudar a su identificación resulta controvertido¹⁰. Se han publicado diferentes cuestionarios que cuentan con traducciones al español¹¹⁻¹⁵, sin embargo, suele haber insuficiente información sobre el procedimiento de adaptación transcultural, a pesar de que la International Society for Pharmacoeconomics and Outcomes Research (ISPOR) especificó estas cuestiones metodológicas¹⁶.

En este contexto, el grupo Family Practice Depression and Multimorbidity (FPDM) del European General Practice Research Network (EGPRN) se planteó seleccionar una herramienta para detectar la depresión en pacientes adultos que permitiera su comparación en el ámbito de la AP de diferentes países europeos¹⁷.

Inicialmente, se efectuó una revisión sistemática con el fin de identificar las escalas validadas para el cribado y la detección de la depresión en adultos y ancianos que hubiesen sido comparadas con una entrevista psiquiátrica basada en la clasificación DSM, con resultados cuantitativos, y que hubiesen contado con la participación de profesionales de

AP. Como resultado de la revisión sistemática, se seleccionaron 7 escalas diferentes¹⁸.

En segundo lugar, la escala Hopkins Symptom Checklist-25 (HSCL-25)¹⁹ fue elegida por consenso. Entre los criterios cuantitativos, se analizaron la validez, la fiabilidad y la reproducibilidad. Se analizó la validez predictiva, es decir, la capacidad de detectar correctamente a pacientes con depresión (sensibilidad) y a los que no la presentan (especificidad). La fiabilidad hace referencia a la precisión de las mediciones. Para valorar la ergonomía, se atendió a la autoadministración de la escala, al grado de facilidad de su cumplimentación y a la sencillez de su interpretación²⁰.

Puesto que la expresión emocional está intrínsecamente ligada al entorno lingüístico y cultural de los pacientes, parece relevante ofrecerles un cuestionario escrito en su lengua materna. Para poder utilizar la escala HSCL-25 en el entorno de AP de nuestro país, es necesaria su traducción y posterior adaptación a la cultura española, garantizando la equivalencia con la versión original. Este proceso se ha realizado simultáneamente en 13 idiomas de los equipos integrantes del grupo FPDM del EGPRN.

El objetivo de este estudio es traducir y adaptar transculturalmente la escala HSCL-25 al español, catalán y gallego.

Material y métodos

La escala HSCL-25¹⁹ es una versión acortada que proviene de la escala Hopkins Symptom Checklist-58 (HSCL-58), diseñada por Derogatis. Es un cuestionario de 25 ítems con 2 dimensiones conocidas: la primera de ellas corresponde a los ítems del 1 al 10 que hacen referencia a síntomas de ansiedad; la segunda dimensión, comprendida por los ítems 11 a 25, es relativa a síntomas de depresión. Cada ítem se responde según una escala tipo Likert de 1 a 4: «en absoluto» (1), «un poco» (2), «bastante» (3) o «mucho» (4). Se calculan 2 puntuaciones: la puntuación total (media de 25 ítems) y la de depresión (media de 15 ítems de la subescala de depresión). La puntuación final se calcula sumando las puntuaciones de cada ítem y dividiendo la suma entre el número de ítems, obteniendo una puntuación numérica entre 1 y 4. El diagnóstico de sospecha de depresión se considera si la puntuación total es $\geq 1,75$ ^{21,22}. Se necesitan entre 5 y 10 min para completar la escala²³.

La traducción y la adaptación transcultural se basó en las recomendaciones de la ISPOR y se diseñó un protocolo común para todo el grupo FPDM²⁴. El presente estudio se realizó en las Unidades de Investigación de AP de Barcelona y Vigo. Se constituyeron equipos de 4 investigadores para el gallego y el catalán, que participaron conjuntamente en la traducción al español. Se tradujeron los ítems de la escala y las categorías de respuesta, así como las instrucciones necesarias para su uso.

Traducción directa

La escala fue traducida al gallego y al español por una graduada en Traducción e Interpretación y al catalán por un licenciado en Filología inglesa.

Estudio piloto con metodología Delphi

La traducción realizada fue analizada con metodología Delphi. Se realizó una selección a conveniencia de MF de diferentes equipos de AP de Cataluña y Galicia que hubieran colaborado en estudios de Salud Mental. Se solicitó evaluación a 134 expertos: 73 para el español, 28 para el catalán y 45 para el gallego. Los profesionales debían tener nivel medio de lengua inglesa, actividad clínica e investigadora en AP, y acreditación de publicaciones en lengua inglesa o actividad como tutores. Se solicitó, mediante encuesta electrónica, que emitiesen su grado de acuerdo sobre la adecuación de la traducción respecto a la escala original, puntuando entre 1 (totalmente en desacuerdo) y 9 (totalmente de acuerdo) cada ítem, incluyendo las observaciones que considerasen pertinentes. Las puntuaciones inferiores o iguales a 3 indicaron que la traducción no se consideró apropiada; entre 4 y 6 se estimaron dudosas, y entre 7 y 9, apropiadas. Se definió como traducción consensuada aquella a la que un porcentaje mínimo del 70% de MF le otorgaba una puntuación igual o superior a 7. De no alcanzar este valor umbral, se planteaba una segunda traducción y una nueva ronda Delphi hasta obtener consenso.

Se analizaron de forma paralela las respuestas en los 3 idiomas. Se calculó la mediana para el grado de adecuación de la traducción, así como el grado de consenso alcanzado para cada ítem.

Retrotraducción

Realizada por una filóloga para el español y el gallego, y por un filólogo para el catalán. Ninguno de ellos había participado en la fase inicial ni conocía la escala original.

Clasificación de equivalencias

Se clasificaron las equivalencias de las traducciones como semánticas, idiomáticas, experienciales y conceptuales²⁵. Existe equivalencia semántica cuando hay correspondencia de significado entre el cuestionario en la lengua de partida y su versión correspondiente en la lengua final²⁶. Por equivalencias idiomáticas se entienden aquellas expresiones idiomáticas y coloquiales, que muchas veces entrañan dificultades a la hora de traducir, puesto que no cuentan con un equivalente fácil en la lengua materna. Las equivalencias experienciales tienen que ver con aspectos socioculturales y pueden diferir considerablemente con respecto a la cultura de la lengua del cuestionario original. Finalmente, en las equivalencias conceptuales se deberá atender a si las palabras del cuestionario original denotan el mismo concepto en ambas lenguas. Este proceso fue realizado por otra filóloga que no había trabajado en las fases previas.

Armonización europea

Se armonizaron las traducciones en todas las lenguas participantes del grupo FPDM y se compararon las retrotraducciones con la versión original. En aquellos ítems en los que había discrepancia entre las diferentes versiones, se valoró si estas eran culturales o semánticas.

Tabla 1 Características de los médicos de familia participantes en la revisión de la traducción con metodología Delphi

	Español (n = 31)		Gallego (n = 20)		Catalán (n = 22)	
	n	%	n	%	n	%
Sexo						
Mujer	11	34,37	6	30,0	12	54,5
Hombre	20	65,62	14	70,0	10	45,5
Años de experiencia como médico/a						
< 10	3	9,37	3	15,0	5	22,7
10-20	12	37,5	3	15,0	11	50,0
> 20	16	53,12	14	70,0	6	27,3
Años de experiencia como investigador/a						
< 10	15	46,87	10	50,0	14	63,6
10-20	14	43,75	7	35,0	6	27,3
> 20	2	9,375	3	15,0	2	9,1
Años de experiencia como tutor/a						
< 10	19	59,37	12	60,0	16	72,7
10-20	10	31,25	5	25,0	4	18,2
> 20	2	9,37	3	15,0	2	9,1

Entrevista cognitiva a pacientes

Se entrevistó a 10 pacientes en cada idioma, mayores de 50 años, captados por profesionales de AP en días sucesivos. Los pacientes debían ser nativos del idioma de la escala. Se preguntó para cada ítem: «¿Tiene alguna dificultad para entender esta pregunta?», «Si tiene alguna dificultad, ¿cómo lo diría usted?». Los pacientes debían responder sí o no a la primera pregunta. En caso de presentar alguna dificultad, podían escribir una redacción alternativa del ítem en cuestión.

Resultados

Traducción directa

Se realizó una versión para cada idioma (español, catalán y gallego).

Estudio piloto con metodología Delphi

Respondieron 31 MF al español, 22 al catalán y 20 al gallego. Algunos participaron tanto en la traducción al español como en la traducción a la lengua de uso habitual en su entorno geográfico. En la [tabla 1](#) se describen las características de los participantes.

Los participantes en la traducción al español contaron con una media de 23 años de experiencia como MF (rango intercuartílico [RIC] = 19-26), 10 años como tutores (RIC = 5-20) y 9 como investigadores (RIC = 1-15). Los participantes catalanes fueron los más jóvenes, con una experiencia médica e investigadora también inferior.

La traducción al español y al catalán alcanzó el consenso para todos los ítems en la primera ronda. La traducción al gallego requirió una segunda ronda para consensuar los ítems 6 (*ten tremores*), 10 (*séntese inqueda*) y 25

(*alteración do sono*). En la [tabla 2](#) se detallan los porcentajes de respuesta considerados aceptables para la traducción de cada ítem en cada uno de los idiomas.

Retrotraducción

La retrotraducción fue similar a la versión original en todos los idiomas. En gallego, el ítem que más difirió del original fue el 9 (*feeling panic*), traducido como *sente pánico* y retrotraducido como *having panic attacks*; en catalán fue el 18 (*thinking of ending one's life*), traducido como *pensaments de suïcidi* y retrotraducido como *thoughts of suicide*. Por último, en español fue el ítem 5 (*heart racing*), traducido como *palpitaciones* y retrotraducido como *palpitations*, el que más divergió del original.

Clasificación de equivalencias

Todas las opciones fueron semánticamente equivalentes; aunque en una lengua se utilice una expresión adjetival (*inqueda* en el ítem 10 del gallego) y en otra un sustantivo (*inquietud* en español), ya que se trata de equivalencias de sentido.

Armonización europea

Los miembros del grupo FPDM discutieron y resolvieron por consenso cualquier posible ambigüedad y discrepancia, tanto de la escala como de las instrucciones. En estas reuniones se modificó la traducción directa de *distress* en las 3 lenguas presentadas en este estudio. *Distress* se tradujo inicialmente como *angustia/angustia/angoixa*, pero en la versión final se optó por *malestar psicológico/psicológico/psicològic*. Este término, sin embargo, no forma parte del cuestionario, sino de las aclaraciones necesarias para su interpretación. En el catalán se cambió la

Tabla 2 Porcentaje de respuestas con puntuación ≥ 7 sobre la adecuación de la traducción de los ítems en cada idioma, según la metodología Delphi

	Español, %	Gallego, %	Catalán, %
Choose the best answer for how you felt over the past week	96,8	100,0	95,5
1. Being scared for no reason	80,7	80,0	95,5
2. Feeling fearful	90,3	90,0	90,9
3. Faintness	87,1	80,0	100,0
4. Nervousness	96,7	80,0	95,5
5. Heart racing	83,9	80,0	90,9
6. Trembling	87,1	91,3 ^a	100,0
7. Feeling tense	96,8	90,0	100,0
8. Headache	93,5	90,0	100,0
9. Feeling panic	83,8	80,0	95,5
10. Feeling restless	96,8	86,9 ^a	90,9
11. Feeling low in energy	83,9	80,0	86,4
12. Blaming oneself	77,4	80,0	95,5
13. Crying easily	90,4	95,0	95,5
14. Losing sexual interest	77,5	70,0	90,9
15. Feeling lonely	96,8	100,0	100,0
16. Feeling hopeless	87,1	70,0	100,0
17. Feeling blue	93,6	100,0	95,5
18. Thinking of ending one's life	77,5	85,0	90,9
19. Feeling trapped	87,2	90,0	100,0
20. Worrying too much	90,3	85,0	95,5
21. Feeling no interest	83,9	85,0	90,9
22. Feeling that everything is an effort	90,4	90,0	100,0
23. Worthless feeling	96,7	95,0	95,5
24. Poor appetite	93,6	80,0	100,0
25. Sleep disturbance	87,1	86,9 [*]	90,9
‘‘A little’’	87,1	85,0	100,0
‘‘Not at all’’	93,5	75,0	100,0
‘‘Extremely’’	93,6	85,0	100,0
‘‘Quite a bit’’	93,6	85,0	100,0

^a Porcentajes obtenidos en la segunda ronda Delphi con 23 participantes. Se considera la traducción aceptable si obtiene una puntuación ≥ 7 puntos

traducción de *pensaments de suicidi por pensa en treure's la vida*.

Entrevista cognitiva a pacientes

Se utilizó una muestra de 10 pacientes para cada idioma, las características de los pacientes participantes se muestran en la [tabla 3](#). En la entrevista cognitiva no se detectaron dificultades en la comprensión de los ítems que justificaran la modificación de la escala. Expresaron alguna dificultad de interpretación al menos 2 personas en el ítem 11 de la versión española, los ítems 11, 18 y 21 de la escala catalana y el 3 de la versión gallega.

Finalmente, se flexionaron aquellos sustantivos y adjetivos que admiten morfemas de género. En la [tabla 4](#) se expone la traducción final en los 3 idiomas.

Discusión

Se ha obtenido una versión adaptada al español, catalán y gallego de la escala HSCL-25 para medir síntomas de depresión.

El diagnóstico de depresión es fundamentalmente clínico y basado en los criterios DSM. Sin embargo, los cuestionarios son herramientas de cribado que ayudan a la identificación y medida de la intensidad de síntomas, a veces difíciles de reconocer a través de la anamnesis. A pesar de que existen otros instrumentos validados en español para un fin similar, esta escala se ha considerado mejor en cuanto a su comodidad para utilizarla en el contexto de AP debido a la facilidad para llevarla a cabo durante la consulta habitual del MF, así como por sus adecuadas propiedades psicométricas^{18,20}. En la versión francesa, se ha objetivado una sensibilidad del 59,1% y una especificidad del 91,4%²³. La escala HSCL-25 evalúa síntomas de depresión y ansiedad, esto supone una ventaja ya que es frecuente encontrar síntomas de ansiedad en pacientes diagnosticados de depresión y a que muchas veces ansiedad y depresión se presentan como comorbilidades asociadas entre sí²⁷. Asimismo, se trata de una herramienta fácil de comprender para los pacientes, aspecto fundamental dado que es un cuestionario autoadministrado. Por este motivo, se decidió traducirla a 3 idiomas oficiales de nuestro país, realizando también la adaptación cultural. Este paso es indispensable para poder

Tabla 3 Participantes y resultados de las entrevistas cognitivas

	Español (n = 10)		Gallego (n = 10)		Catalán (n = 10)	
	n	%	n	%	n	%
Sexo						
Mujer	7	70,0	6	60,0	6	60,0
Hombre	3	30,0	4	40,0	4	40,0
Edad						
Media (DE)	67,4	(11,0)	61,2	(10,5)	70,4	(10,7)
Ítems que presentan alguna dificultad para ser entendidos por al menos 2 pacientes						
	Ítem 11		Ítem 3		Ítem 11 Ítem 18 Ítem 21	

Tabla 4 Traducción final de la escala Hopkins Symptom Checklist-25 (HSCL-25) por idioma

Original	Español	Gallego	Catalán
Choose the best answer for how you felt over the past week	Elija la respuesta que mejor describa cómo se ha sentido durante la semana pasada	Escollo a resposta que mellor describa como se sentiu durante a semana pasada:	Triï la millor resposta per indicar com s'ha sentit en la darrera setmana
1. Being scared for no reason	Se asusta sin motivo	Asústase sen motivo	Estar espantat/espantada sense motiu aparent
2. Feeling fearful	Siente miedo	Ten medo	Sentir por
3. Faintness	Debilidad	Debilidade	Debilitat
4. Nervousness	Nerviosismo	Nerviosismo	Nerviosisme
5. Heart racing	Palpitaciones	Palpitaciïns	Cor accelerat
6. Trembling	Tiembla	Ten tremores	Tremola
7. Feeling tense	Se siente tenso/a	Séntese tenso/a	Sentir-se tens/a
8. Headache	Dolor de cabeza	Dor de cabeza	Mal de cap
9. Feeling panic	Siente pánico	Sente pánico	Sensació de pànic
10. Feeling restless	Siente inquietud	Séntese inqueda/a	Sensació d'inquietud
11. Feeling low in energy	Siente que le falta energía	Sente que lle falta enerxía	Sensació de manca d'energia
12. Blaming oneself	Se culpa a sí mismo/a	Cúlpase a si mesmo/a	Culpar-se un/a mateix/a
13. Crying easily	Llora con facilidad	Chora con facilidade	Plora fàcilment
14. Losing sexual interest	Pierde el interés sexual	Perda do apetito sexual	Pèrdua de l'interès sexual
15. Feeling lonely	Se siente solo/a	Séntese só/soa	Sentir-se sol /a
16. Feeling hopeless	Se siente sin esperanza	Séntese sen esperanza	Sentiment de desesperança
17. Feeling blue	Se siente triste	Séntese triste	Sentir-se trist/a
18. Thinking of ending one's life	Piensa en acabar con su vida	Pensa en acabar coa súa vida	Pensa en treure's la vida
19. Feeling trapped	Se siente atrapado/a	Séntese atrapado/a	Sentir-se atrapat/atrapada
20. Worrying too much	Se preocupa en exceso	Preocúpase en exceso	Preocupar-se en excés
21. Feeling no interest	No siente interés por nada	Non sente interese por nada	Sentiment de manca d'interès
22. Feeling that everything is an effort	Siente que todo le cuesta un esfuerzo	Sente que todo lle supón un esforzo	Sentir que tot és un esforç
23. Worthless feeling	Se siente inútil	Séntese inútil	Sentir-se inútil
24. Poor appetite	Poco apetito	Poco apetito	Pèrdua de la gana
25. Sleep disturbance	Problemas para dormir	Alteraciïns do sono	Alteració de la son

Tabla 4 (continuación)

Original	Español	Gallego	Catalán
The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.	La puntuación HSCL-25 se basa en un cuestionario autocumplimentado con lápiz y papel, de 25 preguntas sobre la presencia y la intensidad de ansiedad y síntomas depresivos en la última semana	A puntuación HSCL-25 baséase nun cuestionario cuberto con lapis e papel, de 25 preguntas sobre a presenza e a intensidade de ansiedade e síntomas depresivos na última semana.	L'escala HSCL-25 es basa en un qüestionari auto administrat de 25 preguntes, sobre la presència i la intensitat de símptomes d'ansietat i depressió en la darrera setmana.
Participants answer to one of 4 categories for each item on a four-point scale ranging from 1 to 4	Los participantes responden una de 4 categorías para cada ítem, en una escala de 4 puntos que van desde 1 a 4	Os/as participantes responden unha de catro categorías para cada ítem, nunha escala de catro puntos que van desde 1 a 4.	Els participants responen a una de les quatre categories per a cada ítem en una escala de quatre punts que va de l'1 al 4.
“Not at all” “A little” “Extremely” “Quite a bit”	En absoluto Un poco Bastante Mucho	En absoluto Un pouco Bastante Moito	Gens Una mica Bastant Molt
The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1,00 and 4,00). It is often used as the measure of distress	La puntuación del HSCL-25 se calcula dividiendo la puntuación total (sumando la puntuación de todas las preguntas) entre el número de respuestas (varía entre 1,00 y 4,00). Se usa habitualmente para medir el malestar psicológico	A puntuación do HSCL-25 calcúlase dividiendo a puntuación total (a suma de todas as preguntas) entre o número de respostas (cuxa puntuación oscila entre 1,00 e 4,00). Úsase de forma habitual para medir o nivel de malestar psicolóxico	La puntuació total del HSCL-25 es calcula dividint la suma de la puntuació dels diferents ítems pel número d'ítems contestats. El resultat total oscil·la entre 1,00 i 4,00. Aquesta escala sovint s'utilitza com a mesura de malestar psicològic.
The patient is considered as a “probable psychiatric case” if the mean rating on the HSCL-25 is 1,55.	El paciente se considera un «probable caso psiquiátrico» si el valor medio del HSCL-25 es $\geq 1,55$	Considérase que o/a paciente é un «caso psiquiátrico probable» se o valor medio do HSCL-25 é $\geq 1,55$	El/La pacient és considerat/considerada com a «probable cas psiquiàtric» si la qualificació mitjana del HSCL-25 és $\geq 1,55$
A cut-off value of 1,75 is generally used for diagnosis of major depression defined as “a case, in need of treatment”. This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender	Por lo general se usa un valor de corte de $\geq 1,75$ para el diagnóstico de depresión mayor, definida como «un caso que necesita tratamiento». Este valor de corte se considera un predictor válido de un trastorno mental, evaluado de forma independiente mediante entrevista clínica, aunque depende en parte del diagnóstico y el género	Polo xeral, úsase un valor de corte $\geq 1,75$ para diagnosticar a depresión maior, definida como “un caso que precisa tratamento”. Este valor de corte recoméndase como un predictor válido dun trastorno mental, avaliado independentemente por medio de entrevistas clínicas, aínda que depende en parte do diagnóstico e do xénero	Generalment s'utilitza un punt de tall $\geq 1,75$ per al diagnòstic de la depressió maior i es defineix com “cas que precisa de tractament”. Es recomana aquest punt de tall com un predictor vàlid de trastorn mental com ho seria l'avaluació independent per entrevista clínica, depenent en part del diagnòstic i del gènere

HSCL-25: Hopkins Symptom Checklist-25.

analizar las propiedades psicométricas y validar posteriormente los instrumentos para su uso en un contexto concreto, tal y como se ha realizado en otros idiomas y en diferentes contextos^{22,23,28-30}. Por tanto, antes de poder utilizarla en nuestro entorno, se requiere su validación para comprobar su efectividad.

Las recomendaciones ISPOR deben equilibrar la rigurosidad asegurando que la traducción sea comprensible para sus usuarios y a su vez evitar criterios estrictos de difícil cumplimiento en situaciones del mundo real¹⁶. Los MF son conocedores del lenguaje de sus pacientes, que utilizan muchas veces términos y expresiones dialectales, sobre todo en comunidades mayoritariamente bilingües. Por este motivo, se optó por seleccionar un número importante de MF en un paso intermedio de la traducción, para trabajar con metodología Delphi.

La selección de los expertos se presenta como una de las principales limitaciones en los métodos de consenso debido a la posibilidad de sesgo en su elección³¹. Se pretendió minimizar este efecto con los criterios de inclusión. El grupo europeo sugería incluir «profesores de universidad de AP», puesto que se entiende que estos cuentan con la experiencia necesaria para el diagnóstico de la depresión y dominan la lengua inglesa. En España, considerando la escasa atención que recibe la Medicina Familiar y Comunitaria (MFyC) en las Facultades de Medicina, se optó por seleccionar tutores de MFyC o clínicos con gran experiencia y publicaciones en revistas con impacto, pretendiendo que se aproximasen en la mayor medida al referente establecido. Se escogió su actividad como tutor de residentes especialistas en MFyC, asegurando así su formación continua y actualización en el campo de investigación y docencia.

La convergencia en la retrotraducción se explica por la proximidad lingüística (lenguas latinas), por la brevedad de los ítems y por el escaso número de preguntas.

Como puntos fuertes de este proyecto cabe destacar que se siguió rigurosamente la metodología ISPOR y que se tuvo en cuenta el ámbito de trabajo de AP como el futuro emplazamiento de utilización de la escala. Se consideró la perspectiva de los profesionales que la interpretarán en el futuro y se llevó a cabo una prueba con una muestra de pacientes en cada idioma. La entrevista cognitiva demostró que los potenciales usuarios de la escala comprenden su uso y la redacción de los ítems, y que no muestran incomodidad frente a ninguno de ellos. Otros estudios similares también han utilizado la entrevista cognitiva para este fin^{32,33}.

Finalmente, cabe resaltar que este estudio forma parte de un proyecto internacional cuya metodología en la traducción ha sido común con otros idiomas europeos; por ello, su desarrollo y su posterior validación permitirán obtener datos comparables de muestras de población con diferencias culturales. Una de las limitaciones de este proyecto es que no hubo ningún integrante del grupo FPDM de habla euskera; a pesar de que el proyecto estuvo abierto, no se consiguió incluir un nodo en el País Vasco para obtener una traducción de la escala en vasco.

Implicaciones para la práctica clínica y conclusiones

La escala HSCL-25 ha sido traducida al español, al catalán y al gallego, y las traducciones son equivalentes con la versión original inglesa. Los MF que han participado en el proceso de traducción y los pacientes de AP la han considerado aceptable y fácil de utilizar. Se realizará la validación de la escala en una fase posterior para poder generalizar su uso en AP.

Lo conocido sobre el tema

- La depresión es un problema de primer orden en la población general.
- La escala HSCL-25 se considera útil para el cribado de depresión en el entorno de Atención Primaria y es fácilmente comprensible para los pacientes.

Qué aporta este estudio

- Versión traducida y adaptada transculturalmente en español, catalán y gallego de la escala HSCL-25 para medir depresión, a partir de la versión original en inglés.

Financiación

Este trabajo recibió financiación de la Sociedad Española de Medicina de Familia y Comunitaria (semFYC) al resultar ganador de una ayuda para la realización de tesis doctorales Isabel Fernández 2019. Este proyecto ha recibido una beca de la Fundación Instituto Universitario de Investigación en Atención Primaria de Salud Jordi Gol i Gurina (IDIAP Jordi Gol 7Z19/019) para la capacitación en investigación y realización del doctorado en Atención Primaria y otra beca de la cátedra de la Universidad Autónoma de Barcelona-Novartis de investigación en Medicina Familiar y Comunitaria, ambas en la convocatoria 2019. Estas ayudas han sido otorgadas a la autora M. Rodríguez-Barragán.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

A los médicos de familia participantes en la metodología Delphi.

Al grupo Family Practice Depression and Multimorbidity.

A los pacientes participantes de las entrevistas cognitivas y a los profesionales que las hicieron posibles.

A Eva Peguero Rodríguez, por la revisión de la redacción del artículo.

Bibliografía

1. Bromet E, Andrade L, Hwang I, Sampson N, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med.* 2011;9:90.
2. Ministerio de Sanidad Consumo y Bienestar Social. Encuesta Nacional de Salud ENSE, España 2017. Serie informes monográficos. *Salud Mental.* 2017.
3. Serrano-Blanco A, Palao DJ, Luciano JV, Pinto-Meza A, Luján L, Fernández A, et al. Prevalence of mental disorders in primary care: Results from the diagnosis and treatment of mental disorders in primary care study (DASMAP). *Soc Psychiatry Psychiatr Epidemiol.* 2010;45:201–10.
4. Cano Vindel A, Martín Salguero J, Mae Wood C, Dongil E, Miguel Latorre J. La depresión en atención primaria: prevalencia, diagnóstico y tratamiento. *Papeles del Psicólogo.* 2012;33:2–11.
5. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, et al. Prevalence of common mental disorders in general practice attendees across Europe. *Br J Psychiatry.* 2008;192:362–7.
6. Portellano-Ortiz C, Garre-Olmo J, Calvó-Perxas L, Conde-Sala JL. Depression and associated variables in people over 50 years in Spain. *Rev Psiquiatr Salud Ment.* 2018;11:216–26.
7. Garrido-Elustondo S, Reneses B, Navalón A, Martín O, Ramos I, Fuentes M. Capacidad de detección de patología psiquiátrica por el médico de familia. *Aten Primaria.* 2016;48:449–57.
8. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: Meta-analysis and meta-regression of predictors. *Br J Gen Pract.* 2011;61:72–80.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5.* Arlington, United States: American Psychiatric Publishing; 2013.
10. Guías de Práctica Clínica en el SNS: Avalia-t 2013/06. Guía de práctica clínica sobre el manejo de la depresión en el adulto. 2014.
11. Reuland DS, Cherrington A, Watkins GS, Bradford DW, Blanco RA, Gaynes BN. Diagnostic accuracy of Spanish language depression-screening instruments. *Ann Fam Med.* 2009;7:455–62.
12. Díez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer R. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosom Med.* 2001;63:679–86.
13. Sanz J, García-Vera M, Espinosa R, Fortún M. Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3 propiedades psicométricas en pacientes con trastornos psicológicos. *Clínica y Salud.* 2005;16:121–42.
14. Montón C, Pérez Echeverría MJ, Campos R, García Campayo J, Lobo A. Anxiety scales and Goldberg's depression: An efficient interview guide for the detection of psychologic distress. *Aten Primaria.* 1993;12:345–9.
15. Aragonès Benaiges E, Masdèu Montalà R, Cando Guasch G, Coll Borràs G. Diagnostic validity of Zung's self-rating depression scale on primary care patients. *Actas Españolas Psiquiatr.* 2001;29:310–6.
16. Wild D, Grove A, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of Good practice for the translation and cultural adaptation process for Patient-Reported Outcomes (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Heal.* 2005;8:94–104.
17. Nabbe P, le Reste J, Chiron B, le Prielec A, Munoz M, Claveria A, et al. The FPDM (family practice depression and Multimorbidity) Study: Project for systematic review of literature to find tools for depression diagnosis used in primary care. *Eur Gen Pract Res. into Chronic Dis., Niza.* 2011:71.
18. Nabbe P, Le Reste J, Chiron B, Le Prielec A, Munoz M, Claveria A, et al. The FPDM (family practice depression and Multimorbidity) Study: Project for systematic review of literature to find tools for depression diagnosis used in primary care. In: *Proceedings of the 72nd European General Practice Research Network Meeting.* Niza. 2011:25.
19. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth E, Covi L. The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behav Sci.* 1974;19:1–15.
20. Nabbe P, Le Reste JY, Guillou-Landreat M, Beck-Robert E, Assenova R, Lazic D, et al. One consensual depression diagnosis tool to serve many countries: A challenge! A RAND/UCLA methodology. *BMC Res Notes.* 2018;11:4.
21. Veijola J, Jokelainen J, Läksy K, Kantojärvi L, Kokkonen P, Järvelin MR, et al. The Hopkins Symptom Checklist-25 in screening DSM-III-R axis-I disorders. *Nord J Psychiatry.* 2003;57:119–23.
22. Glaesmer H, Braehler E, Grande G, Hinz A, Petermann F, Rompel M. The German Version of the Hopkins Symptoms Checklist-25 (HSCL-25) —Factorial structure, psychometric properties, and population-based norms. *Compr Psychiatry.* 2014;55:396–403.
23. Nabbe P, Le Reste JY, Guillou-Landreat M, Gatineau F, le Floch B, Montier T, et al. The French version of the HSCL-25 has now been validated for use in primary care. *PLoS One.* 2019;14:e0214804.
24. Nabbe P, Le Reste J, le Floch B, Sowinska A, Czachowski S, Doer C, et al. FPDM (Family Practice Depression and Multimorbidity): The Hopkins Symptoms Checklist-25 items (HSCL-25), completed translation in 10 European languages. In: *Proceedings of 81st European General Practice Research Network Meeting.* Edirne. 2015:66.
25. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976).* 2000;25:3186–91.
26. Congost Maestre N. El lenguaje de las Ciencias de la Salud. Los cuestionarios de salud y su traducción del inglés al español [Tesis doctoral]. Alicante: Universidad de Alicante; 2010.
27. Martinussen LJ, Šaltyté Benth J, Almdahl IS, Borza T, Selbæk G, Mcpherson B, et al. The effect of anxiety on cognition in older adult inpatients with depression: Results from a multicenter observational study. *Heliyon.* 2019;5:e02235.
28. Lundin A, Hallgren M, Forsell Y. The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden. *J Affect Disord.* 2015;183:247–52.
29. Al-Turkait FA, Ohaeri JU, el-Abbasi AHM, Naguy A. Relationship between symptoms of anxiety and depression in a sample of arab college students using the Hopkins Symptom Checklist 25. *Psychopathology.* 2011;44:230–41.
30. Lee B, Kaaya SF, Mbwambo JK, Smith-Fawzi MC, Leshabari MT. Detecting depressive disorder with the Hopkins Symptom Checklist-25 in Tanzania. *Int J Soc Psychiatry.* 2008;54:7–20.
31. Steurer J. The Delphi method: An efficient procedure to generate knowledge. *Skeletal Radiol.* 2011;40:959–61.
32. Garcimartin P, Pardo-Cladelas Y, Verdú-Rotellar JM, Delgado-Hito P, Astals-Vizcaino M, Comin-Colet J. Transcultural adaptation into Spanish of the Patient empowerment in long-term conditions questionnaire. *Aten Primaria.* 2019;51:24–31.
33. González-Bueno J, Calvo-Cidoncha E, Sevilla-Sánchez D, Espauella-Panicot J, Codina-Jané C, Santos-Ramos B. Spanish translation and cross-cultural adaptation of the ARMS-scale for measuring Medication adherence in polypathological patients. *Aten Primaria.* 2017;49:459–64.



Article

Validation and Psychometric Properties of the Spanish Version of the Hopkins Symptom Checklist-25 Scale for Depression Detection in Primary Care

María Rodríguez-Barragán ^{1,2,3,*} , María Isabel Fernández-San-Martín ^{3,4}, Ana Clavería-Fontán ^{5,6} , Susana Aldecoa-Landesa ^{5,7}, Marc Casajuana-Closas ^{2,3} , Joan Llobera ^{6,8,9}, Bárbara Oliván-Blázquez ^{6,10} and Eva Peguero-Rodríguez ^{3,11,12}

- ¹ Primary Health Centre La Mina, Gerència Territorial de Barcelona, Institut Català de la Salut, 08930 Barcelona, Spain
- ² Faculty of Medicine, Autonomous University of Barcelona, 08193 Bellaterra, Spain; mcasajuana@idiapjgol.info
- ³ Institut Universitari d'Investigació en Atenció Primària Jordi Gol, 08007 Barcelona, Spain; mifsanmartin.bcn.ics@gencat.cat (M.I.F.-S.-M.); pegueroeva@gmail.com (E.P.-R.)
- ⁴ Gerència Territorial de Barcelona, Institut Català de la Salut, 08007 Barcelona, Spain
- ⁵ I-Saúde Group, Galicia Sur Health Research Institute SERGAS-UVIGO, Servicio Galego de Saúde, 36201 Galicia, Spain; anaclaveriaf@gmail.com (A.C.-F.); susialdecoa@gmail.com (S.A.-L.)
- ⁶ Primary Care Prevention and Health Promotion Research Network RedIAPP, 28001 Madrid, Spain; jllobera@ibsalut.caib.es (J.L.); bolivan@unizar.es (B.O.-B.)
- ⁷ Primary Health Centre Beiramar, Área de Xestión Integrada de Vigo, SERGAS, 36201 Vigo, Spain
- ⁸ Institut d'Investigació Sanitària Illes Balears, 07120 Palma, Spain
- ⁹ Primary Care Research Unit of Mallorca, Balearic Islands Health Service, 07120 Palma, Spain
- ¹⁰ Department of Psychology and Sociology, University of Zaragoza, 50009 Zaragoza, Spain
- ¹¹ Primary Health Centre El Castell, Institut Català de la Salut, 08860 Barcelona, Spain
- ¹² Faculty of Medicine, University of Barcelona, 08007 Barcelona, Spain
- * Correspondence: marodriguez.bcn.ics@gencat.cat; Tel.: +34-933-811-593



Citation: Rodríguez-Barragán, M.; Fernández-San-Martín, M.I.; Clavería-Fontán, A.; Aldecoa-Landesa, S.; Casajuana-Closas, M.; Llobera, J.; Oliván-Blázquez, B.; Peguero-Rodríguez, E. Validation and Psychometric Properties of the Spanish Version of the Hopkins Symptom Checklist-25 Scale for Depression Detection in Primary Care. *Int. J. Environ. Res. Public Health* **2021**, *18*, 7843. <https://doi.org/10.3390/ijerph18157843>

Academic Editor: Rosa Angela Fabio

Received: 11 June 2021

Accepted: 22 July 2021

Published: 24 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Depression constitutes a major public health problem due to its high prevalence and difficulty in diagnosis. The Hopkins Symptom Checklist-25 (HSCL-25) scale has been identified as valid, reproducible, effective, and easy to use in primary care (PC). The purpose of the study was to assess the psychometric properties of the HSCL-25 and validate its Spanish version. A multicenter cross-sectional study was carried out at six PC centers in Spain. Validity and reliability were assessed against the structured Composite International Diagnostic Interview (CIDI). Out of the 790 patients, 769 completed the HSCL-25; 738 answered all the items. Global Cronbach's alpha was 0.92 (0.88 as calculated for the depression dimension and 0.83 for the anxiety one). Confirmatory factor analysis (CFA) showed one global factor and two correlated factors with a correlation of 0.84. Area under the curve (AUC) was 0.89 (CI 95%, 0.86–0.93%). For a 1.75 cutoff point, sensibility was 88.1% (CI 95%, 77.1–95.1%) and specificity was 76.7% (CI 95%, 73.3–79.8%). The Spanish version of the HSCL-25 has a high response percentage, validity, and reliability and is well-accepted by PC patients.

Keywords: depression; depressive disorder; primary healthcare; family practice; general practitioners; anxiety; questionnaires; psychometrics

1. Introduction

Depression is a common condition among adults and can lead to harmful consequences. Worldwide, it is considered the third cause of years lost to disability [1], with a prevalence that increased by 17.8% between 2005 and 2015 [2]. In Europe, studies carried out in primary care (PC) settings have reported an incidence from 9.6% [3] to 20.2% [4]. The prevalence of depression in Spain is higher than the European mean and is associated

with a negative perception of physical health, the presence of two or more difficulties in daily living activities, female gender [5], and some physical comorbidities [6].

Several instruments have been designed to screen mental disorders. As a collaborative international project, the Family Practice Depression and Multimorbidity (FPDM) group from the European General Practice Research Network (EGPRN) aimed to select a questionnaire to detect depression symptoms in PC patients [7]. Firstly, a systematic review of validated scales for screening and diagnosis of depression in adults was performed. Scales that had been compared to a psychiatric interview based on the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria with quantitative results and with participation of PC professionals were analyzed [8]. As a result of this systematic review, seven scales were identified: Geriatric Depression Scale of five items (GDS-5), Geriatric Depression Scale of 15 items (GDS-15), Geriatric Depression Scale of 30 items (GDS-30), Hospital Anxiety Depression Scale (HADS), Center for Epidemiologic Studies Depression Scale-Revised (CESD-R), Physical Symptom Checklist of 51 items (PSC-51), and Hopkins Symptom Checklist of 25 items (HSCL-25).

Secondly, the HSCL-25 was selected by consensus [9]. Validity, efficacy, and reproducibility were analyzed as quantitative criteria. Characteristics such as being a self-administered questionnaire, easiness of completion for patients, and the simplicity of its interpretation were taken into account to assess the ergonomics. The HSCL-25 is suitable for use in PC because of its high validity and reliability; moreover, its ergonomics make it easy to use for patients [9]. It is a self-report questionnaire designed to measure psychological distress based on the SCL-90 [10], a longer checklist designed by Derogatis et al. The full version of the SCL-90 covers nine symptom dimensions, with 25 items belonging to the anxiety and depression ones.

Thirdly, the questionnaire was translated into 13 European languages [11], including the Spanish version [12]. The translation and adaptation process consisted of an initial forward translation, a pilot study based on the Delphi methodology with the participation of family doctors, and a back translation. Comprehension analysis was carried out through cognitive debriefing in a sample of PC patients. At the last step, transcultural harmonization was performed simultaneously with other versions of the scale in different European languages [11].

Finally, validation of the different versions is in process, the French version has already been validated [13] and the Croatian one is under way.

Instruments should be tested and validated in different languages, cultures, settings, and populations in order to make comparisons and establish efficacy. The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) [14] initiative has developed criteria to evaluate the measurement properties of outcome measurement instruments. In addition, a considerable number of studies have assessed the HSCL-25 psychometric properties in various populations [15–19], including PC patients [20,21]. PC is the gateway to the healthcare system for most of the Spanish population. It is the ideal setting to study the prevalence of the most common diseases.

The purpose of this study was to assess the HSCL-25 psychometric properties and validate the scale's Spanish version in a PC population.

2. Methods

2.1. Study Design

A cross-sectional multicenter design was used. The participants were patients attending primary healthcare centers (PHC) in Aragon (1), the Balearic Islands (1), and Galicia (4) taking part in the EIRA study [22]. Ethical approval was given by the Ethics Committee of Institut Universitari d'Investigació en Atenció Primària (IDIAP) Jordi Gol (reference number P16/025).

2.2. Participants

The selection criteria were those employed in the EIRA study. Eligible participants were patients aged between 45 and 75 years who had two or more of the following unhealthy behaviors: tobacco use, low adherence to the Mediterranean dietary pattern, and insufficient physical activity. Exclusion criteria were advanced serious illness, cognitive impairment, dependence in basic everyday activities, severe mental illness, inclusion in a long-term home healthcare program, treatment for cancer, end-of-life care, or no plan to reside in the area during the intervention period.

2.3. Recruitment and Sample Size

Recruitment was made by consecutive sampling of patients meeting the selection criteria and attending the PHC for any reason. The recruitment period took 6 months during 2017.

The COSMIN guide [23] was followed to calculate the sample size. It states that seven completed questionnaires are needed per each item of the scale and that at least 100 completed questionnaires are required to assess psychometric properties. As the HCSL-25 has 25 items, and taking into account a 10% possibility of missing values, 193 patients were needed to complete the questionnaire.

In order to estimate the sample size required to compare the HSCL-25 with the Composite International Diagnostic Interview (CIDI), the receiver operating curve (ROC) and the corresponding area under the curve (AUC) were calculated with the BIOSOFT application (<http://www.biosoft.hacettepe.edu.tr/easyROC/>, accessed on 31 January 2021) employing the following parameters:

1. A 0.742 AUC to identify cases of depression (according to Nettelbladt et al. [20]).
2. A type I error of 5% and a power of 95% were selected. Thus, 87 cases and 174 controls were needed (new AUC test: 0.80, standard AUC test: 0.74; case/control ratio: 2).

Taking into account that the estimated prevalence in PC is 16.3% [4], 533 patients were required to complete the scale to obtain 87 cases.

To evaluate test–retest reliability, the same considerations and a 20% possibility of missing values were taken into account, 26 patients were needed to reach an acceptable correlation coefficient of 0.7 [24]. All the included patients were invited to participate in the telephonic retest.

2.4. Variables

Sociodemographic data (sex, age, nationality, marital status, current employment, and education level) were gathered from the participants. They were asked to complete the self-administered HSCL-25 questionnaire and other forms related to the EIRA study. Afterwards, trained professionals blinded to the HSCL-25 results conducted the CIDI with all the participants. Training consisted of a global presentation of the procedure of the interview, the reading question by question, role-playing with the interviewers, and resolution of doubtful situations. Retest of the HSCL-25 was telephonic to facilitate the viability; it was carried out between 1 and 3 months later.

2.5. Hopkins Symptom Checklist-25 (HSCL-25)

The HSCL-25 is a self-administered questionnaire that takes from five to ten minutes to complete [13]. It consists of 25 items on a four-point Likert scale: 1 = “Not at all,” 2 = “A little,” 3 = “Quite a bit,” 4 = “Extremely”. The tool has two well-known dimensions: items 1 to 10 belong to the anxiety dimension, whereas items 11 to 25 constitute the depression one. The HSCL-25 score is calculated by dividing the total score of items by the number of items answered, so the final score can range from 1 to 4. A cutoff value of 1.75 is generally used for diagnosis of major depression, defined as “a case in need of treatment”. This cutoff point is recommended as a valid predictor of mental disorder [15,17,25]. Our study was carried out using the Spanish version of the HSCL-25 obtained by means of translation and cultural adaption of the original English version [12].

2.6. Composite International Diagnostic Interview (CIDI)

The CIDI is a standardized structured diagnostic interview created by the World Health Organization (WHO) according to the DSM-IV and International Classification of Diseases (ICD-10) definitions and criteria. Used by trained interviewers for mental disorder assessment in the general population [26], it has demonstrated high validity and reliability [27]. Whilst the original CIDI was in English, it has been adapted into and validated for many languages using a common procedure overseen by the WHO [28]. Questions related to depression symptoms can be found in section E of the CIDI. In this study, it was considered the gold standard to assess the HSCL-25.

2.7. Patient Health Questionnaire (PHQ)

The PHQ is a well-known self-administered questionnaire used for common mental disorders. The PHQ-9 is the depression module in which each of the nine items is rated with a Likert scale that ranges from 0 to 3 [29]. The total score can vary from 0 to 27. Scores of 15 or more indicate major depression. For this study, the validated Spanish version [30] was employed.

2.8. Statistical Analysis

Analysis was conducted using STATA version 15 (manufacturer StataCorp LLC, Texas, USA).

2.8.1. Missing Data

The missing values for scale item responses were imputed with the mean of the responses to the rest of the scale items of each individual (the participant's most representative value). The subjects with less than 50% response were excluded. The same imputation was carried out for the retest values.

2.8.2. Responding Process and Item Analysis

An analysis of the responding process was performed, looking for patterns of non-response and frequency response distribution of the items by category and sex. The discriminatory capacity of the items was assessed by comparing the two extreme groups. The discrimination index (DI) of each item was calculated by the mean difference of each group. Given that the response options have four possible categories, the DI could vary from -3 to $+3$.

2.8.3. Internal Structure

Confirmatory factor analysis (CFA) was carried out based on the structure of the original English version. The factorial loads for the two models with only one factor and for the two correlated ones (anxiety and depression) were calculated. The robust maximum likelihood mean adjusted method was employed to carry out factorial analysis of the standardized values. To evaluate the estimated model fit, the absolute fit index was calculated with chi-squared distribution. Given that this value may be affected by the sample size, complementary indices were employed, including the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the coefficient of determination (CD). In addition, comparative indices such as the comparative fit index (CFI) and the Tucker–Lewis fit index (TLI) were employed.

2.8.4. Criterion Validity

Criterion validity was measured by calculating the ROC curve for the HSCL-25 scale in comparison with the gold-standard CIDI. The AUC was estimated with 95% CI. Sensitivity, specificity, positive and negative predictive values, Youden index, and the best cutoff point were also assessed.

Concordance with the PHQ-9 was measured with the Pearson correlation coefficient and the prevalence- and bias-adjusted kappa, taking into account cutoff points of 1.75 and 15 for the HSCL-25 and PHQ-9, respectively.

2.8.5. Internal Consistency

The contribution of the items to the internal consistency was analyzed with indicators based on correlation (homogeneity), covariance (Cronbach's alpha coefficient), and regression (R^2). The total Cronbach's alpha and one for each of the two subscales were calculated. The value ≥ 0.7 was considered adequate [24].

2.8.6. Test–Retest Reliability

Test–retest reliability was assessed by calculating the intraclass correlation coefficient (ICC) by the use of the mean of two evaluations (test and retest), absolute agreement, and a two-way mixed-effects model.

3. Results

3.1. Participants

A total of 790 patients were selected for the HSCL-25 and 768 patients completed it (97.2% response rate). The participants' mean age was 58.4 years (± 8.2) without significant gender differences; 54.4% were women. Table 1 depicts the sociodemographic characteristics of the sample. Women and men differed in marital status and current employment.

Table 1. Sociodemographic characteristics of the sample.

Sociodemographic Variables	Male		Female		Total		<i>p</i>
	n	%	n	%	n	%	
	350	45.6	417	54.4	767		
Marital status							
Married/with a partner	259	74.0	291	69.8	550	71.7	0.001
Single	31	8.9	25	6.0	56	7.3	
Separated/divorced	48	13.7	55	13.2	103	13.4	
Widow(er)	11	3.1	46	11.3	57	7.4	
Others (religious . . .)	1	0.3	0	0	1	0.1	
Education							
Secondary or higher	178	50.9	190	45.6	368	48.0	0.153
Primary or lower	172	49.1	226	54.2	398	51.9	
No response	0	0	1	0.2	1	0.1	
Current employment							
Employed	153	43.7	160	38.4	313	40.8	<0.001
Housewife	2	0.6	109	26.1	111	14.5	
Unemployed	45	12.9	38	9.1	83	10.8	
Retired	128	36.6	81	19.4	209	27.3	
Others (student, sick leave, disability)	20	5.7	28	6.7	48	6.3	
No response	2	0.6	1	0.2	3	0.4	

3.2. Responding Process and Item Analysis

Of the 23 participants excluded from the analysis (2.9%), 22 did not answer any of the items and one only responded to 12 of the 25 items. Thirty participants (3.8%) left between one and five items blank; these missing values were imputed. No non-response patterns were found; Supplementary Materials Table S1 shows the non-response patterns in detail.

