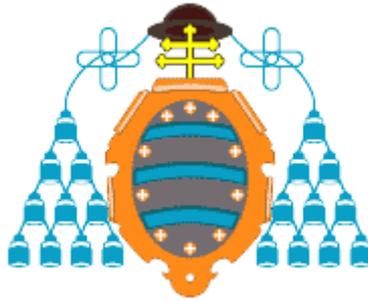


Universidad de Oviedo

PROGRAMA DE DOCTORADO DE PSICOLOGÍA

**“DESARROLLO Y VALIDACIÓN DE INSTRUMENTOS
ÚTILES PARA LA INVESTIGACIÓN DE RESULTADOS
EN SALUD Y LA PRÁCTICA CLÍNICA: APLICACIÓN DE
LA TEORÍA CLÁSICA DE LOS TEST”**

D. Jesús Cuervo Uría



**“DESARROLLO Y VALIDACIÓN DE INSTRUMENTOS
ÚTILES PARA LA INVESTIGACIÓN DE RESULTADOS
EN SALUD Y LA PRÁCTICA CLÍNICA: APLICACIÓN DE
LA TEORÍA CLÁSICA DE LOS TEST”**

Tesis Doctoral

Programa de Doctorado de Psicología

Doctorando: D. Jesús Cuervo Uría

Director: Dr. D. Eduardo García Cueto

Co-Director: Dr. D. Pablo Rebollo Álvarez

Departamento de Psicología

Universidad de Oviedo

Curso 2014-2015



RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español/Otro Idioma: DESARROLLO Y VALIDACIÓN DE INSTRUMENTOS ÚTILES PARA LA INVESTIGACIÓN DE RESULTADOS EN SALUD Y LA PRÁCTICA CLÍNICA: APLICACIÓN DE LA TEORÍA CLÁSICA DE LOS TEST	Inglés: DEVELOPMENT AND VALIDATION OF USEFUL INSTRUMENTS FOR HEALTH OUTCOMES RESEARCH AND CLINICAL PRACTICE: APPLICATION OF THE CLASSICAL TEST THEORY
2.- Autor	
Nombre: JESÚS CUERVO URÍA	DNI/Pasaporte/NIE:
Programa de Doctorado: PSICOLOGÍA	
Órgano responsable: DEPARTAMENTO DE PSICOLOGÍA, UNIVERSIDAD DE OVIEDO	

RESUMEN (en español)

En el ámbito de la Investigación de Resultados en Salud, se aplican, cada vez más, instrumentos validados que incorporan la perspectiva del paciente en la evaluación de aspectos clave de las tecnologías sanitarias. La tesis doctoral aborda el proceso psicométrico de validación de tres medidas que evalúan: la satisfacción de los pacientes crónicos en tratamiento farmacológico, el impacto de los fármacos antipsicóticos en la Calidad de Vida Relacionada con la Salud (CVRS) de personas con esquizofrenia o trastorno bipolar y, por último, la gravedad del Trastorno Depresivo Mayor (TDM). La primera investigación, centrada en el desarrollo del cuestionario SAT-Q, se llevó a cabo en 4 centros de Atención Primaria (AP). A partir de la revisión de la literatura y el trabajo en paneles con expertos y pacientes, se consiguió una versión inicial que fue sometida a validación. A tal fin, se entrevistó a 202 pacientes con enfermedad crónica y que llevaban al menos 1 año en tratamiento. Mediante análisis factorial y el índice de discriminación se obtuvo una versión de 18 ítems tipo Likert que evaluaban las áreas más importantes del modelo teórico subyacente: 2 genéricas (satisfacción con el tratamiento farmacológico y la atención sanitaria), y 4 específicas (efectividad, efectos adversos, comodidad o conveniencia y olvidos). El SAT-Q es un cuestionario de fácil cumplimentación, con adecuadas propiedades psicométricas básicas: consistencia interna (α de Cronbach) de las dimensiones aceptable ($>0,6-0,7$) y coeficientes de correlación intraclase (CCI) moderados-altos (con la excepción de eventos adversos). Se encontró una asociación significativa (prueba de Spearman-rho-) entre sus puntuaciones (con la excepción de olvidos) y los Componentes Sumarios Físico y Mental del cuestionario genérico de CVRS SF-12.