The mean score of the items, the response percentages for each category, and the DI are depicted in Table 2. Item 20 "Worrying too much" with a mean of 2.14 had the highest global rating; it was followed by item 4 "Nervousness" (mean = 2.03). In contrast,

item 18 “Thinking of ending one’s life” (mean = 1.09) had the lowest value, followed by item 9 “Feeling panic” (mean = 1.17). Women scored higher in all the items. The greatest difference between genders was observed in item 14 “Losing sexual interest” with a statistically significant difference of 0.73. The item that varied the least between genders was item 24 “Poor appetite,” and it was non-significantly different.

With respect to response frequency distribution, 60.9% of the responses were found in the lowest response category “Not at all,” with the rating of 1. A floor effect was observed in item 18 “Thinking of ending one’s life,” with 93.6% of responses in the lowest category. None of the items presented a ceiling effect.

The discrimination capacities of the items all showed a positive DI. Item 4 “Nervousness” discriminated the best with the DI of 1.43. In contrast, the item with the worst discrimination values was item 18 “Thinking of taking one’s life” (DI = 0.22).

3.3. Internal Structure: Confirmatory Factorial Analysis

The Satorra–Bentler comparative fit index was significant. Globally, the other indices showed that the proposed one-factor and two-correlated-factor models were reasonably acceptable. Table 3 depicts the fit indices for each model.

Table 4 shows the factor loading for each model and correlation in the two-factor model. All the factor loadings were positive and statistically significant ($p < 0.001$) and ≥ 0.30 . Only item 24 “Poor appetite” had a loading below 0.4. With respect to the two correlated factors, the standardized values ranged from 0.3 for item 24 and 0.84 for item 17 “Feeling blue,” both in the depression category.

3.4. Criterion Validity: Relationship with the Gold-Standard CIDI

Of the 767 patients who completed the HSCL-25 scale, 736 also participated in the CIDI interview (96.0%). The 31 patients who did not take part in the interview were excluded from the following analysis. According to the CIDI, the global depression prevalence was 8.0% (CI 95%, 6.2–10.2%): 4.7% (CI 95%, 2.7–7.5%) in men and 10.8% (CI 95%, 8.0–14.4%) in women. With respect to the HSCL-25, the global prevalence was 28.5% (CI 95%, 25.3–31.9%) for the 1.75 cutoff point. Table 5 shows the different indices and values globally and by sex. The differing optimum cutoff points for women (1.76) and men (1.84) are noteworthy. Sensitivity was similar for both genders whilst specificity was better in men. The global ROC curves are depicted in Figure 1, by gender—in Figure 2. The global AUC was 0.892 (CI 95%, 0.856–0.928); in the gender analysis, it was greater in men. The optimum cutoff point for the Spanish version of the HSCL-25 was 1.76, with the Youden index of 64.8%.

3.5. Criterion Validity: Relationship with PHQ-9 External Criteria

The HSCL-25 scale and the PHQ-9 were completed by 761 patients. The Pearson coefficient for the values of both scales was 0.780 (CI 95%, 0.750–0.806). Considering both variables as categorical with cutoff points of 1.75 and 15 for the HSCL-25 and the PHQ-9, respectively, the prevalence- and bias-adjusted kappa (PABAK) value was 0.553, with the global agreement of 77.7% (CI 95%, 74.6–80.5%).

Table 2. Item analysis: missing values, response distribution, mean score, and discrimination index.

Item	Response Values (n = 790)											Score		p	DI			
	Missing Values		Not at All		A Little		Quite a Bit		Extremely		Male (n = 350)		Female (n = 417)			Total (n = 767)		
	%		%		%		%		%		Mean	SD	Mean			SD	Mean	SD
1. Being scared for no reason	2.8	75.6	18.1	4.8	1.3	1.14	0.41	1.46	0.73	1.32	0.63	<0.001	0.61					
2. Feeling fearful	3.3	67.5	25.0	5.7	1.7	1.27	0.51	1.54	0.77	1.42	0.68	<0.001	0.77					
3. Faintness	3.2	54.4	35.1	8.5	2.1	1.47	0.65	1.68	0.79	1.58	0.73	<0.001	1.01					
4. Nervousness	2.9	31.8	40.0	21.1	7.0	1.81	0.84	2.23	0.91	2.03	0.90	<0.001	1.43					
5. Heart racing	3.2	62.1	29.3	6.3	2.4	1.36	0.60	1.59	0.79	1.49	0.72	<0.001	0.84					
6. Trembling	3.2	76.5	18.0	3.9	1.6	1.28	0.58	1.33	0.65	1.31	0.62	0.253	0.59					
7. Feeling tense	3.4	42.5	39.6	12.7	5.2	1.64	0.75	1.94	0.91	1.81	0.85	<0.001	1.26					
8. Headache	3.3	59.3	26.7	9.9	4.0	1.43	0.72	1.72	0.89	1.59	0.83	<0.001	0.77					
9. Feeling panic	3.2	87.0	9.7	2.5	0.9	1.13	0.43	1.21	0.55	1.17	0.50	0.046	0.39					
10. Feeling restless	2.9	40.8	44.5	11.9	2.9	1.67	0.72	1.85	0.80	1.77	0.77	<0.001	1.10					
11. Feeling low in energy	3.0	37.6	40.8	15.1	6.5	1.70	0.75	2.08	0.95	1.91	0.88	<0.001	1.34					
12. Blaming oneself	3.0	61.2	27.5	8.0	3.4	1.43	0.65	1.62	0.87	1.54	0.78	<0.001	0.94					
13. Crying easily	3.3	48.2	31.0	12.8	8.0	1.57	0.80	2.00	1.01	1.80	0.94	<0.001	1.04					
14. Losing sexual interest	3.3	48.9	24.0	14.2	12.9	1.51	0.82	2.24	1.14	1.91	1.07	<0.001	1.29					
15. Feeling lonely	2.9	65.6	23.3	6.1	5.0	1.32	0.66	1.66	0.90	1.50	0.82	<0.001	1.02					
16. Feeling hopeless	3.3	75.9	16.0	4.7	3.4	1.23	0.57	1.46	0.82	1.36	0.73	<0.001	0.82					
17. Feeling blue	3.2	52.3	34.4	8.6	4.7	1.40	0.66	1.87	0.89	1.66	0.82	<0.001	1.31					
18. Thinking of ending one's life	3.2	93.6	4.4	1.7	0.3	1.06	0.31	1.11	0.40	1.09	0.36	0.041	0.22					
19. Feeling trapped	3.0	72.9	17.3	6.8	3.0	1.32	0.65	1.47	0.81	1.40	0.75	<0.001	0.88					
20. Worrying too much	3.2	27.9	39.0	24.6	8.5	1.95	0.83	2.29	0.96	2.14	0.92	<0.001	1.29					
21. Feeling no interest	3.5	79.1	14.6	4.3	2.0	1.20	0.51	1.37	0.72	1.29	0.64	<0.001	0.69					
22. Feeling that everything is an effort	2.9	48.6	36.9	9.8	4.7	1.56	0.67	1.83	0.92	1.71	0.83	<0.001	1.18					
23. Worthless feeling	2.8	81.4	12.1	4.7	1.8	1.17	0.51	1.35	0.71	1.27	0.63	<0.001	0.59					
24. Poor appetite	2.9	83.3	11.2	3.7	1.8	1.23	0.58	1.25	0.63	1.24	0.60	0.723	0.43					
25. Sleep disturbance	2.9	47.5	28.2	13.6	10.8	1.74	0.93	2.00	1.07	1.88	1.01	<0.001	1.21					
Total	3.1	60.9	25.9	9.0	4.2	1.42	0.64	1.69	0.82	1.57	0.76	<0.001						

DI: discrimination index.

Table 3. Fit indices in the studied factor models.

Model	X ² _SB (df)	p	CFI_SB	TLI_SB	RMSEA (90% CI)	p Close	RMSEA_SB	SRMR	CD
One factor	1600.3 (275)	<0.001	0.828	0.812	0.079 (0.076–0.083)	<0.001	0.061	0.059	0.931
Two correlated factors	899.5 (274)	<0.001	0.862	0.849	0.072 (0.068–0.076)	<0.001	0.055	0.055	0.970

X²_SB: Satorra–Bentler chi-squared statistic; df: degrees of freedom; CFI_SB: Satorra–Bentler comparative fit index; TLI_SB: Satorra–Bentler Tucker–Lewis fit index; RMSEA: root mean square error of approximation; RMSEA_SB: Satorra–Bentler root mean square error of approximation; SRMR: standardized root mean square residual, CD: coefficient of determination.

Table 4. Confirmatory factorial analysis: factor loading values and correlation between two depression and anxiety factors.

Item	One Factor	Two Correlated Factors	
		Anxiety	Depression
1. Being scared for no reason	0.45	0.49	
2. Feeling fearful	0.52	0.56	
3. Faintness	0.60	0.59	
4. Nervousness	0.67	0.75	
5. Heart racing	0.53	0.57	
6. Trembling	0.43	0.48	
7. Feeling tense	0.68	0.74	
8. Headache	0.39	0.41	
9. Feeling panic	0.40	0.44	
10. Feeling restless	0.65	0.70	
11. Feeling low in energy	0.67		0.66
12. Blaming oneself	0.57		0.58
13. Crying easily	0.51		0.50
14. Losing sexual interest	0.51		0.52
15. Feeling lonely	0.65		0.68
16. Feeling hopeless	0.61		0.65
17. Feeling blue	0.82		0.84
18. Thinking of ending one's life	0.43		0.44
19. Feeling trapped	0.60		0.62
20. Worrying too much	0.58		0.56
21. Feeling no interest	0.62		0.65
22. Feeling that everything is an effort	0.68		0.69
23. Worthless feeling	0.53		0.56
24. Poor appetite	0.30		0.30
25. Sleep disturbance	0.48		0.47
Factor correlation			0.84

Table 5. Optimum cutoff points, global and gender-specific sensitivity, specificity, positive and negative predictive values in the HSCL-25 scale.

Calculated for	Index	Male (n = 341)			Female (n = 395)			Total (n = 736)		
		Value	CI 95%, Lower Limit	Upper Limit	Value	CI 95%, Lower Limit	Upper Limit	Value	CI 95%, Lower Limit	Upper Limit
Original cutoff, 1.75	Sensitivity	87.5	61.7	98.4	88.4	74.9	96.1	88.1	77.1	95.1
	Specificity	86.8	82.6	90.3	67.3	62.2	72.2	76.7	73.3	79.8
	PPV	24.6	14.1	37.8	18.3	12.5	25.4	24.8	19.1	31.2
	NPV	99.3	97.5	99.9	97.9	95.2	99.3	98.7	97.3	99.5
Optimal cutoff point	AUC	0.960	0.927	0.994	0.837	0.781	0.893	0.892	0.856	0.928
	Optimal cutoff point	1.84			1.76			1.76		
	Sensitivity	87.5	61.7	98.4	83.7	69.3	93.2	84.7	73.0	92.8
	Specificity	91.7	88.1	94.5	69.3	64.2	74.1	78.7	75.5	81.8
Optimal cutoff point	PPV	34.1	20.1	50.6	25.0	18.2	32.9	25.8	19.8	32.5
	NPV	99.3	97.6	99.9	97.2	94.3	98.9	98.3	96.9	99.2

AUC: area under the curve. PPV: positive predictive value. NPV: negative predictive value.

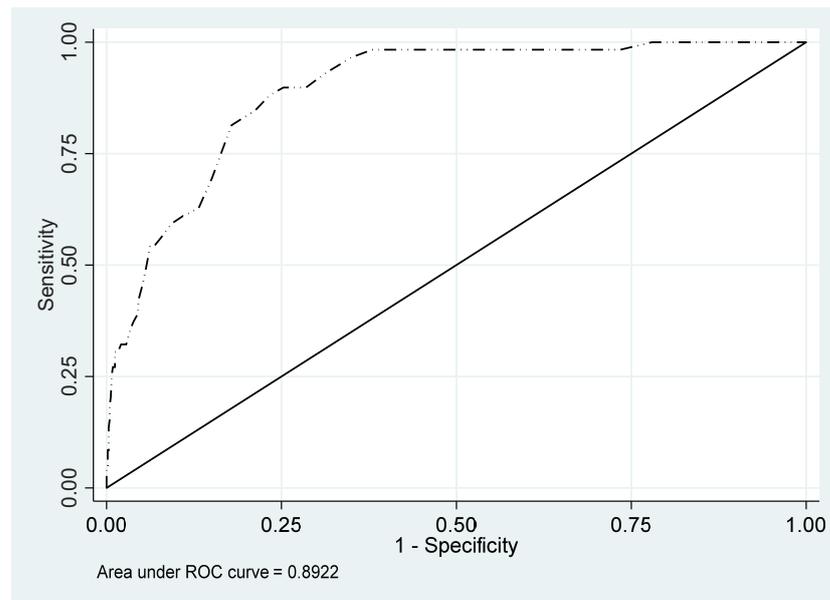


Figure 1. Total ROC (receiver operating curve) and AUC

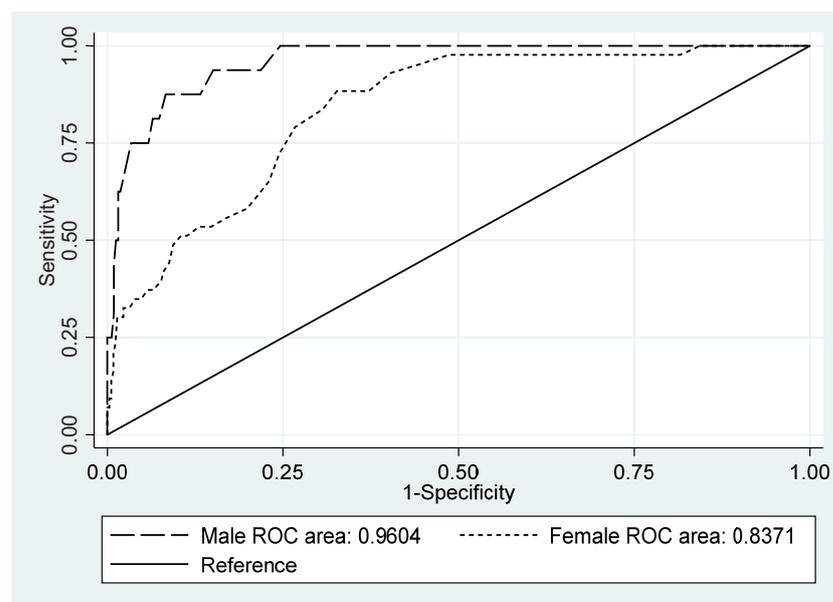


Figure 2. ROC curves and AUC by gender

3.6. Reliability: Internal Consistency and Test–Retest Reliability

Intercorrelation between the items can be observed in Table S2 of the Supplementary Materials. All the correlations were positive and of a moderate magnitude. The correlation mean was 0.362, with the standard deviation (SD) of 0.210. The intercorrelation range was 0.090–0.621. The correlation means and SD for the anxiety and depression subscales were 0.450 (SD, 0.278) and 0.419 (SD, 0.243), respectively.

Cronbach's alpha coefficient for the total values and for each subscale, the total item correlations and determination coefficients (R^2) are depicted in Table 6. The value obtained for the coefficient without the item was also calculated as shown in the middle column in Table 6. All the results were lower with the exception of item 24 "Poor appetite," the elimination of which resulted in an increase in the global coefficient from the 4th decimal. Assessing the item-total correlation and the R^2 of this item resulted in lower values in both cases as the item was the least consistent one.

Table 6. Cronbach's alpha coefficient of each item, globally and by subscales.

Item	Item-Total Correlation	Cronbach's Alpha	R ²
1. Being scared for no reason	0.4466	0.9148	0.3697
2. Feeling fearful	0.5140	0.9137	0.4154
3. Faintness	0.5724	0.9127	0.4640
4. Nervousness	0.6545	0.9109	0.5250
5. Heart racing	0.5172	0.9136	0.3657
6. Trembling	0.4226	0.9151	0.3413
7. Feeling tense	0.6536	0.9109	0.5220
8. Headache	0.3760	0.9163	0.2462
9. Feeling panic	0.3907	0.9156	0.2939
10. Feeling restless	0.6206	0.9117	0.4805
11. Feeling low in energy	0.6472	0.9110	0.5439
12. Blaming oneself	0.5454	0.9131	0.3385
13. Crying easily	0.4881	0.9145	0.2909
14. Losing sexual interest	0.4832	0.9153	0.2866
15. Feeling lonely	0.5973	0.9121	0.4857
16. Feeling hopeless	0.5728	0.9127	0.4842
17. Feeling blue	0.7799	0.9085	0.6677
18. Thinking of ending one's life	0.4033	0.9159	0.2747
19. Feeling trapped	0.5577	0.9129	0.4085
20. Worrying too much	0.5592	0.9129	0.3770
21. Feeling no interest	0.5752	0.9129	0.4460
22. Feeling that everything is an effort	0.6432	0.9112	0.5494
23. Worthless feeling	0.4995	0.9140	0.3552
24. Poor appetite	0.3065	0.9167	0.1568
25. Sleep disturbance	0.4747	0.9152	0.2776
Total		0.9166	
Anxiety subscale (items 1–10)		0.8306	
Depression subscale (items 11–25)		0.8784	

R²: determination coefficient.

The most homogeneous item was 17 “Feeling blue;” when eliminated, the internal consistency of the scale decreased to the lowest value; this item presented the highest item-total correlation and R².

The telephone retest was completed by 94 participants. Test–retest reliability was 0.92 (CI 95%, 0.87–0.95), calculated with CCI.

4. Discussion

A major finding of our study is that the Spanish version of the HSCL-25 is an instrument with good acceptability and high response rate for PC patients. Its reliability in measuring depression is robust and presents considerable sensitivity and specificity when compared to the CIDI interview. The CFA demonstrated that the Spanish version is similar to the original English one.

For most of the Spanish population, PC consultations are the gateway to the healthcare system. Due to the high prevalence of depression [5], it is crucial that easy to use viable tools are available for the PC environment. As the HSCL-25 meets such characteristics [9], awareness of its psychometric properties is relevant, in particular, of those items that most contribute to detecting symptoms and thus permit discrimination between the healthy populations and the potentially depressed ones. In addition, PC professionals should be informed of the reliability of the scale and its sensitivity and specificity values which are key in order to establish its diagnostic utility.

The study participants were PC patients aged 45–75 years who had taken part in the more extensive EIRA study [22]. Whilst this implied a restricted age range, which might signify a limitation, the sample was considered sufficiently representative of such individuals. The sample size was greater than the minimum required for the analysis

according to the COSMIN guidelines [23], which are taken as reference in the field of psychometry. The statistical analysis was carried out based on the same recommendations. The content validity of the Spanish version of the HSCL-25 had been previously evaluated when it had been translated and transculturally adapted to Spanish and other official languages of the country [12].

With respect to item analysis, a considerable percentage of responses was available, and no definite pattern was observed. As a consequence, the questionnaire appears to be widely accepted by PC patients without any items which may cause discomfort or difficulty in understanding. As the study was carried out with patients attending the PHC for any reason, a high percentage of low-rating responses for the categories was expected. In addition, a floor effect was foreseen for item 18 “Thinking of ending one’s life” which concerned suicidal ideation. Taking into account the definition of depression according to the DSM-5 [31], it is not surprising that the item that best discriminated between the healthy population and the one with depressive symptoms referred to sadness. Item 17 “Feeling blue” was shown to be the most homogenous in all the analyses, with the highest correlation compared to the other scale items. It presented the highest coefficient of determination (that is to say, it could be predicted from the rest of the items) and most contributed to augmenting internal global consistency.

Regarding analysis of the scale’s factorial structure, this was performed with the CFA as the HSCL-25 has been widely studied with one single factor or two correlated ones even though other models have been proposed [15,32,33]. The fit indices for both models were acceptable, and the results indicated moderately elevated factorial loads. In the study of the two-factor model, there was a factorial correlation of 0.84 which indicated that the depression and anxiety dimensions strongly correlated in a positive manner. Such a figure is higher than that detected in previous studies [15]. The correlation is understandable as symptoms of anxiety are often observed in patients diagnosed with depression; moreover, anxiety and depression are frequently found to be associated comorbidities [34,35].

In other studies which compared the HSCL-25 scale with structured psychiatric interviews, a subsample of participants was selected for the latter to facilitate viability [13,16,17,20]. A strength of our work is that all the participants who responded to the HSCL-25 scale also took part in the structured CIDI interview imparted by trained professionals. We obtained 736 patients who fully answered both the scale and the gold-standard CIDI. Validity criteria were considerable, the global AUC was 0.890 (higher in men than in women). The global sensitivity and specificity by gender were elevated. The former was greater than that found in previous studies [13,16,17,20,25], whilst the latter was similar to the 73% reported by Nettelbladt et al. [20] and the 78% observed by Lundin et al. [16], both in Swedish populations. Other authors have described higher values [13,17]. In spite of the augmented number of false positives obtained, in a similar manner to other studies [36], the negative predictive value was greater than 97% for both genders. Such a finding indicates that the scale is a good tool for depression screening. With respect to the optimum cutoff point, both the global figure and the one for women were very similar to the 1.75 proposed in the original version and employed in other studies [21,37]. Nevertheless, 1.84 for men was higher, and contrasted with the findings of other authors where the cutoff point was greater for women [25]. When contrasting the total rating of the HSCL-25 scale with that of the PHQ-9 [29,38] of depression, an elevated correlation was obtained, and the PABAK was acceptable. Such an analysis reinforces the elevated criteria validity found.

For the one-factor HSCL-25 scale, Cronbach’s alpha coefficient was 0.92, similar to the 0.93 obtained in the French version [13]. Nunally et al. established the critical level of reliability at 0.70; they stated, however, that for the key individual decisions, such as the diagnosis of depression, reliability should be raised to 0.90 [39]. Cronbach’s alpha for the subscales of depression and anxiety taken separately was greater than 0.80. Such findings demonstrate the strong reliability of the scale to measure depression, especially when employed as a single dimension instrument. The test–retest reliability was greater than 0.90, higher than that observed in other studies [18], which indicates that the ratings

are stable over time. The time interval between the baseline interview and the retest was considered adequate, the test conditions—acceptable, in spite of the retest being carried out by telephone to avoid overwhelming the participants.

Other shortened versions of the scale with five and 10 items [40,41] presenting acceptable reliability have been proposed. They could be of use, taking into account the characteristics of the PC environment. These studies have been performed in other languages and it might be of interest to translate them into Spanish.

Our findings indicate that, in the future, the Spanish version of the HSCL-25 scale could be employed as a diagnostic tool for depression in PC consultations. Our study has taken place within the framework of a European project [8,9], in which a common methodology has been used for the translation and adaptation of different languages. We believe, therefore, that the HSCL-25 is a good tool to carry out research concerning the prevalence of depression at the European level once the various language versions have been validated.

5. Conclusions

The Spanish version of the Hopkins Symptom Checklist-25 is well-accepted by patients and shows high validity and reliability to detect depression symptoms in primary care. It has a similar factorial structure to the original English version and can be used in daily practice and for research.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijerph18157843/s1>. Table S1: Participation and non-response patterns, Table S2: Correlation between the items.

Author Contributions: Conceptualization, M.I.F.-S.-M., A.C.-F., J.L., S.A.-L. and B.O.-B.; methodology, M.R.-B., M.C.-C., J.L., S.A.-L., B.O.-B., M.I.F.-S.-M. and A.C.-F.; formal analysis, M.C.-C., M.R.-B.; investigation, M.R.-B., M.I.F.-S.-M., A.C.-F. and E.P.-R.; resources, A.C.-F., M.C.-C., J.L., S.A.-L. and B.O.-B.; data curation, M.C.-C.; writing—original draft preparation, M.R.-B.; writing—review and editing, M.I.F.-S.-M., A.C.-F. and E.P.-R.; supervision, M.I.F.-S.-M., A.C.-F. and E.P.-R.; project administration, M.C.-C., J.L., S.A.-L. and B.O.-B.; funding acquisition, A.C.-F., M.R.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This study received a predoctoral grant from the University Institute of Research in Primary Care IDIAP Jordi Gol (Spain) and another grant from the Autonomous University of Barcelona (UAB)–Novartis for research in Family and Community Medicine (reference 7Z19/019), both in 2019. The study obtained funding from the Spanish Society of Family Medicine (semFYC) as it was awarded a predoctoral grant, “Isabel Fernández 2019”.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki. It was approved by the Ethics Research Committee of the IDIAP Jordi Gol University Institute of Research (P16/025).

Informed Consent Statement: Informed consent was obtained from all the subjects involved in the study.

Data Availability Statement: All the principal investigators had access to the complete database. The datasets generated and analyzed during the study are available from the corresponding author.

Conflicts of Interest: All authors declare that they have no competing interests.

References

1. James, S.L.; Abate, D.; Abate, K.H.; Abay, S.M.; Abbafati, C.; Abbasi, N.; Abbastabar, H.; Abd-Allah, F.; Abdela, J.; Abdelalim, A.; et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2018**, *392*, 1789–1858. [[CrossRef](#)]
2. Vos, T.; Allen, C.; Arora, M.; Barber, R.M.; Bhutta, Z.A.; Brown, A.; Carter, A.; Casey, D.C.; Charlson, F.J.; Chen, A.Z.; et al. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: A systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **2016**, *388*, 1545–1602. [[CrossRef](#)]

3. Serrano-Blanco, A.; Palao, D.; Luciano, J.V.; Pinto-Meza, A.; Lujan-Barroso, L.; Fernandez, A.; Roura, P.; Bertsch, J.; Mercader, M.; Haro, J.M. Prevalence of mental disorders in primary care: Results from the diagnosis and treatment of mental disorders in primary care study (DASMAP). *Soc. Psychiatry Psychiatr. Epidemiol.* **2009**, *45*, 201–210. [[CrossRef](#)] [[PubMed](#)]
4. King, M.; Nazareth, I.; Levy, G.; Walker, C.; Morris, R.; Weich, S.; Bellón-Saameño, J.; Moreno, B.; Švab, I.; Rotar, D.; et al. Prevalence of common mental disorders in general practice attendees across Europe. *Br. J. Psychiatry* **2008**, *192*, 362–367. [[CrossRef](#)]
5. Portellano-Ortiz, C.; Garre-Olmo, J.; Calvó-Perxas, L.; Conde-Sala, J.L. Depression and associated variables in people over 50 years in Spain. *Rev. Psiquiatr. Salud Ment.* **2018**, *11*, 216–226. [[CrossRef](#)] [[PubMed](#)]
6. Gabilondo, A.; Vilagut, G.; Pinto-Meza, A.; Haro, J.M.; Alonso, J. Comorbidity of major depressive episode and chronic physical conditions in Spain, a country with low prevalence of depression. *Gen. Hosp. Psychiatry* **2012**, *34*, 510–517. [[CrossRef](#)] [[PubMed](#)]
7. Nabbe, P.; Le Reste, J.; Chiron, B.; Le Prielec, A.; Munoz, M.; Claveria, A. The FPDM (family practice depression and Multimorbidity) Study: Project for systematic review of literature to find tools for depression diagnosis used in primary care. In Proceedings of the 72nd European General Practice Research Network Meeting, Niza, France, 16 May 2011; p. 25.
8. Nabbe, P.; Le Reste, J.; Guillou-Landreat, M.; Perez, M.M.; Argyriadou, S.; Claveria, A.; Fernandez-San-Martin, M.I.; Czachowski, S.; Lingner, H.; Lygidakis, C.; et al. Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review. *Eur. Psychiatry* **2017**, *39*, 99–105. [[CrossRef](#)] [[PubMed](#)]
9. Nabbe, P.; Le Reste, J.Y.; Guillou-Landreat, M.; Beck-Robert, E.; Assenova, R.; Lazic, D.; Czachowski, S.; Stojanović-Špehar, S.; Hasanagic, M.; Lingner, H.; et al. One consensual depression diagnosis tool to serve many countries: A challenge! A RAND/UCLA methodology. *BMC Res. Notes* **2018**, *11*, 1–8. [[CrossRef](#)] [[PubMed](#)]
10. Derogatis, L.R.; Lipman, R.S.; Rickels, K.; Uhlenhuth, E.H.; Covi, L. The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Syst. Res. Behav. Sci.* **1974**, *19*, 1–15. [[CrossRef](#)]
11. Nabbe, P.; Le Reste, J.; le Floch, B.; Sowinska, A.; Czachowski, S.; Doer, C. FPDM (Family Practice Depression and Multimorbidity): The Hopkins Symptoms Checklist-25 items (HSCL-25), completed translation in 10 European languages. In Proceedings of the 81st European General Practice Research Network Meeting, Edirne, Turkey, 17 October 2015; p. 66.
12. Claveria, A.; Rodríguez-Barragán, M.; Fernández-San-Martín, M.I.; Nabbe, P.; Le Reste, J.Y.; Miguéns-Blanco, I.; Cossio-Gil, Y. Translation and cross-cultural adaptation into Spanish, Catalan and Galician of the Hopkins Symptom Checklist-25 to identify depression in Primary Care. *Aten Primaria* **2020**, *52*, 539–547. [[CrossRef](#)] [[PubMed](#)]
13. Nabbe, P.; Le Reste, J.Y.; Guillou-Landreat, M.; Gatineau, F.; Le Floch, B.; Montier, T.; Van Marwijk, H.; Van Royen, P. The French version of the HSCL-25 has now been validated for use in primary care. *PLoS ONE* **2019**, *14*, e0214804. [[CrossRef](#)] [[PubMed](#)]
14. Mokkink, L.B.; De Vet, H.C.W.; Prinsen, C.A.C.; Patrick, D.L.; Alonso, J.; Bouter, L.; Terwee, C.B. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual. Life Res.* **2017**, *27*, 1171–1179. [[CrossRef](#)] [[PubMed](#)]
15. Glaesmer, H.; Braehler, E.; Grande, G.; Hinz, A.; Petermann, F.; Romppel, M. The German Version of the Hopkins Symptoms Checklist-25 (HSCL-25)—Factorial structure, psychometric properties, and population-based norms. *Compr. Psychiatry* **2014**, *55*, 396–403. [[CrossRef](#)] [[PubMed](#)]
16. Lundin, A.; Hallgren, M.; Forsell, Y. The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden. *J. Affect. Disord.* **2015**, *183*, 247–252. [[CrossRef](#)] [[PubMed](#)]
17. Veijola, J.; Jokelainen, J.; Läksy, K.; Kantojärvi, L.; Kokkonen, P.; Järvelin, M.-R.; Joukamaa, M. The Hopkins Symptom Checklist-25 in screening DSM-III-R axis-I disorders. *Nord. J. Psychiatry* **2003**, *57*, 119–123. [[CrossRef](#)] [[PubMed](#)]
18. Lee, B.; Kaaya, S.; Mbwambo, J.K.; Smith-Fawzi, M.C.; Leshabari, M.T. Detecting Depressive Disorder With the Hopkins Symptom Checklist-25 in Tanzania. *Int. J. Soc. Psychiatry* **2008**, *54*, 7–20. [[CrossRef](#)] [[PubMed](#)]
19. Al-Turkait, F.A.; Ohaeri, J.U.; El-Abbasi, A.-H.M.; Naguy, A. Relationship between Symptoms of Anxiety and Depression in a Sample of Arab College Students Using the Hopkins Symptom Checklist 25. *Psychopathology* **2011**, *44*, 230–241. [[CrossRef](#)] [[PubMed](#)]
20. Nettelbladt, P.; Hansson, L.; Stefansson, C.G.; Borgquist, L.; Nordström, G. Test characteristics of the Hopkins Symptom Checklist-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Soc. Psychiatry Psychiatr. Epidemiol.* **1993**, *28*, 130–133. [[CrossRef](#)] [[PubMed](#)]
21. Fröjd, K.; Håkansson, A.; Karlsson, I. The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. *Int. J. Geriatr. Psychiatry* **2004**, *19*, 386–390. [[CrossRef](#)] [[PubMed](#)]
22. Zabaleta-del-Olmo, E.; Pombo, H.; Pons-Vigués, M.; Casajuana-Closas, M.; Pujol-Ribera, E.; López-Jiménez, T.; Cabezas-Peña, C.; Martín-Borràs, C.; Serrano-Blanco, A.; Rubio-Valera, M.; et al. Complex multiple risk intervention to promote healthy behaviours in people between 45 to 75 years attended in primary health care (EIRA study): Study protocol for a hybrid trial. *BMC Public Health* **2018**, *18*, 874.
23. Mokkink, L.B.; Prinsen, C.A.C.; Patrick, D.L.; Alonso, J.; Bouter, L.M.; de Vet, H.C.W.; Terwee, C.B. COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments. 2019. Available online: www.cosmin.nl (accessed on 19 April 2021).
24. Terwee, C.B.; Bot, S.D.; de Boer, M.; van der Windt, D.A.; Knol, D.L.; Dekker, J.; Bouter, L.; de Vet, H.C. Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* **2007**, *60*, 34–42. [[CrossRef](#)] [[PubMed](#)]

25. Sandanger, I.; Moum, T.; Ingebrigtsen, G.; Dalgard, O.S.; Sørensen, T.; Bruusgaard, D. Concordance between symptom screening and diagnostic procedure: The Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview I. *Soc. Psychiatry Psychiatr. Epidemiol.* **1998**, *33*, 345–354. [[CrossRef](#)]
26. Kessler, R.C.; Üstün, T.B. The World Mental Health (WMH) Survey Initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int. J. Methods Psychiatr. Res.* **2004**, *13*, 93–121. [[CrossRef](#)] [[PubMed](#)]
27. Haro, J.M.; Arbabzadeh-Bouchez, S.; Brugha, T.S.; de Girolamo, G.; Guyer, M.E.; Jin, R.; Lepine, J.P.; Mazzi, F.; Reneses, B.; Vilagut, G.; et al. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *Int. J. Methods Psychiatr. Res.* **2006**, *15*, 167–180. [[CrossRef](#)] [[PubMed](#)]
28. Harkness, J.; Pennell, B.E.; Villar, A.; Gebler, N.; Aguilar-Gaxiola, S.; Bilgen, I. Translation procedures and translation assessment in the World Mental Health Survey Initiative. In *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*; Kessler, R.C., Üstün, T., Eds.; Cambridge University Press: New York, NY, USA; World Health Organization: Geneva, Switzerland, 2008; pp. 91–113.
29. Kroenke, K. The Phq-9. *J. Gen. Intern. Med.* **2001**, *16*, 606–613. [[CrossRef](#)]
30. Diez-Quevedo, C.; Rangil, T.; Sanchez-Planell, L.; Kroenke, K.; Spitzer, R.L. Validation and Utility of the Patient Health Questionnaire in Diagnosing Mental Disorders in 1003 General Hospital Spanish Inpatients. *Psychosom. Med.* **2001**, *63*, 679–686. [[CrossRef](#)] [[PubMed](#)]
31. American Psychiatric Association. *Manual Diagnóstico y Estadístico de los Trastornos Mentales DSM-5*, 5th ed.; Editorial Médica Panamericana: Madrid, Spain, 2014; 1100p.
32. Skogen, J.C.; Øverland, S.; Smith, O.R.; Aarø, L.E. The factor structure of the Hopkins Symptoms Checklist (HSCL-25) in a student population: A cautionary tale. *Scand. J. Public Health* **2017**, *45*, 357–365. [[CrossRef](#)] [[PubMed](#)]
33. Ashaba, S.; Kakuhikire, B.; Vořechovská, D.; Perkins, J.M.; Cooper-Vince, C.E.; Maling, S.; Bangsberg, D.R.; Tsai, A.C. Reliability, Validity, and Factor Structure of the Hopkins Symptom Checklist-25: Population-Based Study of Persons Living with HIV in Rural Uganda. *AIDS Behav.* **2017**, *22*, 1467–1474. [[CrossRef](#)] [[PubMed](#)]
34. Bendixen, A.B.; Engedal, K. Anxiety among older psychiatric patients: A hidden comorbidity? *Ageing Ment. Health* **2016**, *20*, 1131–1138. [[CrossRef](#)]
35. Martinussen, L.J.; Benth, J.Š.; Almdahl, I.S.; Borza, T.; Selbæk, G.; Mcpherson, B.; Korsnes, M.S. The effect of anxiety on cognition in older adult inpatients with depression: Results from a multicenter observational study. *Heliyon* **2019**, *5*, e02235. [[CrossRef](#)] [[PubMed](#)]
36. Bantjes, J.; Kagee, A.; Saal, W. The Utility of the Hopkins Symptom Checklist as a Trans-Diagnostic Screening Instrument for Common Mental Disorders Among Persons Seeking HIV Testing. *AIDS Behav.* **2019**, *24*, 629–636. [[CrossRef](#)] [[PubMed](#)]
37. Mattisson, C.; Bogren, M.; Horstmann, V. Correspondence between clinical diagnoses of depressive and anxiety disorders and diagnostic screening via the Hopkins Symptom Check List-25 in the Lundby Study. *Nord. J. Psychiatry* **2012**, *67*, 204–213. [[CrossRef](#)] [[PubMed](#)]
38. Reuland, D.S.; Cherrington, A.; Watkins, G.S.; Bradford, D.W.; Blanco, R.A.; Gaynes, B.N. Diagnostic Accuracy of Spanish Language Depression-Screening Instruments. *Ann. Fam. Med.* **2009**, *7*, 455–462. [[CrossRef](#)] [[PubMed](#)]
39. Nunnally, J.C. *Psychometric Theory*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1978.
40. Schmalbach, B.; Zenger, M.; Tibubos, A.N.; Kliem, S.; Petrowski, K.; Brähler, E. Psychometric Properties of Two Brief Versions of the Hopkins Symptom Checklist: HSCL-5 and HSCL-10. *Assessment* **2019**, *28*, 617–631. [[CrossRef](#)] [[PubMed](#)]
41. Strand, B.H.; Dalgard, O.S.; Tambs, K.; Rognerud, M. Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nord. J. Psychiatry* **2003**, *57*, 113–118. [[CrossRef](#)] [[PubMed](#)]

Material suplementario artículo 2:

Table S1. Participation and non-response patterns.

Number of Missing Items	Missing for Which Variables?	<i>n</i> (%)
25	12345 67890 12345 67890 12345	22 (2.78)
13	_234_ _8_0 123_5 ___0 12_5	1 (0.13)
3	_23_ _____ 6_____	1 (0.13)
1	_2_____	2 (0.25)
5	_3_ 6_8 _____ 8_ 1_	1 (0.13)
5	_____ 5 6789_ _____	1 (0.13)
3	_____ 5 6_____ 8_____	1 (0.13)
2	_____ 5 _____ 9_____	1 (0.13)
2	_____ 7 _____ 7_____	1 (0.13)
1	_____ 7_____	3 (0.38)
1	_____ 8_____	1 (0.13)
1	_____ 9_____	2 (0.25)
2	_____ _____ 1_____ 1_____	1 (0.13)
1	_____ _____ 2_____	1 (0.13)
2	_____ _____ 34_____	1 (0.13)
1	_____ _____ 3_____	2 (0.25)
2	_____ _____ 4_ 7_____	1 (0.13)
1	_____ _____ 4_____	2 (0.25)
2	_____ _____ 67_____	1 (0.13)
2	_____ _____ 6_____ 1_____	1 (0.13)
1	_____ _____ 6_____	1 (0.13)
1	_____ _____ 8_____	1 (0.13)
2	_____ _____ 90_____	1 (0.13)
1	_____ _____ 0_____	1 (0.13)
2	_____ _____ _____ 1_ 4_	1 (0.13)
1	_____ _____ _____ 1_____	1 (0.13)
0	_____ _____ _____ _____	737 (93.29)
	Total	790 (100)

The first column refers to how many items are unanswered. The second one specifies which items are left as missing, ordered by units and dozens. Finally, the far right column shows how many participants performed the concrete non-response pattern, e.g. 3 participants left item 7 as a missing value

Item 21	0.239	0.262	0.328	0.338	0.292	0.154	0.336	0.193	0.182	0.363	0.432	0.358	0.282	0.349	0.479	0.493	0.560	0.289	0.377	0.311	1.000			
Item 22	0.251	0.302	0.547	0.409	0.335	0.274	0.434	0.236	0.159	0.367	0.621	0.375	0.274	0.394	0.433	0.413	0.534	0.231	0.406	0.372	0.519	1.000		
Item 23	0.173	0.252	0.304	0.243	0.183	0.226	0.313	0.148	0.216	0.237	0.340	0.335	0.244	0.286	0.414	0.370	0.435	0.323	0.402	0.333	0.389	0.451	1.000	
Item 24	0.092	0.151	0.220	0.215	0.183	0.222	0.139	0.128	0.191	0.191	0.184	0.150	0.158	0.191	0.175	0.195	0.202	0.146	0.190	0.248	0.204	0.209	0.158	1.000
Item 25	0.222	0.239	0.275	0.369	0.324	0.199	0.357	0.335	0.158	0.304	0.344	0.251	0.288	0.226	0.293	0.304	0.379	0.187	0.251	0.287	0.246	0.293	0.208	1.000



OPEN ACCESS

EDITED BY

Simon Ching Lam,
Tung Wah College, Hong Kong SAR,
China

REVIEWED BY

Emma Yun Zhi Huang,
Macau University of Science
and Technology, Macau SAR, China
Fernanda Ávila,
Fluminense Federal University, Brazil

*CORRESPONDENCE

Ana Clavería
✉ anaclaveria@gmail.com

SPECIALTY SECTION

This article was submitted to
Family Medicine and Primary Care,
a section of the journal
Frontiers in Medicine

RECEIVED 08 August 2022

ACCEPTED 21 December 2022

PUBLISHED 09 January 2023

CITATION

Rodríguez-Barragán M,
Fernández-San-Martín MI, Clavería A,
Le Reste JY, Nabbe P, Motrico E,
Gómez-Gómez I and
Peguero-Rodríguez E (2023)
Measuring depression in Primary
Health Care in Spain: Psychometric
properties and diagnostic accuracy
of HSCL-5 and HSCL-10.
Front. Med. 9:1014340.
doi: 10.3389/fmed.2022.1014340

COPYRIGHT

© 2023 Rodríguez-Barragán,
Fernández-San-Martín, Clavería, Le
Reste, Nabbe, Motrico, Gómez-Gómez
and Peguero-Rodríguez. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Measuring depression in Primary Health Care in Spain: Psychometric properties and diagnostic accuracy of HSCL-5 and HSCL-10

María Rodríguez-Barragán^{1,2,3},
María Isabel Fernández-San-Martín^{2,4}, Ana Clavería^{5,6,7*},
Jean Yves Le Reste⁸, Patrice Nabbe⁸, Emma Motrico^{7,9},
Irene Gómez-Gómez^{7,9} and Eva Peguero-Rodríguez^{2,10,11}

¹Primary Health Centre La Mina, Gerència Territorial d'Atenció Primària de Barcelona, Institut Català de la Salut, Sant Adrià de Besòs, Barcelona, Spain, ²Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain, ³Faculty of Medicine, Department of Pediatrics, Obstetrics, Gynecology and Preventive Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain, ⁴Gerència Territorial d'Atenció Primària de Barcelona, Institut Català de la Salut, Barcelona, Spain, ⁵I-Saúde Group, South Galicia Health Research Institute (IISGS-Servicio Gallego de Salud (SERGAS)), Vigo, Spain, ⁶Vigo Health Area, Servicio Gallego de Salud (SERGAS), Vigo, Spain, ⁷Network for Research on Chronicity, Primary Care and Health Promotion (RICAPPS), Vigo, Spain, ⁸Department of General Practice, ER 7479 SPURBO Soins Primaires, Santé Publique, Registre des Cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France, ⁹Department of Psychology, Universidad Loyola, Andalucía, Spain, ¹⁰Primary Health Centre El Castell, Gerència Territorial d'Atenció Primària Metropolitana Sud, Institut Català de la Salut, Castelldefels, Barcelona, Spain, ¹¹Departamento de Ciencias Clínicas, Facultad de Medicina, Universidad de Barcelona, Barcelona, Spain

Background: Depression has a high prevalence among European countries. Several instruments have been designed to assess its symptoms in different populations. The Hopkins Symptom Checklist 25 (HSCL-25) scale has been identified as valid, reproducible, effective, and easy to use. There are short versions of this scale that could be useful in Primary Care (PC) settings, but their psychometric properties are unknown.

Aim: To assess in PC patients the psychometric properties and diagnostic accuracy of the Spanish version of the HSCL-10 and the HSCL-5 consisting of 10 and 5 items, respectively.

Methods: A multicenter, cross-sectional study was carried out at six PC centers in Spain. The HSCL-25 was administered to outpatients aged 45–75 who also participated in the structured Composite International Diagnostic Interview (CIDI). HSCL-10 and HSCL-5 were assessed and compared to HSCL-25 regarding total score correlation, internal consistency, and criterion validity against the gold-standard CIDI. This is a methodological study from a secondary data analysis and the primary data has been previously published.

Results: Out of 790 patients, 767 completed the HSCL-25 and 736 the CIDI interview (96.0%). Cronbach's Alpha was 0.84 for HSCL-10 and 0.77 for HSCL-5. The known-group method and confirmatory factor analysis were acceptable for the establishment of construct validity. Sensitivity was 79.7% (CI95%, 67.7–88.0%) for HSCL-10, and 78.0% (CI95%, 65.9–86.6%) for HSCL-5, whereas specificity was 83% (CI95%, 80.0–85.7%) for HSCL-10, and 72.8% (CI95%, 69.3–76.0%) for HSCL-5. Area under the curve against CIDI was 0.88 (CI95%, 0.84–0.92%) for HSCL-10, and 0.85 (CI95%, 0.81–0.89%) for HSCL-5. Optimum cutoff point calculated with Youden Index was 1.90 for the HSCL-10 and 1.80 for the HSCL-5.

Conclusion: HSCL-10 and HSCL-5 are reliable and valid tools to detect depression symptoms and can be used in PC settings.

KEYWORDS

depression, Hopkins Symptom Checklist, Primary Health Care (MeSH), patient reported outcome measures (MeSH), diagnostic accuracy, psychometric properties

Introduction

Mental health is an issue of increasing concern in Spain and other European countries and represents a considerable percentage of Primary Care (PC) consultations. The clinical diagnosis of mental disorders is based on the symptoms of the Diagnostic and Statistical Manual of Mental Disorders (DSM). In addition, there are many questionnaires available which assess mental health diseases in populations and are frequently employed in PC (1). Such instruments are used for the purposes of screening and symptom detection to aid diagnosis.

The United States Preventive Services Task Force (USPSTF) recommends universal screening for depression in the general adult population (2) as it has been shown to reduce symptoms and improve quality of life and functional status (3). Screening should, however, be combined with adequate systems to ensure accurate diagnosis, effective treatment, and appropriate follow-up.

Up to 75% of patients with depression are treated exclusively in PC (4). Nevertheless, Family Doctors (FD) report difficulties in the identification of mental health disorders (5) and consider that evaluative tools permitting rapid diagnosis with a limited number of items are required (6). The issue of time is particularly relevant within the context of PC given its complexity and limited consultation schedules. Moreover, FD have been found to feel more confident in detecting and managing depression when employing questionnaires (7).

The ideal questionnaire should be reliable, valid, ergonomic, and easy to use for both patients and healthcare professionals. It should be validated according to the Consensus-based Standards for the Selection of Health Measurement Instruments

(COSMIN) (8) for the language and population they are directed at.

The Hopkins Symptom Checklist 25 (HSCL-25) (9) is a widely used, self-administered screening tool for depression and psychological distress; it is also employed for research purposes. It has been shown to be valid and reliable for different populations (10–14), useful in PC (15), and it is available in several languages (16) including Spanish. The Spanish version of HSCL-25 was translated, culturally adapted, and validated following the COSMIN (17, 18).

The HSCL-25 is a short version of the Symptom Checklist 90-Revised (SCL-90-R) (19). There are, however, other questionnaires with fewer items (20) which are of interest due to their brevity and reduced completion time. Shorter versions are useful as they can save time in complex settings such as the PC context. The HSCL-10 and HSCL-5 were developed by selecting 10 and 5 items, respectively, due to their strong correlation with the HSCL-25 mean score (19, 21). Items of the HSCL-10 and HSCL-5 are included in the longer HSCL-25. Both the HSCL-10 and the HSCL-5 have shown a strong mean score correlation with respect to the HSCL-25 and high reliability (Cronbach's Alpha Coefficient >0.80) (22). Both short versions, especially the HSCL-10, have been used in different populations such as adolescents (23–25), industry workers (26), population surveys (22, 27), patients with alcohol use disorder (28), refugees (29), and also to measure symptoms of depression and anxiety due to the COVID pandemic (30–32). Their psychometric properties recommend them for clinical use as screening and symptom assessment tools and for research purposes (25, 27, 33).

The aim of this article is to report the psychometric properties and diagnostic accuracy of the Spanish versions of

the HSCL-5 and HSCL-10 for their use as rapid and accessible depression screening instruments in PC.

Materials and methods

Data collection, study population, and variables

The present study was based on data from a cross-sectional multicenter study designed to validate the HSCL-25 in a Spanish PC population. This is a methodological study from a secondary data analysis and the primary data has been previously published (18). Participants were patients attending six Spanish Primary Health Centers (PHC) taking part in the EIRA study (34, 35). Inclusion criteria were to be aged 45–75 years and presenting two or more of the following: Smoking, low adherence to the Mediterranean dietary pattern, and insufficient physical activity. Exclusion criteria were: Advanced serious illness, cognitive impairment, dependence in basic everyday activities, severe mental illness, unable to attend the PHC, under treatment for cancer or in end-of-life care, or planning to travel during the intervention period.

Participants were recruited by consecutive sampling of patients attending the PHC for any reason during a 6-month-period in 2017. They were asked to complete sociodemographic data (gender, age, nationality, marital status, current employment, and education level) and the self-administered HSCL-25 questionnaire (and other questionnaires/forms related to the EIRA study). Afterward, trained professionals, blinded to the HSCL-25 score, conducted the gold-standard CIDI interview with all participants.

Hopkins symptom checklist-25 (HSCL-25)

The HSCL-25 is a widely used, self-administered questionnaire designed to measure anxiety and depression symptoms (9, 11) and takes 5–10 min to complete (13). It consists of 10 and 15 items belonging to the anxiety and depression dimensions, respectively. The items are answered on a four-point Likert-like scale: 1 = “Not at all;” 2 = “A little;” 3 = “Quite a bit;” 4 = “Extremely.” The average score, ranging from 1 to 4, is calculated by dividing the total score by the number of the items. A cutoff value of 1.75 is generally used for major depression diagnosis, as it is considered a valid predictor of mental disorder (10, 13, 36).

Items belonging to HSCL-25, HSCL-10, and HSCL-5 are shown in [Table 1](#). The corresponding cutoffs points are 1.85 for the HSCL-10 and 2.00 for the HSCL-5 (22). The Spanish version of the HSCL-25 was used in this study (17).

Composite international diagnostic interview (CIDI)

The CIDI is a well-known, standardized interview designed by the World Health Organization (WHO) based on the fourth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and the International Classification of Diseases-10 (ICD-10) criteria (37). It is administered by trained interviewers and available in different languages (38). For this study, section E (questions referring to depression) of the Spanish version was used. The CIDI was conducted by trained psychologists.

As the diagnose of depression is a clinical interview performed by a trained professional and conducted using the DSM criteria, the CIDI is considered the gold-standard in the present study.

Ethical considerations

The study was developed according to national and international legislation (the Declaration of Helsinki and latest versions). The protocol was evaluated by the IDIAP Jordi Gol Ethical Research Committee (approval number: P16/025) and by the corresponding regional governments. Written consent was obtained from the participants, and the questionnaires were codified with an identification number to protect anonymity and confidentiality.

Statistical analysis

Analysis was conducted with STATA version 15. Missing values for the HSCL-25 were replaced with the individual mean for the rest of the items. Subjects with $\geq 50\%$ missing items were excluded.

Total score was calculated for HSCL-25, HSCL-10, and HSCL-5 for the total population and in relation to gender and age categories, and by the following sociodemographic groups: marital status, education level, and current employment.

Reliability of the HSCL-10 and HSCL-5 was analyzed by calculating the Cronbach's Alpha Coefficient and for each of the two depression and anxiety subscales. A value of ≥ 0.7 was considered adequate (39). Cronbach's Alpha without the item was also calculated to assess the contribution of each item to the internal consistency of two versions.

Construct validity was measured with the known-groups method by comparing the total score of the HSCL-10 and the HSCL-5 by gender. The total score was expected to be significantly higher among women (40, 41). Independent sample *t*-test was performed, a significant result ($p < 0.05$) was considered satisfactory (42). Confirmatory factor analysis (CFA) was conducted to assess the structural validity. To

TABLE 1 Items belonging to HSCL-25, HSCL-10, and HSCL-5 scale.

Item in HSCL-25	HSCL-10	HSCL-5	
1. Being scared for no reason	X		Anxiety dimension
2. Feeling fearful	X	X	
3. Faintness	X		
4. Nervousness		X	
5. Heart racing			
6. Trembling			
7. Feeling tense	X		
8. Headache			
9. Feeling panic			
10. Feeling restless			
11. Feeling low in energy			Depression dimension
12. Blaming oneself	X		
13. Crying easily			
14. Losing sexual interest			
15. Feeling lonely			
16. Feeling hopeless	X	X	
17. Feeling blue	X	X	
18. Thinking of ending one's life			
19. Feeling trapped			
20. Worrying too much		X	
21. Feeling no interest			
22. Feeling that everything is an effort	X		
23. Worthless feeling	X		
24. Poor appetite			
25. Sleep disturbance	X		

Items belonging to depression and anxiety dimensions.

evaluate the estimated model fit, the absolute fit index was calculated with chi-squared distribution. Given that this value may be affected by the sample size, complementary indices were employed: The root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the comparative fit index (CFI), and the Tucker–Lewis fit index (TLI) (43, 44). Cutoff values considered adequate were: SRMR < 0.05, RMSEA ≤ 0.08, CFI > 0.90, and TLI > 0.90 (45).

Criterion validity and diagnostic accuracy were measured by calculating the ROC curve for the HSCL-10 and HSCL-5 scale in comparison with the gold-standard CIDI (8). The area under the curve (AUC) was estimated with 95% confidence interval (CI95%). Best cutoff points for the study population and by gender were calculated with the Youden Index for both HSCL-10 and HSCL-5. Youden Index is defined as “Sensitivity + Specificity − 1,” it is a value that indicates the validity of the instrument for a specific cutoff point (46). Sensitivity and specificity were assessed as measures

of internal validity; positive and negative predictive values were also calculated. Both the sensitivity and specificity of a screening test should be greater than 0.70 (42). For these calculations, cutoffs those proposed by Strand et al. were followed: 1.85 for the HSCL-10 and 2.00 for the HSCL-5 (22). Other authors have employed the same cutoff points (47).

Results

Participants

From a total of 790 patients, 767 completed the HSCL-25 (97.1% response rate). Participants' mean age was 58.4 years (± 8.2), 54.4% were women, and there were no significant gender differences among age categories. Table 2 shows the mean score and standard deviation (SD) of the three HSCL

TABLE 2 HSCL-25, HSCL-10, and HSCL-5 scale mean scores and standard deviation (SD) in relation to age, gender, marital status, education, and current employment.

	HSCL-25					HSCL-10			HSCL-5		
	<i>n</i>	%	Mean	SD	<i>P</i> -value	Mean	SD	<i>P</i> -value	Mean	SD	<i>P</i> -value
Total	767		1.57	0.45		1.55	0.50		1.72	0.59	
Gender											
Men	350	45.6	1.42	0.35	<0.001	1.41	0.39	<0.001	1.53	0.46	<0.001
Women	417	54.4	1.69	0.49		1.68	0.55		1.88	0.64	
Age											
45–54	298	38.7	1.63	0.47	0.006	1.62	0.52	0.005	1.71	0.61	0.060
55–64	265	34.7	1.55	0.49		1.54	0.51		1.71	0.61	
65–75	204	26.6	1.50	0.40		1.47	0.44		1.64	0.54	
Marital status											
Married/with a partner	550	71.9	1.54	0.43	0.028	1.52	0.47	0.029	1.69	0.56	0.084
Single	57	7.3	1.58	0.44		1.58	0.48		1.73	0.59	
7-810-11 Separated or divorced	103	13.3	1.68	0.48		1.69	0.56		1.82	0.67	
Widow (er)	57	7.5	1.64	0.51		1.60	0.61		1.86	0.65	
Education											
Primary or lower	398	51.8	1.57	0.42	0.855	1.54	0.48	0.697	1.74	0.58	0.333
Secondary or higher	368	48.2	1.56	0.47		1.56	0.52		1.70	0.60	
Current employment											
Employed	313	41.0	1.55	0.43	0.547	1.54	0.48	0.774	1.84	0.63	0.746
Housewife	111	14.5	1.64	0.45		1.61	0.51		1.82	0.60	
Unemployed	83	10.9	1.65	0.46		1.65	0.55		1.82	0.60	
Retired	209	27.3	1.45	0.40		1.43	0.44		1.57	0.54	
Others (student, sick leave, and disability)	48	6.3	1.88	0.53		1.88	0.58		1.98	0.58	

versions in relation to the sociodemographic characteristics of the sample.

There were statistically significant differences in total scores for the three versions regarding gender. Women scored higher with a minimum difference >0.25 points. There were also statistically significant differences in the total scores of the HSCL-25 and HSCL-10, but not the HSCL-5, with respect to age and marital status. No differences were observed regarding education level or current occupation.

Reliability: internal consistency

Cronbach's Alpha was 0.8417 and 0.7712 for the HSCL-10 and HSCL-5, respectively. When analyzing the two dimensions separately, this value was higher for the depression dimension than the anxiety one. These values, and the value of the coefficient without the item, are depicted in [Table 3](#).

The most consistent item for the HSCL-10 was 17 "Feeling blue" followed by 22 "Feeling that everything is an effort."

The least consistent was 25 "Sleep disturbance" followed by 1 "Being scared for no reason." Item 25 "Sleep disturbance" had a Cronbach's Alpha without the item of 0.8424, as a result, this item worsened the reliability of this version, as this value was above 0.8417. In the HSCL-5, item 17 "Feeling blue" was also the most consistent whilst two "Feeling fearful" was the least consistent although without affecting reliability.

Construct validity: Known-groups method and confirmatory factor analysis (CFA)

The known-group method analysis showed that the total score of the HSCL-10 indicated that women had significantly higher scores (mean = 1.68; SD 0.03) than men (mean = 1.41; SD 0.02; $t = 7.76$; $p < 0.001$). Results were in the same direction with the HSCL-5, total score was significantly higher in women (mean = 1.88; SD 0.03) than in men (mean = 1.53; SD 0.02; $t = 8.51$; $p < 0.001$).

TABLE 3 Cronbach's alpha coefficient without the item and for total values.

Item	HSCL-25	HSCL-10	HSCL-5
1. Being scared for no reason	0.9148	0.8359	
2. Feeling fearful	0.9137	0.8307	0.7580
3. Faintness	0.9127	0.8268	
4. Nervousness	0.9109		0.7156
5. Heart racing	0.9136		
6. Trembling	0.9151		
7. Feeling tense	0.9109	0.8205	
8. Headache	0.9163		
9. Feeling panic	0.9156		
10. Feeling restless	0.9117		
11. Feeling low in energy	0.9110		
12. Blaming oneself	0.9131	0.8279	
13. Crying easily	0.9145		
14. Losing sexual interest	0.9153		
15. Feeling lonely	0.9121		
16. Feeling hopeless	0.9127	0.8261	0.7427
17. Feeling blue	0.9085	0.8080	0.6820
18. Thinking of ending one's life	0.9159		
19. Feeling trapped	0.9129		
20. Worrying too much	0.9129		0.7399
21. Feeling no interest	0.9129		
22. Feeling that everything is an effort	0.9112	0.8188	
23. Worthless feeling	0.9140	0.8315	
24. Poor appetite	0.9167		
25. Sleep disturbance	0.9152	0.8424	
Total	0.9166	0.8417	0.7712
Anxiety subscale (items 1–10)	0.8306	0.6859	0.5404
Depression subscale (items 11–25)	0.8784	0.7804	0.7010

Table 4 shows the results of the CFA: the factor loading for each model and correlation in the two-factor models. All the factor loadings were positive, statistically significant ($p < 0.001$), and above 0.30. In fact, all factor loadings in the different versions were above 0.45. The range of loadings was 0.45–0.81 for the one factor HSCL-10, 0.50–0.83 for the one factor HSCL-5, 0.47–0.82 for the two correlated factor HSCL-10, and 0.51–0.85 for the two correlated factor HSCL-5. Item 17 “Feeling blue” was the item with the highest factor loadings in all the models analyzed. When analyzing the models with two correlated factors, a strong correlation between the two-factors of depression and anxiety was observed for both the HSCL-10 and the HSCL-5.

The goodness-of-fit indices in the studied factor models can be consulted in [Supplementary Table 1](#) of the [Supplementary materials](#). Globally, the indices showed

that the HSCL-10 and the HSCL-5 do not have a stable factor structure.

Criterion validity and diagnostic accuracy: Relation of HSCL-10 and HSCL-5 with gold-standard CIDI

Of the 767 participants who completed the HSCL-25, 736 also took part in the CIDI interview (96.0%). Depression prevalence varied depending on the questionnaire employed (Table 5). Prevalence measured with the HSCL-10 was similar to that obtained with the full HSCL-25 version and higher than the value obtained with the HSCL-5.

Table 6 shows sensitivity, specificity, positive, and negative predictive values for the total number of participants and by

TABLE 4 Confirmatory factorial analysis: Factor loading values and correlation between two anxiety and depression factors.

	One factor		Two correlated factors			
	HSCL-10	HSCL-5	HSCL-10		HSCL-5	
			Anxiety	Depression	Anxiety	Depression
Item 1	0.45		0.49			
Item 2	0.52	0.50	0.56		0.54	
Item 3	0.60		0.62			
Item 4		0.64			0.71	
Item 7	0.66		0.69			
Item 12	0.57			0.58		
Item 16	0.62	0.63		0.64		0.64
Item 17	0.81	0.83		0.82		0.85
Item 20		0.58				0.57
Item 22	0.69			0.68		
Item 23	0.54			0.55		
Item 25	0.47			0.47		
Factor correlation			0.90		0.87	

TABLE 5 Prevalence of depression according to CIDI, HSCL-25, HSCL-10, and HSCL-5, total and by gender.

	Male (n = 341)	Female (n = 395)	Total (n = 736)
CIDI	4.7 (CI95% 2.7–7.5)	10.9 (CI95% 8.0–14.4)	8.0 (CI95% 6.2–10.2)
HSCL-25 (cutoff = 1.75)	16.7 (CI95% 12.9–21.1)	38.7 (CI95% 33.9–43.7)	28.5 (CI95% 25.3–31.9)
HSCL-10 (cutoff = 1.85)	12.6 (CI95% 9.3–16.6)	30.1 (CI95% 25.6–34.9)	22.0 (CI95% 19.1–25.2)
HSCL-5 (cutoff = 2.00)	19.4 (CI95% 15.3–24.0)	41.5 (CI95% 36.6–46.6)	31.3 (CI95% 27.9–34.7)

gender. Sensitivity was similar for both genders, especially in the HSCL-10, whilst specificity was better in men. Negative predictive values were >95% for both versions, these values were for the total population and by gender. All values were higher for the HSCL-10 than for the HSCL-5.

HSCL10 vs. CIDI

The AUC between the HSCL-10 and the CIDI was 0.877 (CI95% 0.836–0.919). In the gender analysis it was greater in men with an AUC of 0.943 (CI95% 0.897–0.989) compared to women who had an AUC of 0.825 (CI95% 0.765–0.886). The ROC curve is depicted in [Figure 1](#) and by gender in the [Supplementary Figure 1](#).

HSCL5 vs CIDI

The AUC between the HSCL-5 and the CIDI was 0.853 (CI95% 0.812–0.894). In the gender analysis it was greater in men with an AUC of 0.918 (CI95% 0.859–0.977) than in women who had an AUC of 0.795 (CI95% 0.734–0.855). The ROC curve

is depicted in [Figure 2](#) and by gender in the [Supplementary Figure 2](#).

The optimum cutoff points for the study population were calculated with the Youden Index. For the HSCL-10 it was 1.90 for the total population and for both genders. For the HSCL-5 the optimum cutoff point was 1.80 for the total population and for women, and 2.00 for men. A table including sensitivity, specificity, positive, and negative predictive values using optimal cutoff points is available in the [Supplementary Table 2](#).

Discussion

This study is the first to analyze the psychometric properties and diagnostic accuracy of the Spanish versions of the HSCL-10 and HSCL-5, the results obtained now allow the use of both scales as screening tools for depression in the PC setting in Spain. The results demonstrated that in the Spanish PC population, the HSCL-10 presents high reliability and validity. The HSCL-5 also showed acceptable psychometric properties although slightly worse than the HSCL-10. Both scales showed adequate sensitivity and specificity when compared to the semistructured clinical interview CIDI conducted by trained

TABLE 6 Sensitivity, specificity, positive and negative predictive values, total and by gender in the HSCL-10, and the HSCL-5.

		Male (n = 341)			Female (n = 395)			Total (n = 736)		
		Index	CI 95% lower limit	CI 95% upper limit	Index	CI 95% lower limit	CI 95% upper limit	Index	CI 95% lower limit	CI 95% upper limit
HSCL-10 cutoff point = 1.85	Sensitivity	81.3	57.0	93.4	79.1	64.8	88.6	79.7	67.7	88.0
	Specificity	90.8	87.1	93.5	75.9	71.1	80.0	83.0	80.0	85.7
	PPV	30.2	18.6	45.1	28.6	21.2	37.3	29.0	22.6	36.4
	NPV	99.0	97.1	99.7	96.7	93.9	98.3	97.9	96.4	98.8
HSCL-5 cutoff point = 2.00	Sensitivity	81.3	57.0	93.4	76.7	62.3	86.8	78.0	65.9	86.6
	Specificity	83.7	79.3	87.3	62.8	57.6	67.7	72.8	69.3	76.0
	PPV	19.7	11.9	30.8	20.1	14.7	26.9	20.0	15.3	25.6
	NPV	98.9	96.8	99.6	95.7	92.2	97.6	97.4	95.7	98.5

PPV, positive predictive value; NPV, negative predictive value.

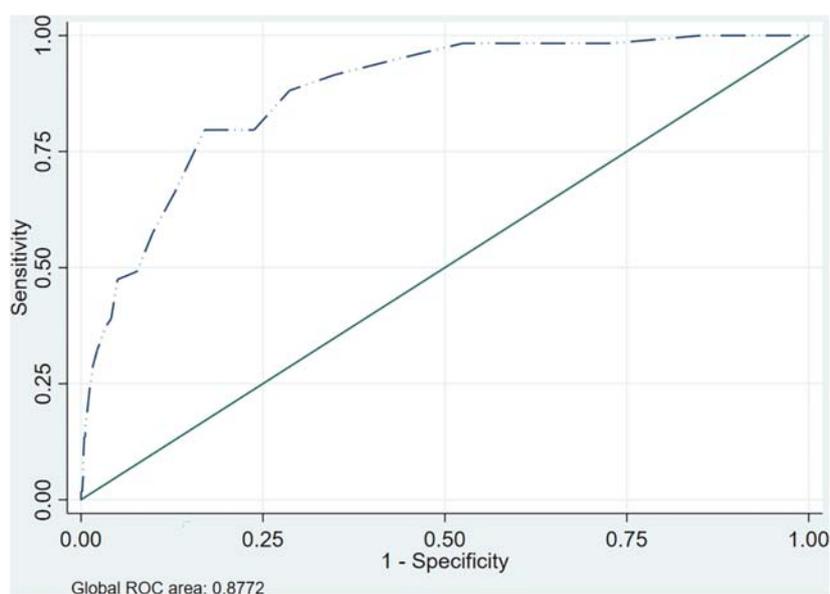


FIGURE 1
ROC curve and AUC HSCL-10 vs. CIDI.

professionals. The optimal cutoff points obtained were very close to those proposed by other authors.

The PC setting is ideal for the detection, diagnosis, and investigation of chronic, highly prevalent disorders as it allows an early study of such pathologies (48). With respect to depression, it is widespread in Spanish and European countries (40) yet patients frequently consult their FD for other motives than their mood disorders (6).

There are many questionnaires that can be employed as screening tools to detect depression at all its stages (49–52), some of which have been validated within the PC setting (1). Moreover, shortened versions have been proposed in

order to provide a similar diagnostic value that saves time for both patient and healthcare professional (20, 53, 54). Recently, the HSCL-10 and HSCL-5 have shown acceptable reliability and validity (22, 27), particularly the HSCL-10 (25, 33, 47).

With respect to reliability, Cronbach's Alpha results of 0.84 and 0.77 were obtained for the HSCL-10 and HSCL-5, respectively. As both were above 0.7, they are considered acceptable (55). Such findings are similar to those obtained by other authors (33) and, as occurred in other studies (19, 22, 27), the reliability of the two short versions was lower than the HSCL-25.

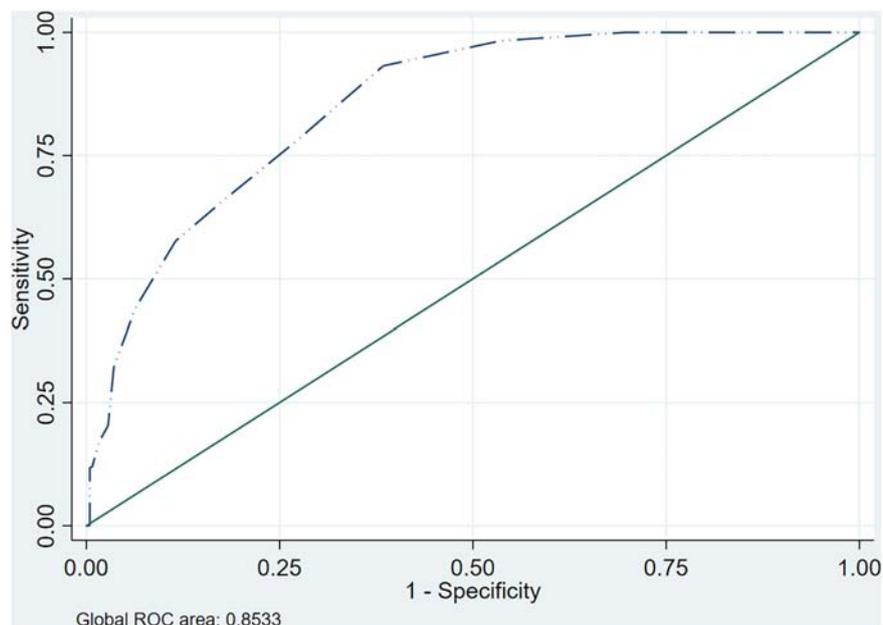


FIGURE 2
ROC curve and AUC HSCL-5 vs. CIDI.

In the full 25-item version, 17 “Feeling blue” was the most consistent (18). This item which asks about sadness, a basic characteristic in patients with depression, is included in the HSCL-10 and the HSCL-5 and was also the most consistent in the two short versions. Once removed, reliability diminished to the minimum as can be seen in Table 3. The next most consistent items in the HSCL-25 were four “Nervousness” and seven “Feeling tense.” The former was included in the 5-item version and the latter in the 10-item one. Item 24 was the least consistent in the HSCL-25 and is not present in either of the short versions. The following least consistent items in the HSCL-25 were 8 “Headache” and 18 “Thinking of ending one’s life,” neither of which is included in the two short versions.

When analyzing Cronbach’s Alpha without the item in the short versions, it was observed that 25 “Sleep disturbance” worsened HSCL-10 reliability, that is to say, by eliminating this item reliability improved. Such a finding concurs with that reported by Kleppang et al. who employed the 10-item version with adolescents in Norway (25). The other items contributed to good reliability in both the HSCL-10 and HSCL-5.

Regarding analysis of the scale’s factorial structure, this was performed with the CFA as the HSCL-25 has been widely studied with one single factor or two correlated ones even though other models have been proposed (10). The HSCL-10 and the HSCL-5 maintain the same factor structure (27). The goodness-of-fit indices for the studied models were not optimal, showing that the data did not fit the hypothesized factor structure of one factor and two correlated factors. The factor structure was unstable for both scales, this is a limitation of

our study. Further validation studies should be done to assess other alternative models of the factor structure of the HSCL-10 and the HSCL-5. By examining the factor loading of 10 and 5 items, respectively, all the items were significantly loaded to the hypothesized construct, and all factor loadings were statistically significant, positive, and above 0.45. Item 17 “Feeling blue” had the highest factor loadings in all models tested and for both the HSCL-10 and the HSCL-5. In the study of the two-factor models, there was a factorial correlation of 0.90 between the dimensions in the HSCL-10 and of 0.87 in the HSCL-5, these findings indicated that the depression and anxiety dimensions strongly correlated in a positive manner. The correlation is expectable as anxiety and depression are frequently found to be associated comorbidities (56).

The calculation of the total score is done in the same way in the different versions of the HSCL, by dividing the total score by the corresponding number of items answered. With respect to the total score, in comparison with the 25-item long version, the mean score of the HSCL-25 was 1.57, very similar to the 1.55 obtained with the HSCL-10. The mean score increased to 1.72 with the five-item version. The means were significantly greater in women for all three versions, this is unsurprising considering that depression is more prevalent in the female gender (40, 41). Significant differences reported in mean scores according to age and marital status for the 25 and 10-item versions were lost in the 5-item one.

With respect to prevalence, it was higher in the HSCL-5 followed by the HSCL-10 and then the HSCL-25. All three versions and the gold-standard CIDI showed greater prevalence

in women than men. Employing the cutoffs corresponding to each version (22), sensitivity was similar for both genders whilst specificity was greater for men in the two short versions. Findings that concur with those reported for the HSCL-25 (18). In a study comparing the HSCL-10 with the CIDI (33), slightly higher results were reported with respect to sensitivity, although a different cutoff was employed. Another multicenter study conducted in General Practice in Norway and Denmark using the HSCL-10 and the CIDI obtained similar results to ours in terms of sensitivity and specificity (23).

Cutoffs can be based on previous studies, cutoffs used in clinical practice, cutoffs recommended by clinical practice guidelines, or cutoffs recommended by the original authors (57). There is very little literature on the appropriate cutoff point for the HSCL-10 and the HSCL-5, Strand et al. recommend 1.85 for the HSCL-10 and 2.00 for the HSCL-5 (22). By interpreting the results from ROC curves, the accuracy for various cutoffs was explored, and optimal cutoff values were obtained considering the maximum value of Youden Index. The optimum cutoff that we calculated for the study population is very close to that in the literature (22). Therefore, we considered that those of 1.85 and 2.00 for the HSCL-10 and HSCL-5, respectively, were appropriate to use showing adequate sensitivity, specificity, and AUC. The negative predictive values were >90% for the two short versions and both genders whilst the positive predictive values were low. Such a finding reinforces the need to complete the diagnosis of depression through a clinical interview.

Other authors have analyzed validity between the HSCL-25 and the two short versions with ROC curves (22). As we had the semi-structured CIDI interview for all our participants, we contrasted it against the two short versions with ROC curves. This is one of the strengths of the present study as sometimes there is a lack of gold-standard and the full version is used as a reference to assess validity. In other studies, only those participants who have a positive result on the scale and a small sample of those with a negative one, undergo the clinical interview of reference. The AUC was 0.88 for the HSCL-10 and 0.85 for the HSCL-5, both above the 0.75 cutoff considered to be of clinical utility, and greater than the 0.80 which confers a “good” classification (≥ 0.90 is considered “excellent”) in terms of discriminative properties of the diagnostic accuracy (58). A study carried out by Haavet et al. (33) also obtained an AUC of 0.88 when comparing the HSCL-10 with the CIDI, other studies with similar methodology have obtained a lower AUC (23). For both short versions the AUC was greater in men, thus the probability of accuracy in diagnosis in males is greater. Such a gender difference has also been reported for the longer HSCL-25 (18).

The main limitation of our study is that the psychometric properties of the two short versions were evaluated based on the responses to the 25-item scale as performed by other authors (22, 27). The study population came from the EIRA study (34, 35) and were patients aged 45–75 years presenting an unhealthy

behavior. Adults often engage in two or more unhealthy behaviors simultaneously, the co-occurrence of unhealthy diet with insufficient physical activity ranges between 47 and 54%, unhealthy diet with smoking between 23 and 28%, and insufficient physical activity with smoking between 8 and 20% (59). Unhealthy diet can be associated to depression (60). Despite the limited age range and the selection criteria in the present study, the authors consider that the results are transferable given the large number of participants, moreover, they had attended the PHC for any motive, consequently, the sample is sufficiently representative.

When deciding which questionnaire to use, there are a few relevant factors to be taken into account, including the population involved and the setting. A balance must be achieved between psychometric properties and such pragmatic characteristics as self-administration, number of items, simplicity/interpretability of scores, and accessibility (61). Scales with a reduced number of items may be the best option as they are practical and feasible within the PC setting without excessively losing reliability and validity.

These findings indicate that the HSCL-10 and HSCL-5 questionnaires show adequate reliability and validity in order to be employed in PC to detect and evaluate depressive symptoms. With such a short number of items they are timesaving and facilitate the detection of cases of depression that could otherwise go unnoticed.

Conclusion

The Spanish versions of the HSCL-10 and HSCL-5, especially the HSCL-10, are reliable and valid tools to detect depressive symptoms and can be used in Primary Care settings.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Clinical Research Ethics Committee of the University Institute Foundation for Research into Primary Health Care Jordi Gol from Barcelona (reference: P16/025). All the participants gave their written informed consent, and the trial was conducted in accordance with the ethical standards set out in the Declaration of Helsinki. Data were anonymously processed and were used only for the study objectives. Personal data confidentiality was ensured as set out in applicable regulations. Access to medical records was

gained as indicated in the legislation currently in force. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MR-B, MF-S-M, AC, JL, and PN: conceptualization. EM and IG-G: data collection. MR-B, MF-S-M, and AC: formal analysis. MR-B, MF-S-M, and EP-R: writing—original draft preparation. All authors have reviewed and agreed to the published version of the manuscript.

Funding

This study was supported by the Carlos III Health Institute, the Spanish Ministry of Economy and Competitiveness via a health research grant (PI15/00114, PI15/00565, PI15/00762, PI15/01072, PI15/00896, PI15/01412, PI15/01151, PI15/00519, and PI15/01133) through the Research Network in Preventive Activities and Health Promotion in Primary Care (redIAPP), (RD12/0005/0001, RD16/0007/0001, RD16/0007/0002, RD16/0007/0003, RD16/0007/0004, RD16/0007/0005, RD16/0007/0006, RD16/0007/0008, RD16/0007/0009, RD16/0007/0010, RD16/0007/0012, RD16/0007/0013, and RD16/0007/0015), the European Union ERDF funds, (European Regional Development Fund) and the Health Department (SLT002/16/00112) of the Generalitat de Catalunya. This study received a predoctoral grant from the University Institute of Research in Primary Care IDIAP Jordi Gol (Spain) and another grant from the Autonomous University of Barcelona (UAB)–Novartis for research in Family and Community Medicine (reference: 7Z19/019), both in 2019. The study obtained funding from the Spanish Society of Family Medicine (semFYC) as it was awarded a predoctoral grant, “Isabel Fernández.” The project also received a research grant from the Carlos III Institute of Health, Ministry of Science and Innovation (Spain) co-funded with European Union—NextGenerationEU funds, through the

References

- Nabbe P, Le Reste J, Guillou-Landreat M, Munoz Perez M, Argyriadou S, Claveria A, et al. Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review. *Eur Psychiatry*. (2017) 39:99–105. doi: 10.1016/j.eurpsy.2016.08.004
- Siu A, Bibbins-Domingo K, Grossman D, Baumann L, Davidson K, Ebell M, et al. Screening for depression in adults US preventive services task force recommendation statement. *JAMA J Am Med Assoc*. (2016) 315:380–7. doi: 10.1001/jama.2015.18392
- Smithson S, Pignone M. Screening adults for depression in primary care. *Med Clin North Am*. (2017) 101:807–21. doi: 10.1016/j.mcna.2017.03.010
- Kroenke K. Depression screening and management in primary care. *Fam Pract*. (2018) 35:1–3.
- Garrido-Elustondo S, Reneses B, Navalón A, Martín O, Ramos I, Fuentes M. Capacidad de detección de patología psiquiátrica por el médico de familia. *Aten Primaria*. (2016) 48:449–57. doi: 10.1016/j.aprim.2015.09.009
- Latorre Postigo J, Navarro Bravo B, Parra Delgado M, Salguero J, Mae Wood C, Cano Vindel A. Evaluación e intervención de los problemas de ansiedad y depresión en atención primaria: un problema sin resolver [Assessment and intervention of Anxiety and Depression disorders in Primary Care: an unsolved problem]. *Rev Clin Med Fam*. (2012) 5:37–45. doi: 10.4321/s1699-695x2012000100007

Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), with reference RD21/0022. This study received funding from Novartis. The funder was not involved in the study design, collection, analysis, interpretation of data, and the writing of this article or the decision to submit it for publication.

Acknowledgments

We thank to the Family Practice Depression and Multimorbidity group (FPDM) of the European General Practice Research Network (EGPRN). To the REDIAPP network, family physicians, other professionals, and the patients who participated in the study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.1014340/full#supplementary-material>

7. Tavabie J, Tavabie O. Improving care in depression: qualitative study investigating the effects of using a mental health questionnaire. *Qual Prim Care*. (2009) 17:251–61.
8. Mokkink L, Prinsen C, Patrick D, Alonso J, Bouter LM, de Vet H, et al. *COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments*. La Maddalena: COSMIN (2019). p. 1–32.
9. Derogatis L, Lipman R, Rickels K, Uhlenhuth E, Covi L. The Hopkins symptom checklist (HSCL): a self-report symptom inventory. *Behav Sci*. (1974) 19:1–15.
10. Glaesmer H, Braehler E, Grande G, Hinz A, Petermann F, Romppel M. The German version of the Hopkins symptoms checklist-25 (HSCL-25) - factorial structure, psychometric properties, and population-based norms. *Compr Psychiatry*. (2014) 55:396–403. doi: 10.1016/j.comppsy.2013.08.020
11. Nettelbladt P, Hansson L, Stefansson C, Borgquist L, Nordström G. Test characteristics of the Hopkins symptom check list-25 (HSCL-25) in Sweden, using the present state examination (PSE-9) as a caseness criterion. *Soc Psychiatr Psychiatr Epidemiol*. (1993) 28:130–3. doi: 10.1007/BF00801743
12. Lundin A, Hallgren M, Forsell Y. The validity of the symptom checklist depression and anxiety subscales: a general population study in Sweden. *J Affect Disord*. (2015) 183:247–52. doi: 10.1016/j.jad.2015.05.024
13. Nabbe P, Le Reste J, Guillou-Landreat M, Gatineau F, Le Floch B, Montier T, et al. The French version of the HSCL-25 has now been validated for use in primary care. *PLoS One*. (2019) 14:e0214804. doi: 10.1371/journal.pone.0214804
14. Kamacooko O, Bagiire D, Kasujja F, Mirembe M, Seeley J, King R. Prevalence of probable depression and factors associated with mean Hopkins symptom checklist (HSCL) depression score among young women at high risk aged 15–24 years in Kampala, Uganda. *PLoS One*. (2022) 17:e0270544. doi: 10.1371/JOURNAL.PONE.0270544
15. Nabbe P, Le Reste J, Guillou-Landreat M, Beck-Robert E, Assenova R, Lazic D, et al. One consensual depression diagnosis tool to serve many countries: a challenge! A RAND/UCLA methodology. *BMC Res Notes*. (2018) 11:4. doi: 10.1186/s13104-017-3111-x
16. Roncone R, Gone J, Barlati S, Nabbe P, Reste L, Lazic K, et al. Nine forward-backward translations of the Hopkins symptom checklist-25 with cultural checks. *Front Psychiatry*. (2021) 1:688154. doi: 10.3389/fpsy.2021.688154
17. Clavería A, Rodríguez-Barragán M, Fernández-San-Martín M, Nabbe P, Le Reste J, Miguéns-Blanco I, et al. Translation and cross-cultural adaptation into Spanish, Catalan and Galician of the Hopkins symptom checklist-25 to identify depression in primary care. *Aten Primaria*. (2020) 52:539–47. doi: 10.1016/j.aprim.2020.05.017
18. Rodríguez-Barragán M, Fernández-San-Martín M, Clavería-Fontán A, Aldecoa-Landesa S, Casajuana-Closas M, Llobera J, et al. Validation and psychometric properties of the Spanish version of the Hopkins symptom checklist-25 scale for depression detection in primary care. *Int J Environ Res Public Health*. (2021) 18:7843. doi: 10.3390/ijerph18157843
19. Müller J, Postert C, Beyer T, Furniss T, Achtergarde S. Comparison of eleven short versions of the symptom checklist 90-revised (SCL-90-R) for use in the assessment of general psychopathology. *J Psychopathol Behav Assess*. (2009) 32:246–54. doi: 10.1007/s10862-009-9141-5
20. Tambs K, Røysamb E. Selection of questions to short-form versions of original psychometric instruments in MoBa*. *Nor Epidemiol*. (2014) 24:195–201.
21. Tambs K, Moum T. How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatr Scand*. (1993) 87:364–7.
22. Strand B, Dalgard O, Tambs K, Rognerud M. Measuring the mental health status of the Norwegian population: a comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nord J Psychiatry*. (2003) 57:113–8. doi: 10.1080/08039480310000932
23. Sirpal M, Haugen W, Sparle K, Haavet O. Validation study of HSCL-10, HSCL-6, WHO-5 and 3-key questions in 14-16 year ethnic minority adolescents. *BMC Fam Pract*. (2016) 17:7. doi: 10.1186/s12875-016-0405-3
24. Søberg Finbråten H, Kleppang A, Steigen A. Using the Rasch measurement theory to assess the psychometric properties of the Hopkins symptom checklist-10 in adolescents. *Health Qual Life Outcomes*. (2021) 19:248. doi: 10.1186/s12955-021-01884-9
25. Kleppang A, Hagquist C. The psychometric properties of the Hopkins symptom checklist-10: a Rasch analysis based on adolescent data from Norway. *Fam Pract*. (2016) 33:740–5. doi: 10.1093/FAMPRA/CMW091
26. Ljoså C, Tyssen R, Lau B. Mental distress among shift workers in Norwegian offshore petroleum industry - relative influence of individual and psychosocial work factors. *Scand J Work Environ Heal*. (2011) 37:551–5. doi: 10.5271/sjweh.3191
27. Schmalbach B, Zenger M, Tibubos A, Kliem S, Petrowski K, Brähler E. Psychometric properties of two brief versions of the Hopkins symptom checklist: HSCL-5 and HSCL-10. *Assessment*. (2019) 28:617–31. doi: 10.1177/1073191119860910
28. Lien I, Bolstad I, Lien L, Bramness J. Screening for depression in patients in treatment for alcohol use disorder using the beck depression inventory-II and the Hopkins symptom checklist-10. *Psychiatry Res*. (2022) 308:114363. doi: 10.1016/J.PSYCHRES.2021.114363
29. Nesterko Y, Jäckle D, Friedrich M, Holzapfel L, Glaesmer H. Factors predicting symptoms of somatization, depression, anxiety, post-traumatic stress disorder, self-rated mental and physical health among recently arrived refugees in Germany. *Confl Health*. (2020) 14:44. doi: 10.1186/s13031-020-00291-z
30. Harris S, Sandal G. COVID-19 and psychological distress in Norway: the role of trust in the healthcare system. *Scand J Public Health*. (2021) 49:96–103. doi: 10.1177/1403494820971512
31. Karlsson Stafseth S, Skogstad L, Raeder J, Strand Hovland I, Hovde H, Ekeberg Ø, et al. Symptoms of anxiety, depression, and post-traumatic stress disorder in health care personnel in Norwegian ICUs during the first wave of the COVID-19 pandemic, a prospective, observational cross-sectional study. *Int J Environ Res Public Heal*. (2022) 19:7010. doi: 10.3390/ijerph19127010
32. Frounfelker R, Li Z, Santavica T, Miconi D, Rousseau C. Latent class analysis of COVID-19 experiences, social distancing, and mental health. *Am J Orthopsychiatry*. (2022) 92:121–32. doi: 10.1037/ORT0000593
33. Haavet O, Sirpal M, Haugen W, Christensen K. Diagnosis of depressed young people in primary health care—a validation of HSCL-10. *Fam Pract*. (2011) 28:233–7. doi: 10.1093/fampra/cm078
34. Zabaleta-Del-Olmo E, Pombo H, Pons-Vigués M, Casajuana-Closas M, Pujol-Ribera E, López-Jiménez T, et al. Complex multiple risk intervention to promote healthy behaviours in people between 45 to 75 years attended in primary health care (EIRA study): study protocol for a hybrid trial. *BMC Public Health*. (2018) 18:874. doi: 10.1186/s12889-018-5805-y
35. Represas-Carrera F, Couso-Viana S, Méndez-López F, Masluk B, Magallón-Botaya R, Recio-Rodríguez J, et al. Effectiveness of a multicomponent intervention in primary care that addresses patients with diabetes mellitus with two or more unhealthy habits, such as diet, physical activity or smoking: multicenter randomized cluster trial (EIRA Study). *Int J Environ Res Public Heal Artic Public Heal*. (2021) 18:5788. doi: 10.3390/ijerph18115788
36. Veijola J, Jokelainen J, Läksy K, Kantojärvi L, Kokkonen P, Järvelin M, et al. The Hopkins symptom checklist-25 in screening DSM-III-R axis-I disorders. *Nord J Psychiatry*. (2003) 57:119–23. doi: 10.1080/08039480310000941
37. Kessler R, Bedirhan Üstün T. The World mental health (WMH) survey initiative version of the World Health Organization (WHO) composite international diagnostic interview (CIDI). *Int J Methods Psychiatr Res*. (2004) 13:93–121.
38. Harkness J, Pennell B, Villar A, Gebler N, Aguilar-Gaxiola S, Bilgen I. Translation procedures and translation assessment in the World Mental Health survey initiative. In: Kessler RC, Bedirhan T editors. *The WHO World Mental Health Surveys: Global Perspectives on the Epidemiology of Mental Disorders*. New York, NY: Cambridge University Press, World Health Organization (2008). p. 91–113.
39. Terwee C, Bot S, de Boer M, van der Windt D, Knol D, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. (2007) 60:34–42. doi: 10.1016/j.jclinepi.2006.03.012
40. Ferrari A. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990-2019: a systematic analysis for the Global burden of disease study 2019. *Lancet Psychiatry*. (2022) 9:137–50. doi: 10.1016/S2215-0366(21)00395-3
41. Portellano-Ortiz C, Garre-Olmo J, Calvo-Perxas L, Conde-Sala J. Depresión y variables asociadas en personas mayores de 50 años en España. *Rev Psiquiatr y Salud Ment*. (2018) 11:216–26. doi: 10.1016/j.rpsm.2016.10.003
42. Huang E, Cheung J, Liu J, Kwan R, Lam S. Groningen frailty indicator-Chinese (GFI-C) for pre-frailty and frailty assessment among older people living in communities: psychometric properties and diagnostic accuracy. *BMC Geriatr*. (2022) 22:788. doi: 10.1186/s12877-022-03437-1
43. Jackson D, Gillaspay J, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychol Methods*. (2009) 14:6–23. doi: 10.1037/a0014694
44. de Vet H, Terwee C, Mokkink L, Knol D. *Measurement in Medicine. Practical Guides to Biostatistics and Epidemiology*. Cambridge: Cambridge University Press, World Health Organization (2018). 338 p.