El segundo trabajo, presentó la adaptación a España del Tool (*Tolerability and quality of Life Questionnaire*), un cuestionario de 8 ítems (Likert) que mide el impacto de los antipsicóticos en los pacientes esquizofrénicos o con trastorno bipolar. Tras la adaptación lingüística y de contenido, se entrevistó a 242 pacientes en 60 centros de salud mental. Se analizó la factibilidad, fiabilidad (consistencia interna y test-retest) y validez (constructo, convergente y de criterio). El tiempo medio necesario para cumplimentarlo fue inferior a 5 minutos. La fiabilidad fue adecuada (α de Cronbach = 0,76 y CCI = 0,90). El análisis factorial mostró una única dimensión principal y la



asociación (ρ) con medidas genéricas de CVRS y específicas fue significativa y moderada-alta. Finalmente, fue capaz de discriminar entre estados de gravedad clínica (U de Mann-Whitney, $p < 0,001$).

Por último, se adaptó la escala CUDOS (*Clinically Useful Depression Outcome Scale*), una medida breve (18 ítems tipo Likert) que cubre el espectro de síntomas del TDM y además recoge información de CVRS. Tras el análisis del contenido y su validación lingüística, se llevó a cabo en AP un estudio multicéntrico el que se evaluaron 338 pacientes con diagnóstico de TDM (criterios DSM-IV TR). Se aplicaron también: escala de Depresión de Hamilton (HAMD-17), SOFAS (Social and Occupational Functioning Assessment Scale), SF-36 (Sumarios Físico-PCS- y Mental-MCS-) y medidas de gravedad global de los síntomas. El tiempo medio de cumplimentación fue de 4,47 minutos. Los efectos techo/suelo se encontraron en menos del 1% de los casos y el α de Cronbach= 0,88. Los coeficientes de correlación de Pearson fueron moderados-altos con la SOFAS, HAMD-17 y MCS; $p < 0,001$. La escala CUDOS discriminó adecuadamente entre los diferentes de gravedad ($p < 0,03$, Mann-Whitney con corrección de Bonferroni).

En conclusión, los estudios presentados mantienen un nexo común en cuanto a los principios y procedimientos psicométricos aplicados. Además, permiten poner a disposición de los profesionales sanitarios, tres nuevas medidas, útiles para la medición de constructos importantes en AP y en consultas de salud mental.

RESUMEN (en Inglés)

Validated instruments are being increasingly used within the framework of the Health Outcomes Research to incorporate patients' perspective into the evaluation of health technologies. This thesis presents the psychometric validation processes of three measures addressing the following constructs: the satisfaction of chronic patients undergoing a pharmacological treatment, the impact of neuroleptics on the Health-Related Quality of Life (HRQoL) of schizophrenic and bipolar patients and, finally, the severity of the Major Depressive Disorder (MDD).

The first research, focused on the development of the SAT-Q questionnaire, was carried out in 4 Primary Care (PC) centers. A preliminary version of the questionnaire was obtained from an extensive literature review and experts' and patients' panels. Chronic patients ($n=202$), undergoing a pharmacological treatment for at least one year, were interviewed during the validation study. Factorial analysis and the discrimination index were applied to reach a new version with 18 items (Likert type), covering the fundamental domains of the underlying theoretical model: 2 general domains (satisfaction with pharmacological treatment and the health service) and 4 specific (effectiveness, adverse events, ease of use and oversights). The SAT-Q is an easy to respond questionnaire with adequate psychometric proprieties: internal consistency (Cronbach's α) of domains $> 0.6-0.7$ and with moderate-high intraclass correlation coefficients (ICC), with the exception of adverse events domain. A statistically significant association (Spearman ρ) was found between SAT-Q dimensions (excluding oversights) and mental and physical summary components of the SF-12. The second publication described the adaptation into Spain of the Tool (*Tolerability and quality of Life Questionnaire*) which comprises 8 Likert type items. This instrument measures the impact of neuroleptics on schizophrenic and bipolar patients. Following