45. Viladrich C, Doval E, Penelo E. *Fiabilidad y Validez*. 12th ed. Bellaterra: Universitat Autònoma de Barcelona Laboratori d'Estadística Aplicada i de Modelització (2019).
46. Brehaut, E, Neupane D, Levis B, Wu Y, Sun Y, Ioannidis J, et al. Optimal[™] cutoff selection in studies of depression screening tool accuracy using the PHQ-9, EPDS, or HADS-D: a meta-research study. *Int J Methods Psychiatr Res.* (2022):e1956. doi: 10.1002/MPR.1956
47. Søgaard A, Bjelland I, Tell G, Røysamb E. A comparison of the CONOR mental health index to the HSCL-10 and HADS. *Nor Epidemiol.* (2003) 13:279–84.
48. Violán Fors C, Odriozola G, Zabaleta-del-Olmo E, Moral E. La investigación en atención primaria como área de conocimiento. Informe SESPAS 2012. *Gac Sanit.* (2012) 26:76–81. doi: 10.1016/j.gaceta.2011.05.018
49. Reuland D, Cherrington A, Watkins G, Bradford D, Blanco R, Gaynes B. Diagnostic accuracy of Spanish language depression-screening instruments. *Ann Fam Med.* (2009) 7:455–62. doi: 10.1370/afm.981
50. Sanz J, Garcia-Vera M, Espinosa R, Fortún M. Adaptación española del inventario para la depresión de beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. *Clín Salud.* (2005) 16:121–42.
51. Fernández-San Martín M, Andrade C, Molina J, Muñoz P, Carretero B, Rodríguez M, et al. Validation of the Spanish version of the geriatric depression scale (GDS) in primary care. *Int J Geriatr Psychiatry.* (2002) 17:279–87. doi: 10.1002/gps.588
52. Kakemam E, Navvabi E, Albelbeisi A, Saedikia F, Rouhi A, Majidi S. Psychometric properties of the Persian version of depression anxiety stress scale-21 items (DASS-21) in a sample of health professionals: a cross-sectional study. *BMC Health Serv Res.* (2021) 22:111. doi: 10.1186/s12913-022-07514-4
53. Kroenke K, Spitzer R, Williams J, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics.* (2009) 50:613–21. doi: 10.1016/S0033-3182(09)70864-3
54. Petrowski K, Zenger M, Schmalbach B, Bastianon C, Strauss B. Psychometric properties and validation of the English version Giessen subjective complaints list (GBB-8). *BMC Psychol.* (2022) 10:60. doi: 10.1186/s40359-022-00741-8
55. Nunnally J. *Psychometric Theory*. 2nd ed. New York, NY: McGraw-Hill (1978).
56. Martinussen L, Šaltytė Benth J, Almdahl I, Borza T, Selbæk G, Mcpherson B, et al. The effect of anxiety on cognition in older adult inpatients with depression: results from a multicenter observational study. *Heliyon.* (2019) 5:e02235. doi: 10.1016/j.heliyon.2019.e02235
57. Cohen J, Korevaar D, Altman D, Bruns D, Gatsonis C, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* (2016) 6:e012799. doi: 10.1136/bmjopen-2016-02016
58. Polo T, Miot H. Aplicações da curva ROC em estudos clínicos e experimentais. *J Vasc Bras.* (2020) 19:e20200186. doi: 10.1590/1677-5449.200186
59. Meader N, King K, Moe-Byrne T, Wright K, Graham H, Petticrew M, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. *BMC Public Health.* (2016) 16:657. doi: 10.1186/s12889-016-3373-6
60. Oliván-Blázquez B, Aguilar-Latorre A, Motrico E, Gómez-Gómez I, Sarasa-Bosque C, Sanchez-Perez A, et al. The Relationship between adherence to the mediterranean diet, intake of specific foods and depression in an adult population (45–75 Years) in primary health care. A cross-sectional descriptive study. *Nutrients.* (2021) 13:2724. doi: 10.3390/nu13082724
61. Kroenke K, Monahan P, Kean J. Pragmatic characteristics of patient-reported outcome measures are important for use in clinical practice. *J Clin Epidemiol.* (2016) 68:1085–92. doi: 10.1016/j.jclinepi.2015.03.023

Material suplementario artículo 3:

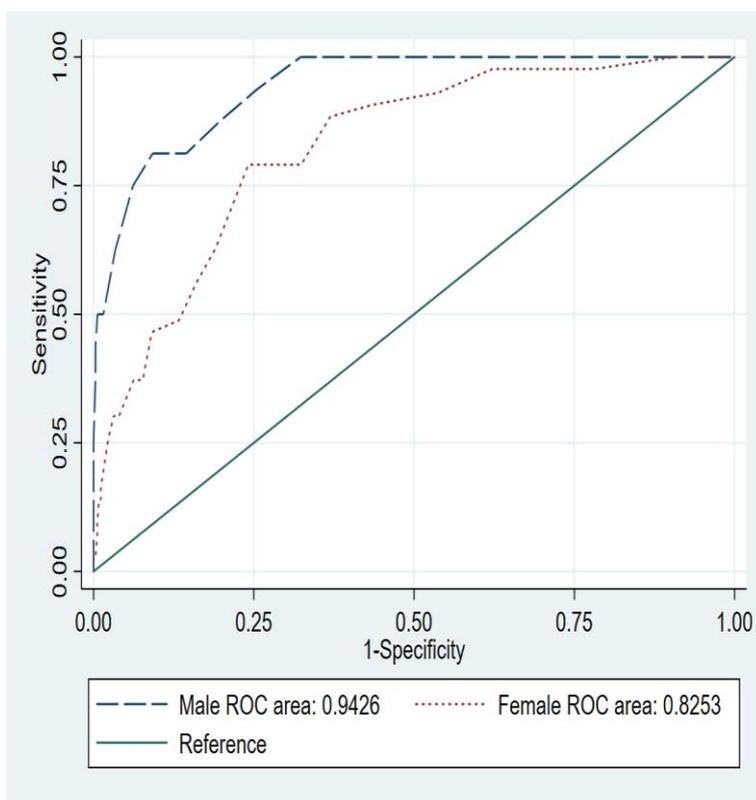


Figure 1. ROC curve and AUC HSCL-10 vs. CIDI by gender

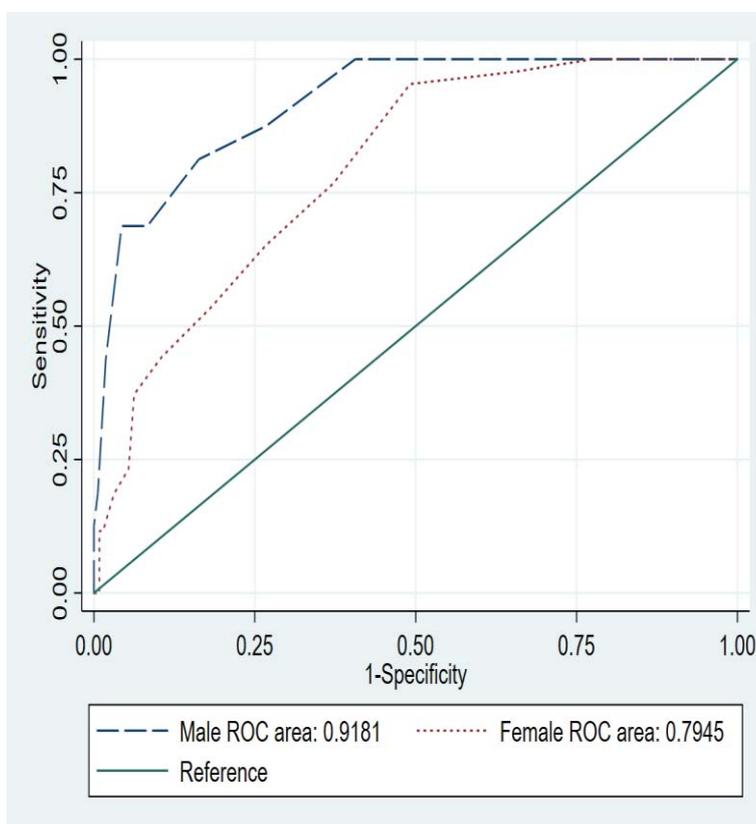


Figure 2. ROC curve and AUC HSCL-5 vs. CIDI by gender

Table S1. Goodness-of-fit indices in studied factor models

Model		χ^2 (df)	p	CFI	TLI	RMSEA (90% CI)	SRMR
One factor	HSCL-10	291 (35)	<0.001	0.89	0.86	0.098 (0.087 - 0.108)	0.056
	HSCL-5	53 (5)	<0.001	0.95	0.90	0.112 (0.086 - 0.140)	0.041
Two correlated factors	HSCL-10	269 (34)	<0.001	0.90	0.87	0.095 (0.084 - 0.206)	0.054
	HSCL-5	42 (4)	<0.001	0.96	0.90	0.112 (0.083 - 0.144)	0.039

χ^2 : chi-square; df: degrees of freedom, CFI: Comparative Fit Index; TLI: Tucker-Lewis Fit Index; RMSEA: Root Mean Square Error of Approximation; SRMR: Standardized Root Mean Square Residual

Table S2. Sensitivity, specificity, positive and negative predictive values, total and by gender in the HSCL-10 and the HSCL-5 using optimal cutoff points calculated by the Youden Index.

		Male (n=341)			Female (n=395)			Total (n=736)		
		Index	CI95% lower limit	CI95% upper limit	Index	CI95% lower limit	CI95% upper limit	Index	CI95% lower limit	CI95% upper limit
Optimal cutoff point		1.90			1.90			1.90		
HSCL-10	Sensitivity	81.3	57.0	93.4	79.1	64.8	88.6	79.7	67.7	88.0
	Specificity	90.8	87.1	93.5	75.9	71.1	80.0	83.0	80.0	85.7
	PPV	30.2	18.6	45.1	28.6	21.2	37.3	29.0	22.6	36.4
	NPV	99.0	97.1	99.7	96.7	93.9	98.3	97.9	96.4	98.8
Optimal cutoff point		2.00			1.80			1.80		
HSCL-5	Sensitivity	81.3	57.0	93.4	95.3	84.5	98.7	93.2	83.8	97.3
	Specificity	83.7	79.3	87.3	50.9	45.6	56.0	61.6	57.9	65.2
	PPV	19.7	11.9	30.8	19.2	14.4	25.0	17.5	13.7	22.0
	NPV	98.9	96.8	99.6	98.9	96.1	99.7	99.0	97.6	99.6

PPV: positive predictive value; NPV: negative predictive value.

7

DISCUSIÓN

7 Discusión

7.1 Resumen de los resultados principales

Los resultados de la tesis doctoral se han presentado en los artículos publicados en revistas científicas indexadas, adjuntados en el apartado anterior. A continuación, se realiza un resumen de los resultados más destacados de cada artículo.

7.1.1 Artículo 1

En este artículo se describe el proceso de traducción y adaptación transcultural de la HSCL-25 del inglés al español, catalán y gallego.

- La traducción directa fue evaluada por MF con metodología Delphi en la que participaron 31 MF para la traducción española, 22 para la catalana y 20 para la gallega. Los panelistas participantes en la traducción al español tenían una media de 23 años de experiencia como MF (rango intercuartílico [RIC] = 19 - 26), 10 años como tutores de MFyC (RIC = 5 - 20) y 9 como investigadores (RIC = 1 - 15).
- La traducción al español y al catalán alcanzó el consenso para todos los ítems en la primera ronda Delphi, mientras que la traducción al gallego requirió una segunda ronda.
- La retrotraducción de todos los idiomas fue similar a la versión original. Todas las opciones fueron semánticamente equivalentes.
- Se realizó una armonización europea de las versiones finales de cada idioma por consenso, llevada a cabo por miembros del grupo investigador del FPDM del EGPRN.
- 10 pacientes de AP participaron en la entrevista cognitiva para cada idioma, sin que se modificara la redacción de ningún ítem. Se demostró la comprensibilidad de la HSCL-25 en español por parte de los usuarios finales.
- Se flexionaron los sustantivos y adjetivos que admiten morfemas de género.
- Tras este proceso se obtuvo una versión traducida y adaptada culturalmente de la HSCL-25 en español, en catalán y en gallego.

7.1.2 Artículo 2

En este artículo se analizan la validez y la fiabilidad de las puntuaciones de la HSCL-25 en español en población de AP. Se utiliza la entrevista clínica estandarizada CIDI como *gold standard* para el diagnóstico de la depresión.

- La tasa de respuesta de la HSCL-25 fue de 97,2%. Respondieron 738 pacientes a la totalidad de los ítems. No se observaron patrones de no respuesta específicos.
- La puntuación media fue 1,57. Las mujeres puntuaron más alto en todos los ítems y también en la puntuación total, siendo esta diferencia estadísticamente significativa ($p < 0,05$) en 23 de los 25 ítems y en la puntuación total.
- Se confirmó la estructura interna de un solo factor y de dos factores correlacionados, siendo esta última la que mostró mejor ajuste. Se obtuvo una correlación elevada de 0,84 en el modelo de dos factores (ansiedad y depresión) correlacionados. Todas las cargas factoriales fueron $\geq 0,3$ y estadísticamente significativas ($p < 0,001$) en el constructo esperado.
- La puntuación total de la HSCL-25 se correlacionó de forma elevada y positiva con la escala de depresión PHQ9, con una correlación de 0,78.
- La validez de criterio fue buena al comparar la puntuación total de la HSCL-25 con el resultado de la CIDI. Se obtuvo un AUC de 0,89, siendo mayor en los hombres que en las mujeres. La sensibilidad fue de 88,1% y la especificidad fue de 76,7%. El VPP fue de 24,8% y el VPN fue de 98,7%.
- El punto de corte óptimo, obtenido mediante el cálculo del índice de Youden, fue de 1,76 de forma global y para las mujeres, mientras que fue más elevado para los hombres (1,84). El punto de corte global fue muy cercano al de 1,75 que proponen los autores originales.
- La fiabilidad fue muy elevada. La HSCL-25 mostró un valor alfa de Cronbach de 0,92 (0,88 para la dimensión de depresión y 0,82 para la dimensión de ansiedad). El análisis test-retest mostró que las puntuaciones se mantuvieron estables en el tiempo, con un CCI de 0,92.
- El ítem 17 “Se siente triste” fue el más consistente en todos los análisis. Por el contrario, el menos consistente fue el 24 “Poco apetito”, que restó fiabilidad al global de la escala.
- La HSCL-25 en español muestra adecuada validez y fiabilidad en los pacientes de AP de España.

7.1.3 Artículo 3

En este artículo se analizan la validez y la fiabilidad de la HSCL-10 y HSCL-5 a partir de las respuestas obtenidas en la HSCL-25. Igual que en el segundo artículo, se utiliza la entrevista clínica estandarizada CIDI como *gold standard* para el diagnóstico de depresión.

- La puntuación media de la HSCL-10 fue 1,55 y la de la HSCL-5 fue 1,72. La HSCL-10 mostró una puntuación media más cercana a la de 1,57 obtenida con la HSCL-25. Se mantuvo en ambas versiones breves la significación estadística ($p < 0,001$) de la puntuación mayor observada en las mujeres en la HSCL-25 respecto a los hombres.
- El método de la comparación de la puntuación en subgrupos como comprobación de hipótesis para el análisis de la validez de constructo mostró una diferencia de puntuación estadísticamente significativa ($p < 0,001$) por sexos, siendo mayor en las mujeres tanto en la HSCL-10 como en la HSCL-5.
- La validez estructural analizada mediante el CFA mostró índices de ajuste aceptables, pero no óptimos, mostrando inestabilidad en la estructura interna de la HSCL-10 y la HSCL-5. El modelo de dos factores correlacionados se ajustó mejor en ambas versiones. Todas las cargas factoriales fueron $\geq 0,45$ y estadísticamente significativas ($p < 0,001$) en el constructo esperado.
- La validez de criterio fue adecuada, obteniendo un AUC de 0,88 en la HSCL-10 y de 0,85 en la HSCL-5 al compararlas con la entrevista CIDI.
- La precisión diagnóstica fue adecuada al comparar la puntuación total con el resultado obtenido de la escala CIDI, especialmente para HSCL-10. Se obtuvo una sensibilidad de 79,9% y especificidad de 83,0% para la HSCL-10, y una sensibilidad de 78,0% y una especificidad de 72,8% para la HSCL-5. Se obtuvo un VVP de 29,0% en la HSCL-10 y de 20,0% en la HSCL-5, y un VPN de 97,9% en la HSCL-10 y de 97,4% en la HSCL-5.
- Los puntos de corte óptimos, obtenidos mediante el cálculo del índice de Youden, fueron 1,90 para la HSCL-10 y 1,80 para la HSCL-5, cercanos a los propuestos en la literatura disponible.
- A menor número de ítems en la escala HSCL, menor fue la fiabilidad, aunque fue adecuada tanto para HSCL-10 como para HSCL-5, con un valor alfa de Cronbach de 0,84 y 0,77, respectivamente.
- La HSCL-10 y HSCL-5 presentan adecuada validez y fiabilidad para el cribado de la depresión en las consultas de AP, siendo una buena alternativa a escalas más largas como la HSCL-25.

7.2 Discusión por objetivos y comparación con otros artículos

Se ha redactado una discusión de los resultados por objetivos específicos, para dar respuesta al objetivo principal que es “Evaluar la validez y fiabilidad de la HSCL-25 en español para el cribado de la depresión en pacientes de Atención Primaria en España”.

Objetivo 1. Describir el proceso de traducción y adaptación transcultural de la HSCL-25 para su uso en Atención Primaria en España.

Los síntomas de la esfera mental se describen y se interpretan de manera diferente en cada cultura²⁴⁵ porque la expresión emocional está intrínsecamente ligada al entorno lingüístico y cultural de los pacientes. Esto hace necesario utilizar términos o preguntas sustancialmente diferentes en distintos países para evaluar los mismos síntomas²⁴⁶, lo que lleva a la traducción y adaptación de cuestionarios y otras herramientas para poder ofrecerlos a los pacientes en su lengua materna.

Las versiones en idiomas diferentes o las adaptaciones culturales no son una mera traducción realizada por el equipo investigador, sino que requieren de un proceso riguroso para asegurar la equivalencia conceptual (en contraposición a la equivalencia literal que sería la traducción palabra por palabra)²⁴⁶. Es un paso necesario previo al análisis de las propiedades psicométricas en el nuevo contexto. Varias guías y documentos, entre los que destacan la COSMIN^{181,192} y la ISPOR²⁰¹, proponen una serie de pasos que incluyen siempre un mínimo de dos traducciones independientes y una retrotraducción.

La traducción y adaptación de la HSCL-25 al español se realizó simultáneamente con el resto de los idiomas de los miembros del equipo investigador europeo. Este proceso fue liderado por los investigadores franceses, quienes velaron por el seguimiento del protocolo y la homogeneidad en cada uno de los pasos²²⁹. En el primer artículo del compendio de esta tesis doctoral se describe el proceso que se siguió.

El primer paso fue la traducción realizada por una traductora titulada en Traducción e Interpretación que tenía como lengua materna el español. Este paso fue supervisado por un MF del equipo investigador. En lugar de realizar dos traducciones directas tal y como proponen las guías, en el protocolo conjunto del equipo europeo se optó por realizar una traducción seguida de un procedimiento Delphi en el que participaron MF de los países participantes analizando la traducción obtenida.

Se escogió este método sistemático por ser un modo de alcanzar consenso de forma rigurosa, ya que permite²²⁹:

- El anonimato de los participantes para garantizar la fiabilidad de las respuestas y evitar la contaminación.
- La iteración, que da la posibilidad a los participantes de afinar sus opiniones a la luz del progreso del trabajo en grupo.
- La agregación estadística de las respuestas del grupo para permitir un análisis cuantitativo y cualitativo de los datos.

La selección de los panelistas de un método Delphi se presenta como una de las principales limitaciones en los métodos de consenso debido a la posibilidad de sesgo en la elección de los participantes²⁰⁸. Se pretendió minimizar este efecto con los criterios de inclusión decididos por el grupo FPDM. El grupo europeo sugirió incluir MF que fueran profesores de universidad, puesto que se entiende que estos cuentan con la experiencia necesaria para el diagnóstico de la depresión y dominan la lengua inglesa. En España, considerando el escaso número de MF en las Facultades de Medicina, se optó por seleccionar tutores de MFyC o clínicos con gran experiencia y publicaciones en revistas con impacto, asegurando así su formación continua y actualización en los campos de la investigación y de la docencia, pretendiendo que se aproximasen en la mayor medida al referente establecido.

La metodología Delphi se utiliza con fines muy diferentes en la investigación científica, especialmente para lograr una mayor comprensión de una realidad a partir de diferentes perspectivas y para llegar a acuerdos en temas de interés de los que no se dispone de información concluyente²⁴⁷. En cuanto al uso de este método en el ámbito de las escalas de medida, suele utilizarse para la selección de las variables adecuadas para medir un constructo²⁴⁸, para la selección de los ítems en el desarrollo de un nuevo instrumento^{205,249} y para la evaluación de la validez de contenido (relevancia y representatividad de los ítems)²⁵⁰. Es infrecuente el uso del método Delphi en las fases de traducción y adaptación de escalas porque se trata de una técnica laboriosa que alarga el proceso, pero se utiliza algunas veces como método de consenso formal tras la traducción directa¹⁹³ así como en otros momentos durante el proceso de adaptación transcultural²⁵¹.

Asegurar la comprensibilidad de los PROMs por los usuarios finales tiene una importancia capital. Si los individuos de la población diana presentan dificultades para responder una escala autoadministrada, todos los esfuerzos puestos en su desarrollo habrán sido en vano. Un método sencillo y sin coste para comprobar la comprensibilidad de los ítems antes de probar el PROM

con la población final puede ser solicitar a compañeros de profesión que respondan a los ítems. Este método puede revelar algunos problemas que se pueden resolver en una fase previa²⁵². Los MF son conocedores del lenguaje de sus pacientes, quienes utilizan muchas veces términos y expresiones dialectales para referirse a sus problemas de salud, sobre todo en comunidades mayoritariamente bilingües. Por este motivo, se optó por implicar a los MF en un paso intermedio de la traducción para participar con metodología Delphi.

La definición de consenso que se utiliza en los estudios con metodología Delphi publicados es discrepante y no siempre se detalla²⁵³. La más comúnmente utilizada es el porcentaje de acuerdo basado en un punto de corte predefinido o una medida de tendencia central, que se seleccionan por el equipo investigador en función de los objetivos del estudio²⁰⁴. En el presente estudio se definió el consenso a priori como el 70% de respuestas mayores de 7 puntos en una escala numérica de 1 a 9. Se escogió un criterio exigente con el fin de obtener traducciones adecuadas y adaptadas al contexto de AP.

La versión en español obtenida de la traducción directa, que se puede consultar en el ANEXO 5, consiguió consenso en la primera ronda Delphi tanto para la traducción de los ítems, como para la traducción de las cuatro categorías de respuesta y las normas escritas para los pacientes. También se obtuvo consenso en la primera ronda en la traducción al catalán, búlgaro y polaco²²⁹. La traducción alemana¹³⁷, ya disponible previamente, se analizó también mediante el método Delphi, con consenso en la primera ronda. Las traducciones al gallego, francés, italiano, croata y griego requirieron una segunda propuesta y otra ronda Delphi para alcanzar consenso²²⁹.

Tras la versión traducida consensuada por los MF panelistas, se realizó la retrotraducción al inglés por parte de una filóloga cuya lengua materna era el inglés y que no conocía la versión original ni había participado en el proceso previo, tal y como indican los estándares¹⁸¹. La retrotraducción se comparó con la versión original sin encontrar diferencias conceptuales. El ítem 5 "*Heart Racing*" de la versión original, traducido como "Palpitaciones" y retrotraducido como "*Palpitations*", fue el que más divergió del original. La retrotraducción obtenida se puede consultar en el mismo ANEXO 5. No se realizaron modificaciones tras este paso. Las retrotraducciones del catalán y el gallego fueron similares a la obtenida en español, lo cual no sorprende debido a la proximidad lingüística de las tres lenguas, todas ellas lenguas romances.

Hay que tener en cuenta que la versión original está escrita en inglés. El inglés es un idioma que tiene un léxico muy amplio. Esto puede conllevar que las distinciones que se hacen en inglés no puedan igualarse en una o más lenguas de destino. También puede ocurrir lo contrario con respecto a ciertos ámbitos de distinciones léxicas o gramaticales, en los que la lengua de origen

puede no especificar suficientes detalles necesarios para la traducción a una lengua de destino determinada²⁴⁶. Por este motivo se realizó una clasificación de equivalencias con las traducciones obtenidas del español, catalán y gallego. Este paso no forma parte de los procesos habituales de traducción, pero se consideró oportuno añadirlo para analizar y comparar las tres versiones. El hecho de hacerlo de forma simultánea permitió realizar este análisis para resolver posibles diferencias en los resultados de las tres lenguas. Este proceso lo realizó una tercera filóloga que no había participado en las fases previas. Todas las opciones en las tres lenguas fueron semánticamente equivalentes; aunque en una lengua se utilice una expresión adjetival (“*inquedo*” en el ítem 10 del gallego) y en otra un sustantivo (“*inquietud*” en español).

Tras este paso, se realizó un proceso conjunto con el resto del equipo europeo en el que se analizaron todas las versiones traducidas con el fin de comprobar su equivalencia. En este paso de armonización transcultural se modificó la traducción directa de “*distress*”, este término se tradujo inicialmente como “*angustia*”, pero en la versión final se optó por “*malestar psicológico*”. Este término, sin embargo, no forma parte del cuestionario, sino de las aclaraciones necesarias para su interpretación.

Una vez obtenida la traducción, se realizó una entrevista cognitiva con un grupo de pacientes de AP para comprobar la comprensibilidad y aceptabilidad en la población final antes de realizar un estudio de mayor envergadura para analizar sus propiedades psicométricas. Las guías recomiendan involucrar a participantes de la población diana en el estudio piloto, ya que solamente ellos/as pueden juzgar la comprensibilidad y relevancia de los ítems¹⁸¹. La entrevista cognitiva debería ser un estándar en el proceso de desarrollo de cualquier instrumento de medida²⁵², la ISPOR recomienda realizarla en 5-8 pacientes de la población diana²⁰¹. En el presente estudio se realizó la entrevista cognitiva con 10 pacientes de AP. No se detectaron dificultades en la comprensión de los ítems que justificaran modificar la redacción de ninguno de ellos. Este paso demostró que los futuros usuarios de la escala comprenden su uso, la redacción de los ítems y que no muestran incomodidad frente a ninguno de ellos. Este es un aspecto fundamental dado que es un cuestionario autoadministrado.

Finalmente se flexionaron los sustantivos y adjetivos de los 4 ítems que admiten morfemas de género, dada la importancia creciente que ha adquirido este aspecto en nuestros tiempos. Así, por ejemplo, el ítem 15 en la versión final en español es “*Se siente solo/a*” en lugar de “*Se siente solo*”. Este aspecto no tiene relevancia en otros idiomas en los que el género no modifica la redacción como el inglés, pero sí que es relevante en español. Algunos cuestionarios disponibles en español están redactados en masculino como el PHQ⁹⁷, mientras que otros como el HARS y

la MADRS están redactados de forma flexionada por género¹¹¹. El equipo investigador tuvo este aspecto en consideración en el último paso de la traducción.

A la hora de utilizar escalas en contextos diferentes al de la versión original, muchos estudios realizan análisis de la validez y la fiabilidad minuciosos en los nuevos contextos, pero no todos detallan el proceso seguido para obtener la versión que se administra a la nueva población diana. Los estudios que publican este proceso lo hacen con un proceso similar al llevado a cabo en el presente estudio²⁵⁴⁻²⁵⁹. Todos ellos incluyen la entrevista cognitiva con personas de la población diana y la mayoría de ellos incluyen revisiones de las versiones por expertos o miembros del equipo investigador, aunque no de forma estructurada como se hace en el presente estudio a través de la técnica Delphi.

Cabe aclarar que en el contexto de este estudio se hace referencia al idioma de España como “español” y este término puede llevar a confusión dado que también se habla el español en gran parte de América, Filipinas, Guinea Ecuatorial y otros lugares del mundo²⁶⁰. El término “castellano” como la lengua que se habla en España, se utiliza especialmente cuando se quiere distinguir de alguna otra lengua vernácula de España²⁶⁰. El término “castellano”, traducido al inglés como “*Castilian*” no se conoce fuera del ámbito de España. Dado que este proyecto es de carácter internacional y además se han publicado sus resultados en revistas internacionales en inglés, se ha utilizado el término “español”, tal y como hacen la mayoría de los estudios que han traducido otras herramientas a nuestra lengua^{104,121,261}.

Durante el diseño del proyecto se consideró realizar la traducción también a los idiomas cooficiales catalán y gallego, dado que en el equipo de investigación participaron investigadores de Cataluña y Galicia liderados por dos de las directoras de esta tesis doctoral, la Dra. Fernández y la Dra. Clavería, respectivamente. A pesar de que se ofreció colaboración, no se consiguieron integrantes de habla euskera y por este motivo no se realizó la traducción a este idioma, también cooficial en nuestro país.

Objetivo 2. Analizar el proceso de respuesta de los pacientes de Atención Primaria a la HSCL-25 en español.

El análisis del proceso de respuesta consiste en evaluar de qué manera se relacionan los ítems entre sí y con la puntuación que representa el constructo evaluado. Las versiones adaptadas deben proporcionar información de las características psicométricas de los ítems para facilitar el análisis de equivalencia estadística respecto a la versión original¹⁸².

Se obtuvo un porcentaje de respuesta de la HSCL-25 de 97,2% que se consideró muy elevado. El hecho de incluir la HSCL-25 en el cuaderno de recogida de datos del proyecto EIRA²³⁰ facilitó este buen resultado.

Antes de la imputación de los valores perdidos, conviene realizar un análisis descriptivo de los ítems para comprobar si las respuestas siguen el patrón esperado. Se analizaron las respuestas en blanco en búsqueda de patrones de no respuesta para comprobar que la ausencia de respuestas no siguió ningún patrón sistemático. Las respuestas en blanco pueden revelar problemas en los ítems como dificultad de comprensión, incomodidad, falta de identificación con esa cuestión concreta, etc¹⁸¹. 737 de los 768 pacientes que respondieron la HSCL-25 lo hicieron sin dejar ningún ítem en blanco. Las repeticiones en las no respuestas tuvieron un máximo de 3 individuos (0,38% de los participantes) que dejaron en blanco el ítem 7 “Se siente tenso/a”. No se encontraron patrones de no respuesta específicos tal y como se observa en la Tabla S1 del Material Suplementario del segundo artículo de este compendio.

Teniendo en cuenta los 22 pacientes que no respondieron a ningún ítem, las respuestas en blanco no superaron el 3,5% de los pacientes en ninguno de los ítems. Se consideró un porcentaje de valores perdidos aceptable, similar al obtenido por otros autores¹⁵⁴ y lejano al 15% considerado inaceptable¹⁸¹. El elevado porcentaje de respuesta y los escasos valores perdidos demostraron la gran aceptabilidad de la versión en español de la HSCL-25 en sus destinatarios finales, los pacientes de AP.

Siguiendo con el análisis de las respuestas, se confeccionó una tabla (Tabla 2 del segundo artículo) detallada con el porcentaje de respuestas en cada categoría posible en la escala de Likert del 1 al 4 de cada ítem. Este método permite detectar si en algún ítem se acumulan la mayoría de las respuestas en una categoría concreta de forma inesperada. Asimismo, permite evaluar si existe efecto suelo o efecto techo, que ocurren cuando se acumula un porcentaje elevado de respuestas de un ítem en la categoría de respuestas más baja o en la más alta, respectivamente. En este estudio se utilizó la HSCL-25 en población que acudió a la consulta de su MF por cualquier motivo, por lo que no es de extrañar que en muchos de los ítems la mayoría de las respuestas se situaran en la categoría más baja que corresponde a “1 = En absoluto”, ya que los pacientes no consultaban necesariamente por problemas de salud mental. Este hecho ocurre en otros estudios que utilizan la HSCL-25¹⁴⁴. Cabe destacar el efecto suelo en el ítem 18 “Piensa en acabar con su vida” que pregunta por los deseos de morir, el 93,6% de los pacientes respondieron con una puntuación de 1 de forma esperable. Otros estudios que ofrecen este dato muestran un porcentaje superior al 90% de respuestas en la categoría más baja para este ítem^{144,147}. No se encontró efecto techo en ninguno de los ítems. El ítem que presentó más

respuestas en la categoría “4 = Mucho” fue el 14 “Pierde el interés sexual”, con un 12,9% de las respuestas; en otros estudios el ítem con más respuestas en la categoría superior fue el 25 “Problemas para dormir”¹⁴⁴ o el 20 “Se preocupa en exceso”¹⁴⁷. Los efectos techo y suelo ocurren frecuentemente cuando se aplica un PROM existente a una población distinta, que presenta la enfermedad en un grado diferente a la población original¹⁸¹.

Se analizaron las puntuaciones medias obtenidas en cada ítem de forma global y por sexo. La puntuación media global fue de 1,57 puntos (1,42 en los hombres y 1,69 en las mujeres). El ítem con una puntuación más alta fue el 20 “Se preocupa en exceso”, con una media de 2,14. Por el contrario, el que mostró una media de puntuación inferior fue el 18 “Piensa en acabar con su vida” con una media de 1,09.

Es conocido y se ha comentado anteriormente en esta tesis doctoral que las mujeres presentan depresión más frecuentemente que los hombres. Las mujeres puntuaron más alto en todos los ítems, lo que es congruente con la bibliografía disponible^{32,41}, siendo esta diferencia estadísticamente significativa en 23 de los 25 ítems y en la puntuación total. Otros estudios que utilizan la HSCL-25 también obtienen puntuaciones globales más elevadas entre las mujeres^{144,148,150}. El ítem que presentó una diferencia mayor en cuanto a puntuación media entre hombres y mujeres fue el 14 “Pierde el interés sexual”, con una diferencia estadísticamente significativa de 0,73 puntos mayor en las mujeres. La disfunción sexual es más frecuente entre las mujeres que entre los hombres con depresión, especialmente en aquellas mayores de 45 años y con síntomas somáticos²⁶². Este ítem se omite en algunas versiones de la HSCL-25 (pasando a ser HSCL-24) por cuestiones culturales²⁶³. Por el contrario, el ítem 24 “Poco apetito” fue el que presentó una diferencia menor y no significativa en cuanto a puntuación media por sexos.

Pocos estudios que utilizan la HSCL-25 detallan el análisis pormenorizado de los ítems. La versión alemana realizada con población general y publicada por Glaesmer et al.¹³⁷ muestra un análisis similar al realizado en el presente estudio, aunque no detalla las diferencias de puntuaciones por sexos. En la población alemana, el ítem con la media más elevada fue el 8 “Dolor de cabeza”, con una media de 1,63; mientras que el ítem con una media menor fue el mismo que en el presente estudio, el ítem 18 “Piensa en acabar con su vida” con una media de 1,07 muy similar a la obtenida de 1,09. Este ítem también es el reportado con una media más baja en el estudio original de Derogatis et al. con la versión SCL-58 en pacientes psiquiátricos¹³⁰. El estudio de Mattison et al.¹⁴⁴ ofrece los porcentajes de respuestas de la HSCL-25 en cada categoría de la

escala de Likert, aunque agrupa los porcentajes de las dos categorías superiores al reportar los resultados.

La capacidad de discriminación de los ítems es un aspecto muy interesante que no siempre recibe la atención suficiente ni se presenta de forma habitual en los artículos de publicación de resultados. El ID basado en la comparación de las respuestas a los ítems de grupos extremos es uno de los propuestos por Likert y consiste en comparar el comportamiento respecto a cada ítem de dos grupos de personas con puntuaciones totales extremas¹⁸². Se refiere a la capacidad que tienen los ítems de clasificar correctamente en sus categorías de respuesta a la población sana y a la población enferma. Da información de cómo se comporta cada ítem en un contexto concreto. En este estudio se calculó el ID comparando la puntuación en cada ítem del tercio de participantes con la puntuación total más alta menos la puntuación de ese ítem del tercio de los participantes con la puntuación total más baja. Las personas con puntuaciones totales elevadas en teoría escogen las categorías más altas a la hora de responder a los ítems y, al contrario, la población sana con puntuaciones totales bajas, suelen marcar las categorías de respuesta más baja en cada ítem. Todos los ID fueron positivos, por lo que todos discriminan en el sentido esperable en menor o mayor medida. El ítem que mostró un ID más elevado fue el 4 “Nerviosismo”. No se encontró este dato en ninguna publicación de resultados de la HSCL-25 para poder realizar comparaciones, sin embargo, un estudio muestra que el ítem que mejor discriminó entre los pacientes que tenían depresión según la entrevista realizada por psiquiatras y los pacientes que tenían otro diagnóstico distinto del de depresión (fobia, ansiedad, trastorno somatomorfo, etc.) fue el ítem 3 “Debilidad”¹⁴⁴. Un ID negativo significaría que un ítem presenta respuestas en la categoría más alta en la población sana o lo contrario. Ningún ID fue negativo en el presente estudio. Valores del ID cercanos a 0 denotan respuestas muy similares en los dos grupos de puntuaciones, por lo que no son satisfactorios¹⁸² e indican que dicho ítem no discrimina adecuadamente. Esto ocurre en el ítem 18 “Piensa en acabar con su vida” debido a que la mayoría de las respuestas se agruparon en la misma categoría, por lo que una respuesta de 1 a este ítem en población de AP contribuye poco a la puntuación total para discernir entre posibles casos.

Los resultados que muestran la diferencia de puntuaciones y el ID de cada ítem por sexo se presentaron por la doctoranda y las tres directoras de esta tesis doctoral en el XLI congreso de la semFYC de 2021²⁶⁴.

Objetivo 3. Analizar la validez de constructo de la HSCL-25 en español.

La validez de constructo evalúa si las puntuaciones obtenidas son coherentes con las hipótesis disponibles, en este caso basadas en el supuesto de que la HSCL-25 mide de forma válida el constructo depresión en la población de AP de España.

En cuanto a la validez estructural, se condujo el CFA para evaluar la estructura interna de la HSCL-25, ya que se trata de una escala ampliamente estudiada y utilizada. Según la bibliografía original^{130,134,135}, la HSCL-25 presenta una estructura conocida de un único factor o de dos factores correlacionados, a pesar de que se han propuesto otros modelos por algunos autores^{137,147,148,150}. En el presente estudio se realizó el CFA para evaluar los dos modelos de un único factor y de dos factores correlacionados. El modelo de dos factores correlacionados está compuesto por el factor ansiedad (ítems 1 al 10) y por el factor depresión (ítems 11 al 25)^{135,150} y contempla que ambos factores están correlacionados. No se evaluaron otros modelos alternativos propuestos y analizados por otros autores, ya que se consideró evaluar únicamente los propuestos por los autores originales^{130,134,135} y porque no estaba entre los objetivos de esta tesis doctoral hacer un estudio exhaustivo de la estructura interna de la HSCL-25. Tampoco se condujo un EFA dado que se trata de un estudio de validación en el que la estructura de la HSCL-25 es clara y se ha estudiado ampliamente.

Los índices de ajuste de ambos modelos obtenidos en el presente estudio fueron razonablemente aceptables, si bien muestran que la estructura de la HSCL-25 es inestable y acepta más de un modelo. En general, es complicado obtener buenos índices de ajuste cuando se aplica el CFA a nivel de los ítems²⁶⁵. Frecuentemente ocurre que el CFA, mucho más restrictivo que el EFA, no permite aceptar la estructura interna que en repetidos análisis encontraron los análisis exploratorios, siendo complicado replicar la estructura factorial incluso en la misma población²⁶⁶.

En el presente estudio, el modelo de dos factores correlacionados mostró mejor ajuste que el de un único factor, lo que concuerda con otros estudios¹⁵⁰. Este resultado sugiere que el desarrollo de los síntomas de ansiedad y depresión de la población de AP se produce como dos entidades distintas²⁶⁷ pero que dichos síntomas están correlacionados. Se obtuvo una correlación factorial de 0,84, lo que indica que ambas dimensiones depresión y ansiedad están fuertemente correlacionadas entre sí y de forma positiva. Esta cifra es similar a la obtenida en algunos estudios^{147,151} y mayor a la detectada en otros¹³⁷. Esta correlación es comprensible, ya que es frecuente encontrar síntomas de ansiedad en pacientes diagnosticados de depresión y

muchas veces ansiedad y depresión se presentan como comorbilidades asociadas entre sí^{85,268-270}.

Los resultados obtenidos indicaron cargas factoriales de valores moderados-elevados, todas ellas $\geq 0,3$ y estadísticamente significativas, lo que indica que cada ítem se correlacionó de forma adecuada con el factor esperado. El ítem con una carga factorial mayor en ambos modelos fue el 17 “Se siente triste”. Este ítem también obtuvo la carga factorial más elevada en la versión francesa en el modelo de un solo factor¹⁴⁰, a pesar de que el análisis llevado a cabo por los investigadores franceses fue el análisis por componentes principales en lugar del CFA. Otros estudios también revelan que este ítem es el que presenta mayor carga factorial en el modelo de un factor¹⁴⁷ y en el de dos factores¹⁵¹. Por el contrario, el ítem con menor carga factorial fue el 24 “Poco apetito” en ambos modelos; resultado que diverge del de otros estudios.

La comprobación de hipótesis que se realiza para el análisis de validez de constructo incluye la comparación de las puntuaciones de la HSCL-25 frente a las puntuaciones de otros instrumentos de medida del mismo constructo. COSMIN considera como “muy buena” esta medida si se realiza en más de 100 pacientes y se obtienen ambas medidas en el mismo momento²¹³. En el presente estudio, la totalidad de los pacientes fueron evaluados con la escala PHQ-9, que también evalúa el constructo de la depresión⁹⁶. Ambas escalas fueron cumplimentadas en el día de la inclusión al estudio, ya que estaban incluidas en el cuaderno de recogida de datos²³⁰. Se demostró una validez convergente con la PHQ-9 ya que la correlación entre ambas puntuaciones fue de 0,78, considerada adecuada²³⁷.

Es frecuente el análisis de correlación con otros PROMs que miden el mismo constructo en la evaluación de la validez. En un artículo se analizó la correlación de las subescalas de depresión y ansiedad de la versión española de la HSCL-25 con las versiones PHQ-2, PHQ-8 y PHQ-9 con los mismos datos que el presente estudio, procedentes de los datos EIRA¹⁷⁹. La subescala de depresión mostró una correlación de 0,73 con la PHQ-9, mientras que la correlación para la subescala de ansiedad fue de 0,61. En otros estudios la HSCL-25 ha demostrado una correlación $> 0,7$ con la versión PHQ-4^{137,155} y $> 0,8$ con la *Brief Symptom Inventory* de 18 ítems (BSI-18), que mide depresión, ansiedad y somatización^{137,155}. En otro estudio se describe una correlación de 0,78 de la HSCL-25 con el *Mental Health Index* de 5 ítems (MHI-5)¹⁵⁴.

La validez de constructo incluye la validez transcultural, que en el caso del presente estudio hace referencia a si el funcionamiento de los ítems de la HSCL-25 en español refleja el funcionamiento de los ítems en la versión original inglesa. Este tema se ha tratado en los objetivos específicos 1 y 2 comentados previamente. La versión española mostró las correlaciones esperadas con el

constructo de la depresión y los ítems discriminaron de forma adecuada en la población con puntuación más alta.

Objetivo 4. Analizar la validez de criterio de la HSCL-25 en español.

La validez de criterio en el presente estudio se refiere al grado en que las puntuaciones obtenidas en la HSCL-25 son un reflejo adecuado del *gold standard* de la medida de la depresión.

Un PROM es una medida que únicamente puede ser reportada por el paciente, por lo que, estrictamente hablando, no dispone de un *gold standard*. La única excepción sería en la comparación de una versión corta de una escala si se dispone de la medida de la versión completa²¹³. La validez de tipo concurrente se utiliza cuando se dispone de una medida extremadamente convincente pero poco práctica, considerada *gold standard*, que se enfrenta a una medida más práctica que quiere resultar convincente¹⁸⁰. En el caso de la depresión, la entrevista clínica basada en los criterios DSM-5 es la referencia para el diagnóstico. Se trata de una herramienta insustituible pero que requiere tiempo y habilidades por parte del profesional, por lo que disponer de una herramienta ágil de ayuda al diagnóstico puede suponer una ventaja en el entorno de complejidad de la AP. Se utilizó la CIDI como *gold standard* para evaluar la validez de criterio en el presente estudio con el fin de disponer de una medida homogénea en todos los pacientes participantes.

Para analizar la validez de criterio de tipo concurrente, todos los pacientes fueron invitados a ser entrevistados mediante la CIDI en el mismo momento que cumplimentaron la HSCL-25, tal y como sugieren las guías COSMIN¹⁸¹. Por cuestiones de factibilidad, es frecuente encontrar estudios en los que solo una pequeña parte de la muestra, entre aquellos que han resultado con una puntuación positiva en el PROM, recibe la medición de la prueba de referencia^{136,138,140,151}. En el presente estudio, sin embargo, el 96% de los participantes que cumplimentaron la HSCL-25 accedieron a responder a la CIDI.

La medida para dicho análisis fue el AUC obtenida en la curva ROC, ya que este es el método de elección en el caso de que la variable del PROM sea cuantitativa y la de la variable *gold standard* sea dicotómica²¹³, como es el caso del presente estudio ya que la HSCL-25 aporta una puntuación numérica y la CIDI clasifica al paciente “con depresión” o “sin depresión”. Se obtuvo un valor de AUC de 0,892, mayor en hombres (AUC 0,960) que en mujeres (AUC 0,837), lo que indica que la HSCL-25 tiene una probabilidad cercana al 90% de clasificar de forma correcta a los pacientes en relación con la CIDI, siendo esta probabilidad mayor en los hombres, hecho que coincide con otros estudios^{139,144}. El valor del AUC global fue discretamente superior al obtenido en otros

estudios que contrastan la HSCL-25 frente a la CIDI¹³⁹, la entrevista SCAN¹⁵¹, la entrevista SCID¹⁴⁹ o frente a entrevistas clínicas realizadas por psiquiatras¹⁴⁴.

Por otro lado, dado que la HSCL-25 tiene establecidos unos puntos de corte concretos, también se puede utilizar como variable dicotómica clasificando a los pacientes en sanos o posibles casos de depresión. En este caso, la sensibilidad y la especificidad son los métodos preferidos para la evaluación de la validez de criterio²¹³. Los puntos de corte de la HSCL-25 utilizados en la literatura son: caso probable de depresión si la puntuación media de la escala es $\geq 1,55$ y caso psiquiátrico que requiere tratamiento si la puntuación media es $\geq 1,75$ ¹³⁶⁻¹³⁹. El punto de corte de 1,75 es más restrictivo y se recomienda como predictor válido de trastorno mental^{136,137,139}.

La sensibilidad y especificidad globales y por sexos fueron elevadas utilizando el punto de corte de 1,75; se obtuvo una sensibilidad de 88,1% y una especificidad de 76,7%. Los valores de sensibilidad fueron similares en ambos sexos, mientras que la especificidad fue más elevada en los hombres. La sensibilidad global fue mayor a la encontrada en estudios previos^{136,138-140,146,151}, mientras que la especificidad fue similar a la de 73% reportada por Nettelbladt et al.¹³⁸ y la de 78% del estudio de Lundin et al.¹⁵¹, ambas realizadas en población de Suecia. Otros estudios han obtenido un valor de especificidad más elevada^{136,140,144}. A pesar del elevado número de falsos positivos que se obtuvieron, hecho que ocurre también en otros estudios^{149,151}, el valor predictivo negativo fue superior a 97% en ambos sexos para el punto de corte de 1,75. El punto de corte de 1,55 obtiene mayor número de falsos positivos^{136,144}, por lo que se consideró no incluirlo en los análisis. El elevado VPN, junto con una buena sensibilidad, apoya la idea de que la escala es una buena herramienta para un primer cribado de la depresión que debe ir seguido de una entrevista clínica en caso de una puntuación superior al punto de corte para confirmar o descartar el diagnóstico. Este proceso pueden llevarlo a cabo los propios MF, quienes en caso de confirmar el diagnóstico de depresión pueden iniciar el tratamiento y el seguimiento en la mayoría de los casos. Por el contrario, una puntuación por debajo del punto de corte tiene una elevada probabilidad de tratarse de un paciente que no presenta depresión en el momento de cumplimentar la HSCL-25, por lo que puede ser de utilidad en las consultas de los MF. Se debe tener en cuenta la diferencia de prevalencia de la depresión encontrada entre ambas medidas, ya que fue de 28,5% en la HSCL-25 y de 8,0% en la CIDI. Este aspecto influye en el bajo VPP calculado y obliga a los MF a completar el diagnóstico mediante la entrevista clínica.

El concepto de validez de criterio de tipo concurrente en este contexto se solapa con el de exactitud diagnóstica, lo que conlleva que algunos autores utilicen un término u otro para referirse al AUC, la sensibilidad y la especificidad cuando evalúan PROMs frente a pruebas de referencia. Según las directrices *Standards for Reporting Diagnostic accuracy studies (STARD)*²⁷¹,

un estudio de exactitud diagnóstica evalúa la capacidad de una o más pruebas médicas para clasificar correctamente a los participantes como portadores de una enfermedad concreta. Las directrices STARD indican que cuando pueden definirse múltiples puntos de corte de una prueba diagnóstica, los autores pueden informar de una curva ROC que represente gráficamente la combinación de sensibilidad y especificidad para cada posible punto de corte de positividad de la prueba. El AUC informa en un único valor numérico sobre la precisión diagnóstica global de la prueba que se evalúa. Según esta definición, es cierto que los términos AUC, sensibilidad y especificidad se agrupan también bajo el término "precisión diagnóstica".

Al analizar el mejor punto de corte obtenido mediante el cálculo del índice de Youden tal y como realizan otros estudios¹⁵¹, el punto de corte global y para mujeres fue de 1,76, muy similar al de 1,75 propuesto en la versión original y utilizado en otros estudios^{144,145}. Este también fue el punto de corte óptimo en las mujeres encontrado por Sandanger et al. al comparar la HSCL-25 frente a la CIDI¹³⁹. Otros estudios obtienen valores estadísticos de sensibilidad y especificidad similares para el punto de corte de 1,75 y el de 2,00¹⁵¹. En el presente estudio, el punto de corte fue más elevado en los hombres, siendo de 1,84, al contrario de lo encontrado en otros estudios en que el punto de corte fue mayor en las mujeres¹³⁹. Este punto de corte ofrece mejor especificidad y valor predictivo positivo que el punto de corte original de 1,75 en los hombres, a pesar de que no cambia la sensibilidad ni el valor predictivo negativo. A pesar de que los hombres presentan puntuaciones más bajas que las mujeres en todos los ítems de la HSCL-25, los resultados indican que precisan de una puntuación más elevada que ellas para ser considerados posibles pacientes con depresión, concretamente de tres puntos más que las mujeres en la suma de las respuestas (antes de dividir la puntuación total entre el número de ítems).

Objetivo 5. Analizar la fiabilidad de la HSCL-25 en español.

El análisis de la fiabilidad en este estudio evalúa el grado en que la puntuación obtenida en la HSCL-25 está libre de errores de medición. Dicho de otra manera, si un mismo paciente obtendría la misma puntuación en repetidas ocasiones siempre y cuando no se modificase el estado de dicho paciente.

Existen diferentes formas de analizar la fiabilidad. El coeficiente alfa de Cronbach es una buena medida de fiabilidad de consistencia interna en PROMs unidimensionales, pero si se trata de PROMs multidimensionales cuyos factores presentan una correlación moderada entre 0,2 y 0,4 el alfa de Cronbach puede infravalorar la consistencia interna del conjunto²⁷². La HSCL-25 se

comporta de forma unidimensional o con dos factores con una correlación elevada de 0,84, por lo que el valor alfa de Cronbach obtenido es un buen reflejo de la fiabilidad de la HSCL-25.