the linguistic and content adaptation, 242 patients were surveyed in 60 mental health centers. The psychometric proprieties tested were: feasibility, reliability (internal consistency and test-retest) and validity (construct, convergent and criterion). The mean time required to answer the TooL was lower than 5 minutes. Reliability was fairly good (Cronbach's $\alpha = 0.76$ and ICC = 0.90). A principal single domain structure was confirmed by means of factorial analysis. Moderate to high and significant associations (ρ) with generic and specific instruments of HRQoL were found. Finally, the TooL questionnaire was able to capture the differences in a range of diverse clinical status according to severity (Mann-Whitney's U, $p < 0.001$).

The third manuscript addresses the adaptation of the CUDOS (*Clinically Useful Depression Outcome Scale*), a brief scale (18 Likert type items) covering the full range of MDD symptoms and generic information about HRQoL. After a review of its content and its linguistic validation, a multicenter study was carried out. A total of 338 patients diagnosed of MDD (DSM IV-TR criteria) were included. The measures applied were: Hamilton depression scale (HAMD-17), SOFAS (Social and Occupational Functioning Assessment Scale), SF-36 (Physical-PCS- and Mental-MCS- summary components) and measures of global clinical severity. The completion time mean was 4.47 minutes. Ceiling and floor effects were found in less than 1% of valid cases. Cronbach's α was 0.88. Pearson correlation coefficients with SOFAS, HAMD-17 and MCS were moderate-high, $p < 0.001$. CUDOS scale was able to distinguish between levels of severity ($p < 0.03$, Mann-Whitney's U with Bonferroni's correction).

In conclusion, the above studies are closely related in terms of the psychometric principles considered and the procedures applied. Furthermore, from these publications, three new measures are now available for other health professionals to test relevant constructs in PC and mental health centers.

SR. DIRECTOR DE DEPARTAMENTO DE _____ /
SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN _____

*A Jimena, al gran Leo y la esperada Lena,
mi vida.*

*A mis padres, hermanos (de sangre o no), al recién llegado Jake y demás familia, por
seguir dándome su corazón, hombro y educación, aún a mis años.*

A Llar, por cuidar de todos ellos con infinita paciencia mientras sesteas.

Sin duda, como escribía Lord Byron,

tienes lo mejor del hombre y ninguno de sus defectos.

*A los amigos y amigas que, a pesar de los tiempos que corren, siempre provocan una
sonrisa y me hacen mirar el futuro confiado de su presencia.*

A todos los que una vez estuvieron aquí y siempre seguirán estando, ...

...Gracias.

PRESENTACIÓN

DOCTORANDO

Nombre: Jesús

Apellidos: Cuervo Uría

D.N.I.: 09437254D

PROGRAMA DE DOCTORADO: Psicología.

DEPARTAMENTO: Psicología. Universidad de Oviedo.

TÍTULO DEL PROYECTO DE TESIS DOCTORAL: Desarrollo y validación de instrumentos útiles para la investigación de resultados en salud y la práctica clínica: Aplicación de la Teoría Clásica de los Test.

DIRECTORES:

Nombre: Dr. D. Eduardo García Cueto

D.N.I.: 71611317N

Cargo: Catedrático de Psicometría. Facultad de Psicología. Universidad de Oviedo.

Nombre: Dr. D. Pablo Rebollo Álvarez

D.N.I.: 09388387V

Cargo: Director científico BAP Health Outcomes, LASER Analytica, Oviedo.