El coeficiente de fiabilidad alfa de Cronbach obtenido en el presente estudio para un solo factor fue de 0,92, similar al de 0,93 alcanzado con la versión en francés¹⁴⁰. Este valor fue superior a 0,80 al medirlo de forma separada para cada una de las subescalas depresión (0,88) y de ansiedad (0,83). La versión original SCL-58 de Derogatis et al. reporta un valor alfa de Cronbach muy similar de 0,86 en la subescala depresión y 0,84 subescala ansiedad¹³⁰. Otros autores obtienen valores cercanos al obtenido en el presente estudio, que varían entre 0,80 y 0,94^{133,137,146,151}.

Nunnally sitúa el nivel crítico de fiabilidad en 0,70; pero establece que para decisiones de importancia individual, como sería el caso de un diagnóstico de depresión, la fiabilidad debería aumentar a 0,90²¹⁰. Los resultados objetivaron una excelente fiabilidad de la puntuación de la HSCL-25 para medir depresión en la población de AP, especialmente al utilizarla con la puntuación obtenida con el conjunto de los 25 ítems.

La aportación de cada ítem a la consistencia interna de la puntuación total puede medirse con indicadores basados en la correlación (homogeneidad), en la covariancia (alfa de Cronbach sin el ítem) y en la regresión (coeficiente de determinación, R^2)^{182,209}. La media de correlación entre los ítems fue 0,362, con un rango que osciló entre 0,090 y 0,621, como se refleja en la Tabla S2 del material suplementario del segundo artículo de esta tesis doctoral. Se consideran adecuados valores de correlación entre-ítem de entre 0,2 y 0,5¹⁸¹. Valores de correlación muy elevados indican ítems redundantes, mientras que valores muy bajos indican ítems poco relacionados con el resto de las puntuaciones y que por lo tanto podrían no ser adecuados. La mayor parte de los ítems obtuvieron valores adecuados en el presente estudio. Es infrecuente que los autores aporten la tabla de correlación entre ítems en las publicaciones, no se ha encontrado ningún artículo con el que poder comparar este dato.

El valor del alfa de Cronbach sin el ítem hace referencia al valor que tendría la fiabilidad de la escala en el caso de que ese ítem fuera eliminado. Es otra forma de analizar la contribución de cada ítem a la consistencia interna a partir de la variancia común entre ellos. Si los ítems no tuvieran información en común sus covariancias valdrían 0 y el alfa de Cronbach tomaría el valor 0, indicando que los ítems no son consistentes. Si un ítem consistente con los demás se elimina de la escala, la consistencia interna global sin ese ítem debería disminuir¹⁸² y viceversa. Por consiguiente, valores alfa de Cronbach sin el ítem superiores al valor alfa de Cronbach total de la escala indican ítems inconsistentes, que en caso de ser eliminados conllevarían una mejor

fiabilidad de la escala. Esto ocurre en el ítem 24 “Poco apetito”, que tiene un valor alfa de Cronbach sin el ítem de 0,9167 (superior al del alfa de Cronbach de la escala total de 0,9166). Este matiz se aprecia a partir del cuarto decimal, por lo que es habitual publicar estos valores con tantos decimales.

La correlación de cada ítem con la puntuación total de la escala también es un método para detectar ítems no homogéneos. Se basa en el supuesto de que, si un ítem es consistente con el resto, deberá existir cierta correlación entre ese ítem y todos los demás. Al estar el ítem incluido en la puntuación total de dicha correlación, se debe corregir el valor calculándolo a partir de la suma de todos los ítems de la escala menos el ítem evaluado¹⁸². Las recomendaciones sobre cómo interpretar las correlaciones ítem-total corregido no son unánimes, los valores que se consideran mínimos se sitúan entre 0,3 y 0,4 dependiendo de los autores²⁴⁰.

Si un ítem es consistente con el resto de los ítems de la escala, su puntuación podría predecirse a partir de ellos. Otra forma de medir la contribución de un ítem a la consistencia interna es mediante una regresión múltiple con ese ítem como variable dependiente y el resto de ítems como variables independientes (predictoras)¹⁸². El coeficiente de determinación (R^2) es la proporción de la variancia del ítem analizado explicada por el resto de los ítems. Cuanto más alto sea el valor, más homogéneo es el ítem.

El ítem 17 “Se siente triste” mostró ser el más homogéneo en todos los análisis, siendo el que presentó una correlación más elevada con la puntuación del resto de ítems de la escala, un coeficiente de determinación más elevado y el que más contribuyó a aumentar la consistencia interna global. Este resultado coincide con el encontrado en otro estudio que publica los datos de la correlación ítem-total corregida¹³⁷. Por el contrario, el ítem menos homogéneo fue el 24 “Poco apetito”. Si bien en la depresión el apetito suele estar disminuido²⁸, es posible que haya un aumento de la ingesta que conlleve una ganancia ponderal durante el episodio depresivo¹⁸. La fiabilidad de la HSCL-25 en español sería mejor en caso de eliminar este ítem.

Se analizó también la fiabilidad test-retest mediante el cálculo del ICC, tal y como sugieren los estándares²¹³. Este valor fue de 0,92 y superior al observado en otros estudios¹⁴¹, lo que indica que las puntuaciones son estables en el tiempo. La duración del episodio depresivo es un aspecto esencial en el diagnóstico de la depresión y que está incorporado en los criterios diagnósticos del DSM-5¹⁸. La puntuación de la HSCL-25 tiende a mantenerse estable en un periodo de 4 semanas¹³⁵, por lo que se decidió administrar el retest en el periodo aproximado de un mes. El intervalo de tiempo entre la entrevista basal y el retest se consideró adecuado y

las condiciones de aplicación aceptables, a pesar de que el retest se realizó por teléfono para evitar la sobrecarga de los participantes.

Objetivo 6. Evaluar la validez y fiabilidad de las escalas breves Hopkins Symptom Checklist-10 (HSCL-10) y Hopkins Symptom Checklist-5 (HSCL-5) como alternativas a la HSCL-25.

Existen múltiples cuestionarios para evaluar la depresión que pueden utilizarse como herramientas de cribado para la detección de casos de todos los espectros de gravedad^{97,101,120}, varios de ellos han sido validados en el contexto de la AP¹²⁹. Se han propuesto versiones cortas de algunos cuestionarios con el fin de ofrecer una utilidad diagnóstica similar a las versiones largas pero con ahorro de tiempo tanto para el paciente como para el profesional^{27,152,273}.

La escala HSCL-25 es una de las versiones cortas de la versión original SCL-90-R¹³³. Sin embargo, en los últimos años se han propuesto otras versiones todavía más breves, la HSCL-5 y la HSCL-10, que han mostrado una fiabilidad y validez aceptables^{154,155}, especialmente la de 10 ítems^{158,165,166}. Se consideró analizar las propiedades psicométricas de estas dos versiones en español en la población de AP, dado que por su brevedad pueden resultar más fácilmente manejables en la consulta de los MF. No se ha encontrado en la bibliografía ningún artículo que haya utilizado estas versiones con población adulta de AP en ningún idioma. Se realizó un análisis prácticamente idéntico al llevado a cabo con la HSCL-25.

La puntuación media de la HSCL-25 fue de 1,57, muy similar a la de 1,55 obtenida con la escala HSCL-10. Esta media aumentó a 1,72 al utilizar la versión de 5 ítems. La media fue superior a la obtenida con las mismas versiones en población general alemana¹⁵⁵. Se analizó la puntuación media por subgrupos teniendo en cuenta las características sociodemográficas de la población encontrando medias muy cercanas entre la HSCL-25 y la HSCL-10, y algo superiores en la HSCL-5. Igual que con la HSCL-25, las medias de puntuación fueron mayores en mujeres de forma significativa. Se encontraron diferencias significativas en la puntuación media de los pacientes en función de la edad y del estado civil en las versiones de 25 y 10 ítems, dicha significación se perdió en la versión de 5 ítems. Con este primer análisis se intuye una semejanza de puntuaciones entre la HSCL-10 y la HSCL-25 que no se mantiene de la misma forma en la HSCL-5.

Se condujo el CFA para evaluar la validez estructural dentro del análisis de la validez de constructo. Los ítems de la HSCL-10 y la HSCL-5 mantienen la misma estructura factorial de la HSCL-25, es decir, los ítems forman parte de las mismas dimensiones que en la HSCL-25¹⁵⁵. Se analizaron los mismos modelos que en la HSCL-25 de un factor y de dos factores correlacionados.

Todos los ítems presentaron cargas factoriales $\geq 0,45$ estadísticamente significativas en el factor que representaba el constructo esperado.

Los índices de bondad de ajuste de los modelos estudiados no fueron óptimos, mostrando que la estructura factorial era inestable para ambas la HSCL-10 y la HSCL-5. Se debe tener en cuenta que los pacientes respondieron a los 25 ítems y que este análisis debería completarse en el futuro con las respuestas a las versiones de 10 y 5 ítems propiamente para evaluar otros modelos alternativos de la estructura factorial. El modelo de dos factores correlacionados se ajustó mejor en ambas versiones, especialmente en la HSCL-5. Otro estudio que analiza la estructura interna de ambas también a partir de las respuestas a la HSCL-25 mostró que ambos modelos de uno y dos factores son viables en la HSCL-5 y que el de dos factores se ajusta mejor en la HSCL-10¹⁵⁵.

El ítem 17 "Se siente triste" tuvo la mayor carga factorial en todos los modelos analizados de ambas versiones. La correlación entre los dos factores en los modelos estudiados fue de 0,90 entre las dimensiones de la HSCL-10 y de 0,87 en la HSCL-5. Igual que sucedió en la HSCL-25, estos resultados indicaron que las dimensiones de depresión y ansiedad se correlacionaron de forma positiva.

A sugerencia de una revisora de la revista en la que se publicó el tercer artículo del compendio de esta tesis doctoral, se utilizó el método de la comparación de la puntuación en subgrupos como comprobación de hipótesis para incluirlo en el análisis de la validez de constructo. Se analizaron las diferencias de puntuación obtenidas en ambos sexos, y se encontró una diferencia de puntuación entre ambos grupos, siendo mayor y estadísticamente significativa en las mujeres, tal y como era esperable teniendo en cuenta la literatura disponible³². Este análisis no se había realizado previamente en la HSCL-25.

Para el análisis de la validez de criterio no se analizó la correlación de puntuaciones entre las versiones breves y la HSCL-25 dado que los datos se obtuvieron a partir de las mismas respuestas de la HSCL-25, por lo que es evidente que las puntuaciones mostrarían una correlación muy elevada. Otros estudios han demostrado una correlación elevada de las puntuaciones de ambas con la HSCL-25 y entre la HSCL-10 y la HSCL-5^{133,154,155}.

Dado que se disponía del resultado de la CIDI en todos los pacientes, se utilizó como *gold standard* para evaluar la validez de criterio mediante el AUC de las curvas ROC. Otros estudios han evaluado la validez de criterio mediante curvas ROC entre la versión de 25 ítems enfrentada a las versiones cortas¹⁵⁴. El AUC fue de 0,88 para la HSCL-10 y de 0,85 para la HSCL-5, ambos superiores a 0,75 considerado el valor mínimo de utilidad clínica²¹⁴ y por encima de 0,80, valor

que confiere una exactitud de clasificación buena ($\geq 0,90$ se considera excelente). El estudio llevado a cabo por Haavet et al.¹⁶⁵ también obtuvo un AUC de 0,88 al comparar la HSCL-10 frente a CIDI. Igual que con la HSCL-25, el AUC fue mayor en hombres que en mujeres.

La prevalencia de depresión fue menor según la HSCL-10 que con la HSCL-5 y la HSCL-25. Las tres versiones y el *gold standard* CIDI mostraron una prevalencia superior en mujeres que en hombres. Utilizando los puntos de corte correspondientes a cada versión¹⁵⁴, la sensibilidad fue similar en ambos géneros, mientras que la especificidad fue mejor en los hombres, igual que ocurrió con la HSCL-25. Los valores de sensibilidad y especificidad fueron mejores para la HSCL-10 que para la HSCL-5, siendo la sensibilidad de 79,9% en la HSCL-10 y de 78,0% en la HSCL-5, y la especificidad de 83,0% en la HSCL-10 y de 72,8% en la HSCL-5. Otro estudio que compara la HSCL-10 con la CIDI¹⁶⁵ muestra resultados ligeramente superiores en cuanto a sensibilidad, aunque utilizan un punto de corte distinto. El valor predictivo negativo fue superior a 97% en ambas escalas, igual que ocurrió con la HSCL-25.

En el presente estudio, el punto de corte óptimo calculado con el índice de Youden fue muy cercano al propuesto en la bibliografía consultada de 1,85 para la HSCL-10 y 2,00 para HSCL-5¹⁵⁴. Estos puntos de corte han sido utilizados por otros autores en el análisis de estas escalas¹⁵⁸, aunque otros han propuesto unos puntos de corte diferentes al utilizar estas versiones con adolescentes¹⁶⁵. Dado que los puntos de corte obtenidos fueron muy cercanos a los ya utilizados por otros autores, se optó por utilizar los mismos para permitir la comparabilidad con otros estudios.

El valor alfa de Cronbach depende del número total de ítems de forma directamente proporcional¹⁸¹. Es normal que al utilizar versiones reducidas de un PROM que contienen un menor número de ítems, la fiabilidad disminuya. Este hecho se debe a la fórmula del coeficiente, en cuyo numerador está el número de ítems elevado al cuadrado (k^2)¹⁸⁰. Un estudio que compara 11 versiones diferentes de la SCL mostró que el valor alfa de Cronbach es mayor en las versiones más extensas¹³³.

Se obtuvo un valor alfa de Cronbach de 0,84 para la HSCL-10 y de 0,77 para la HSCL-5. Los valores de fiabilidad de ambas versiones superaron el valor de 0,7 considerado como aceptable²¹⁰. Este valor es muy similar al obtenido en otros estudios^{133,165,166} y, tal y como es esperable y sucede también en otros estudios^{133,154,155}, la fiabilidad de las versiones cortas fue inferior a la HSCL-25.

En el apartado previo se ha argumentado que el ítem 17 “Se siente triste” fue el más consistente de la HSCL-25 en todos los análisis. Este ítem está incluido tanto en la HSCL-10 como en la HSCL-5 y también fue el más consistente en ambas y en caso de eliminarlo de las versiones cortas la

fiabilidad disminuiría al mínimo. En el CFA también fue el ítem con más peso en su carga factorial en todas las versiones. En la HSCL-25, los siguientes ítems más consistentes fueron el 4 “Nerviosismo” y el 7 “Se siente tenso/a”; el primero está incluido en la versión de 5 ítems y el segundo en la de 10 ítems. Por el contrario, los tres ítems más inconsistentes de la HSCL-25, el ítem 24 “Poco apetito”, el ítem 8 “Dolor de cabeza” y el ítem 18 “Piensa en acabar con su vida” no están incluidos en ninguna de las dos. Este es un aspecto relevante, dado que los ítems más consistentes en las versiones largas son los que deben permanecer en el caso de eliminar ítems para acortar la extensión de un PROM.

Al analizar el coeficiente alfa de Cronbach sin el ítem, se observó que el ítem 25 “Problemas para dormir” empeora la fiabilidad de la HSCL-10, es decir, que la fiabilidad mejoraría en caso de eliminar este ítem. Este hallazgo coincide con el observado por Kleppang et al. al utilizar la versión de 10 ítems con adolescentes en Noruega¹⁵⁸. El resto de los ítems contribuyeron a una buena fiabilidad tanto en la HSCL-10 como en la HSCL-5.

A modo de resumen, tanto la HSCL-10 como la HSCL-5 muestran valores de validez y fiabilidad adecuados para el cribado de la depresión en pacientes de AP de España, siendo mejores los valores de la HSCL-10.

7.3 Fortalezas

La fortaleza principal de esta tesis doctoral es que este es el primer estudio que obtiene una traducción de la HSCL-25 en español en la que se ha evaluado la validez y la fiabilidad de sus puntuaciones en España, concretamente en el entorno de AP. Asimismo, es el primer estudio publicado que utiliza la HSCL-10 y la HSCL-5 con población adulta de AP.

Otro aspecto relevante es que forma parte de un proyecto europeo internacional en el que colaboran diferentes países utilizando una metodología común en el contexto de la AP. España ha sido el segundo país en completar las cuatro etapas del estudio planteado, por detrás de los investigadores principales del proyecto en Francia¹⁴⁰. Este esfuerzo para fortalecer las redes de investigación en AP es la base que ha de proporcionar la evidencia científica en este ámbito, tan diferente a veces del contexto hospitalario o del resto de especialidades médicas.

Desde la fase 1 de este estudio en la que se llevó a cabo la traducción y adaptación de la HSCL-25 al español, se tuvo en cuenta el ámbito de trabajo de AP como futuro emplazamiento de utilización de la escala. Se consideró la perspectiva de los profesionales que la interpretarán en

el futuro y se llevó a cabo una prueba piloto con una muestra de pacientes de AP captados en la consulta de sus MF.

Cabe destacar que, como alternativa a las traducciones directas recomendadas en las guías para traducir escalas, en este proyecto se hizo una única traducción directa realizada por una licenciada seguida de un proceso de consenso formal con metodología Delphi en la que los panelistas fueron MF.

Adicionalmente, es relevante el elevado tamaño muestral y la heterogeneidad de la población incluida en el análisis de la validez y la fiabilidad en la fase 2, ya que se trata de pacientes de AP reclutados en diferentes CS rurales y urbanos de diferentes comunidades autónomas de España, lo que aumenta la representatividad de la población de referencia.

El escaso número de respuestas en blanco es otro punto fuerte del estudio. De los 790 pacientes que aceptaron participar, 22 no respondieron a ninguno de los ítems de la HSCL-25 y un individuo respondió a solamente 12 de los 25 ítems. Estos individuos fueron eliminados del análisis ya que se consideró necesaria la respuesta a como mínimo la mitad de los ítems (13 ítems respondidos) para poder imputar las respuestas en blanco. Las guías COSMIN aconsejan un número total ≥ 200 participantes de la población diana para evaluar un PROM conocido en un nuevo contexto, teniendo en cuenta los valores perdidos esperados²¹³. Este estudio supera con creces este requisito ya que, tras eliminar los 23 individuos, se obtuvo una muestra de 767 individuos. Se imputaron los valores perdidos de 30 de los individuos que tenían entre 1 y 5 ítems en blanco. Como método de imputación se escogió el que daba la puntuación más representativa del individuo, que consiste en la media de las respuestas del propio individuo al resto de los ítems. En contraposición al valor más representativo de la muestra, que sería la media de la puntuación del ítem en blanco por el resto de los participantes.

Los participantes fueron evaluados por una unidad externa y todos los instrumentos se administraron en el mismo día, tanto la HSCL-25, como la PHQ-9 y la CIDI. Es destacable que todos los pacientes fueron entrevistados mediante la CIDI (*gold standard*). Este aspecto ya se ha comentado en la discusión por objetivos y es una de las principales fortalezas de este estudio, ya que es frecuente encontrar estudios en los que solo una pequeña parte de la muestra recibe la medición de la prueba de referencia^{136,138,140,151}.

7.4 Limitaciones

La limitación principal de este trabajo fue que los pacientes fueron reclutados en el contexto del estudio EIRA²³⁰ siguiendo los criterios de selección establecidos en dicho proyecto. Los

participantes del estudio fueron pacientes visitados en AP de entre 45 y 75 años con dos o más hábitos de vida no saludables. A pesar de que estos criterios de selección pueden parecer algo restrictivos, es frecuente que los adultos presenten dos o más comportamientos poco saludables simultáneamente²⁷⁴, además, las conductas de salud evaluadas en el proyecto EIRA²³⁰ (tabaquismo, baja adherencia a la dieta mediterránea y escasa actividad física) son las más comunes entre los pacientes AP²⁷⁵. Por este motivo, y teniendo en cuenta el gran tamaño muestral, se consideró que los participantes eran representativos de la población española que consulta en AP. La dieta poco saludable puede asociarse a la depresión²⁷⁶, y aunque esto podría haber aumentado la probabilidad de que los pacientes sufrieran depresión, la prevalencia de depresión encontrada fue similar a la esperada⁷.

Como *gold standard* del diagnóstico de la depresión se utilizó la sección E de la CIDI que incluye únicamente las preguntas que conducen a dicho diagnóstico. A pesar de que la HSCL-25 está formada por las dos dimensiones ansiedad y depresión, sus puntuaciones pretenden medir el constructo de la depresión. La estructura de la HSCL-25, que combina los síntomas de ansiedad y depresión, la hace adecuada para la práctica diaria en AP, donde es frecuente encontrar ansiedad y depresión de forma solapada¹⁴⁰.

Por lo que se refiere al análisis de la validez y la fiabilidad de la HSCL-10 y la HSCL-5, la principal limitación es que estas se evaluaron a partir de las respuestas de la escala de 25 ítems. Para realizar un análisis adecuado se deberían administrar estas versiones sin el resto de los ítems, ya que esto puede modificar las respuestas de los pacientes que autocumplimentan la escala. Por cuestiones de factibilidad se decidió realizar el análisis de este modo, tal y como han hecho en otros estudios^{154,155}. A pesar de ello, se considera que los resultados son válidos dado el elevado número de participantes.

Finalmente, se quiere mencionar que gran parte de la bibliografía utilizada en esta tesis doctoral es antigua, algunos artículos citados datan de los años 1960-1970. Este hecho se debe a que se han citado algunos autores clásicos de la psicometría y a que varias escalas se desarrollaron en esos años y se han citado las versiones originales.

7.5 Implicaciones de los resultados y líneas futuras

El diagnóstico de la depresión es un proceso complejo que requiere de la habilidad del clínico y de la disponibilidad de tiempo en la consulta de AP, muchas veces limitado en la práctica clínica real. Los pacientes con depresión que acuden a visitarse con su MF a menudo presentan síntomas inespecíficos o quejas somáticas como fatiga, cefalea, problemas de sueño o dolor. En

estos casos, utilizar un cuestionario puede ser de utilidad ya que las preguntas concretas en forma de ítems pueden ayudar a los pacientes a comprender la depresión de una forma menos abstracta²⁷⁷ y a poner palabras a sus síntomas.

La entrevista clínica es insustituible en el diagnóstico final de la depresión, sin embargo, la probabilidad de diagnosticarla de forma temprana mejora con el uso de instrumentos de cribado y de seguimiento que han mostrado validez⁸² y existe evidencia convincente de que estos pueden identificar los casos de depresión con precisión⁹². A pesar de ello, sus puntuaciones se deben interpretar con cautela, teniendo en cuenta los diferentes puntos de corte para evitar el sobrediagnóstico y el sobretreatmento^{78,87} y nunca deben sustituir al juicio clínico. Los profesionales deben considerar el contexto de los síntomas y comprender la complejidad²⁷⁸ de la relación entre los problemas psicológicos, físicos y sociales así como sus variaciones temporales^{78,279}. No cabe ninguna duda de que los MF tienen habilidades a la hora de manejar la complejidad, ya que se mueven en ese terreno diariamente^{280,281}.

Un estudio cualitativo reciente que pregunta a los MF sus opiniones acerca del uso de herramientas de evaluación de la depresión en AP²⁸² muestra que las consideran útiles y destacan como ventajas el fomento de la identificación de casos no detectados, el alivio de la carga de trabajo diaria y la posibilidad de tener un canal de comunicación más amplio para llegar a más pacientes. Por otro lado, señalan como aspectos negativos que el cribado puede amenazar la relación médico/a-paciente con algunas preguntas que pueden incomodar a los pacientes.

Algunos artículos revelan la preocupación de los MF por el uso de guías de práctica clínica, criterios diagnósticos y escalas realizadas únicamente por psiquiatras sin tener en cuenta el punto de vista de la AP y que por lo tanto, hacen que no sean transferibles a la práctica diaria de la AP^{78,82,283}. Es imprescindible que las escalas clínicas estén correctamente adaptadas al entorno cultural donde serán utilizadas, estén disponibles en la lengua materna de los pacientes y que dispongan de unas propiedades psicométricas robustas en cada contexto¹¹¹. Es importante comprobar la validez y la fiabilidad en AP de las herramientas comúnmente utilizadas en hospitales antes de extender su uso en AP.

Entre los beneficios de la detección precoz de la depresión hay pruebas convincentes de que los programas de cribado de la depresión en AP o entornos comparables producen una mejora de los resultados en salud⁹². Por otro lado, no existe evidencia consistente sobre los perjuicios del cribado de la depresión y en caso de existir, son de pequeña magnitud⁹², ya que un resultado positivo conlleva una anamnesis específica, sin pruebas invasivas ni tratamientos inadecuados. En contraposición, existen otros cribados controvertidos y sin evidencia sólida que pueden

conllevar perjuicios para los pacientes, como es el caso de la biopsia de próstata en pacientes con PSA elevado²⁸⁴.

El término cribado es un concepto epidemiológico que consiste en realizar una prueba o examen médico con el fin de detectar una patología. Por otro lado, el término detección hace referencia a descubrir algo que no era patente²⁶⁰ y es un concepto más clínico que posiblemente encaja mejor en el contexto de la consulta de los MF en AP. En la evaluación de la depresión en AP, el uso de cuestionarios puede ofrecer un primer cribado. La HSCL-25 ha demostrado tener elevada validez y fiabilidad como herramienta de cribado de la depresión que, utilizada en consulta por los MF, puede ayudar a detectar casos que pasarían desapercibidos. El elevado valor predictivo negativo de 97%, que se mantiene en el mismo valor en la HSCL-10 y en la HSCL-5, ofrece la posibilidad de descartar depresión en los pacientes con un resultado negativo con el punto de corte correspondiente. En los casos de un resultado positivo, es preciso completar la evaluación de los síntomas de depresión para establecer o descartar el diagnóstico, siguiendo los criterios del DSM-5.

Los análisis que se han realizado y se han presentado en los artículos publicados, acumulan pruebas que confirman la validez y la fiabilidad de las puntuaciones obtenidas en la HSCL-25 en pacientes de AP. No se ha empleado la expresión “escala válida y fiable” por ser una terminología desaconsejada^{180,181}, ya que los análisis que se hacen son con el objetivo de aportar evidencia a la validez y fiabilidad de las puntuaciones, y no de la escala en sí, en un contexto concreto. Las pruebas que puedan realizarse en el futuro con esta versión en español de la HSCL-25 seguirán aportando evidencia a la validez y la fiabilidad de sus puntuaciones en diferentes contextos. En otros países se han utilizado la HSCL-25 y la HSCL-10 en adolescentes con buenos resultados^{146,156–158,165}; aunque también sería interesante evaluarlo en ancianos, así como en grupos específicos de pacientes, por ejemplo.

En España, está extendido el uso de la versión española de la SCL-90-R como medida de malestar psicológico en estudios de diferentes problemáticas. En la bibliografía de los años comprendidos entre 2020 y 2023 se ha utilizado en pacientes con epilepsia resistente²⁸⁵, prisioneros con problemas académicos²⁸⁶, personas con disforia de género²⁸⁷, pacientes en tratamiento por ludopatía por apuestas relacionadas con eventos deportivos²⁸⁸, pacientes afectados de patologías concretas durante el confinamiento por la pandemia por la COVID-19²⁸⁹, así como para estudiar la relación de la salud mental de las mujeres con la exposición a áreas verdes²⁸⁹ o la relación con el cortisol y otras medidas de estrés²⁹⁰. Incluso un estudio ha utilizado únicamente las subescalas

de depresión y ansiedad de la SCL-90-R en español para correlacionar sus puntuaciones con la de la escala de ansiedad ante la muerte²⁹¹.

A la hora de decidir qué cuestionario utilizar, varios factores son relevantes y hay que tener en cuenta la población y el entorno sobre el que se va a administrar. COSMIN¹⁹² aconseja seleccionar el cuestionario a partir de revisiones sistemáticas disponibles o realizar una en caso de que no exista previamente, tal y como se realizó por el equipo investigador europeo con anterioridad al inicio de esta tesis doctoral. Se debe encontrar un balance adecuado entre las propiedades psicométricas y las características pragmáticas como la autoadministración, el número de ítems, la sencillez y la interpretabilidad de las puntuaciones²⁹². Es frecuente que los cuadernos de recogida de datos incluyan múltiples variables y escalas, lo que hace que sean extensos y puedan provocar fatiga en los participantes y sesgos que provoquen inexactitud en los datos²⁹³. Por este motivo es importante seleccionar escalas breves, a poder ser autoadministradas, cuyas puntuaciones sean válidas y fiables para medir el constructo deseado. La HSCL-25 contiene menos ítems que la SCL-90-R y puede ser una buena alternativa para la medida de la depresión en estudios que se realizan en España en diferentes contextos. También pueden considerarse la HSCL-10 y la HSCL-5 por ser más breves, sin renunciar en exceso a la fiabilidad y la validez. Sería adecuado realizar el análisis de la validez y la fiabilidad de la HSCL-25 en catalán y gallego, así como realizar la traducción y adaptación al euskera para poder posteriormente analizar también las propiedades psicométricas.

La HSCL-25 fue escogida tras una revisión sistemática teniendo en cuenta el uso concreto en la consulta de AP, como herramienta de ayuda al diagnóstico de la depresión para los MF. En las fases previas a esta tesis doctoral se evaluó la validez de contenido por parte de los MF integrantes del equipo europeo. En el consenso que condujo a la selección de la HSCL-25, se preguntó a los MF concretamente si les gustaba esta herramienta (el ítem concreto que debían evaluar fue *I like this tool*) para su uso en la consulta de AP. Esto es una muestra de su adecuada validez aparente. Asimismo, los pacientes confirmaron la comprensión y la aceptabilidad de la HSCL-25 en el estudio piloto y posteriormente se ha confirmado que es una herramienta adecuada en este contexto con los resultados publicados en esta tesis doctoral.

Se trata de una herramienta para evaluar la depresión en el momento presente, ya que pregunta por los síntomas presentados durante la semana previa, similar al DSM-5 que contempla en sus criterios diagnósticos los síntomas de las dos semanas anteriores¹⁸. Varias guías de práctica clínica ponen énfasis en la importancia de evaluar el riesgo de suicidio ante un paciente que padece depresión^{21,22}, por lo que es pertinente que los cuestionarios contengan un ítem que pregunte por este aspecto, como es el caso de la PHQ-9⁹⁶ y la HSCL-25, dada su gran relevancia

clínica y en la vida del paciente. La HSCL-10 y la HSCL-5 no incluyen el ítem sobre la ideación autolítica, por lo que se debe tener en cuenta en caso de obtener una puntuación por encima del punto de corte para recabar este dato a partir de la anamnesis posterior.

Respecto a las diferencias en las puntuaciones en cuestión de género, se encontraron de forma esperable puntuaciones más elevadas en las mujeres, esto fue así en todos los ítems, siendo estadísticamente significativo en 23 de los 25 ítems y en la puntuación total. Al analizar las respuestas, los ítems que discriminaron mejor entre hombres y mujeres fueron distintos. En los resultados presentados en un congreso nacional²⁶⁴ se analizó este aspecto en detalle. Los ítems 4 “Nerviosismo”, 11 “Siente que le falta energía” y 17 “Se siente triste” fueron los que discriminaron mejor de forma global, ordenados de mayor a menor ID. En las mujeres, fueron el ítem 4 “Nerviosismo”, el 14 “Pierde el interés sexual” y el 11 “Siente que le falta energía”. En los hombres fueron el ítem 4 “Nerviosismo”, el 17 “Se siente triste” y en tercer lugar hubo tres ítems con el mismo valor ID: 7 “Se siente tenso/a”, 11 “Siente que le falta energía” y 25 “Problemas para dormir”. Asimismo, el AUC de la curva ROC mostró un valor más elevado entre los hombres, coincidiendo con otras publicaciones^{139,144} y mostrando que la puntuación global clasifica mejor a los hombres respecto al diagnóstico de depresión. Por otro lado, el punto de corte óptimo que maximiza la sensibilidad y la especificidad fue 0,09 puntos más elevado en los hombres que en las mujeres; al contrario de lo reportado en otro estudio¹³⁹. Todos estos resultados llevan a cuestionarse si se debiesen utilizar puntos de corte diferentes en función del sexo, y, yendo más allá, si algunos ítems debieran ser diferentes entre hombres y mujeres o tener un peso ponderado a la hora de calcular la puntuación total. Otras escalas utilizadas en salud tienen en cuenta la cuestión del sexo en sus puntuaciones, como por ejemplo en la estimación del riesgo coronario²⁹⁴, o la escala CHA2DS2-VASC utilizada para la estimación del riesgo de tromboembolismo en fibrilación o flúter auricular no valvular²⁹⁵. La escala de cribado de consumo de alcohol AUDIT C contempla un punto de corte diferente en función del sexo²⁹⁶. Sin embargo, no se han encontrado escalas o PROMs para la evaluación de la depresión que consideren el sexo a la hora de calcular las puntuaciones. Se considera que son necesarios futuros estudios en esta línea para dar luz a este aspecto que puede ser relevante a la hora de utilizar los PROMs para medir la depresión en la práctica clínica.

También merece cuestionarse si se debiesen utilizar puntos de corte distintos en función del contexto, especialmente cabe preguntarse si el punto de corte en la evaluación de la depresión debe ser el mismo en AP, en los Centros de Salud Mental de Adultos (CSMA) y en los pacientes que hacen seguimiento hospitalario en el servicio de psiquiatría. La prevalencia de enfermedades es mayor en el hospital que en AP, ya que la AP supone un filtro que retiene la

mayoría de la población sana o de menor gravedad⁷⁸. Puesto que el uso de la HSCL-25 no es intrusivo es lógico pensar que el punto de corte en AP debe ser relativamente bajo con el fin de evitar falsos negativos, aun a riesgo de tener un mayor número de falsos positivos. La posterior anamnesis basada en los criterios de DSM-5 realizada por el MF deberá ser capaz de discriminar los falsos positivos de los verdaderos positivos. Los diagnósticos que se confirmen serán tratados por los profesionales de la propia AP, salvo en los casos de mayor gravedad o los que no respondan al tratamiento, que serán derivados a la atención especializada. Por el contrario, es habitual que otros especialistas diferentes de los MF evalúen a pacientes con depresión de mayor gravedad. Por todo ello, sería factible establecer un punto de corte diferente para los casos que puedan requerir esta atención especializada. En este sentido, existe una clara necesidad de estandarizar las medidas evaluadas en los estudios de valoración de las propiedades psicométricas para poder comparar las distintas herramientas y estudios en cada contexto. Esto permitirá a los investigadores y a los clínicos poder hacer recomendaciones sólidas para su uso como herramientas de evaluación de la depresión en el contexto específico de AP⁹⁸, que pueden ser diferentes a las del entorno hospitalario. En el futuro se podría evaluar la validez y la fiabilidad de la HSCL-25 en el seguimiento especializado por psiquiatras con el fin de comparar los resultados con los obtenidos en AP.

A partir del desarrollo de esta tesis doctoral, se dispone de la HSCL-25 en español, que ha mostrado validez y fiabilidad en sus puntuaciones para la evaluación de la depresión en AP.

8

CONCLUSIONES

8 Conclusiones

1. La HSCL-25 en español resultado de este proyecto fue conceptualmente equivalente a la versión original inglesa de la HSCL-25 tras el proceso de traducción y adaptación transcultural.
2. Hubo consenso de los médicos de familia sobre la comprensión y adecuación de la traducción de la HSCL-25 en español.
3. Los pacientes de AP mostraron buena comprensión de los ítems de la HSCL-25 en español y buena aceptabilidad de todos sus ítems y el conjunto de la escala.
4. Todos los ítems de la HSCL-25 puntuaron en el sentido esperado entre los grupos con puntuaciones totales extremas, lo que indica una buena discriminación de los ítems.
5. En el análisis de la validez de constructo, la HSCL-25 en español reprodujo la estructura interna de un solo factor y de dos factores (ansiedad y depresión) correlacionados, siendo el segundo modelo el que mostró mejor ajuste. Ambas dimensiones depresión y ansiedad estaban fuertemente correlacionadas entre sí y de forma positiva.
6. La puntuación de la HSCL-25 mostró buena correlación con la puntuación de la escala de síntomas de depresión PHQ-9, lo que reforzó la validez de constructo.
7. La HSCL-25 mostró buena validez de criterio frente al diagnóstico de depresión obtenido mediante la entrevista estructurada CIDI, con una probabilidad del 89% de clasificar de forma correcta a los pacientes.
8. La HSCL-25 mostró una elevada capacidad de descartar a personas que no padecen depresión para el punto de corte establecido en la versión original.
9. El ítem 17 “Se siente triste” fue el más consistente, es decir, el que más contribuyó a mejorar el resultado de la fiabilidad en todos los análisis.
10. La consistencia interna de la HSCL-25 fue excelente, lo que refleja una homogeneidad elevada entre sus ítems.
11. Las puntuaciones obtenidas en la HSCL-25 se mantuvieron estables en un intervalo de tiempo en el que se consideró que los pacientes permanecían sin cambios respecto al diagnóstico de la depresión.
12. Las puntuaciones en las versiones HSCL-10 y HSCL-5 presentaron una validez y fiabilidad adecuadas, por lo que pueden ser utilizadas en el entorno de AP con el objetivo de detectar y evaluar síntomas de depresión en los pacientes que acuden a la consulta.
13. Las puntuaciones de la HSCL-25 mostraron adecuada validez y fiabilidad para el cribado de la depresión en el contexto de la consulta de AP.

9

BIBLIOGRAFÍA

9 Bibliografía

1. Bolívar, Bonaventura. Investigación biomédica : Papel de la atención primaria. *Med Clin (Barc)*. 2006;126(16):614–5.
2. Hummers-Pradier E, Beyer M, Chevallier P, Eilat-Tsanani S, Lionis C, Peremans L, et al. The research agenda for general practice/family medicine and primary health care in Europe. Part 1. background and methodology 1. In: *European Journal of General Practice*. 2009.
3. Guiscafré H, Salmerón J. Historia natural y curso clínico de la enfermedad. In: McGraw-Hill, editor. *Epidemiología clínica*. 2ª. México: Interamericana; 1994. p. 117–30.
4. Violán Fors C, Odriozola GG, Zabaleta-del-Olmo E, Moral EG. La investigación en atención primaria como área de conocimiento. Informe SESPAS 2012. *Gac Sanit*. 2012;26(S):76–81.
5. Mant PD, Mar PC Del, Glasziou P, Knottnerus A, Wallace PP, Weel PC Van. The state of primary-care research. *Lancet*. 2004;364:1004–6.
6. Camarelles Guillem F. Putting prevention into practice. *Aten Primaria*. 2022 Oct 1;54.
7. Equipo de trabajo del Sistema Nacional de Salud M de S. Informe Anual del Sistema Nacional de Salud 2020-2021 [Internet]. Informes, estudios e investigación 2022, Ministerio de Sanidad. Madrid; 2022. Available from: https://www.sanidad.gob.es/estadEstudios/estadisticas/sisInfSanSNS/tablasEstadisticas/InfAnualSNS2020_21/INFORME_ANUAL_2020_21.pdf
8. Basora Gallisà J, Berenguera Ossó A. La financiación de la investigación en Atención Primaria de Salud, quid farem? 2022;15(3):130–1.
9. Martín Álvarez R, Tranche Iparraguirre S. Sobrediagnóstico, cuando las personas reciben un diagnóstico que no necesitan. Vol. 48, *Atención Primaria*. Elsevier Doyma; 2016. p. 619–20.
10. Fernández Fernández I. ¿Investigación en atención primaria? Existen dificultades. *Atención Primaria*. 2003;31(5):281–4.
11. Red de Investigación en Actividades Preventivas y Promoción de la Salud [Internet]. Available from: <https://www.rediapp.org/>
12. Red de Investigación en Cronicidad, Atención Primaria y Prevención y Promoción de la Salud [Internet]. Available from: <https://www.ricapps.es/>
13. Vich Pérez P. Razones para investigar en atención primaria. *Semergen*. 2007;33(3):111–2.
14. European General Practice Research Network [Internet]. Available from: <https://www.egprn.org/page/about>
15. World Organization of Family Doctors (WONCA). Wonca Global Family Doctor [Internet]. Available from: globalfamilydoctor.com/Home.aspx
16. Nabbe P, Le Reste J, Chiron B, Le Prielec A, Muñoz M, Claveria A, et al. The FPDM (family practice depression and Multimorbidity) Study: Project for systematic review of literature to find tools for depression diagnosis used in primary care. In: *Proceedings of the 72nd European General Practice Research Network Meeting*. Niza; 2011. p. 71.
17. Jameson J, Fauci A, Kasper D, Hauser S, Longo, DL Loscalzo J, Eds. *Trastornos psiquiátricos*. In: *Harrison Manual de Medicina*, 20e. McGraw Hill; 2020.
18. American Psychiatric Association. *Manual diagnóstico y estadístico de los trastornos*

mentales DSM-5. 5th ed. Madrid: Editorial Médica Panamericana; 2014. 1100 p.

19. World Health Organization. Depresión [Internet]. Available from: <https://www.who.int/es/news-room/fact-sheets/detail/depression>
20. Barragán N, Blanch J, Campiñez M. Trastornos de salud mental. In: Elsevier, editor. Atención primaria Problemas de salud en la consulta de Medicina de Familia-. 8ª. Barcelona; 2019. p. 897–925.
21. Álvarez Ariza M, Atienza Merino G, Ávila González MJ, González García A, Guitián Rodríguez D, de las Heras Liñero E, et al. Guía de Práctica Clínica sobre el Manejo de la Depresión en el Adulto [Internet]. 2014. Available from: <https://portal.guiasalud.es/gpc/depresion-adulto/>
22. Alonso J, Álvarez Ariza M, Atienza Merino G, Castro Bouzas M, Bugarín González R, de las Heras Liñero ME, et al. Revisión de la Guía de Práctica Clínica sobre el manejo de la Depresión en el adulto (2014) del Programa de GPC en el SNS [Internet]. 2023. Available from: https://portal.guiasalud.es/wp-content/uploads/2023/05/gpc_534_depresion_adulto_avaliat_revisio_2023.pdf
23. Roca M, Aragonès E. Abordaje compartido de la depresión. Consenso multidisciplinar [Internet]. Euromedice, Ediciones Médicas. Badalona; 2018. Available from: https://www.semfy.com/wp-content/uploads/2018/05/Consenso_depresion.pdf
24. World Health Organization (WHO). CIE-10 Chapter V Mental and behavioural disorders (F00-F99). 2019.
25. American Psychiatric Association. DSM-IV. Diagnostic and Statistical Manual of Mental Disorders. Washington; 1994.
26. Rodríguez-Testal JF, Cristina Senín-Calderón, Perona-Garcelán S. From DSM-IV-TR to DSM-5: Analysis of some changes. *Int J Clin Heal Psychol*. 2014;14(3):221–31.
27. Kessler R, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA J Am Med Assoc*. 2003;289(23):3095–105.
28. García Parajuá P, Mónica Magariños M. Guía de bolsillo de la clasificación CIE-10. Clasificación de los trastornos mentales y del comportamiento con glosarios y criterios diagnósticos de investigación. Madrid: Editorial Médica Panamericana; 2000.
29. Oliván-Blázquez B, Rubio-Aranda E, García-Sanz O, Magallón-Botaya R. Correlation between diagnosis of depression and symptoms. *Actas Españolas Psiquiatr*. 2016;44(2):55–63.
30. Mellado Fernández ML. Depresión. In: Guía de actuación en Atención Primaria. 5th ed. Barcelona: Sociedad Española de Medicina de Familia y Comunitaria; 2023. p. 478–86.
31. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, et al. The Long-term Natural History of the Weekly Symptomatic Status of Bipolar I...: Universitat Autònoma de Barcelona. *Arch Gen Psychiatry*. 2002;59(6):530–7.
32. Ferrari A. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry*. 2022;9:137–50.
33. World Health Organization (WHO). Depresión [Internet]. 2023. Available from: <https://www.who.int/es/news-room/fact-sheets/detail/depression>
34. Malhi GS, Mann JJ. Depression. *Lancet*. 2018;392(10161):2299–312.

35. Ferrari AJ, Somerville AJ, Baxter AJ, Norman R, Patten SB, Vos T, et al. Global variation in the prevalence and incidence of major depressive disorder: A systematic review of the epidemiological literature. *Psychol Med*. 2013;43(3):471–81.
36. Gogorcena MA. The adoption of the ICD-10-ES codification rules or how to make of necessity a virtue. *Rev Calid Asist*. 2017 Jan 1;32(1):1–5.
37. Cano Vindel A, Martín Salguero J, Mae Wood C, Dongil E, Miguel Latorre J. La Depresión En Atención Primaria: Prevalencia, Diagnóstico Y Tratamiento Depression in Primary Care: Prevalence, Diagnosis and Treatment. *Papeles del Psicólogo*. 2012;33(1):2–11.
38. King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, et al. Prevalence of common mental disorders in general practice attendees across Europe. *Br J Psychiatry*. 2008;192:362–7.
39. Serrano-Blanco A, Palao DJ, Luciano J V., Pinto-Meza A, Luján L, Fernández A, et al. Prevalence of mental disorders in primary care: Results from the diagnosis and treatment of mental disorders in primary care study (DASMAP). *Soc Psychiatry Psychiatr Epidemiol*. 2010;45:201–10.
40. Walrave R, Beerten SG, Mamouris P, Coteur K, Van Nuland M, Van Pottelbergh G, et al. Trends in the epidemiology of depression and comorbidities from 2000 to 2019 in Belgium. *BMC Prim Care*. 2022;23(1):1–12.
41. Portellano-Ortiz C, Garre-Olmo J, Calvó-Perxas L, Conde-Sala JL. Depression and associated variables in people over 50 years in Spain. *Rev Psiquiatr Salud Ment*. 2018;11(4):216–26.
42. Bromet E, Andrade L, Hwang I, Sampson N, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*. 2011;9(90).
43. Ministerio de Sanidad Consumo y Bienestar Social. Encuesta Nacional de Salud ENSE, España 2017. Serie informes monográficos. Salud mental. [Internet]. 2017. Available from: <https://www.mscbs.gob.es/estadEstudios/estadisticas/encuestaNacional/encuesta2017.htm>
44. Bentué-Martínez C, Rodrigues M, García-Foncillas López R, Llorente González JM, Zúñiga-Antón M. Socio-Economic Development and Mental Health: Case Study of the Spanish Region of Aragon (2010–20). *Front Psychol*. 2022;13(June):1–10.
45. Gabilondo A, Vilagut G, Pinto-Meza A, Haro JM, Alonso J. Comorbidity of major depressive episode and chronic physical conditions in Spain, a country with low prevalence of depression. *Gen Hosp Psychiatry*. 2012;34(5):510–7.
46. Yasaei R, Peterson E, Saadabadi A. Chronic Pain Syndrome. In: StatPearls [Internet]. reasure Island (FL): StatPearls Publishing; 2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK470523/>
47. Haro JM, Palacín C, Vilagut G, Martínez M, Bernal M, Luque I, et al. Prevalence of mental disorders and associated factors: Results from the ESEMeD-Spain study. *Med Clin (Barc)*. 2006;126(12):445–51.
48. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. *Lancet*. 2018;392(10157):1553–98.
49. Melchior M, Chastang JF, Head J, Goldberg M, Zins M, Nabi H, et al. Socioeconomic position predicts long-term depression trajectory: A 13-year follow-up of the GAZEL cohort study. *Mol Psychiatry*. 2013;18(1):112–21.

50. Bellón JÁ, Moreno-Küstner B, Torres-González F, Montón-Franco C, Gildegómez-Barragán MJ, Sánchez-Celaya M, et al. Predicting the onset and persistence of episodes of depression in primary health care. The predictD-Spain study: Methodology. *BMC Public Health*. 2008;8:1–9.
51. Woody CA, Ferrari AJ, Siskind DJ, Whiteford HA, Harris MG. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. *J Affect Disord*. 2017;219:86–92.
52. Chandra PS, Satyanarayana VA. Gender disadvantage and common mental disorders in women. *Int Rev Psychiatry*. 2010;22(5):513–24.
53. Abbafati C, Abbas KM, Abbasi-Kangevari M, Abd-Allah F, Abdelalim A, Abdollahi M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396(10258):1204–22.
54. World Health Organization. Global Health Estimates: Life expectancy and leading causes of death and disability [Internet]. Global health estimates: Leading causes of DALYs 2019. Available from: <https://www.who.int/data/gho/data/themes/mortality-and-global-health-estimates/global-health-estimates-leading-causes-of-dalys>
55. Cuijpers P, Stringaris A, Wolpert M. Treatment outcomes for depression: challenges and opportunities. *The Lancet Psychiatry*. 2020;7(11):925–7.
56. Lépine JP, Briley M. The increasing burden of depression. *Neuropsychiatr Dis Treat*. 2011;
57. Fazel S, Runeson B, Ropper AH. Suicide. *N Engl J Med*. 2020;382(3):266–74.
58. Naghavi M. Global, regional, and national burden of suicide mortality 1990 to 2016: systematic analysis for the Global Burden of Disease Study 2016. *BMJ*. 2019;364:l94.
59. Harmer B, Lee S, Duong T vi H, Saadabadi A. Suicidal Ideation. In: Treasure Island (FL): StatPearls Publishing, editor. StatPearls [Internet] [Internet]. 2023. Available from: <https://pubmed.ncbi.nlm.nih.gov/33351435/>
60. Fernández Domínguez MJ. Intento de suicidio. In: Guía de actuación en Atención Primaria. 5th ed. Barcelona: Sociedad Española de Medicina de Familia y Comunitaria; 2023. p. 487–92.
61. Coppens E, Van Audenhove C, Scheerder G, Arensman E, Coffey C, Costa S, et al. Public attitudes toward depression and help-seeking in four European countries baseline survey prior to the OSPI-Europe intervention. *J Affect Disord*. 2013;150(2):320–9.
62. Alonso J, Buron A, Bruffaerts R, He Y, Posada-Villa J, Lepine J-P, et al. Association of perceived stigma and mood and anxiety disorders: results from the World Mental Health Surveys. *Acta Psychiatr Scand*. 2008;118:305–14.
63. Lahera G, Andrade-González N, Gasull V, Pagés-Lluyot JR, Roca M. The public perception of depression in Spain. *An Sist Sanit Navar*. 2019;42(1):31–9.
64. López M, Laviana M, Fernández L, López A, Rodríguez AM, Aparicio A. La lucha contra el estigma y la discriminación en salud mental: Una estrategia compleja basada en la información disponible. Vol. 28, *Revista de la Asociación Española de Neuropsiquiatría*. 2008.
65. Arrarás JI, Manrique E. How depression and its treatment are perceived. Vol. 42, *Anales del Sistema Sanitario de Navarra*. Gobierno de Navarra; 2019. p. 5–8.
66. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al.

- Undertreatment of people with major depressive disorder in 21 countries. *Br J Psychiatry*. 2017;210(2):119–24.
67. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, et al. Treatment gap for anxiety disorders is global: Results of the World Mental Health Surveys in 21 countries. *Depress Anxiety*. 2018;35(3):195–208.
 68. Degenhardt L, Glantz M, Evans-Lacko S, Sadikova E, Sampson N, Thornicroft G, et al. Estimating treatment coverage for people with substance use disorders: an analysis of data from the World Mental Health Surveys. *World Psychiatry*. 2017;16(3):299–307.
 69. Park LT, Zarate CA. Depression in the Primary Care Setting. *N Engl J Med*. 2019;380(6):559–68.
 70. Latorre Postigo JM, Navarro Bravo B, Parra Delgado M, Salguero JM, Mae Wood C, Cano Vindel A. Evaluación e intervención de los problemas de Ansiedad y Depresión en Atención Primaria: un problema sin resolver [Assessment and intervention of Anxiety and Depression disorders in Primary Care: An unsolved problem]. *Rev Clínica Med Fam*. 2012;5(1):37–45.
 71. Lee ML, Yano EM, Wang MM, Simon BF, Rubenstein L V. What patient population does visit-based sampling in primary care settings represent? *Med Care*. 2002;40(9):761–70.
 72. Aragonès E, Piñol JL, Labad A, Folch S, Mèlich N. Detection and management of depressive disorders in primary care in Spain. *Int J Psychiatry Med*. 2004;34(4):331–43.
 73. Frank RG, Huskamp HA, Pincus HA. Aligning Incentives in the Treatment of Depression in Primary Care With Evidence-Based Practice. *Psychiatr Serv*. 2003;
 74. Mark TL, Levit KR, Buck JA. Psychotropic Drug Prescriptions by Medical Speciality. *Psychiatr Serv*. 2009;60(9):1167.
 75. Ahmedani BK, Stewart C, Simon GE, Lynch F, Lu CY, Waitzfelder BE, et al. Racial/Ethnic differences in health care visits made before suicide attempt across the United States. *Med Care*. 2015;53(5):430–5.
 76. Axencia de Avaliación de Tecnoloxías Sanitarias de Galicia., Plan de Calidad para el Sistema Nacional de Salud (España), España. Ministerio de Sanidad PS e I. Guía de práctica clínica de prevención y tratamiento de la conducta suicida : versión resumida. Ministerio de Sanidad; 2012.
 77. Agencia Gallega para la Gestión del Conocimiento en Salud A. Grupo de Trabajo de revisión de la Guía de Práctica Clínica de prevención y tratamiento de la conducta suicida 2012. Revisión de la guía de práctica clínica de prevención y tratamiento de la conducta suicida (2012) del programa de GPC en el SNS.Ministerio [Internet]. Ministerio de Sanidad. 2020. Available from: https://portal.guiasalud.es/wpcontent/uploads/2020/09/gpc_481_conducta_suicida_avaliat_revision.pdf.
 78. Dowrick C. Reasons to be cheerful? Reflections on GPs' responses to depression. *Br J Gen Pract*. 2009;59(566):636–7.
 79. Garrido-Elustondo S, Reneses B, Navalón A, Martín O, Ramos I, Fuentes M. Capacidad de detección de patología psiquiátrica por el médico de familia. *Aten Primaria*. 2016;48(7):449–57.
 80. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: Meta-analysis and meta-regression of predictors. *Br J Gen Pract*. 2011;61(583):72–80.

81. Thompson C, Ostler K, Peveler RC. Dimensional perspective on the recognition of depressive symptoms in primary care: The Hampshire depression project 3. *Prim Care Companion J Clin Psychiatry*. 2001;3(5):229.
82. Campagne DM. Diagnostica la depresión antes. *Med Fam Semer*. 2018;44(4):270–5.
83. Faisal-Cury A, Ziebold C, Rodrigues DM de O, Matijasevich A. Depression underdiagnosis: Prevalence and associated factors. A population-based study. *J Psychiatr Res*. 2022;151(April):157–65.
84. Pelletier L, O'Donnell S, Dykxhoorn J, McRae L, Patten SB. Under-diagnosis of mood disorders in Canada. *Epidemiol Psychiatr Sci*. 2017;26(4):414–23.
85. Iglesias García C, López García P, Ayuso Mateos JL, García JÁ, Bobes J. Detección de ansiedad y depresión en Atención Primaria: utilidad de 2 escalas breves adaptadas a los nuevos criterios CIE-11-AP. *Rev Psiquiatr Salud Ment*. 2021;14(4):196–201.
86. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. 2009;374(9690):609–19.
87. Dowrick C, Frances A. Medicalising unhappiness: New classification of depression risks more patients being put on drug treatment from which they will not benefit. *BMJ*. 2013;347(December):1–5.
88. Problemas del sistema nervioso, de conducta y de relación. Depresión Mayor. *Actual en Med Fam*. 2020;16(10):572–80.
89. Aragonès E, Piñol JL, Labad A. The overdiagnosis of depression in non-depressed patients in primary care. *Fam Pract*. 2006;23(3):363–8.
90. Siu AL, Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, Ebell M, et al. Screening for Depression in Adults US Preventive Services Task Force Recommendation Statement. *JAMA J Am Med Assoc*. 2016;315(4):380–7.
91. Maurer DM, Raymond TJ, Davis BN. Depression: Screening and diagnosis. *Am Fam Physician*. 2018;98(8):508–15.
92. Barry MJ, Nicholson WK, Silverstein M, Chelmow D, Coker TR, Davidson KW, et al. Screening for Depression and Suicide Risk in Adults. *JAMA*. 2023;329(23):2057.
93. National Institute for Health and Care Excellence (NICE). Depression in adults: screening, treatment and management. In: NICE Guideline, No 222 [Internet]. London; 2022. Available from: <https://www.nice.org.uk/guidance/ng222>
94. Jiao B, Rosen Z, Bellanger M, Belkin G, Muennig P. The cost-effectiveness of PHQ screening and collaborative care for depression in New York City. *PLoS One*. 2017;12(8):1–13.
95. McCarron RM, Vanderlip ER, Rado J. Depression. *Ann Intern Med*. 2016;165(7):ITC49–63.
96. Kroenke K. The Phq-9. *J Gen Intern Med*. 2001;16:606–13.
97. Reuland DS, Cherrington A, Watkins GS, Bradford DW, Blanco RA, Gaynes BN. Diagnostic accuracy of Spanish language depression-screening instruments. *Ann Fam Med*. 2009;7(5):455–62.
98. El-Den S, Chen TF, Gan YL, Wong E, O'Reilly CL. The psychometric properties of depression screening tools in primary healthcare settings: A systematic review. *J Affect Disord*. 2018;225:503–22.
99. Kroenke K, Spitzer RL, Williams JBW. The patient health questionnaire-2: Validity of a two-item depression screener. *Med Care*. 2003;41(11):1284–92.

100. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An Inventory for Measuring Depression. *Arch Gen Psychiatry*. 1961;4:561–71.
101. Sanz J, Garcia-Vera M, Espinosa R, Fortún M. Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 3. Propiedades psicométricas en pacientes con trastornos psicológicos. *Clínica y Salud*. 2005;16(2):121–42.
102. Sanz J, Perdigón AL, Vázquez C. Adaptación española del Inventario para la Depresión de Beck-II (BDI-II): 2. Propiedades psicométricas en población general. *Clin y Salud*. 2003;14(3):249–80.
103. Sanz J, Eugenia M, Vázquez C. The spanish adaptation of Beck's Depression Inventory-II (BDI-II): 2. Psychometric properties in the student population. Vol. 29, *Análisis y modificación de conducta*. 2003. p. 239–88.
104. Sanz J, García-Vera MP. Rendimiento diagnóstico y estructura factorial del inventario de depresión de beck-II (BDI-II). *An Psicol*. 2013;29(1):66–75.
105. Beck A, Steer R, Brown G. BDI-II. Beck Depression Inventory-Second Edition manual. San Antonio, TX: Psychological Corporation; 1996.
106. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56–63.
107. Ramos-Brieva J, Cordero Villafafila A. Validación de la versión castellana de la escala Hamilton para la depresión. *Actas Luso Esp Neurol Psiquiatr Cienc Afines*. 1986;14:324–34.
108. Bobes J, Bulbena A, Luque A, Dal-Ré R, Ballesteros J, Ibarra N. Evaluación psicométrica comparativa de las versiones en español de 6, 17 y 21 ítems de la Escala de valoración de Hamilton para la evaluación de la depresión. *Med Clin (Barc)*. 2003;120(18):693–700.
109. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–9.
110. Cunningham JL, Wernroth L, Von Knorring L, Berglund L, Ekselius L. Agreement between physicians' and patients' ratings on the Montgomery-Åsberg Depression Rating Scale. *J Affect Disord*. 2011;135(1–3):148–53.
111. Lobo A, Chamorro L, Luque A, Dal-Ré R, Badia X, Baró E, et al. Validation of the Spanish versions of the Montgomery-Asberg Depression and Hamilton Anxiety Rating Scales. *Med Clin (Barc)*. 2002;118(13):493–9.
112. Zung WWK. Self-Rating Depression. *Arch Gen Psychiatry*. 1965;
113. Conde V, Escribá J, Izquierdo J. Evaluación estadística y adaptación castellana de la escala autoaplicada para la depresión de Zung. *Arch Neurobiol (Madr)*. 1970;33:185–206.
114. Aragonès Benaiges E, Masdèu Montalà R, Cando Guasch G, Coll Borràs G. Diagnostic validity of Zung's self-rating depression scale on primary care patients. *Actas Españolas Psiquiatr*. 2001;29(5):310–6.
115. Radolff LS. The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas*. 1977;1(3):385–401.
116. Ruiz-Grosso P, Loret de Mola C, Vega-Dienstmaier JM, Arevalo JM, Chavez K, Vilela A, et al. Validation of the Spanish Center for Epidemiological Studies Depression and Zung Self-Rating Depression Scales: A Comparative Validation Study. *PLoS One*. 2012;7(10).
117. Vázquez FL, Blanco V, López M. An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. *Psychiatry Res*. 2007;149(1–3):247–52.

118. Yesavage J. Geriatric Depression scale. *Psychopharmacol Bull.* 1988;24(4):709–11.
119. Martínez de la Iglesia J, Onis M, Dueñas R, Albert C, Aguado C, Luque R. Versión española del cuestionario de Yesavage abreviado (GDS) para el despistaje de depresión en mayores de 65 años: adaptación y validación. *Rev Med Fam y comunitaria.* 2002;12(10):620–30.
120. Fernández-San Martín MI, Andrade C, Molina J, Muñoz PE, Carretero B, Rodríguez M, et al. Validation of the Spanish version of the geriatric depression scale (GDS) in primary care. *Int J Geriatr Psychiatry.* 2002;17(3):279–87.
121. Ortega Orcos R, Salinero Fort MA, Kazemzadeh Khajoui A, Vidal Aparicio S, De Dios Del Valle R. Validación de la versión española de 5 y 15 ítems de la Escala de Depresión Geriátrica en personas mayores en Atención Primaria. *Rev Clin Esp.* 2007;207(11):559–62.
122. Lovibond S, Lovibond P. *Manual for the Depression Anxiety & Stress Scales.* 2nd ed. Sydney: Psychology Foundation of Australia; 1995.
123. Daza P, Novy DM, Stanley MA, Averill P. The depression anxiety stress scale-21: Spanish translation and validation with a Hispanic sample. *J Psychopathol Behav Assess.* 2002;24(3):195–205.
124. Bados A, Solanas A, Andrés R. Psychometric properties of the Spanish version of Depression, Anxiety and Stress Scales (DASS). *Psicothema.* 2005;17(4):679–83.
125. Zigmond AS, Snalth RP. The Hospital Anxiety and Depression Scale. *Acta psychiatr. scand.* [revista en Internet] 2014 [acceso 28 de noviembre de 2019]; 64(5): 361-370. *Acta Psychiatr Scand.* 1983;67(6):361–70.
126. Carmen Terol-Cantero M, Cabrera-Perona V, Martín-Aragón M. Hospital Anxiety and Depression Scale (HADS) review in Spanish samples. *An Psicol.* 2015;31(2):494–503.
127. Terol M, López-Roig S, Rodríguez-Marín J, Martín-Aragón M, Pastor M, Reig M. Propiedades psicométricas de la Escala Hospitalaria de Ansiedad y Estrés (HAD) en población española. *Ansiedad y Estrés.* 2007;13(2–3):163–76.
128. Nabbe P, Le Reste JY, Guillou-Landreat M, Beck-Robert E, Assenova R, Lazic D, et al. One consensual depression diagnosis tool to serve many countries: A challenge! A RAND/UCLA methodology. *BMC Res Notes.* 2018;11:4.
129. Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, et al. Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review. *Eur Psychiatry.* 2017;39:99–105.
130. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth E, Covi L. The Hopkins Symptom Checklist (HSCL): A Self-Report Symptom Inventory. *Behav Sci.* 1974;19:1–15.
131. Derogatis LR, Rickels K, Rock AF. The SCL 90 and the MMPI: a step in the validation of a new self report scale. *Br J Psychiatry.* 1976;128(3):280–9.
132. De las Cuevas C, González de Rivera J, Benitez M, Monterrey A, Rodriguez Pulido F, Gracia Marco R. Análisis factorial de la versión española del SCL-90-R en la población general. *An Psiquitria.* 1991;7(3):93–6.
133. Müller JM, Postert C, Beyer T, Furniss T, Achtergarde S. Comparison of eleven short versions of the symptom checklist 90-Revised (SCL-90-R) for use in the assessment of general psychopathology. *J Psychopathol Behav Assess.* 2009 Jun;32(2):246–54.
134. Hesbacher P, Rickels K, Morris R, Newman H, Rosenfeld H. Psychiatric illness in family practice. *J Clin Psychiatry.* 1980;41(1):6–10.
135. Winokur A, Winokur DF, Rickels K, Cox DS. Symptoms of emotional distress in a family

- planning service: Stability over a four-week period. *Br J Psychiatry*. 1984;144:395–9.
136. Veijola J, Jokelainen J, Läsky K, Kantojärvi L, Kokkonen P, Järvelin MR, et al. The Hopkins Symptom Checklist-25 in screening DSM-III-R axis-I disorders. *Nord J Psychiatry*. 2003;57:119–23.
 137. Glaesmer H, Braehler E, Grande G, Hinz A, Petermann F, Rompell M. The German Version of the Hopkins Symptoms Checklist-25 (HSCL-25) - Factorial structure, psychometric properties, and population-based norms. *Compr Psychiatry*. 2014;55:396–403.
 138. Nettelblatt P, Hansson L, Stefansson CG, Borgquist L, Nordström G. Test characteristics of the Hopkins Symptom Check List-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Soc Psychiatry Psychiatr Epidemiol*. 1993;28:130–3.
 139. Sandanger I, Moum T, Ingebrigtsen G, Dalgard OS, Sørensen T, Bruusgaard D. Concordance between symptom screening and diagnostic procedure: The Hopkins Symptom Checklist-25 and the Composite International Diagnostic Interview I. *Soc Psychiatry Psychiatr Epidemiol*. 1998;33:345–54.
 140. Nabbe P, Le Reste JY, Guillou-Landreat M, Gatineau F, Le Floch B, Montier T, et al. The French version of the HSCL-25 has now been validated for use in primary care. *PLoS One*. 2019 Apr 1;14(4):e0214804.
 141. Lee B, Kaaya SF, Mbwambo JK, Smith-Fawzi MC, Leshabari MT. Detecting depressive disorder with the Hopkins Symptom Checklist-25 in Tanzania. *Int J Soc Psychiatry*. 2008;54(1):7–20.
 142. Nils Johan Lavik, Edvard Hauff Ø. The use of self-reports in psychiatric studies of traumatized refugees: Validation and analysis of HSCL-25. *Nord J Psychiatry*. 1999 Jan 12;53(1):17–20.
 143. Mollica R, Wyshak G, de Marneffe D, Khuon F, Lavelle J. Indochinese versions of the Hopkins Symptom Checklist-25: a screening instrument for the psychiatric care of refugees. *Am J Psychiatry*. 1987 Apr;144(4):497–500.
 144. Mattisson C, Bogren M, Horstmann V. Correspondence between clinical diagnoses of depressive and anxiety disorders and diagnostic screening via the Hopkins Symptom Check List-25 in the Lundby Study. *Nord J Psychiatry*. 2013;67(3):204–13.
 145. Fröjd K, Håkansson A, Karlsson I. The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. *Int J Geriatr Psychiatry*. 2004;19:386–90.
 146. Jakobsen M, Meyer DeMott MA, Heir T. Validity of screening for psychiatric disorders in unaccompanied minor asylum seekers: Use of computer-based assessment. *Transcult Psychiatry*. 2017 Dec 1;54(5–6):611–25.
 147. Skogen JC, Øverland S, Smith ORF, Aarø LE. The factor structure of the Hopkins Symptoms Checklist (HSCL-25) in a student population: A cautionary tale. *Scand J Public Health*. 2017 Jun 1;45(4):357–65.
 148. Ashaba S, Kakuhikire B, Vořechovská D, Perkins JM, Cooper-Vince CE, Maling S, et al. Reliability, Validity, and Factor Structure of the Hopkins Symptom Checklist-25: Population-Based Study of Persons Living with HIV in Rural Uganda. *AIDS Behav*. 2018 May 1;22(5):1467–74.
 149. Bantjes J, Kagee A, Saal W. The Utility of the Hopkins Symptom Checklist as a Trans-Diagnostic Screening Instrument for Common Mental Disorders Among Persons Seeking

- HIV Testing. *AIDS Behav.* 2020 Feb 1;24(2):629–36.
150. Al-Turkait FA, Ohaeri JU, El-Abbasi AHM, Naguy A. Relationship between symptoms of anxiety and depression in a sample of arab college students using the Hopkins Symptom Checklist 25. *Psychopathology.* 2011 May;44(4):230–41.
 151. Lundin A, Hallgren M, Forsell Y. The validity of the symptom checklist depression and anxiety subscales: A general population study in Sweden. *J Affect Disord.* 2015;183:247–52.
 152. Tambs K, Røysamb E. Selection of questions to short-form versions of original psychometric instruments in MoBa *. *Nor Epidemiol.* 2014;24(2):195–201.
 153. Tambs K, Moum T. How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatr Scand.* 1993;87(5):364–7.
 154. Strand BH, Dalgard OS, Tambs K, Rognerud M. Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nord J Psychiatry.* 2003;57(2):113–8.
 155. Schmalbach B, Zenger M, Tibubos AN, Kliem S, Petrowski K, Brähler E. Psychometric Properties of Two Brief Versions of the Hopkins Symptom Checklist: HSCL-5 and HSCL-10. *Assessment.* 2019 Mar 1;28(2):617–31.
 156. Sirpal MK, Haugen W, Sparle K, Haavet OR. Validation study of HSCL-10, HSCL-6, WHO-5 and 3-key questions in 14-16 year ethnic minority adolescents. *BMC Fam Pract.* 2016 Jan 27;17(1).
 157. Sjøberg Finbråten H, Kleppang AL, Steigen AM. Using the Rasch measurement theory to assess the psychometric properties of the Hopkins Symptom Checklist-10 in adolescents. *Health Qual Life Outcomes.* 2021;19.
 158. Kleppang AL, Hagquist C. The psychometric properties of the Hopkins Symptom Checklist-10: A Rasch analysis based on adolescent data from Norway. *Fam Pract.* 2016 Dec 1;33(6):740–5.
 159. Ljoså CH, Tyssen R, Lau B. Mental distress among shift workers in Norwegian offshore petroleum industry - Relative influence of individual and psychosocial work factors. *Scand J Work Environ Heal.* 2011;37(6):551–5.
 160. Lien IA, Bolstad I, Lien L, Bramness JG. Screening for depression in patients in treatment for alcohol use disorder using the Beck Depression Inventory-II and the Hopkins Symptom Checklist-10. *Psychiatry Res.* 2022 Feb 1;308.
 161. Nesterko Y, Jäckle D, Friedrich M, Holzapfel L, Glaesmer H. Factors predicting symptoms of somatization, depression, anxiety, post-traumatic stress disorder, self-rated mental and physical health among recently arrived refugees in Germany. *Confl Health.* 2020;14(44).
 162. Harris SM, Sandal GM. COVID-19 and psychological distress in Norway: The role of trust in the healthcare system. *Scand J Public Health.* 2021;49:96–103.
 163. Karlsson Stafseth S, Skogstad L, Raeder J, Strand Hovland I, Hovde H, Ekeberg Ø, et al. Symptoms of Anxiety, Depression, and Post-Traumatic Stress Disorder in Health Care Personnel in Norwegian ICUs during the First Wave of the COVID-19 Pandemic, a Prospective, Observational Cross-Sectional Study. *Int J Environ Res Public Heal.* 2022;
 164. Frounfelker RL, Li ZY, Santavicca T, Miconi D, Rousseau C. Latent class analysis of COVID-19 experiences, social distancing, and mental health. *Am J Orthopsychiatry.* 2022;92(1):121–32.

165. Haavet OR, Sirpal MK, Haugen W, Christensen KS. Diagnosis of depressed young people in primary health care—a validation of HSCL-10. *Fam Pract.* 2011 Apr 1;28(2):233–7.
166. Sjøgaard AJ, Bjelland I, Tell GS, Røysamb E. A comparison of the CONOR Mental Health Index to the HSCL-10 and HADS. *Nor Epidemiol.* 2003;13(2):279–84.
167. Kessler RC, Bedirhan Üstün T. The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res.* 2004;13(2):93–121.
168. Wing J, Babor T, Brugha T, Burke J, Cooper J, Giel R, et al. SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry.* 1990;47(6):589–93.
169. Robins LN. The Composite International Diagnostic Interview. *Arch Gen Psychiatry.* 1988;45(12):1069.
170. Haro J, Arbabzadeh-Bouchez S, Brugha TS, De Girolamo G, Guyer ME, Jin R, et al. Concordance of the Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO World Mental Health Surveys. *Int J Methods Psychiatr Res.* 2006;15(4):167–180.
171. Harkness J, Pennell B-E, Villar A, Gebler N, Aguilar-Gaxiola S, Bilgen I. Translation procedures and translation assessment in the World Mental Health Survey Initiative. In: *The WHO World Mental Health Surveys: global perspectives on the epidemiology of mental disorders.* New York: Cambridge University Press, World Health Organization; 2008. p. 91–113.
172. Alonso J, Ferrer M, Romera B, Vilagut G, Angermeyer M, Bernert S, et al. The European Study of the Epidemiology of Mental Disorders (ESEMeD / MHEDEA 2000) Project : rationale and methods. *Int J Methods Psychiatr Res.* 2003;11(2):55–67.
173. Wittchen HU. Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *J Psychiatr Res.* 1994;28(1):57–84.
174. Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier D. Cross-cultural feasibility, reliability and sources of variance of the composite international diagnostic interview (CIDI). *Br J Psychiatry.* 1991;159(NOV.):645–53.
175. Wu Y, Levis B, Ioannidis JPA, Benedetti A, Thombs BD, Sun Y, et al. Probability of major depression classification based on the SCID, CIDI, and MINI diagnostic interviews: A synthesis of three individual participant data meta-analyses. *Psychother Psychosom.* 2020;90(1):28–40.
176. Maske UE, Busch MA, Jacobi F, Beesdo-Baum K, Seiffert I, Wittchen HU, et al. Current major depressive syndrome measured with the Patient Health Questionnaire-9 (PHQ-9) and the Composite International Diagnostic Interview (CIDI): Results from a cross-sectional population-based study of adults in Germany. *BMC Psychiatry.* 2015;15(1):1–7.
177. Robison J, Gruman C, Gaztambide S, Blank K. Screening for depression in middle-aged and older Puerto Rican primary care patients. *Journals Gerontol - Ser A Biol Sci Med Sci.* 2002;57(5):308–14.
178. Errazuriz A, Beltrán R, Torres R, Passi-Solar A. The Validity and Reliability of the PHQ-9 and PHQ-2 on Screening for Major Depression in Spanish Speaking Immigrants in Chile: A Cross-Sectional Study. *Int J Environ Res Public Health.* 2022;19(21):1–13.
179. Gómez-Gómez I, Benítez I, Bellón J, Moreno-Peral P, Oliván-Blázquez B, Clavería A, et al. Utility of PHQ-2, PHQ-8 and PHQ-9 for detecting major depression in primary health care: a validation study in Spain. *Psychol Med.* 2022;1–11.

180. Viladrich C, Doval E, Penelo E. *Medición: Fiabilidad y Validez*. 12th ed. Bellaterra: Laboratori d'Estadística Aplicada i de Modelització (Universitat Autònoma de Barcelona); 2019.
181. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine. Practical Guides to Biostatistics and Epidemiology*. Cambridge: Cambridge University Press, World Health Organization; 2018. 338 p.
182. Doval E, Viladrich C. *Desarrollo y adaptación de cuestionarios en el ámbito de la Salud*. 12th ed. Bellaterra: Laboratori d'Estadística Aplicada i de Modelització (Universitat Autònoma de Barcelona); 2019.
183. Alonso J. La medida de la calidad de vida relacionada con la salud en la investigación y la práctica clínica. *Gac Sanit*. 2000;14(2):163–7.
184. Herdman M, Badia X, Berra S. EuroQol-5D: a simple alternative for measuring health-related quality of life in primary care. *Aten Primaria*. 2001;28(6):425–30.
185. Ferrer-Peña R, Gil-Martínez A, Pardo-Montero J, Jiménez-Penick V, Gallego-Izquierdo T, La Touche R. Adaptación y validación de la Escala de gradación del dolor crónico al español. *Reumatol Clin*. 2016;12(3):130–8.
186. Ordóñez-Carrasco JL, Cuadrado Guirado I, Rojas Tejada A. Scale of psychological pain: Spanish adaptation of the Psychache Scale in young adults. *Rev Psiquiatr Salud Ment*. 2022;15(3):196–204.
187. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety Disorder. *Arch Intern Med*. 2006 May 22;166(10):1092.
188. Montón C, Pérez Echeverría MJ, Campos R, García Campayo J, Lobo A. Anxiety scales and Goldberg's depression: an efficient interview guide for the detection of psychologic distress. *Aten Primaria*. 1993;12(6):345–9.
189. Cueli Arce M, Santibañez M, Sarabia C, Paras-Bravo P, Gomez M, Alconero-Camarero AR. Transcultural adaptation of the revised caregiving appraisal scale (RCAS) in the Spanish population. *Int J Older People Nurs*. 2023;18(1):1–12.
190. Hernández-Padilla JM, Correa-Casado M, Granero-Molina J, Cortés-Rodríguez AE, Matarín-Jiménez TM, Fernández-Sola C. Psychometric evaluation and cultural adaptation of the Spanish version of the Scale for End-of Life Caregiving Appraisal. *Palliat Support Care*. 2019;17(3):314–21.
191. de la Revilla-Ahumada L, Bailón E, de Dios Luna E, Delgado J, Prados M, Fleitas M. Validation of a functional social support scale for use in the family doctor's office. *Atención primaria*. 1991;8(9):688–92.
192. COSMIN [Internet]. Available from: <https://www.cosmin.nl/>
193. Ortiz-Gutiérrez S, Cruz-Avelar A. Translation and Cross-Cultural Adaptation of Health Assessment Tools. *Actas Dermosifiliogr*. 2018 Apr 1;109(3):202–6.
194. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract* (Print Ed. Accepted for publ... 2011;17(2):268–74.
195. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-Reported Outcomes to Support Medical Product Labeling Claims: FDA Perspective. *Value Heal*. 2007 Nov;10:S125–37.
196. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN

- guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res.* 2018 May 1;27(5):1147–57.
197. Borsboom D, Mellenbergh GJ, Van Heerden J. The Theoretical Status of Latent Variables. Vol. 110, *Psychological Review*. 2003. p. 203–19.
 198. Park CG. Implementing alternative estimation methods to test the construct validity of Likert-scale instruments. *Korean J Women Heal Nurs.* 2023 Jun 30;29(2):85–90.
 199. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine (Phila Pa 1976).* 2000;25(24):3186–91.
 200. María Valderas J, Ferrer M, Alonso J. Instrumentos de medida de calidad de vida relacionada con la salud y de otros resultados percibidos por los pacientes. *Med Clin (Barc).* 2005;125:56–60.
 201. Wild D, Grove A, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaption. *Value Heal.* 2005;8(2):94–104.
 202. Acquadro C, Patrick DL, Eremenco S, Martin ML, Kuliś D, Correia H, et al. Emerging good practices for Translatability Assessment (TA) of Patient-Reported Outcome (PRO) measures. *J Patient-Reported Outcomes.* 2018;2(8).
 203. Muñiz J, Elosua P, Hambleton RK. Directrices para la traducción y adaptación de los tests: Segunda edición. Vol. 25, *Psicothema.* 2013. p. 151–7.
 204. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: How to decide its appropriateness. *World J Methodol.* 2021 Jul 20;11(4):116–29.
 205. Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. *BMC Med Res Methodol.* 2020 Dec 1;20(1).
 206. Powell C. The Delphi technique: myths and realities. *J Adv Nurs.* Submitted. 2003;41(4):376–82.
 207. Green B, Jones M, Hughes D, Williams A. Applying the Delphi technique in a study of GPs' information requirements. *Heal Soc Care Community.* 1999;7(3):198–205.
 208. Steurer J. The Delphi method: an efficient procedure to generate knowledge. *Skeletal Radiol.* 2011;40(8):959–61.
 209. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010 Jul;63(7):737–45.
 210. Nunnally JC. *Psychometric Theory.* 2nd ed. New York: McGraw-Hill; 1978.
 211. Strauss ME, Smith GT. Construct validity: Advances in theory and methodology. Vol. 5, *Annual Review of Clinical Psychology.* 2009. p. 1–25.
 212. Connell J, Carlton J, Grundy A, Taylor Buck E, Keetharuth AD, Ricketts T, et al. The importance of content and face validity in instrument development: lessons learnt from service users when developing the Recovering Quality of Life measure (ReQoL). *Qual Life Res.* 2018 Jul 1;27(7):1893–902.

213. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, De Vet HC, et al. COSMIN Study Design checklist for Patient-reported outcome measurement instruments [Internet]. Amsterdam; 2019. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf
214. Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *Can J Emerg Med*. 2006;8(1):19–20.
215. Polo T, Miot H. Aplicações da curva ROC em estudos clínicos e experimentais. *J Vasc Bras*. 2020;19:e20200186.
216. Martínez Pérez JA, Pérez Martín PS. La curva ROC. *Med Fam Semer*. 2023;49(1).
217. Herrero J. El Análisis Factorial Confirmatorio en el estudio de la Estructura y Estabilidad de los Instrumentos de Evaluación: Un ejemplo con el Cuestionario de Autoestima CA-14. *Psychosoc Interv*. 2010;19(3):289–300.
218. Park JH, Kim JI. Practical Consideration of Factor Analysis for the Assessment of Construct Validity. Vol. 51, *Journal of Korean Academy of Nursing*. Korean Society of Nursing Science; 2021. p. 643–7.
219. Muñiz J, Fonseca-Pedrero E. Diez pasos para la construcción de un test. Vol. 31, *Psicothema*. 2019. p. 7–16.
220. Teresi JA. Different Approaches to Differential Item Functioning in Health Applications Advantages, Disadvantages and Some Neglected Topics. *Med Care*. 2006;44(11):152–70.
221. Muñiz J. Teoría clásica de los tests. Madrid: Pirámide; 1992.
222. Streiner D, Norman G. Health measurement scales: A practical guide to their development and use. 4th ed. Oxford: Oxford University Press; 2008.
223. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007 Jan;60(1):34–42.
224. Nunnally JC, Berstein I. Teoría psicométrica. 3rd ed. México: McGraw-Hill; 1995.
225. American Psychiatric Association, National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 2014.
226. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychol Methods*. 2002;7(2):147–77.
227. Lizán L. Características métricas del Cuestionario de Calidad de Vida Profesional. *Gac Sanit*. 2004;18(6):489–489.
228. Fitch K, Bernstein María SJ, Aguilar D, Burnand B, Lacalle JR, Lázaro P, et al. The RAND/UCLA Appropriateness Method User’s Manual [Internet]. 2001. Available from: <http://www.rand.org>
229. Roncone R, Gone J, Barlati S, Nabbe P, Reste LJ, Lazic KD, et al. Nine Forward-Backward Translations of the Hopkins Symptom Checklist-25 With Cultural Checks. *Front Psychiatry*. 2021;1:688154.
230. Zabaleta-Del-Olmo E, Pombo H, Pons-Vigués M, Casajuana-Closas M, Pujol-Ribera E, López-Jiménez T, et al. Complex multiple risk intervention to promote healthy behaviours in people between 45 to 75 years attended in primary health care (EIRA study): Study protocol for a hybrid trial. *BMC Public Health*. 2018;18:874.

231. Bully P, Sanchez A, Grandes G, Pombo H, Arietalenizbeaskoa MS, Arce V, et al. Metric properties of the “prescribe healthy life” screening questionnaire to detect healthy behaviors: a cross-sectional pilot study. *BMC Public Health*. 2016;16(1):1–11.
232. Marshall AL, Smith BJ, Bauman AE, Kaur S. Reliability and validity of a brief physical activity assessment for use by family doctors. *Br J Sports Med*. 2005;39(5):294–7.
233. Puig-Ribera A, Martín-Cantera C, Puigdomenech E, Real J, Romaguera M, Magdalena-Belio JF, et al. Screening physical activity in family practice: Validity of the Spanish version of a brief physical activity questionnaire. *PLoS One*. 2015;10(9):1–16.
234. Clavería A, Rodríguez-Barragán M, Fernández-San-Martín MI, Nabbe P, Le Reste JY, Miguéns-Blanco I, et al. Translation and cross-cultural adaptation into Spanish, Catalan and Galician of the Hopkins Symptom Checklist-25 to identify depression in Primary Care. *Aten Primaria*. 2020;52(8):539–47.
235. World Health Organization (WHO). Composite International Diagnostic Interview (CIDI). version 2.1. *Jurnal Penelitian Pendidikan Guru Sekolah Dasar*. Geneva, Switzerland: World Health Organization; 1997.
236. Jackson D, Gillaspay J, Purc-Stephenson R. Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations. *Psychol Methods*. 2009;14(1):6–23.
237. Akoglu H. User’s guide to correlation coefficients. *Turkish J Emerg Med*. 2018;18(3):91–3.
238. Brehaut E, Neupane D, Levis B, Wu Y, Sun Y, Ioannidis JPA, et al. “Optimal” cutoff selection in studies of depression screening tool accuracy using the PHQ-9, EPDS, or HADS-D: A meta-research study. *Int J Methods Psychiatr Res*. 2022;32(3):e1956.
239. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biometrical J*. 2005;47(4):458–72.
240. Wang H, Hang Kwok SW, Liu X-L, Wang T, Bressington D, Shen Y, et al. Quality of Life Patient/Cancer Survivor Version in Chinese cancer survivors: a validation study. *Asia-Pacific J Oncol Nurs*. 2023;10(8):100255.
241. Huang EYZ, Cheung J, Liu JYW, Kwan RYC, Lam SC. Groningen Frailty Indicator–Chinese (GFI-C) for pre-frailty and frailty assessment among older people living in communities: psychometric properties and diagnostic accuracy. *BMC Geriatr*. 2022 Dec 1;22(1).
242. Declaración de Helsinki de la AMM - Principios éticos para las investigaciones médicas en seres humanos. WMA - The World Medical Association; 2013.
243. Boletín Oficial del Estado. Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal [Internet]. Available from: <https://www.boe.es/eli/es/lo/2018/12/05/3/con>
244. Boletín Oficial del Estado. Ley Orgánica 3/2018, De 5 De Diciembre, De Protección De Datos Personales Y Garantía De Los Derechos Digitales.
245. Cheng ATA. Case definition and culture: Are people all the same? *Br J Psychiatry*. 2001;178(JULY):1–3.
246. Kessler R, Heeringa S, Pennell B, Mneimneh Z, Chardoul S, Zaslavsky A. Methods of the World Mental Health Surveys [Internet]. 2005. Available from: <https://www.hcp.med.harvard.edu/wmh/>
247. Varela-Ruiz M, Díaz-Bravo L, García-Durán R. Descripción y usos del método Delphi en investigación del área de la salud. *Investig en Educación Médica*. 2012;1(2):90–5.
248. Castilla-Polo F, Ruiz-Rodríguez M del C, Delgado-Marfil C. La técnica Delphi para la

- validación de escalas de medida: las variables innovación y reputación dentro de almazaras cooperativas. *REVESCO Rev Estud Coop.* 2020;136(136):e71852.
249. Fernández-García O, Gil-Llario MD, Ballester-Arnal R. Construction of a Form for Users of the Child Welfare System Based on the Delphi Method. *Children.* 2023 Jun 7;10(6):1026.
 250. López-Bustos-de-las-heras R, Cano-De-la-cuerda R, De-La-Llave-rincón AI, Guerrero-Blázquez A, Plaza-Manzano G, Martín-Casas P. Traducción y adaptación transcultural de la Toe Walking Tool: Herramienta para el cribado de la marcha de puntillas. *Rev Neurol.* 2019;69(3):99–108.
 251. Betancurth Loaiza D, Vélez Álvarez C, Jurado Vargas L. Validación de contenido y adaptación del cuestionario Fantastico por técnica Delphi. *Salud Uninorte.* 2015;31(2):214–27.
 252. Collins D. Pretesting Survey Instruments: An Overview of Cognitive Methods. *Qual life Res.* 2003;12(3):229–38.
 253. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol.* 2014;67(4):401–9.
 254. Onteniente-Gomis MM, Ortiz-Romero PL, Tous Romero F, Salamanca Castro AB, Ortiz de Frutos FJ. Traducción, adaptación transcultural, correlación con otros PRO y validación del cuestionario RECAP sobre el grado de control de la dermatitis atópica. *Actas Dermosifiliogr.* 2023;114(6):488–93.
 255. Salmerón EV, Romero EM, Nogueira CD, Nogueira CD, Romero EM. Traducción y adaptación transcultural al español del cuestionario de calidad de vida y úlceras por presión PU-QOL. *Rev Esp Salud Publica.* 2023;97:e202304032.
 256. Romano-Andrioni B, Martín-Lleixà A, Carrasco-Serrano M, Barba S, Quintela M, Pérez I, et al. Nueva herramienta de cribado nutricional para pacientes hospitalizados con enfermedad renal crónica: traducción, adaptación transcultural del iNUT Renal al castellano y comparación con cuestionarios clásicos. *Nutr Hosp.* 2023;
 257. Lopiz Y, Garríguez-Pérez D, Scarano-Pereira JP, Fuentes Ferrer ME, Arvinus C, Ponz V, et al. The Spanish version of the Constant-Murley Shoulder Score: translation, cultural adaptation, and validity. *J Shoulder Elb Surg.* 2023;32(7):1348–56.
 258. Garcimartin P, Pardo-Cladellas Y, Verdú-Rotellar JM, Delgado-Hito P, Astals-Vizcaino M, Comin-Colet J. Transcultural adaptation into Spanish of the Patient empowerment in long-term conditions questionnaire. *Aten Primaria.* 2019 Jan 1;51(1):24–31.
 259. González-Bueno J, Calvo-Cidoncha E, Sevilla-Sánchez D, Espauella-Panicot J, Codina-Jané C, Santos-Ramos B. Spanish translation and cross-cultural adaptation of the ARMS-scale for measuring medication adherence in poly pathological patients. *Aten Primaria.* 2017 Oct 1;49(8):459–64.
 260. REAL ACADEMIA ESPAÑOLA. Diccionario de la lengua española [Internet]. 23ª ed., [versión 23.6 en línea]. [cited 2023 Aug 28]. Available from: <https://dle.rae.es>
 261. Díez-Quevedo C, Rangil T, Sanchez-Planell L, Kroenke K, Spitzer R. Validation and utility of the patient health questionnaire in diagnosing mental disorders in 1003 general hospital Spanish inpatients. *Psychosom Med.* 2001;63(4):679–86.
 262. Liu X, Feng Z, Galling B, Qi N, Zhu X, Xiao L, et al. Gender specific sexual dysfunction in patients with depression. *Front Psychiatry.* 2023;14(June):1–8.
 263. Ross R, Hess R, Pittman C, Croasmun A, Baird M. Validation of the Hopkins Symptom

- Checklist-25/Nepali Version Among Bhutanese Refugees in the United States. *J Nurs Meas.* 2022;30(1):168–78.
264. Rodríguez-Barragán M, Peguero-Rodríguez E, Fernández-San-Martín MI, Clavería-Fontán A. Diferencias de género en la escala Hopkins Symptom Checklist-25 de depresión. In: XLI Congreso de la semFYC. Palma de Mallorca; 2021.
 265. Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. Vol. 11, *Structural Equation Modeling.* Routledge; 2004. p. 320–41.
 266. van Prooijen JW, van der Kloot WA. Confirmatory analysis of exploratively obtained factor structures. *Educ Psychol Meas.* 2001;61(5):777–92.
 267. Hale WW, Raaijmakers QAW, Muris P, Van Hoof A, Meeus WHJ. One factor or two parallel processes? Comorbidity and development of adolescent anxiety and depressive disorder symptoms. *J Child Psychol Psychiatry.* 2009;50(10):1218–26.
 268. Bendixen AB, Engedal K. Anxiety among older psychiatric patients: a hidden comorbidity? *Aging Ment Heal.* 2016 Nov 1;20(11):1131–8.
 269. Martinussen LJ, Šaltyté Benth J, Almdahl IS, Borza T, Selbæk G, Mcpherson B, et al. The effect of anxiety on cognition in older adult inpatients with depression: results from a multicenter observational study. *Heliyon.* 2019;5.
 270. Charlson F, van Ommeren M, Flaxman A, Cornett J, Whiteford H, Saxena S. New WHO prevalence estimates of mental disorders in conflict settings: a systematic review and meta-analysis. *Lancet.* 2019;394(10194):240–8.
 271. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6:e012799.
 272. Osburn HG. Coefficient alpha and related internal consistency reliability coefficients. *Psychol Methods.* 2000;5(3):343–55.
 273. Kroenke K, Spitzer RL, Williams JBW, Löwe B. An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ–4. *Psychosomatics.* 2009 Nov 1;50(6):613–21.
 274. Meader N, King K, Moe-Byrne T, Wright K, Graham H, Petticrew M, et al. A systematic review on the clustering and co-occurrence of multiple risk behaviours. *BMC Public Health.* 2016;16(1):1–9.
 275. Gómez-Gómez I, Bellón J, Resurrección DM, Cuijpers P, Moreno-Peral P, Rigabert A, et al. Effectiveness of universal multiple-risk lifestyle interventions in reducing depressive symptoms: Systematic review and meta-analysis. *Prev Med (Baltim).* 2020;134:106067.
 276. Oliván-Blázquez B, Aguilar-Latorre A, Motrico E, Gómez-Gómez I, Sarasa-Bosque C, Sanchez-Perez Á, et al. The Relationship between Adherence to the Mediterranean Diet, Intake of Specific Foods and Depression in an Adult Population (45–75 Years) in Primary Health Care. A Cross-Sectional Descriptive Study. *Nutrients.* 2021;13(8):2724.
 277. Kroenke K. Depression screening and management in primary care. *Fam Pract.* 2018;35(1):1–3.
 278. Barragán N. Situaciones complejas en Medicina Familiar. *Actual en Med Fam.* 2022;18(3):122–3.
 279. Gask L, Klinkman M, Fortes S, Dowrick C. Capturing complexity: The case for a new

- classification system for mental disorders in primary care. *Eur Psychiatry*. 2008;23(7):469–76.
280. Tavera-vilchis MA, Ruíz-mauricio LY. Teorías de la complejidad: una mirada desde la Medicina Familiar. *Arch en Med Fam*. 2023;25(1):43–9.
281. Hercberg M, Rodríguez M, Níguez N. Habilidades en comunicación para situaciones complejas. *Actual en Med Fam*. 2022;18(3):137–43.
282. Braunschneider LE, Lehmann M, Magaard JL, Seeralan T, Marx G, Eisele M, et al. GPs' views on the use of depression screening and GP-targeted feedback: a qualitative study. *Qual Life Res*. 2021;30(11):3279–86.
283. Davidsen AS, Fosgerau CF. What is depression? Psychiatrists' and GPs' experiences of diagnosis and the diagnostic process. *Int J Qual Stud Health Well-being*. 2014;9:1–10.
284. Marzo-Castillejo M, Nuin-Villanueva MÁ, Vela-Vallespín C. Recommendations against prostate cancer screening with PSA. *Aten Primaria*. 2012;44(7):377–8.
285. Sobregrau P, Baillès E, Carreño M, Donaire A, Boget T, Setoain X, et al. Psychiatric and psychological assessment of patients with drug-resistant epilepsy and psychogenic nonepileptic seizures (PNES) with no response to previous treatments. *Epilepsy Behav*. 2023;145:109329.
286. Muñoz-López L, López-Torrecillas F, Martín I, Sánchez-Barrera MB, López-Torrecillas M del C, Serrano F. Writing Abilities in Compulsive Prisoners. *Front Psychol*. 2021;12:701941.
287. Castelo-Branco C, RiberaTorres L, Gómez-Gil E, Uribe C, Cañizares S. Psychopathological symptoms in Spanish subjects with gender dysphoria. A cross-sectional study. *Gynecol Endocrinol*. 2021;37(6):534–40.
288. Jiménez-Murcia S, Giménez M, Granero R, López-González H, Gómez-Peña M, Moragas L, et al. Psychopathological status and personality correlates of problem gambling severity in sports bettors undergoing treatment for gambling disorder. *J Behav Addict*. 2022;
289. Santos-Ruiz A, Montero-López E, Ortego-Centeno N, Peralta-Ramírez MI. Efecto del confinamiento por COVID-19 en el estado mental de pacientes con lupus eritematoso sistémico. *Med Clin (Barc)*. 2021;156(8):379–85.
290. Arco Garcia L, Gonzalez Perez R, Henares Extremera A, García Leon MA, Peralta-Ramirez MI. Psychological and Sociodemographic Variables related to Cortisol in Hair in Spanish Healthy Population. *Span J Psychol*. 2020;23:e14, 1–10.
291. Miranda-Díaz S, Hassan BH, Fernández-Alcántara M, García-Caro MP. Attitudes and anxiety towards death in elderly people from 2 different cultures: Spain and Egypt. *Rev Esp Geriatr Gerontol*. 2022;57(3):168–73.
292. Kroenke K, Monahan PO, Kean J. Pragmatic Characteristics of Patient-Reported Outcome Measures are Important for Use in Clinical Practice. *J Clin Epidemiol*. 2016;68(9):1085–92.
293. Choi B, Granero R, Pak A. Catálogo de sesgos o errores en cuestionarios sobre salud. *Rev Costarr Salud Pública*. 2010;19(2):106–18.
294. Marrugat J, Solanas P, D'Agostino R, Sullivan L, Ordovas J, Cordón F, et al. Estimación del riesgo coronario en España mediante la ecuación de Framingham calibrada. *Rev Española Cardiol*. 2003;56(3):253–61.
295. Camm AJ, Kirchhof P, Lip GYH, Schotten U, Savelieva I, Ernst S, et al. Guidelines for the management of atrial fibrillation. Vol. 31, *European Heart Journal*. 2010. p. 2369–429.
296. Gual A, Segura L, Contel M, Heather N, Colom J. Audit-3 and audit-4: Effectiveness of two

short forms of the alcohol use disorders identification test. Alcohol Alcohol. 2002;37(6):591-6.

10

ANEXOS

10 Anexos

ANEXO 1 Grupo FPDM del EGPRN

Argyriadou, S	The Greek Association of General Practitioners (ELEGEIA), Thessaloniki, Greece
Assenova, R	Department of Urology and General Medicine, Faculty of Medicine, Medical University of Plovdiv, Plovdiv, Bulgaria
Clavería, A	South Galicia Health Research Institute , Vigo Health Area, Galician Health Service, Vigo, Spain. Red de Investigación en Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS).
Czachowski, S	Department of Clinical Psychology and Neuropsychology, Nicolaus Copernicus University, Torun, Poland
Fernández-San-Martin, MI	Unitat Docent Atenció Familiar i Comunitària Barcelona ciutat, Gerència Territorial d'Atenció Primària de Barcelona, Institut Català de la Salut. Unitat de Suport a la Recerca, Idiap Jordi Gol.
Guillou-Landreat, M	Department of Addictology, EA 7479 Soins primaires, Santé publique, Registre des cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France
Hasanagic, M	Health Care Studies, University "Djemal Bijedic" Mostar, Bosnia-Herzegovina.
Kasuba Lazic, D	Department of Family Medicine "Andrija Stampar, " School of Public Health, School of Medicine, University of Zagreb, Zagreb, Croatia
Le Floch, B	Department of General Practice, EA 7479 Soins primaires, Registre des cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France
Le Reste, JY	Department of General Practice, EA 7479 Soins primaires, Registre des cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France
Lingner, H	Centre for Public Health and Healthcare, Hannover Medical School, Hanover, Germany.
Lygidakis, C	Department of Behavioral and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg
Montier, T	INSERM, Etablissement Français du Sang, UMR 1078, Génétique, Génomique Fonctionnelle et biotechnologies, Univ Brest, Brest, France
Nabbe, P	Department of General Practice, EA 7479 Soins primaires, Registre des cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France
Rodríguez-Barragán, M	Centro de Atención Primaria La Mina, Gerencia Territorial de Atención Primaria de Barcelona, Institut Català de la Salut, Sant Adrià de Besòs, Barcelona, Spain.
Sowinska, A	Department of Experimental Linguistics, Nicolaus Copernicus University, Torun, Poland.
Stojanović-Špehar, S	Department of Family Medicine "Andrija Stampar, " School of Public Health, School of Medicine, University of Zagreb, Zagreb, Croatia
Van Marwijk, H	Department of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton, Brighton, United Kingdom
Van Royen, P	Department of Family Medicine, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

ANEXO 2 Ítems en las diferentes versiones de la Hopkins Symptom Checklist (HSCL)

Ítems en la HSCL-25	Ítems en la HSCL-10	Ítems en la HSCL-5
1. Se asusta sin motivo	Sí	
2. Siente miedo	Sí	Sí
3. Debilidad	Sí	
4. Nerviosismo		Sí
5. Palpitaciones		
6. Tiembla		
7. Se siente tenso/a	Sí	
8. Dolor de cabeza		
9. Siente pánico		
10. Siente inquietud		
11. Siente que le falta energía		
12. Se culpa a sí mismo/a	Sí	
13. Llorar con facilidad		
14. Pierde el interés sexual		
15. Se siente solo/a	Sí	Sí
16. Se siente sin esperanza	Sí	Sí
17. Se siente triste		
18. Piensa en acabar con su vida		
19. Se siente atrapado/a		
20. Se preocupa en exceso		Sí
21. No siente interés por nada		
22. Siente que todo le cuesta un esfuerzo	Sí	
23. Se siente inútil		
24. Poco apetito	Sí	
25. Problemas para dormir	Sí	

ANEXO 3 Taxonomía COSMIN en inglés

COSMIN definitions of domains, measurement properties, and aspects of measurement properties

Term			Definition
Domain	Measurement property	Aspect of a measurement property	
Reliability			The degree to which the measurement is free from measurement error
Reliability (extended definition)			The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: e.g. using different sets of items from the same health related-patient reported outcomes (HR-PRO) (internal consistency); over time (test-retest); by different persons on the same occasion (inter-rater); or by the same persons (i.e. raters or responders) on different occasions (intra-rater)
	Internal consistency		The degree of the interrelatedness among the items
	Reliability		The proportion of the total variance in the measurements which is due to 'true' [†] differences between patients
	Measurement error		The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Validity			The degree to which an HR-PRO instrument measures the construct(s) it purports to measure
	Content validity		The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured
		Face validity	The degree to which (the items of) an HR-PRO instrument indeed looks as though they are an adequate reflection of the construct to be measured
	Construct validity		The degree to which the scores of an HR-PRO instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the HR-PRO instrument validly measures the construct to be measured
		Structural validity	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured
		Hypotheses testing	Idem construct validity
		Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted HR-PRO instrument are an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
	Criterion validity		The degree to which the scores of an HR-PRO instrument are an adequate reflection of a 'gold standard'
Responsiveness			The ability of an HR-PRO instrument to detect change over time in the construct to be measured
	Responsiveness		Idem responsiveness
Interpretability*			Interpretability is the degree to which one can assign qualitative meaning - that is, clinical or commonly understood connotations - to an instrument's quantitative scores or change in scores.

[†] The word 'true' must be seen in the context of the CTT, which states that any observation is composed of two components – a true score and error associated with the observation. 'True' is the average score that would be obtained if the scale were given an infinite number of times. It refers only to the consistency of the score, and not to its accuracy (ref Streiner & Norman)

* Interpretability is not considered a measurement property, but an important characteristic of a measurement instrument



ANEXO 4

Nine Forward–Backward Translations of the Hopkins Symptom Checklist-25 With Cultural Checks

Patrice Nabbe^{1*}, Jean Yves Le Reste¹, Morgane Guillou-Landreat², Radost Assenova³, Djurdjica Kasuba Lazic⁴, Slawomir Czachowski⁵, Stanislava Stojanović-Špehar⁴, Melida Hasanagic⁶, Heidrun Lingner⁷, Ana Clavería⁸, María Rodríguez-Barragán^{9,10,11}, Agnieszka Sowinska^{12,13}, Stella Argyriadou¹⁴, Charileos Lygidakis¹⁵, Bernard Le Floch¹, Tristan Montier^{16,17}, Harm Van Marwijk¹⁸ and Paul Van Royen^{16,19}

OPEN ACCESS

Edited by:

Rita Roncone,
University of L'Aquila, Italy

Reviewed by:

Joseph Gone,
Harvard University, United States
Stefano Barlati,
University of Brescia, Italy

*Correspondence:

Patrice Nabbe
nabbe.patrice@univ-brest.fr

Specialty section:

This article was submitted to
Social Psychiatry and Psychiatric
Rehabilitation,
a section of the journal
Frontiers in Psychiatry

Received: 31 March 2021

Accepted: 07 July 2021

Published: 12 August 2021

Citation:

Nabbe P, Le Reste JY, Guillou-Landreat M, Assenova R, Kasuba Lazic D, Czachowski S, Stojanović-Špehar S, Hasanagic M, Lingner H, Clavería A, Rodríguez-Barragán M, Sowinska A, Argyriadou S, Lygidakis C, Le Floch B, Montier T, Van Marwijk H and Van Royen P (2021) Nine Forward–Backward Translations of the Hopkins Symptom Checklist-25 With Cultural Checks. *Front. Psychiatry* 12:688154. doi: 10.3389/fpsy.2021.688154

¹ Department of General Practice, EA 7479 Soins primaires, Santé publique, Registre des cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France, ² Department of Addictology, EA 7479 Soins primaires, Santé publique, Registre des cancers de Bretagne Occidentale, Université de Bretagne Occidentale, Brest, France, ³ Department of Urology and General Medicine, Faculty of Medicine, Medical University of Plovdiv, Plovdiv, Bulgaria, ⁴ Department of Family Medicine "Andrija Stampar," School of Public Health, School of Medicine, University of Zagreb, Zagreb, Croatia, ⁵ Department of Clinical Psychology and Neuropsychology, Nicolaus Copernicus University, Torun, Poland, ⁶ Health Care Studies, University "Djemal Bijedic," Mostar, Bosnia and Herzegovina, ⁷ Centre for Public Health and Healthcare, Hannover Medical School, Hanover, Germany, ⁸ Xerencia Xestión Integrada de Vigo, Servizo Galego de Saúde, Instituto de Investigación Sanitaria Galicia-Sur, Red de Investigación en Actividades Preventivas y de Promoción de la Salud, Vigo, Spain, ⁹ Centro de Atención Primaria La Mina, Gerencia Territorial de Atención Primaria de Barcelona, Instituto Catalán de la Salud, Sant Adrià de Besòs, Barcelona, Spain, ¹⁰ Fundación Instituto Universitario de Investigación en Atención Primaria de Salud Jordi Gol i Gurina (IDIAP Jordi Gol), Barcelona, Spain, ¹¹ Universidad Autónoma de Barcelona, Barcelona, Spain, ¹² Department of Experimental Linguistics, Nicolaus Copernicus University, Torun, Poland, ¹³ Escuela de Inglés, Universidad Católica del Norte, Antofagasta, Chile, ¹⁴ The Greek Association of General Practitioners (ELEGEIA), Thessaloniki, Greece, ¹⁵ Department of Behavioral and Cognitive Sciences, University of Luxembourg, Esch-sur-Alzette, Luxembourg, ¹⁶ INSERM, Etablissement Français du Sang, UMR 1078, Génétique, Génomique Fonctionnelle et biotechnologies, Univ Brest, Brest, France, ¹⁷ Service de Génétique Médicale et Biologie de la Reproduction, CHRU de Brest, Brest, France, ¹⁸ Department of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton, Brighton, United Kingdom, ¹⁹ Department of Family Medicine, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

Introduction: The Hopkins Symptom Checklist-25 (HSCL-25) is an effective, reliable, and ergonomic tool that can be used for depression diagnosis and monitoring in daily practice. To allow its broad use by family practice physicians (FPs), it was translated from English into nine European languages (Greek, Polish, Bulgarian, Croatian, Catalan, Galician, Spanish, Italian, and French) and the translation homogeneity was confirmed. This study describes this process.

Methods: First, two translators (an academic translator and an FP researcher) were recruited for the forward translation (FT). A panel of English-speaking FPs that included at least 15 experts (researchers, teachers, and practitioners) was organized in each country to finalize the FT using a Delphi procedure.

Results: One or two Delphi procedure rounds were sufficient for each translation. Then, a different translator, who did not know the original version of the HSCL-25, performed a backward translation in English. An expert panel of linguists compared

the two English versions. Differences were listed and a multicultural consensus group determined whether they were due to linguistic problems or to cultural differences. All versions underwent cultural check.

Conclusion: All nine translations were finalized without altering the original meaning.

Keywords: depression, Hopkins Symptom Checklist-25, depressive disorder, HSCL-25, diagnostic tool

INTRODUCTION

How to manage people with depression in primary care is a growing challenge worldwide. Indeed, Family practice physicians (FPs) are at the frontline, while secondary care services are increasingly under threat (1–4). Depression manifests (for laypersons) itself in various ways: (i) as a syndromic “disorder” in which contextual distress, anxiety, and somatoform disorders overlap; (ii) as a suffering that is difficult to express, acknowledge, and discuss; and (iii) as a long-term condition with subjective and objective features that can be measured (5). Due to these inter-individual variabilities, FPs may experience difficulties in detecting depression and may easily misjudge the symptoms and their intensities, if they do not use formal instruments (6, 7). Moreover, the depression incidence and prevalence rates differ widely in family practice, due to complex contextual variations, differences in healthcare systems, concepts of disorder, objectives, and practices, as well as cultural variations in symptom expression (8, 9). These difficulties may lead to inappropriate care and potential side effects due to drugs’ use as well as public health issues (10–12). A short discussion of the results obtained using a relevant questionnaire is often the first step toward an open dialogue with the patient.

Collaborative primary care mental health models can improve the management of patients with depression. To this aim, the European General Practice Research Network (EGPRN) developed a collaborative research agenda (13). Specifically, the EGPRN adopted a standardized methodology in which European FPs experts from different healthcare systems and who speak different languages and have different cultural references set up an established consensus procedure to identify reliable, standardized, efficient, and ergonomic tools for depression assessment that take into account cultural and linguistic differences (14–17). These tools need to be accepted by both FPs and psychiatrists to improve collaboration (18). They must be feasible in the FP’s surgery, in primary or psychiatric care, and also suitable for research purposes (19). Finally, they must be validated and reliable.

A handbook was developed to guide the selection of a single tool that would be then translated into different languages, using a forward and backward translation procedure (inspired by Brislin’s translation model). This is a consensual procedure that has been used in other cross-cultural studies (20–22). At

each step, the key points and purposes were debated and decided by consensus among the involved European experts. First, a systematic literature review, according to the PRISMA criteria, allowed the identification of seven tools that had been validated against a psychiatric examination using the DSM-IV or DSM-5 major depression criteria (23). Then, a consensus procedure (RAND/UCLA Appropriateness Method) led to the selection of one tool on the basis of its effectiveness, reliability, and ergonomics (24): the self-report Hopkins Symptom Checklist-25 (HSCL-25) (23–26). This is a validated, reliable diagnostic tool to assess (27, 28) the presence and severity of anxiety and depression symptoms during the previous week (29, 30). Its specificity compared with clinical interview is robust: between 0.78 to 0.88, the reliability (Alpha de Cronbach) is between 0.87 to 0.97 (31). The HSCL-25 short length self-administered format is perfectly suited for use in busy primary care settings with many competing demands. It may represent a practical instrument to alert FPs to potentially depressive or anxious symptomatology.

A qualitative procedure with the FP’s involvement was necessary to obtain that were linguistically and culturally equivalent to the original version, ecologically embedded in primary care.

The objective of the present study was to translate the HSCL-25 into the languages of the different team members, without losing homogeneity, and in a language suitable to the primary care context (22, 32).

MATERIALS AND METHODS

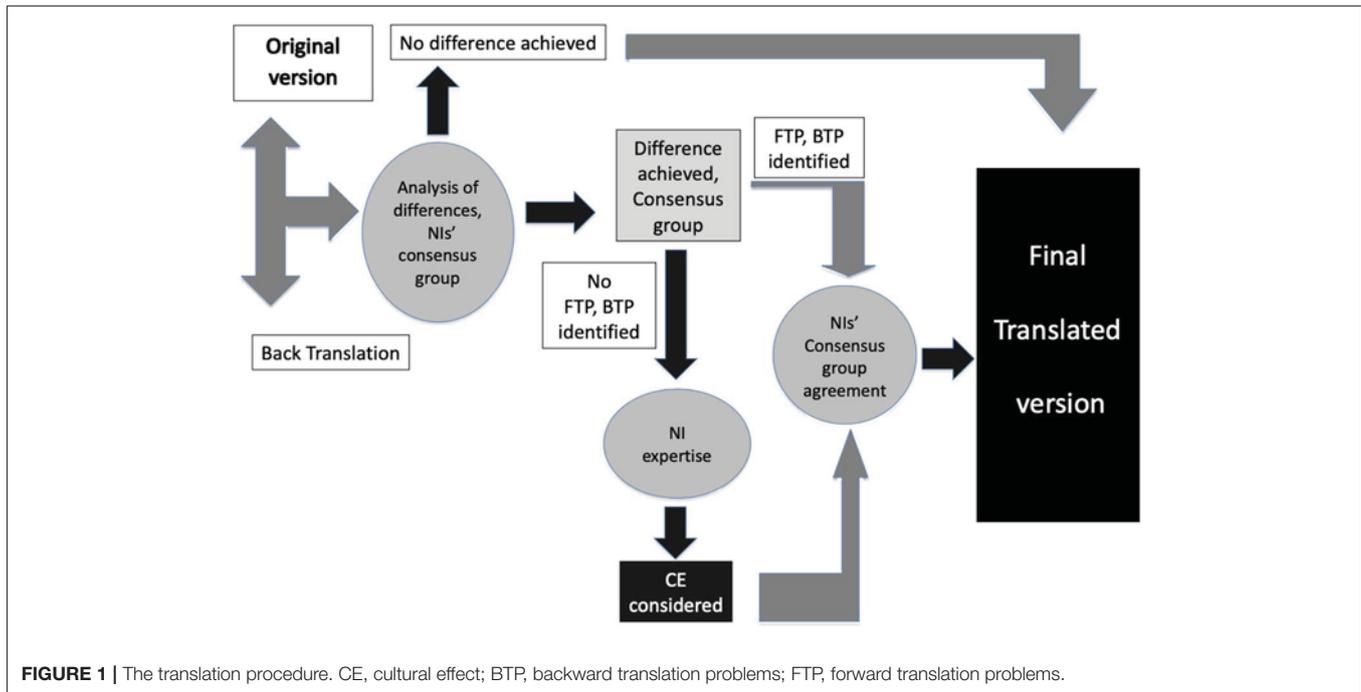
This three-step standardized study included: (i) forward translation (FT), (ii) backward translation (BT), and (iii) cultural check (8, 33, 34) (Figure 1).

The FT was carried out with an incorporated Delphi consensus procedure (35–37). This is a systematic, interactive method that involves a panel of experts using iterative procedures (38) and that allows reaching consensus in a rigorous way (39–41). This process requires:

- Anonymity of participants to ensure response reliability and avoid contamination,
- Iteration, which allows participants to refine their views in the light of the group work progress,
- Feedback control under the investigator’s responsibility,
- Statistical aggregation of the group’s responses to allow a quantitative and qualitative analysis of data (42–45).

The EGPRN French team ensured that this protocol was followed throughout the process. The FT of the different HSCL-25 items had to be validated daily by the expert panel, composed of EGPRN members, all actively involved in the process.

Abbreviations: BT, backward translation; CE, cultural effect; DSM, diagnostic and statistical manual of mental disorders; EGPRN, European general practice research network; FPs, family practice physicians; FT, forward translation; PRISMA, preferred reporting items for systematic reviews and meta-analyses; RAND, research and development; RAND/UCLA, research and development/University of California Los Angeles.



Briefly, for each language, the National Investigators (NI) selected translators knowledgeable about healthcare terminology to organize two translation (FT and BT) teams who were blind to the other team's work. The FT team included one member of the FP research group and one official translator for each country. The BT team involved one (or two) FPs and one official translator (22).

The NIs also recruited a panel of FP experts in their own countries, anonymized the experts' responses, and allocated an identification number for later identification (42). Initially, 20 to 30 experts were recruited per country to secure the presence of at least 15 participants till the project end. The FP experts were selected using the following inclusion criteria: native of their country of residence and speaking their native language, and fluent in English (32). At least half of them had to be involved in teaching and/or research activities. To assess the panel representativeness of their country FPs, the experts provided the following information: sex, practice type, years of practice, and publication record (46).

According to the Brislin's Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures, once the FT was completed, a BT was performed with two goals: (i) to ensure the identification of language issues and (ii) to detect translation problems linked to cultural adaptation issues. Indeed, as translation biases related to cultural aspects of each country were possible, a cultural check was required to ensure homogeneity (17, 20, 33, 34, 47). To this aim, in each country, an FP researcher and a linguist analyzed all BT propositions and compared them with the original HSCL-25 version to establish whether there was any significant difference in terms of meaning. Their report was submitted to a consensus group whose task was to clarify the nature of each FT-BT discrepancy

from three problem areas: (i) BT problems were eliminated if the difference was explained by an incorrect BT; (ii) FT problems were defined as an anomaly in transcribing the original English (semantic/idiomatic differences relative to the original English version); and (iii) cultural effects (CE) were considered validated if there was no linguistic problem with the translation, but the item needed to be modified to be understood by the patients in their own "everyday" language (Figure 1).

This led to a linguistically stable, definitive translation that maintained the HSCL-25 meaning (i.e., structure and question order and method of use) for each involved country.

Ethical request: The EGPRN French team was in charge of checking the volunteering process and confirming the absence of potential conflicts of interest for all participants. The Ethics Committee of the approved the whole process.

The EGPRN French team recruited all NIs and obtained their consent, managed the voluntary participation in the study and produced an absence of conflict-of-interest statement.

Each NI asked participants to sign the informed consent.

RESULTS

NI Panel Description

The NI panel included 11 NIs (including $n = 8$ women) from eight European countries. They were all FPs, EGPRN members, and fluent in English. Ten NIs practiced in urban areas of more than 5,000 inhabitants and one worked in an urban area with 2,000–5,000 inhabitants. Eight had also teaching duties in addition to being researchers (total number of publications by the panel members: 152). The mean number of years of practice and of research were 21.3 and 12.4 years, respectively. In the panel, two NIs were from two distinct cultural regions of coastal

TABLE 1 | National investigators' panel.

Experts	Gender	Country	Academic Status	Number of inhabitants	Practice type	Number of international publications	Years of practice	Years of research
9	F	Bulgaria	Teacher/Researcher	>5,000	FP group practice	9	14	12
7	F	Croatia	Teacher/Researcher	>5,000	Alone	6	20	12
8	F	Croatia	Teacher/Researcher	>5,000	FP group practice	18	30	20
11	M	France	Teacher/Researcher	>5,000	FP group practice	11	20	5
5	F	Germany	Researcher	2,000–5,000	Ceased practicing 2 years previously	19	23	5
10	F	Germany	Researcher	>5,000	FP group practice	4	18	7
3	F	Greece	Teacher/Researcher	>5,000	FP and paramedic group practice	14	30	18
4	M	Italy	Researcher	>5,000	FP group practice	23	7	6
6	M	Poland	Teacher/Researcher	>5,000	FP group practice	20	30	12
2	F	Spain (Cataluña)	Teacher/Researcher	>5,000	FP group practice	13	22	25
1	F	Spain (Galicia)	Teacher/Researcher	>5,000	FP group practice	15	20	14

F, female; M, male; FPs, family practice physicians.

Spain (Catalonia and Galicia), and two were Croats. The other countries were each represented by a single NI (Table 1).

Forward Translation

For the Delphi consensus procedure, 14 (Germany) to 31 experts (Spain) were recruited. In compliance with the selection criteria, they were all FPs and fluent in English. The expert panel included 215 FPs (111 men and 104 women). Among them, 20 worked in a city of <2,000 inhabitants, 36 in a city with 2,000–5,000 inhabitants, and 159 in a city with >5,000 inhabitants. Their clinical experience was analyzed according to years of practice (mean: 16.4 years of experience) (Table 2).

In Poland, Bulgaria, Germany, Spain, and the Catalonia region of Spain, there was only one Delphi round, and two rounds in the other countries. Almost all translation proposals for each item of the HSCL-25 questionnaire were accepted in one round (273/320: 85.3%) (Table 3). The other proposals for which consensus was not reached went through a second round. The NI and the forward official translator synthesized the experts' comments to produce a new translation proposition for the second round.

Some Translation Issues Required a Second Proposal and Another Delphi Round

In Croatian, eleven proposals were rejected in the first round. For example, for item #17 ("Feeling blue"), the first proposal was "Bili ste tužni," which was considered to be too focused on melancholia, and was modified to "Bili ste sjetni" closer to the concept of sadness. All new proposals were accepted during the second round.

As a German version of the HCL-25 was already available, the German NIs proposed that their expert panel would discuss this version, without producing a new FT. All items were accepted in

the first Delphi round. At this step, the German NIs stopped the procedure. No cultural check was performed.

Nine Greek proposals were rejected in the first round. For example, for item #1 ("Being scared for no reason"): the first proposal "Είμαι τρομοκρατημένος χωρίς αιτία" was considered too strong. Consensus was reached on the second proposal: "Είμαι τρομαγμένος χωρίς αιτία." All new proposals were accepted during the second round.

In the French translation, consensus was not reached on 18 proposals in the first round and needed further specification in the second round. For example, for item #25 ("Sleep disturbance"), the first proposal was "Vous n'arrivez pas à dormir" that was modified to "Votre sommeil était perturbé," closer to the English word: "disturbance." All new proposals were accepted during the second round.

In the Italian translation, consensus was not reached on five proposals during the first round. For example, for item #5 ("Heart racing"), the first proposal "Avere tachicardia" was considered too focused on clinical symptoms and was modified to "Sentire il cuore battere veloce," which was more familiar according to the reviewers. All new proposals were accepted during the second round.

In the Spanish Galician translation, consensus was not reached on three proposals in the first round. For example, for item #6 ("Trembling"), the first proposal was "Trema," the present indicative of the verb "Tremar." The second proposal was "Ten tremores" and was accepted in the second round. All new proposals were accepted during the second round.

Backward Translation and Cultural Check

The initial instructions, the 25 items, the quotation and the explanatory sentences were all back-translated into English by the BT team. In total, 36 propositions were analyzed. All BTs were compared linguistically to the original. Differences were noted for

TABLE 2 | Characteristics of each country expert panel.

	N (women)	Practice (mean years)	Number of inhabitants in the practice area			Academic researcher and/or teacher		Number of publications	Participants in the second Delphi round
			<2,000	2,000–5,000	>5,000	Number	Experience (mean, years)		
Bulgaria	22 (13)	20.5	1	5	16	5	5.4	8	No second round
Catalonia	22 (9)	15.7	0	2	20	20	10.5	22	No second round
Croatia	16 (13)	19.2	1	1	14	16	11.5	15	15
France	16 (7)	12.5	1	7	8	15	6.3	11	15
Galicia	20 (6)	22.3	0	0	20	17	13.1	19	20
Germany	14 (8)	16.7	0	3	11	9	10	6	No second round
Greece	26 (13)	10.9	10	9	7	24	5.1	26	15
Italy	18 (6)	17.2	3	2	13	13	14	12	No second round
Poland	30 (18)	11.9	4	6	20	26	13.1	10	No second round
Spain	31 (11)	19.5	0	1	30	27	12	30	No second round
Total	215 (104)	15.55	20	36	178	172	10.1	159	4 Second round

submission to the NIs and the consensus group. Three consensus group meetings were necessary with national feedback between each. The main adaptations, produced as a result of national feedback and the consensus resulting from the cultural check, are described below.

By Languages and Language Groups

Croatia: 8 items were different (2 were BT problems, and 8 required a cultural adaptation).

The main cultural aspect was the use of the present perfect, which is a tense of state and not of action, commonly employed in daily life. Therefore, in items #2, 7, 9, and 10, “feeling” was replaced by “you have been.” Only one item seemed to be stronger than in the original version. Indeed, “Faintness,” was replaced by “Weakness,” but in Croatian this is equivalent to faintness.

Bulgaria: 3 items were different (2 were BT problems, and 1 required a cultural adaptation).

“Feeling low in energy” became “A sense of low energy.” Overall, the Bulgarian translation was the most stable among the three Slavic languages.

Poland: 13 items were different (7 were BT problems, and 6 required a cultural adaptation).

Most problems resulted from a conceptual issue. For instance, in Polish, “Heart racing” became “Palpitations,” “Trembling” became “Tremors,” and “An effort” was translated into “A burden.” “Headache” was translated into “Headaches” in Polish for grammatical reasons.

In all three Slavic languages (Croatian, Bulgarian, and Polish), “Feeling restless” was translated into “Anxiety” because there is no equivalent word to express these ideas. A word-by-word translation, in that case, was impossible.

For the Greek language, the translation was mainly based on an adaptation according to gender. The experts concluded that there was a general CE affecting all parts of the scale. However, no real difference in meaning was detected, and the Greek HSCL-25 scale remained stable relative to the original.

France: 5 items were different (4 were BT problems, and 1 required a cultural adaptation).

For the French scale, the present tense is normally used in everyday language. However, the past tense was used in the FT. In everyday life French, the past tense is considered an older, upper-class language style. Therefore, all tenses were modified. For instance, “Tout était un effort pour vous” became “Tout est un effort pour vous” in the final version.

Italy: 7 items were different (6 were BT problems, and 1 required a cultural adaptation).

In the Italian scale, the male plural form was used because this is the usual way of speaking/writing; the translation had to be modified according to gender.

Spain: 6 items were different (1 was a BT problem, and 5 required a cultural adaptation).

“Feeling no interest” was translated in “No sienta interes por nada” in standard Spanish, and “Worthless feeling” became “Feeling useless.” However, in Standard Spanish, “inutil” means also “worthless.”

Catalonia: 7 items were different (4 were BT problems, and 3 required a cultural adaptation).

Galicia: 5 items were different (1 was a BT problem, and 4 required a cultural adaptation).

In the Galician scale, item #14 “Losing sexual interest,” was translated into “Loss of sexual interest” that expresses a state, and not an action (the original English version); however, the local experts considered it a normal way of speaking/writing in that language.

In the Galician and Catalan translations, “Blame oneself” turned into “Blame yourself” in the BT because the term “oneself” is not commonly employed.

For the Hispanic languages, the translation had to be modified according to gender. The item “Faintness” was translated into “Weakness” (e.g., “Debilidad,” “Debilitat,” and “Debilidad” in standard Spanish, Catalan and Galician respectively). Similarly, the item “Heart racing” was translated into “Palpitations” (i.e., “Palpitaciones” and “Palpitacions” in the standard Spanish and Galician versions).

TABLE 3 | Results of the first Delphi round.

Item/Country	Galicia	Castile	Catalonia	France	Italy	Bulgaria	Croatia	Greece	Germany	Poland
1 Being scared for no reason	C	C	C	C	C	C	C	NC	C	C
2 Feeling fearful	C	C	C	C	C	C	NC	C	C	C
3 Faintness	C	C	C	NC	NC	C	NC	NC	C	C
4 Nervousness	C	C	C	C	C	C	C	C	C	C
5 Heart racing	C	C	C	NC	C	C	C	C	C	C
6 Trembling	NC	C	C	NC	NC	C	C	C	C	C
7 Feeling tense	C	C	C	C	C	C	C	C	C	C
8 Headache	C	C	C	C	C	C	C	C	C	C
9 Feeling panic	C	C	C	NC	C	C	NC	C	C	C
10 Feeling restless	NC	C	C	NC	C	C	NC	C	C	C
11 Feeling low in energy	C	C	C	C	C	C	NC	NC	C	C
12 Blaming oneself	C	C	C	NC	NC	C	C	C	C	C
13 Crying easily	C	C	C	C	C	C	C	NC	C	C
14 Losing sexual interest	C	C	C	NC	C	C	NC	C	C	C
15 Feeling lonely	C	C	C	NC	C	C	NC	C	C	C
16 Feeling hopeless	C	C	C	C	C	C	NC	C	C	C
17 Feeling blue	C	C	C	NC	C	C	NC	C	C	C
18 Thinking of ending one's life	C	C	C	C	C	C	C	NC	C	C
19 Feeling trapped	C	C	C	NC	C	C	C	C	C	C
20 Worrying too much	C	C	C	NC	C	C	NC	NC	C	C
21 Feeling no interest	C	C	C	NC	C	C	NC	NC	C	C
22 Feeling that everything is an effort	C	C	C	C	C	C	C	C	C	C
23 Feelings of worthlessness	C	C	C	NC	C	C	C	NC	C	C
24 Poor appetite	C	C	C	C	C	C	C	NC	C	C
25 Sleep disturbance	NC	C	C	NC	C	C	C	C	C	C
26 Choose the best answer for how you felt over the past week	C	C	C	NC	C	C	C	C	C	C
27 Not at all	C	C	C	C	NC	C	C	C	C	C
28 A little	C	C	C	NC	C	C	C	C	C	C
29 Quite a bit	C	C	C	C	C	C	C	C	C	C
30 Extremely	C	C	C	C	C	C	C	C	C	C
31 The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress. The patient is considered as a "probable psychiatric case" if the mean rating on the HSCL-25 is ≥ 1.55 .	C	C	C	NC	NC	C	C	C	C	C
32 A cut-off value of ≥ 1.75 is generally used for diagnosis of major depression defined as "a case in need of treatment." This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender. The administration time of HSCL 25 is 5–10 min	C	C	C	NC	C	C	C	C	C	C

C, consensus; NC, no consensus.

For All of Languages

Item #17 "Feeling Blue" generated a CE in six of the nine languages. A word-by-word rendition was impossible and required a cultural adaptation.

Items #15 "Feeling lonely," #18 "Thinking of ending one's life," #19 "Feeling trapped" and #25 "Sleep disturbance" remained stable after the BT.

Concerning the scale instructions and the quotation question, the BT was different from the original version in nine items, except the explanation concerning the time required to fill in the scale. Many translation problems were related to "cultural" effects. For example: in French, some terms were replaced by typical expressions commonly employed in questionnaires: e.g., "pencil-and-paper" was

TABLE 4 | Final translation of the HSCL-25 in nine European languages: items 1–25.

HSCL-25 Original version	Greece	Poland	Bulgaria	Croatia	Castile	Catalonia	Galicia	Italy	France
Choose the best answer for how you felt over the past week	Επιλέξτε την καλύτερη απάντηση για το πώς αισθανθήκατε την τελευταία εβδομάδα	Wybierz najlepszą odpowiedź	Изберете отговора, който най-добре описва как сте се чувствали през изминалата седмица	Izaberite jedan odgovor koji najbolje opisuje kako ste se osjećali tijekom prošlog tjedna:	Elija la respuesta que mejor describa cómo se ha sentido durante la semana pasada	Trii la millor resposta per indicar com s'ha sentit en la darrera setmana	Escolla a resposta que mellor describa como se sentiu durante a semana pasada	Scegliere la risposta più adatta su come ti sei sentito/a nell'ultima settimana	Veuillez choisir la réponse qui décrit le mieux comment globalement vous vous sentiez toute la semaine dernière
Being scared for no reason	Είμαι τρομαγμένος/η χωρίς αιτία	Bać sie bez powodu	Чувство за уплаха без причина	Bili ste bezrazložno uplašeni	Se asusta sin motivo	Estar espantat/espantada sense motiu aparent	Asústase sen motivo	Avere paura senza motivo	Vous avez peur sans raison
Feeling fearful	Αισθάνομαι φοβισμένος / η	Poczucie strachu	Чувство за страх	Bojali ste se	Siente miedo	Sentir por	Ten medo	Sentirsi impauriti	Vous vous sentez effrayé
Faintness	Αισθημα λιποθυμιάς	Omdlenia	Отпадналост	Bili ste slabi	Debilitad	Debilitat	Debilitade	Sensazione di mancamento	Vous avez une sensation d'étourdissement
Nervousness	Νευρικότητα	Nerwowość	Нервност	Bili ste nervozni	Nerviosismo	Nerviosisme	Nerviosismo	Esseri nervosi	Vous vous sentez nerveux
Heart racing	Ταχυπαλμία	Kolatanie serca	Сърцебиене	Ubrzano vam je lupalo srce	Palpitaciones	Cor accelerat	Palpitaciós	Sentire il cuore battere veloce	Vous avez l'impression que votre coeur bat anormalement vite
Trembling	Τρεμούλα	Drzenia	Трелерене	Drhtali ste	Tiembra	Tremola	Ten tremores	Tremore	Vous avez la sensation de trembler
Feeling tense	Αισθάνομαι υπερένταση	Poczucie napiecia	Чувство за напрежение	Bili ste napeti	Se siente tenso/a	Sentir-se tens/a	Séntese tenso/a	Sensazione di tensione	Vous vous sentez tendu
Headache	Πονοκέφαλος	Bóle głowy	Главоболне	Bojela vas glava	Dolor de cabeza	Mal de cap	Dor de cabeza	Avere mal di testa	Vous avez des maux de tête
Feeling panic	Αισθάνομαι πανικό	Uzucie paniki	Чувство за паника	Bili ste u panici	Siente pánico	Sensació de pànic	Sente pánico	Sensazione di panico	Vous vous sentez paniqué
Feeling restless	Αισθάνομαι ταραχή	Uzucie niepokoju	Чувство на безпокойство	Bili ste uznemireni	Siente inquietud	Sensació d'inquietud	Séntese inquieto/a	Sensazione di irrequietezza	Vous vous sentez agité

(Continued)

TABLE 4 | Continued

HSCL-25 Original version	Greece	Poland	Bulgaria	Croatia	Castile	Catalonia	Galicia	Italy	France
Feeling low in energy	Αισθάνομαι ότι δεν έχω ενέργεια	Poczucie braku energii	Усещане за понижена енергия	Niste imali dovoljno energije	Siente que le falta energia	Sensació de manca d'energia	Sente que lle falta enerxía	Sentirsi senza energia	Vous manquez d'énergie
Blaming oneself	Κατηγορώ τον εαυτό μου	Obwinianie samego siebie	Самобвинение	Okrivljavali ste se	Se culpa a si mismo/a	Culpar-se un/a mateix/a	Cúlpanse a si mesmo/a	Avere sensi di colpa	Vous ressentez une sensation de culpabilité
Crying easily	Εύκολο κλάμα	Placzliwość	Плачливост	Bili ste plačljivi	Llora con facilidad	Plora fàcilment	Chora con facilidade	Piangere facilmente	Vous pleurez facilement
Losing sexual interest	Απώλεια σεξουαλικού ενδιαφέροντος	Utrata zainteresowań sfera seksualna	Забугота на сексуален интерес	Niste bili zainteresirani za spolni odnos	Pierde el interés sexual	Pèrdua de l'interès sexual	Perda do interese sexual	Perdere l'interesse sessuale	Vous ressentez un désintérêt pour la vie sexuelle
Feeling lonely	Αισθάνομαι μοναξιά	Poczucie osamotnienia	Чувство за самотност	Bili ste usamljem	Se siente solo/a	Sentir-se sol/a	Séntese só/soa	Sentirsi soli	Vous avez une sensation de solitude
Feeling hopeless	Αισθάνομαι απελπισμένος/η	Poczucie beznadziejności	Чувство за безнадежност	Osjećali ste sebeznažno	Se siente sin esperanza	Sentiment de desesperança	Séntese sen esperanza	Sentirsi senza speranza	Vous vous sentez désespéré
Feeling blue	Νωθόλο/η πεσμένος/η	Poczucie przygnębienia	Чувства се нещастен	Bili ste sjetni	Se siente triste	Sentir-se trist/a	Séntese triste	Sentirsi tristi	Vous avez le cafard
Thinking of ending one's life	Σκέφτομαι να δώσω τέλος στη ζωή	Myśli samobójcze	Мисли за самоубийство	Razmišljali ste da si oduzmete život	Piensa en acabar con su vida	Pensa en treure's la vida	Pensa en acabar coa súa vida	Avere pensieri di togliersi la vita	Vous avez pensé à mettre fin à votre vie
Feeling trapped	Αισθάνομαι παγιδευμένος /η	Poczucie uwięzienia	Чувствам се като в капан	Osjećali ste se kao da ste u klopci	Se siente atrapado/a	Sentir-se atrapad/atrapada	Séntese atrapado/a	Sentirsi intrappolati	Vous vous sentez pris au piège
Worrying too much	Ανησυχώ υπερβολικά	Zamartwanie sie	Пригесивам се твърде много	Bili ste previše zabrinuti	Se preoccupa en exceso	Preoccupar-se en excés	Preocupase en exceso	Preoccuparsi troppo	Vous vous inquiétez trop
Feeling no interest	Αισθάνομαι ότι τίποτε δεν είναι ενδιαφέρον	Poczucie braku zainteresowań	Чувство за забуга на интерес	Bez interesa za bilo što	No siente interés por nada	Sentiment de manca d'interès	Non sente interese por nada	Non avere alcun interesse	Plus rien ne vous intéresse
Feeling that everything is an effort	Αισθάνομαι ότι για το χαθε τί χρειάζεται να κάνω προσπάθεια	Poczucie, ze wszystko jest ciężarem	Чувство, че всичко изисква усилие	Sve vam je bilo naporno	Siente que todo le cuesta un esfuerzo	Sentir que tot és un esforç	Sente que todo lle supón un esforzo	Sentire che tutto è uno sforzo	Tout est un effort pour vous
Feelings of Worthlessness	Αισθάνομαι ότι δεν αξίζω τίποτε	Poczucie bezwartościowości	Чувство за безполезност	Osjećali ste se bezvrijedno	Se siente inútil	Sentir-se inútil	Séntese inútil	Sentirsi inutili	Vous avez le sentiment d'être bon à rien
Poor appetite	Μείωση της όρεξης	Slaby apetyt	Лопш апетит	Imali ste slab apetit	poco apetito	Pèrdua de la gana	Poco apetito	Avere poco appetito	Vous avez perdu l'appétit
Sleep disturbance	Διαταραχές ύπνου	Zaburzenia snu	Нарушения на съня	Imali ste problema sa spavanjem	Problemas para dormir	Alteració de la son	Alteración do sono	Disturbi del sonno	Votre sommeil est perturbé

TABLE 5 | Final translation of the HSCL-25 in nine European languages: scale instructions.

Scale instructions original version	Greece	Poland	Bulgaria	Croatia	Spain	Catalonia	Galicia	Italy	France
The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.	Η βαθμολογία του ΗΣΥΛ-25 βασίζεται σε γραπτό ερωτηματολόγιο αυτοαξιολόγησης 25 ερωτήσεων σχετικά με την παρουσία και την ένταση των συμπτωμάτων άγχους και κατάθλιψης κατά την τελευταία εβδομάδα. Οι συμμετέχοντες απαντούν σε μία από τις τέσσερις κατηγορίες για κάθε ερώτημα σε μια λίστα ερώσεων βαθμών με τιμές από 1 μέχρι 4.	Ocena testu HSCL-25 oparta jest na kwestionariuszu 25 pytań, w którym zakreśla się na papierze obecność i nasilenie objawów leku i depresji w ciągu ostatniego tygodnia.	Резултатът от HSCL-25 се основава на самостоятелно попълнен инструмент на хартиен носител, включващ 25 въпроса за наличието и интензивността на симптомите на тревожност и депресия през последната седмица.	HSCL-25 skor sastoji se od 25 pitanja koja se jedinstavno odgovorjavaju olovkom i papiriom, a temelji se na samoprocjeni prisutnosti i intenzitetu anksioznih i depresivnih simptoma tijekom prošlog tjedna.	La puntuación HSCL-25 se basa en un cuestionario auto cumplimentado con lápiz y papel, de 25 preguntas sobre la presencia y la intensidad de ansiedad y síntomas depresivos en la última semana.	L'escala HSCL-25 es basa en un questionari auto administrat de 25 preguntes, sobre la presència i la intensitat de símptomes d'ansietat i depressió en la darrera setmana.	A puntuación HSCL-25 baséase nun cuestionario cumprimentado con lapis e papel, de 25 preguntas sobre a presenza e a intensidade de ansiedade e síntomas depresivos na última semana.	Il punteggio dell' HSCL-25 si basa sulla compilazione di un questionario di autovalutazione in cartaceo ("carta/perma") di 25 domande sulla presenza e intensità di sintomi di ansia e depressione nel corso dell'ultima settimana.	La HSCL-25 est un auto-questionnaire en 25 questions relatives à la présence et à l'intensité des symptômes d'anxiété et de dépression durant toute la semaine dernière.
Participants answer to one of four categories for each item on a four-point scale ranging from 1 to 4		Badani odpowiadają na jedno z czterech możliwych kategorii na skali mierzącej wartości od 1 do 4.	Участниците избират една от категориите за всяка позиция по скала от 1.00 до 4.00.	Ispitanici odgovaraju jednom od četiri kategorija za svako pitanje na skali od 1-4.	Los/ las participantes responden una de cuatro categorías para cada ítem, en una escala de cuatro puntos que van desde 1 a 4.	Els/les participants responen a una de les quatre categories per a cada ítem en una escala de quatre punts que va de l'1 al 4.	Os participantes responden unha de catro categorías para cada ítem, nunha escala de catro puntos que van desde 1 a 4.	I partecipanti rispondono a una delle quattro categorie per ciascun sintomo su una scala di punteggio che va da 1 a 4.	Les participants cotent chaque proposition, sur une échelle en quatre points, cotée de 1 à 4.
1. "Not at all"	Καθόλου	Wcale	Съвсем не	Nimalo	En absoluto	Gens	En absoluto	Per niente	Pas du tout d'accord
2. "A little"	Λίγο	Troche	Незначително	Malo	Un poco	Una mica	Un pouco	Poco	Un peu d'accord
3. "Quite a bit"	Αρκετά	Znacznie	Съвсем малко	Dosta	Bastante	Bastant	Bastante	Abbastanza	Plutôt d'accord
4. "Extremely"	Πάρα πολύ	Bardzo mocno	Извънредно	Jako	Mucho	Molt	Molto	Moltissimo	Complètement d'accord

TABLE 6 | Final translation of the HSCL-25 in nine European languages: general instructions.

Scale instructions original version	Greece	Poland	Bulgaria	Croatia	Spain	Catalonia	Galicia	Italy	France
The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1.00 and 4.00). It is often used as the measure of distress.	Η βαθμολογία του ΗΣΓΛ-25 υπολογίζεται διαφέροντας τη βαθμολογία (αποφραστική βαθμολογία των ερωτημάτων), διά του αριθμού των ερωτημάτων που απαντήθηκαν (χωριζόμενο μεταξύ του 1,00 έως 4,00).	Wynik testu HSCL-25 jest obliczany poprzez podzielenie całkowitej liczby punktów (suma punktów z każdej pozycji testu) przez liczbę pozycji na które udzielono odpowiedzi (w skali od 1 do 4). Często służy on do pomiaru dystres.	HSCL-25 резултатът се изчислява, като се раздели общият брой точки (сбор точки по критерий) на броя на отговорите (критерий (вариращи между 1,00 и 4,00). Той често се използва като мярка за страдание.	Skor HSCL-25 se izračunava dijeljenjem ukupnog zbroja (zbroj skora pojedinih pitanja) s brojem odgovorenih pitanje (raspon od 1,00 do 4,00). Obično se koristi za mjerenje distresa.	La puntuación del HSCL-25 se calcula dividiendo la puntuación total (sumando la puntuación de todos las preguntas) entre el número de respuestas (varia entre 1,00 y 4,00). Se usa habitualmente para medir el malestar psicológico.	La puntuació total del HSCL-25 es calcula dividint la suma de la puntuació dels diferents ítems pel número d'ítems contestats. El resultat total oscil·la entre 1,00 i 4,00. Aquesta escala sovint s'utilitza com a mesura del malestar psicològic.	A puntuación do HSCL-25 calcúlase dividiendo a puntuación total (a suma de todas as preguntas) entre o número de respostas (cuxa puntuación oscila entre 1,00 e 4,00). Úsase de forma habitual para medir o nivel del malestar psicológico.	Il punteggio del HSCL-25 si calcola dividendo il punteggio totale (somma dei punteggi degli elementi) con il numero di elementi (risposti che variano da 1,00 a 4,00). Spesso si usa come misura di ansietà.	Le score du HSCL-25 se calcule en divisant la somme des cotation des propositions par le nombre de réponses reçues. Le résultat final est compris entre 1,00 à 4,00. Il est couramment utilisé pour mesurer la souffrance psychologique.
The patient is considered as a "probable psychiatric case" if the mean rating on the HSCL-25 is ≥ 1.55 .	Ο ασθενής θεωρείται σαν "πιθανό ψυχιατρικό περιστατικό" εάν η μέση βαθμολογία του ΗΣΓΛ-25 είναι $\geq 1,55$.	Pacjenta uważamy za "prawdopodobny przypadek psychiatryczny" jeśli średnia ocena w teście HSCL-25 jest $\geq 1,55$. (wieszka lub równa) 1,55.	Пациент се смята като "вероятно психиатричен случай," ако средната оценка по HSCL-25 е $\geq 1,55$.	Pacijent se smatra kao "vjerojatno psihijatrijskim slučajem" ako je srednja vrijednost na HSCL-25 $\geq 1,55$.	El/la paciente se considera un "probable caso psiquiátrico" si el valor medio del HSCL-25 es $\geq 1,55$.	El/la pacient és considerat/considerada com a "probable cas psiquiàtric" si la qualificació mitjana del HSCL-25 és $\geq 1,55$.	Considérase que o/a paciente é un "caso psiquiátrico probable" se o valor medio do HSCL-25 é $\geq 1,55$.	Il paziente è considerato come un "probabile caso psichiatrico" se il punteggio medio dell'HSCL-25 è $\geq 1,55$.	Le patient est considéré comme « probablement atteint d'un trouble psychiatrique » si le score moyen du HSCL-25 est supérieur ou égal à 1,55.

(Continued)

TABLE 6 | Continued

Scale instructions original version	Greece	Poland	Bulgaria	Croatia	Spain	Catalonia	Galicia	Italy	France
A cut-off value of ≥ 1.75 is generally used for diagnosis of major depression defined as "a case, in need of treatment." This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender.	Το όριο του $\geq 1,75$ γενικά χρησιμοποιείται για τη διάγνωση της μείζονος κατάθλιψης που ορίζεται ως "περίπτωση που χρήζει φαρμάκων." Αυτό το όριο συνίσταται σαν ένας έγκυρος προγνωστικός δείκτης ψυχικής διαταραχής, όπως εκτιμάται ανεξάρτητα από την κλινική εξέταση, η οποία εφαρμόζεται από τον ψυχίατρο και το φύλο.	Wartość graniczna $\geq 1,75$ (większa lub równa) 1,75 ogólnie przyjmuje się w diagnozowaniu ciężkiej depresji, definiowanej jako „przypadek wymagający leczenia.” Wartość ta jest zalecana jako istotny czynnik w przewidywaniu obecności choroby psychicznej, wymagającej jednak niezależnego wywiadu klinicznego i pewnym sensie zależy od rozpoznania i płci.	Границна стойност от 1,75 обикновено се използва за диагностичране на тежка депресия и определя случая като "случай, нуждаещ се от лечение". Тази границна стойност, получена независимо от клиничното интервю и зависеща до определена степен от диагностича и пола, се препоръчва като валиден предиктор за психично разстройство.	Razdjelna točka (cut-off) $\geq 1,75$ se koristi za dijagnozu velikog depresivnog poremećaja i to kao slučaj koji zahtjeva liječenje." Razdjelna točka se preporuča kao validni prediktor mentalnog poremećaja pojednako kao i sama procjena neovisnim kliničkim intervjuom, dijelom ovisan o određenoj dijagnozi i spolu.	Por lo general se usa un valor de corte de $\geq 1,75$ para el diagnóstico de depresión mayor, definida como "un caso que necesita tratamiento." Este valor de corte se considera un predictor válido de un trastorno mental, evaluado de forma independiente mediante entrevista clínica, aunque depende en parte del diagnóstico y el género.	Generalment s'utilitza un punt de tall $\geq 1,75$ per al diagnòstic de la depressió major i es defineix com "cas que precisa de tractament." Es recomana aquest punt de tall com un predictor vàlid de trastorn mental l'avaluació independent per entrevista clínica, dependent en part del diagnòstic i del gènere.	Polo xeral, úsase un valor de corte $\geq 1,75$ para diagnosticar a depresión maior, definida como "un caso que precisa tratamento." Este valor de corte recoméndase como un predictor válido dun trastorno mental, avaliado independentemente por medio de entrevistas clínicas, aínda que depende en parte do diagnóstico e do xénero.	Un cut-off che sia $\geq 1,75$ è normalmente usato per la diagnosi di depressione maggiore definita come "un caso che necessita di trattamento." Questo cut-off è raccomandato come un valido predittore di disordine mentale come valutato in modo indipendente da un colloquio clinico, dipendente in qualche modo dalla diagnosi e dal genere	Un score supérieur ou égal à 1,75 diagnostique généralement une dépression caractérisée et définit « un patient nécessitant un traitement ». Ce seuil est considéré comme un score prédictif valide des troubles mentaux. Il a été évalué de manière indépendante par des études cliniques. Il varie peu quelles que soient les situations diagnostiques et le sexe.
The time of HSCL-25 is 5 to 10 minutes.	Ο χρόνος χορήγησης του HSCL-25 είναι 5 έως 10 λεπτά.	Czas na wykonanie testu HSCL-25 wynosi od 5 do 10 minut.	Времето за провеждане на HSCL-25 е от 5 до 10 минути.	Vrijeme za ispunjavanje HSCL-25 je 5-10 minuta.	El tiempo de administración del HSCL-25 es de 5 a 10 minutos.	El temps d'administració del HSCL-25 és de 5 a 10 minuts.	O tempo de realização do HSCL-25 é de 5 a 10 minutos.	Il tempo di somministrazione dell'HSCL-25 è da 5 a 10 minuti.	Remplir le questionnaire HSCL-25 prend entre 5 et 10 minutes.

translated into “auto questionnaire” and “Not at all” by “Pas du tout d'accord.”

Interestingly, there were translation similarities (often with stronger meanings or medical connotations) not only among languages belonging to the same linguistic group, but also among languages from different groups. The best example concerns item #3 “Faintness” that was translated into “Weakness” in Catalan, Standard Spanish, Galician, and also in Croatian, a term with a more prosaic than medical connotation.

At the end of the cultural analysis, the consensus group finally concluded that the meaning was not changed, and the translation was finalized in all nine languages (see **Tables 4–6**).

DISCUSSION

Using a three-step qualitative procedure, ecologically embedded in primary care, nine consensual translations of the HSCL-25 were obtained that were linguistically and culturally equivalent to the original version, in three language families (Hellenic, Slavic, and Romance). A German version already existed. The aim of this procedure was to meticulously track inconsistencies between local translations that could lead to misinterpretation. This methodical and transcultural validation ensured the transfer of the same content from one language to another and its reliability (17, 47).

The Greek translation remained the most stable, followed by Bulgarian. Item #17, “Feeling blue” was the most challenging to translate, followed by item #3 “Faintness” and item #5 “Heart racing.” Some scales needed adaptations in terms of tense (French, Croatian) and in terms of gender (Greek, Italian, and Hispanic languages).

Research and Teaching Implications

Translation remains the most crucial step in the adoption of an instrument developed in another nation using a different language. Errors in translation may distort the original intent of the instrument, thus compromising its validity and reliability (48). Semantic issues might affect comparability in international studies because the same word is interpreted differently across countries and cultures (49, 50). Moreover, some terms and concepts may not exist in other languages or may have additional connotations that backward translations do not always reveal. Challenges arise not only because of the word-to-word literal translation, but also because of the linguistic form of the language, such as tone and syntax (51).

These nine translations of the HSCL-25 are now linguistically similar, in terms of meaning, compared to the original version. However, they need further testing because this first step is not sufficient to complete the task of translating them and supporting their cross-cultural validity. The external and internal validity of each version has to be evaluated to ensure that their reliability is comparable with that of the original version. This will be achieved through quantitative studies in primary care daily practices (52).

In most European countries, FPs can now use this tool for family practice research studies and for assessing depression

severity in their patients. The use of such a shared tool may have a great impact on the feasibility of future research on depression in primary care. It will facilitate data comparison among European countries and consequently it will allow statistical reviews on depression epidemiology and symptoms throughout Europe. The use of the same instrument can support the conceptualization of the studied phenomenon across different studies, and the findings can then be compared (21).

LIMITATIONS

A key point of this study was the FPs' involvement in the translation to reduce the selection bias and to ensure the sample quality nevertheless as in all formalized expert consensus procedure a selection bias of the experts remained possible. Our experts' sample was constructed purposively and if we did our best to avoid a selection bias it remained possible. As described by many translators when discussing scientific translation work, a “specialist” in the field (e.g., primary care daily practice in this case) should take a last look at the translation (20, 53, 54) and become the main arbiter of the quality of the final translation (55). Thus, specific attention was paid in choosing FP researchers and certified bilingual translators with sufficient knowledge of healthcare terminology a selection bias was still possible.

The cultural control check was as consistent as possible. It involved a careful step-by-step analysis to prevent confusion bias and linguistic problems. The formalized consensus method allowed the gradual evaluation of each item to strengthen the accuracy of the validated translations and designing the end-result. Nevertheless, an information or a confusion bias remained possible. Our results should be interpreted in the light of these limitations.

CONCLUSION

A translation of the HSCL-25 in which homogeneity is ensured is now available for Spain and its culturally distinct regions of Galicia and Catalonia, and also for France, Greece, Italy, Poland, Bulgaria, and Croatia. It is now ready to be tested in actual and representative primary care populations to further validate its test-parameters.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by CPP (Protection of Persons Committee) of the University Hospital of Brest. Reference CPP: CPP Ouest VI 872; Study ID RCB: n°2014-A01790-47. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PN designed the study, collected data, led meetings, drafted the article, and submitted it for publication. JL designed the study, collected data, attended meetings, and reviewed the article. MG-L and BL reviewed the article. RA, DK, SC, MH, HL, AC, MR-B, AS, SA, and CL participated as national investigator. SS-S participated as co-national investigator. TM reviewed the article and gave final approval for the version to be published. HV and PV designed the study, reviewed the article, and gave final approval for the version to be published. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Institute for Health Research (NIHR) Applied Research Collaboration Kent, Surrey, Sussex.

ACKNOWLEDGMENTS

We would like to thank all FPs who participated in the research process throughout Europe and all trainees in Family Medicine from Brest University who participated in the research process and our proof-reader Mrs. Alex Gillman for her accurate translations.

REFERENCES

- King M, Nazareth I, Levy G, Walker C, Morris R, Weich S, et al. Prevalence of common mental disorders in general practice attendees across Europe. *Br J Psychiatry*. (2008) 192:362–7. doi: 10.1192/bjp.bp.107.039966
- Ayuso-Mateos JL, Vázquez-Barquero JL, Dowrick C, Lehtinen V, Dalgard OS, Casey P, et al. Depressive disorders in Europe: prevalence figures from the ODIN study. *Br J Psychiatry*. (2001) 179:308–16. doi: 10.1192/bjp.179.4.308
- Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl*. (2004) 109:21–7. doi: 10.1111/j.1600-0047.2004.00327.x
- Torzsa P, Szeifert L, Dunai K, Kalabay L, Novák M. Diagnosis and therapy of depression in family practice. *Orv Hetil*. (2009). 150:1684–93. doi: 10.1556/oh.2009.28675
- Jorm AF. Mental health literacy. Public knowledge and beliefs about mental disorders. *Br J Psychiatry*. (2000) 177:396–401. doi: 10.1192/bjp.177.5.396
- Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet*. (2009) 374:609–19. doi: 10.1016/S0140-6736(09)60879-5
- Ani C, Bazargan M, Hindman D, Bell D, Farooq MA, Akhanjee L, et al. Depression symptomatology and diagnosis: discordance between patients and physicians in primary care settings. *BMC Fam Pract*. (2008) 9:1. doi: 10.1186/1471-2296-9-1
- Cuellar I, Paniagua FA. *Handbook of Multicultural Mental Health: Assessment and Treatment of Diverse Populations*. San Diego, CA: Academic Press (2000).
- Marsella AJ, Yamada AM. Culture and mental health: An introduction and overview of foundations, concepts, and issues. In: Cuellar I, Paniagua FA, editors. *Handbook of Multicultural Mental Health*. San Diego, CA: Academic Press (2000). p. 3–24.
- Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Psychotropic drug utilization in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. *Acta Psychiatr Scand Suppl*. (2004) 109:38–46. doi: 10.1111/j.1600-0047.2004.00331.x
- Wells KB. Caring for depression in primary care: defining and illustrating the policy context. *J Clin Psychiatry*. (1997) 58(Suppl 1):24–7.
- Schoenbaum M, Unützer J, McCaffrey D, Duan N, Sherbourne C, Wells KB. The effects of primary care depression treatment on patients' clinical status and employment. *Health Serv Res*. (2002) 37:1145–58. doi: 10.1111/1475-6773.01086
- Woltmann E, Grogan-Kaylor A, Perron B, Georges H, Kilbourne AM, Bauer MS. Comparative effectiveness of collaborative chronic care models for mental health conditions across primary, specialty, and behavioral health care settings: systematic review and meta-analysis. *Am J Psychiatry*. (2012) 169:790–804. doi: 10.1176/appi.ajp.2012.11111616
- Lehti A, Hammarström A, Mattsson B. Recognition of depression in people of different cultures: a qualitative study. *BMC Fam Pract*. (2009) 10:53. doi: 10.1186/1471-2296-10-53
- Kirmayer LJ, Robbins JM, Dworkind M, Yaffe MJ. Somatization and the recognition of depression and anxiety in primary care. *Am J Psychiatry*. (1993). 150:734–41. doi: 10.1176/ajp.150.5.734
- Steinert C, Hofmann M, Kruse J, Leichsenring F. The prospective long-term course of adult depression in general practice and the community. A systematic literature review. *J Affect Disord*. (2014) 152, 65–75. doi: 10.1016/j.jad.2013.10.017
- Bullinger M, Anderson R, Cella D, Aaronson N. Developing and evaluating cross-cultural instruments from minimum requirements to optimal models. *Qual Life Res*. (1993) 2:451–9. doi: 10.1007/BF00422219
- Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. *J Biomed Inform*. (2004) 37:193–204. doi: 10.1016/j.jbi.2004.04.004
- Mitchell AJ. Clinical utility of screening for clinical depression and bipolar disorder. *Curr Opin Psychiatry*. (2012) 25:24–31. doi: 10.1097/YCO.0b013e32834de45b
- Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. (2000) 25:3186–91. doi: 10.1097/00007632-200012150-00014
- Brislin RW. Comparative research methodology: cross-cultural studies. *Int J Psychol*. (1976) 11:215–29. doi: 10.1080/00207597608247359
- Jones PS, Lee JW, Phillips LR, Zhang XE, Jaceldo KB. An adaptation of Brislin's translation model for cross-cultural research. *Nurs Res*. (2001) 50:300–4. doi: 10.1097/00006199-200109000-00008
- Nabbe P, Le Reste JY, Guillou-Landreat M, Munoz Perez MA, Argyriadou S, Claveria A, et al. Which DSM validated tools for diagnosing depression are usable in primary care research? A systematic literature review. *Eur Psychiatry*. (2017) 39:99–105. doi: 10.1016/j.eurpsy.2016.08.004
- Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, et al. *The RAND/UCLA Appropriateness Method User's Manual* (2001).
- Fröjd K, Håkansson A, Karlsson I, Fröjd K, Hakansson A. The Hopkins Symptom Checklist-25 is a sensitive case-finder of clinically important depressive states in elderly people in primary care. *Int J Geriatr Psychiatry*. (2004) 19:386–90. doi: 10.1002/gps.1102
- Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci*. (1974) 19:1–15. doi: 10.1002/bs.3830190102
- Nettelbladt P, Hansson L, Stefansson CG, Borgquist L, Nordström G. Test characteristics of the Hopkins Symptom Check List-25 (HSCL-25) in Sweden, using the Present State Examination (PSE-9) as a caseness criterion. *Soc Psychiatry Psychiatr Epidemiol*. (1993) 28:130–3. doi: 10.1007/BF00801743
- Sandanger I, Moum T, Ingebrigtsen G, Dalgard OS, Sørensen T, Bruusgaard D. Concordance between symptom screening and diagnostic procedure: the Hopkins symptom checklist-25 and the composite international diagnostic interview I. *Soc Psychiatry Psychiatr Epidemiol*. (1998) 33:345–54. doi: 10.1007/s001270050064
- Ventevogel P, De Vries G, Scholte WF, Shinwari NR, Faiz H, Nassery R, et al. Properties of the Hopkins symptom checklist-25 (HSCL-25) and the Self-Reporting Questionnaire (SRQ-20) as screening instruments used

- in primary care in Afghanistan. *Soc Psychiatry Psychiatr Epidemiol.* (2007) 42:328–35. doi: 10.1007/s00127-007-0161-8
30. Moum T. Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression. *Soc Indic Res.* (1998) 45:279–318. doi: 10.1023/A:1006958100504
31. Lipman RS, Covi' L, Shapiro AK. THE HOPKINS SYMPTOM CHECKLIST (HSCL) factors derived from the HSCL-90. *J Affect Disord.* (1979) 1:9–24. doi: 10.1016/0165-0327(79)90021-1
32. Spooner D, Pachana N. Ecological validity in neuropsychological assessment: a case for greater consideration in research with neurologically intact populations. *Arch Clin Neuropsychol.* (2006) 21:327–37. doi: 10.1016/j.acn.2006.04.004
33. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. *J Eval Clin Pract.* (2011) 17:268–74. doi: 10.1111/j.1365-2753.2010.01434.x
34. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. *J Adv Nurs.* (2004) 48:175–86. doi: 10.1111/j.1365-2648.2004.03185.x
35. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ Clin Res Ed.* (1995) 311:376–80. doi: 10.1136/bmj.311.7001.376
36. Haute Autorité Santé. *Bases Méthodologiques Pour l'élaboration de Recommandations Professionnelles Par Consensus Formalisé.* HAS (2006).
37. Linstone HA, Turoff M. *The Delphi Method: Techniques And Applications.* Advanced Book Program. Reading, MA: Addison-Wesley Pub. Co. (1975).
38. Hasson F, Keeney S, McKenna H. Research guidelines for the delphi survey technique. *J Adv Nurs.* (2000) 32:1008–15. doi: 10.1046/j.1365-2648.2000.t01-1-01567.x
39. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol.* (2003) 56:1150–6. doi: 10.1016/S0895-4356(03)00211-7
40. Hassan T, Barnett D. Delphi type methodology to develop consensus on the future design of EMS systems in the United Kingdom. *Emerg Med J.* (2002) 19:155–9. doi: 10.1136/emj.19.2.155
41. De Villiers MR, De Villiers PJT, Kent AP. The delphi technique in health sciences education research. *Med Teach.* (2005) 27:639–43. doi: 10.1080/13611260500069947
42. Bourrée F, Michel P, Salmi LR. Consensus methods: review of original methods and their main alternatives used in public health. *Rev Epidemiol Sante Publiq.* (2008) 56:e13–21. doi: 10.1016/j.respe.2008.10.005
43. Powell C. The Delphi technique: myths and realities. *J Adv Nurs.* (2003) 41:376–82. doi: 10.1046/j.1365-2648.2003.02537.x
44. Romm FJ, Hulka BS. Developing criteria for quality of assessment: effect of the Delphi technique. *Health Serv Res.* (1979) 14:309–12.
45. Jamieson S. Likert scales: how to (ab)use them. *Med Educ.* (2004) 38:1217–8. doi: 10.1111/j.1365-2929.2004.02012.x
46. Anadón M, Guillemette F. La recherche qualitative est-elle nécessairement inductive? *Rech Qual.* (2007) 5:26–37.
47. Herdman M, Fox-Rushby J, Badia X. "Equivalence" and the translation and adaptation of health-related quality of life questionnaires. *Qual Life Res.* (1997) 6:237–47.
48. Yu DSF, Lee DTF, Woo J. Issues and challenges of instrument translation. *West J Nurs Res.* (2004) 26:307–320. doi: 10.1177/0193945903260554
49. Daugherty JC, Puente AE, Fasfous AF, Hidalgo-Ruzzante N, Pérez-García M. Diagnostic mistakes of culturally diverse individuals when using North American neuropsychological tests. *Appl Neuropsychol.* (2017) 24:16–22. doi: 10.1080/23279095.2015.1036992
50. Schnohr CW, Gobina I, Santos T, Mazur J, Alikasifoglu M, Välimaa R, et al. Semantics bias in cross-national comparative analyses: is it good or bad to have "fair" health? *Health Qual Life Outcomes.* (2016) 14:70. doi: 10.1186/s12955-016-0469-8
51. Hanrahan D, Sexton P, Hui K, Teitcher J, Sugarman J, London AJ, et al. Linguistic and cultural challenges in communication and translation in usponsored HIV Prevention research in emerging economies. *PLoS ONE.* (2015). 10:e0133394. doi: 10.1371/journal.pone.0133394
52. Nabbe P, Le Reste JY, Guillou-Landreat M, Gatineau F, Le Floch B, Montier T, et al. The French version of the HSCL-25 has now been validated for use in primary care. *PLoS ONE.* (2019) 14:e0214804. doi: 10.1371/journal.pone.0214804
53. Skulmoski GJ, Hartman FT, Krahn J. The delphi method for graduate research. *J Inf Technol Educ.* (2007) 6:1. doi: 10.28945/199
54. Vesga O, Agudelo M, Salazar BE, Rodriguez CA, Zuluaga AF. Generic vancomycin products fail *in vivo* despite being pharmaceutical equivalents of the innovator. *Antimicrob Agents Chemother.* (2010) 54:3271–9. doi: 10.1128/AAC.01044-09
55. Balliu C. L'enseignement de la traduction médicale: pour une nouvelle pragmatique. *Meta J Traduct.* (1994) 39:15–25. doi: 10.7202/001964ar

Author Disclaimer: The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Nabbe, Le Reste, Guillou-Landreat, Assenova, Kasuba Lazić, Czachowski, Stojanović-Špehar, Hasanagic, Lingner, Clavería, Rodríguez-Barragán, Sowinska, Argyriadou, Lygidakis, Le Floch, Montier, Van Marwijk and Van Royen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

ANEXO 5 Versión original, traducción directa y retrotraducción de la HSCL-25

Delphi Translation form for HSCL-25 (to translate into your native language)



Département Universitaire de Médecine Générale

22, avenue Camille Desmoulins CS 93837 – 29238 – Brest CEDEX 3
Tél: 02 98 01 65 52 – fax : 02 98 01 64 74

The translation of the Local group (use your nation's name) is: Castellano

Elija la respuesta que mejor describa cómo se ha sentido durante la semana pasada:

Original: Choose the best answer for how you felt over the past week

Back translation: Choose the best answer to describe how you felt over the last week

1.- Se asusta sin motivo

Original: Being scared for no reason

Back translation: Being scared for no reason

2.- Siente miedo

Original: Feeling fearful

Back translation: Feeling afraid/fearful

3.- Debilidad

Original: Faintness

Back translation: Weakness

4.- Nerviosismo

Original: Nervousness

Back translation: Nervousness

5.- Palpitaciones

Original: Heart racing

Back translation: Palpitations

6.- Tiembla

Original: Trembling

Back translation: Trembling

7.- Se siente tenso

Original: Feeling tense

Back translation: Feeling tense

8.- Dolor de cabeza

Original: Headache

Back translation: Headache

9.- Siente pánico

Original: Feeling panic

Backtranslation: Feeling panic

10.- Siente inquietud

Original: Feeling restless

Back translation: Feeling restless

11.- Siente que le falta energía

Original: Feeling low in energy

Back translation: Feeling a lack of energy

12.- Se culpa a sí mismo

Original: Blaming oneself

Back translation: Blaming oneself

13.- Llora con facilidad

Original: Crying easily

Back translation: Crying easily

14.- Pierde el interés sexual

Original: Losing sexual interest

Back translation: Losing interest in sex

15.- Se siente solo

Original: Feeling lonely

Back translation: Feeling alone

16.- Se siente sin esperanza

Original: Feeling hopeless

Back translation: Feeling hopeless

17.- Se siente triste

Original: Feeling blue

Back translation: Feeling sad

18.- Piensa en acabar con su vida

Original: Thinking of ending one's life

Back translation: Thinking of ending your life

19.- Se siente atrapado.

Original: Feeling trapped

Back translation: Feeling trapped

20.- Se preocupa en exceso

Original: Worrying too much

Back translation: Worrying too much

21.- No siente interés por nada

Original: Feeling no interest

Back translation: Not interested in anything

22.- Siente que todo le cuesta un esfuerzo

Original: Feeling that everything is an effort

Back translation: Feeling that everything is an effort

23.- Se siente inútil

Original: Worthless feeling

Back translation: Feeling useless

24.- Falta de apetito

Original: Poor appetite

Back translation: Poor appetite

25.- Problemas para dormir

Original: Sleep disturbance

Back translation: Problems sleeping

La puntuación HSCL-25 se basa en un cuestionario autocumplimentado con lápiz y papel, de 25 preguntas sobre la presencia y la intensidad de ansiedad y síntomas depresivos en la última semana. Los participantes responden una de cuatro categorías para cada ítem, en una escala de cuatro puntos que van desde 1 a 4.

Original: The HSCL-25 score is based on pencil-and-paper self-report of 25 questions about the presence and intensity of anxiety and depression symptoms over the last week.

Participants answer to one of four categories for each item on a four-point scale ranging from 1 to 4.

Back translation: The HSCL-25 score is based on a survey that is self-administered with pen and paper. It has 25 questions about the presence and intensity of anxiety and depressive symptoms over the last week.

The participants answer one of the four categories for each item on a four-point scale ranging from 1 to 4.

"En absoluto"

Original: "Not at all"

Back translation: Not at all

" Un poco"

Original: "A little"

Back translation: A little

"Bastante"

Original: "Quite a bit":

Back translation: Quite a bit

“Mucho”

Original: “Extremely”

Back translation: A lot

La puntuación del HSCL-25 se calcula dividiendo la puntuación total (sumando la puntuación de todas las preguntas) entre el número de respuestas (varía entre 1,00 y 4,00). Se usa habitualmente para medir el malestar psicológico. El paciente se considera un “probable caso psiquiátrico” si el valor medio del HSCL-25 es $\geq 1,55$. Por lo general se usa un valor de corte de $\geq 1,75$ para el diagnóstico de depresión mayor, definida como “un caso que necesita tratamiento”. Este valor de corte se considera un predictor válido de un trastorno mental, evaluado de forma independiente mediante entrevista clínica, aunque depende en parte del diagnóstico y el género.

El tiempo de administración del HSCL-25 es de 5 a 10 minutos.

Original: The HSCL-25 score is calculated by dividing the total score (sum score of items) by the number of items answered (ranging between 1,00 and 4,00). It is often used as the measure of distress.

The patient is considered as a “probable psychiatric case” if the mean rating on the HSCL-25 is $\geq 1,55$.

A cut-off value of $\geq 1,75$ is generally used for diagnosis of major depression defined as “a case, in need of treatment”. This cut-off point is recommended as a valid predictor of mental disorder as assessed independently by clinical interview, somewhat depending on diagnosis and gender.

The administration time of HSCL 25 is 5 to 10 minutes.

Back translation: The HSCL-25 score is calculated dividing the total score (the sum of the scores of all the questions) by the number of items answered (from 1 to 4). It is usually used to measure the level of anxiety.

The patient is considered a “probable psychiatric case” if the average of the HSCL-25 score is $\geq 1,55$. A cut-off point $\geq 1,75$ is generally used for the diagnosis of major depression, defined as “a case that needs treatment”. This cut-off is considered a valid predictor of mental disorder, assessed in an independent way by a clinical interview, although it depends, in part, on the diagnosis and gender.

The administration time of HSCL-25 is from 5 to 10 minutes.

ANEXO 6 Versión HSCL-25 en español utilizada en la Fase 2

HSCL-25 en español				
Elija la respuesta que mejor describa cómo se ha sentido durante la semana pasada				
	1= En absoluto	2= Un poco	3= Bastante	4 = Mucho
1. Se asusta sin motivo				
2. Siente miedo				
3. Debilidad				
4. Nerviosismo				
5. Palpitaciones				
6. Tiembla				
7. Se siente tenso/a				
8. Dolor de cabeza				
9. Siente pánico				
10. Siente inquietud				
11. Siente que le falta energía				
12. Se culpa a sí mismo/a				
13. Lloro con facilidad				
14. Pierde el interés sexual				
15. Se siente solo/a				
16. Se siente sin esperanza				
17. Se siente triste				
18. Piensa en acabar con su vida				
19. Se siente atrapado/a				
20. Se preocupa en exceso				
21. No siente interés por nada				
22. Siente que todo le cuesta un esfuerzo				
23. Se siente inútil				
24. Poco apetito				
25. Problemas para dormir				

ANEXO 7 Composite International Diagnostic Interview (CIDI)

- E1. Ahora quiero preguntarle sobre periodos en que se ha sentido triste, vacío(a) o deprimido(a). En los últimos 12 meses, ¿alguna vez ha tenido usted dos semanas o más en que casi todos los días se sintió triste, vacío(a) o deprimido(a) la mayor parte del día? No Sí
- E2. En los últimos 12 meses, ¿alguna vez ha tenido 2 semanas o más en que perdió el interés en la mayoría de las cosas, como el trabajo, los pasatiempos y otras cosas que normalmente disfrutaba? No Sí

Solo si alguna de las dos es “afirmativa” es necesario continuar con el cuestionario CIDI.

	I (X) (alguna vez en los últimos 12 meses)	II (Y) (cuando mayoría sx)
SIN ENERGÍA		
E3. Durante un periodo de al menos dos semanas o más, en los últimos 12 meses, cuando se sintió triste, vacío(a) o deprimido(a)/perdió el interés en las cosas, [<u>¿se sintió sin energías o cansado(a) todo el tiempo casi todos los días, aun cuando no había estado trabajando mucho?</u>	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
CAMBIO DE APETITO		
E4. Hablemos de otros problemas que tuvo durante los periodos cuando se sintió deprimido(a)/perdió interés en las cosas/ se sintió cansado(a) todo el tiempo durante dos semanas o más, en los últimos 12 meses. [<u>¿Tenía menos apetito de lo usual casi todos los días?</u>	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E5. Durante uno de esos periodos, [<u>¿perdió peso sin quererlo alrededor de un kilo por semana durante varias semanas?</u>	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí SI SE CODIFICÓ “1” PREGUNTE A. SI CODIFICÓ “0”, PASE AL PRÓXIMO CÓDIGO. ___/___ Kg
SI SE CODIFICÓ “0” EN E5, PASE A E6 A. ¿Cuánto peso perdió? ___/___ Kg	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí ___/___ Kg
E6. Durante uno de esos periodos, [<u>¿tenía mucho más apetito de lo usual todos los días por dos semanas o más?</u>	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
CODIFIQUE –“0” - SOLO POR PERIODO DE RÁPIDO CRECIMIENTO O EMBARAZO		

	I (X) (alguna vez en los últimos 12 meses)	II (Y) (cuando mayoría sx)
<p>E7. Durante uno de esos periodos, [¿su apetito aumentó tanto <u>que subió de peso</u>, hasta un kilo por semana durante varias semanas?</p> <p>CODIFIQUE –“0”- SI SOLO RECUPERÓ LOS KILOS PERDIDOS</p> <p>SI SE CODIFICÓ “0” EN E7, PASE A E8</p> <p>A. ¿Cuánto peso aumentó?</p> <p>___/___ Kg</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p> <p>SI SE CODIFICÓ “1” PREGUNTE A. SI CODIFICÓ “0”, PASE AL PRÓXIMO CÓDIGO.</p> <p>___/___ Kg</p>
PROBLEMAS PARA DORMIR		
<p>E8. Cuando se sentía deprimido(a)/había perdido interés en las cosas/estaba sin energía, en los últimos 12 meses, [¿tuvo <u>problemas para dormir</u> casi todas las noches durante dos semanas o más, tales como dificultad para dormirse, despertarse durante la noche o demasiado temprano?</p> <p>SI SE CODIFICÓ “0” EN E8, PASE A E9</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
<p>E8A. [¿Se despertó <u>por lo menos dos horas antes de lo deseado</u> todos los días por lo menos durante dos semanas?</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
<p>E9. Durante un periodo en que se sintió deprimido(a)/perdió el interés en las cosas/estaba sin energía, [¿<u>dormía demasiado</u> casi todos los días?</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
LENTITUD/INQUIETUD		
<p>E10. Durante uno de esos periodos, en los últimos 12 meses, [¿<u>hablaba o se movía más lento</u> (despacio) de lo normal para usted, casi todos los días, durante al menos dos semanas?</p> <p>SI SE CODIFICÓ “0”, PASE A E11</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
<p>E10A. ¿Notó alguien que usted hablaba o se movía más lento (despacio)?</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
<p>E11. Durante uno de esos periodos, [¿<u>tenía que estar en continuo movimiento</u>, es decir, no podía sentarse quieto(a) y andaba de un lado para otro, o no podía mantener las manos quietas cuando estaba sentado(a)?</p> <p>SI SE CODIFICÓ “0”, PASE A E12</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
<p>E11A. ¿Se dio cuenta alguien de que usted se movía constantemente?</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>	<p><input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí</p>
CULPABILIDAD O PÉRDIDA DE AUTOESTIMA		

	I (X) (alguna vez en los últimos 12 meses)	II (Y) (cuando mayoría sx)
E12. Durante uno de esos periodos, en los últimos 12 meses, [¿sentía que <u>valía poco o nada</u> casi todos los días?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E12A. ¿Se sentía culpable?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
SI SE CODIFICÓ "0", EN E12 Y E12A, PASE A E13		
E12B. ¿Hubo alguna razón en particular para sentirse (que valía poco o nada/culpable)?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E12BXA SOLO PARA COL. I: ANOTE EL EJEMPLO:		
E12BYB SOLO PARA COL. II: ANOTE EL EJEMPLO:		
E12C. ¿SENTÍA EL ENTREVISTADO QUE VALÍA POCO O NADA/CULPABLE <u>SOLAMENTE</u> POR ESTAR DETERIORADO POR LA DEPRESIÓN?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
FALTA DE CONFIANZA EN SÍ MISMO		
E13. Durante uno de esos periodos, en los últimos 12 meses, [¿sentía que <u>no era tan bueno(a) o era inferior a otras personas</u> ?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E14. [¿Tenía <u>tan poca confianza en si mismo(a)</u> que no se atrevía a opinar sobre nada?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
DIFICULTAD PARA PENSAR		
E15. Durante uno de esos periodos, en los últimos 12 meses, [¿tuvo mucha más <u>dificultad para concentrarse de lo acostumbrado</u> ?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E15A. [¿Era incapaz de leer cosas que normalmente le interesaban, o <u>escuchar la radio o mirar televisión o películas que normalmente le gustaban</u> , porque no podía prestarles atención?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E16. [¿Le vinieron sus <u>pensamientos mucho más lento de lo usual</u> o le parecían confusos?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E17. [¿Era <u>incapaz de tomar decisiones</u> sobre cosas que normalmente no le cuesta trabajo decidir?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
PENSAMIENTOS SOBRE LA MUERTE		
E18. Durante uno de esos periodos, en los últimos 12 meses, [¿pensó mucho en la muerte?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E19. [¿Se sintió tan decaído(a) que pensó mucho en <u>suicidarse</u> ?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
SI SE CODIFICÓ "0", PASE A E22.		
E19A. [¿Planeó cómo podría hacerlo?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E20. [¿Intentó suicidarse?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E21. ESTA PREGUNTA FALTA INTENCIONADAMENTE		

	I (X) (alguna vez en los últimos 12 meses)	II (Y) (cuando mayoría sx)
E22. Durante un periodo de dos semanas de sentirse deprimido(a), haber perdido interés/estar sin energía, [en la mayoría de los días, se sintió particularmente mal <u>al momento de levantarse</u> pero a medida que pasaba el día se sentía mejor?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E23. Durante uno de esos periodos, [¿su <u>interés por el sexo fue mucho menos de lo acostumbrado</u> ?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E24. [¿Ha perdido la capacidad de disfrutar las cosas buenas que le pudieran pasar, como ganar algo o que lo feliciten o elogien?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí
E25. ESTA PREGUNTA FALTA INTENCIONADAMENTE		
E26. En los últimos 12 meses, ¿cuánto duró el periodo más largo que ha tenido de ese tipo? SI DURÓ LOS DOCE MESES, CODIFIQUE 48 (Nº DE MESES X 4 = Nº DE SEMANAS). SI EL NÚMERO DE SEMANAS ES 1, FIN DE LA ENTREVISTA.	____/____ SEMANAS	
E26A. Algún periodo de dos semanas o más, en los últimos 12 meses, ¿interfirió seriamente con su capacidad para trabajar, cuidar de su familia o cuidar de su persona?	<input type="checkbox"/> 0. No	<input type="checkbox"/> 1. Sí
E27. INI/REC: ¿Cuándo fue la (primera/última vez), en los últimos 12 meses, que tuvo un periodo de dos semanas o más en el que se sintió triste, vacío(a) o deprimido(a)/perdió el interés/estuvo sin energía y también tuvo alguno de estos problemas como: (SX CODIFICADOS –SI (1)- DE E4-E20)? El tiempo debe codificarse de este modo: 1= en las últimas 2 semanas 2= 2 semanas y < 1mes 3= 1 mes y < 6 meses 4 = 6 meses y menos de un año 5= en los últimos 12 meses 6= hace más de un año. Entonces pregunte: ¿cuántos años tenía usted?	INI: 1 2 3 4 5 6 REC: 1 2 3 4 5 6 EDAD INI: <input type="checkbox"/> <input type="checkbox"/>	
E28. ESTA PREGUNTA FALTA INTENCIONADAMENTE		
E29. En los últimos 12 meses, ¿cuántos periodos distintos ha tenido de dos semanas o más en que se sintió deprimido(a)/perdió interés en las cosas/estuvo sin energía y tuvo algunos problemas de los que hemos hablado? SI SE CODIFICÓ 01, PASE A E30	____/____ EPISODIOS	
E29A. Entre (cualquiera de) esos periodos en que se sintió deprimido(a)/perdió interés en las cosas/estuvo sin energías, ¿se sintió bien durante algunos meses?	<input type="checkbox"/> 0. No (PASE A E30) <input type="checkbox"/> 1. Sí, BIEN ENTRE MEDIO	

	I (X) (alguna vez en los últimos 12 meses)	II (Y) (cuando mayoría sx)
E29B. Entre sus periodos de depresión, ¿era capaz de trabajar y disfrutar de la compañía de otras personas, tal como lo hacía antes de que comenzaran?	<input type="checkbox"/> 0. No (PASE A E30) <input type="checkbox"/> 1. Sí	
E29C. Ese periodo, en los últimos 12 meses, en que se sentía bien y disfrutaba de la compañía de otras personas ¿duró por lo menos dos meses?	<input type="checkbox"/> 0. No (PASE A E30) <input type="checkbox"/> 1. Sí	
E29D. En los últimos 12 meses, ¿cuántos periodos distintos de estar deprimido(a)/sin interés en las cosas/sin energía ha tenido, contando sólo aquellos periodos entre los cuales se sintió bien por dos meses o más?	____/____ EPISODIOS CON INTERVALOS DE 2 MESES	
E30. ¿Ocurrió (este periodo/alguno de estos periodos) de (depresión/pérdida de interés), en los últimos 12 meses, poco después de morir alguna persona allegada a usted?	<input type="checkbox"/> 0. No (PREGUNTE E30B) <input type="checkbox"/> 1. Sí	
E30A. En los últimos 12 meses, ¿ha tenido algún periodo de sentirse deprimido(a)/haber perdido interés junto con estos otros problemas (LEA 3 SX CODIFICADOS 1 DE E3 A E20) en otro momento que no fuesen después de la muerte de una persona allegada a usted? SI ES HOMBRE, O MUJER SIN HIJOS, PASE A E31.	<input type="checkbox"/> 0. NO, SOLO DESPUÉS DE MUERTE (PASE A E31) <input type="checkbox"/> 1. Sí, OTRAS VECES	
E30B. ¿Comenzó este periodo/alguno de estos periodos antes que pasara un mes del nacimiento de su hijo o hija?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	
E31. ¿HA TENIDO EL ENTREVISTADO DOS O MAS PERIODOS SEPARADOS DE DEPRESIÓN (E29D CODIFICADO 2 O MÁS)?	<input type="checkbox"/> 0. No, FIN DE LA ENTREVISTA <input type="checkbox"/> 1. Sí	
E32. Ahora quisiera preguntarle sobre el periodo, en los últimos 12 meses, en que se sentía deprimido(a)/perdió el interés/estuvo sin energías durante al menos dos semanas y tuvo el mayor número de estos problemas de los que hemos mencionado, ¿en qué mes fue? (SI NO PUEDE ELEGIR: entonces escoja uno de los peores periodos de dos semanas)	MES _____ <input type="checkbox"/> <input type="checkbox"/>	
E33. Le voy a preguntar cuál de estos problemas tenía, durante estas dos semanas (MES), cuando tenía la mayor cantidad de estos problemas. En ese periodo, ¿se sentía triste y deprimido(a)?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	
E33A. ¿Perdió interés en casi todo?	<input type="checkbox"/> 0. No <input type="checkbox"/> 1. Sí	
REGRESE A E3, COL. II LEA CADA PREGUNTA CODIFICADA 1 EN LA COL. I Y COMIENZE DESDE [. CODIFIQUE EN LA COL. II		

TARJETA DE SÍNTOMAS DE LA SECCIÓN DE DEPRESIÓN DEL CIDI

- E1 Periodos en que se ha sentido triste, vacío/a o deprimido/a**
- E2 Pérdida de interés en la mayoría de las cosas**
- E3 Se sintió sin energías o cansado/a**
- E4 Tenía menos apetito de lo usual**
- E5 Perdió peso**
- E6 Tenía más apetito de lo usual**
- E7 Subió peso**
- E8 Tuvo problemas para dormir**
- E8A Se despertó al menos dos horas antes de lo deseado**
- E9 Dormía demasiado**
- E10 Hablaba o se movía más lento**
- E11 Tenía que estar en continuo movimiento**
- E12 Sentía que valía poco o nada**
- E12A Se sentía culpable**
- E13 Sentía que no era tan bueno/a o era inferior a otras personas**
- E14 Tenía poca confianza en sí mismo/a**
- E15 Tuvo mucha más dificultad para concentrarse**
- E15A Era incapaz de leer o escuchar radio o mirar televisión**
- E16 Le vinieron sus pensamientos mucho más lento de lo usual**
- E17 Era incapaz de tomar decisiones**
- E18 Pensó mucho en la muerte**
- E19 Se sintió tan decaído/a que pensó mucho en suicidarse**
- E20 Intentó suicidarse**
- E22 Se sintió mal al momento de levantarse**
- E23 Su interés por el sexo fue mucho menos de lo acostumbrado**
- E24 Ha perdido la capacidad de disfrutar las cosas buenas que le pudieran pasar**

ANEXO 8 Formación a los entrevistadores de la CIDI

Contenido del programa de formación de 12 h impartido por una médica de familia:

1. Presentación general de la entrevista CIDI
2. Sección exhaustiva de depresión en CIDI (criterios CIE-10 / DSM-IV)
3. Procedimiento de la entrevista con los pacientes
4. Cómo realizar la lectura de la entrevista pregunta por pregunta
5. Registro de las respuestas
6. Rol-playing simulando entrevistas con los pacientes reales
7. Manejo de las situaciones confusas que se pudieran dar en la práctica real.

Énfasis en el apartado 4. Cómo plantear las preguntas del CIDI

- Todas las preguntas se deben hacer exactamente como están escritas, sin añadir contenido.
- Cuando los entrevistados responden antes de que la pregunta termine, la pregunta debe ser leída nuevamente, poniendo énfasis a la última parte.
- Las preguntas largas pueden leerse fraccionadamente en preguntas más cortas.
- Escuchar cuidadosamente las respuestas.
- Cuando se obtiene una respuesta que indica que el sujeto ha comprendido claramente la intención de la pregunta consideraremos que ha contestado correctamente.
- Los entrevistadores no deberían presuponer, intuir o condicionar las respuestas.

Registro de respuestas:

- Las respuestas a las preguntas se codifican en el margen derecho del folio. La mayor parte de la codificación requiere que el entrevistador encierre en un círculo el número asignado a una posible respuesta.
- Otras respuestas requieren anotar un número, utilizando ceros cuando sea necesario rellenar los recuadros vacíos a la izquierda.
- Codificación de "no sabe": El CIDI recoge de manera diferente las respuestas "Yo no sé" cuando se trata de preguntas sobre síntomas o de otro tipo:
 - Las respuestas "no se" a preguntas sobre síntomas (con codificación horizontal) se interpretan como que el síntoma no ocurrió o fue tan leve que no se recuerda.

- Cuando una respuesta "no se" es la respuesta a otras preguntas, tales como edad de inicio, se deberá estimular al sujeto para que intente recordar con un sondeo como: ¿Podría darme su mejor aproximación?. Si a pesar de ello aún no recuerda, se codificará un 8 si hay espacio para un sólo dígito, 98 si hay dos espacios y 998 si hay tres.

Criterios de selección de los entrevistadores:

- No se requiere experiencia clínica.
- Capaz de leer en voz alta, suave y a modo de conversación.
- Buenas aptitudes administrativo-secretariales.
- Empatía y destreza al escuchar.
- Dispuesto a seguir las instrucciones con precisión.

Criterios para las personas entrevistadas:

- Diseñado para individuos mayores de 18 años.
- Apropiado para personas con variados niveles educacionales, culturales y de inteligencia.
- Leer y escribir es sólo necesario para completar el diagnóstico sobre Deterioro Cognitivo.

ANEXO 9 PHQ-9

CUESTIONARIO PATIENT HEALTH QUESTIONNAIRE (PHQ-9 ITEMS)

Tipo de administración <input type="checkbox"/> Autoadministrado <input type="checkbox"/> Heteroadministrado	
--	--

1. Frecuencia con que le han afectado a la persona participante los siguientes problemas durante las ÚLTIMAS DOS SEMANAS.

	Nunca	Varios días	Más de la mitad de los días	Casi todos los días
	0	1	2	3
a. Poco interés o agrado al hacer las cosas.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Se ha sentido triste, deprimido o desesperado.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Ha tenido problemas para dormir, mantenerse despierto o duerme demasiado.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Se siente cansado o tiene poca energía.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Tiene poco o excesivo apetito.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Se ha sentido mal consigo mismo, ha sentido que usted es un fracaso o ha sentido que se ha fallado a sí mismo o a su familia.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. Ha tenido problemas para concentrarse en actividades como leer el periódico o ver televisión.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
h. Se mueve o habla tan despacio que otras personas pueden darse cuenta o por el contrario, está tan inquieto o intranquilo que da vueltas de un lugar a otro más que de costumbre.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
i. Ha pensado que estaría mejor muerto o ha deseado hacerse daño de alguna forma.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Si puntuó 1,2, o 3 en alguno de los problemas antes indicados en este cuestionario, ¿cuán difícil se le ha hecho cumplir con su trabajo, atender su casa o relacionarse con otras personas debido a estos problemas?

- Nada difícil
- Un poco difícil
- Muy difícil
- Extremadamente difícil