INDICE

PRESENTACIÓN	3
INDICE.....	4
AGRADECIMIENTOS.....	6
ABREVIATURAS	7
RESUMEN EN ESPAÑOL DEL CONTENIDO DE TESIS DOCTORAL.....	9
RESUMEN EN INGLÉS DEL CONTENIDO DE TESIS DOCTORAL.....	12
INTRODUCCIÓN.....	15
Investigación de Resultados en Salud (IRS).....	16
Variables de interés en la IRS	17
Satisfacción del paciente.....	21
Definición y modelos teóricos.....	21
Importancia del estudio de la Satisfacción e Instrumentos de Medida	25
Calidad de Vida Relacionada con la Salud (CVRS).....	28
Definición de CVRS.....	28
Instrumentos de medida de la CVRS	29
Esquizofrenia, Trastorno Bipolar, Tratamiento Antipsicótico y CVRS.....	34
Trastorno Depresivo Mayor y CVRS.....	36
Propiedades psicométricas deseables en los instrumentos de medida en el entorno de la práctica de la atención socio-sanitaria	39
OBJETIVOS E HIPÓTESIS	52
Objetivos.....	52
Generales.....	52
Secundarios	52
Hipótesis	53
Asociadas a los objetivos generales	53
Asociadas a los objetivos secundarios	54
TRABAJOS INCLUIDOS EN EL PROYECTO DE TESIS	55

ARTÍCULO 1 Desarrollo y validación de un cuestionario genérico para la evaluación de la satisfacción de los pacientes crónicos: Cuestionario SAT-Q.....	55
ARTÍCULO 2. Validación de una medida específica para evaluar Calidad de Vida Relacionada con la Salud en pacientes con esquizofrenia y trastorno bipolar: versión Española del cuestionario ToolL.	64
ARTÍCULO 3. Evaluación de la depresión mayor desde la perspectiva del paciente: validación al castellano de la escala CUDOS (Clinically Useful Depression Outcome Scale).	73
RESULTADOS	88
ARTÍCULO 1: Desarrollo y validación cuestionario SAT-Q.....	89
ARTÍCULO 2: Adaptación cultural y validación del cuestionario ToolL.....	91
ARTÍCULO 3. Adaptación cultural y validación de la escala CUDOS.	94
DISCUSIÓN.....	98
ARTÍCULO 1: Desarrollo y validación cuestionario SAT-Q.....	99
ARTÍCULO 2: Adaptación cultural y validación del cuestionario ToolL.....	103
ARTÍCULO 3. Adaptación cultural y validación de la escala CUDOS.	106
CONCLUSIONES.....	109
INFORME FACTOR IMPACTO DE LAS PUBLICACIONES PRESENTADAS EN EL PROYECTO DE TESIS DOCTORAL	111
Atención Primaria: ISSN 0212-6567	112
Annals of General Psychiatry: ISSN 1744859X.....	112
Actas Españolas de Psiquiatría: ISSN 1139-9287.....	113
REFERENCIAS BIBLIOGRÁFICAS	114
ANEXO:.....	143
COMUNICACIONES EN CONGRESOS Y PUBLICACIONES CON LOS INSTRUMENTOS DESARROLLADOS.	143

AGRADECIMIENTOS

Esta tesis doctoral ha sido posible gracias al apoyo profesional y personal de muchas personas. Me gustaría agradecer, en primer lugar, a mis directores en este proyecto, el profesor Dr. D. Eduardo García Cueto, y el director científico de BAP Health Outcomes (LASER Analytica, España), el Dr. D. Pablo Rebollo Álvarez, su ánimo, experiencia y dedicación durante este largo pero ilusionante camino. Aprender y trabajar con vosotros es muy fácil, mi gratitud por todos estos años.

En segundo lugar, quisiera expresar mi profundo agradecimiento a todos los profesionales médicos y compañeros de investigación que son autores también de las publicaciones presentadas, porque desde el primer momento consintieron en la presentación del trabajo, y de forma desinteresada, prestaron su apoyo. He tenido la suerte de trabajar con ellos durante estos años y espero poder seguir haciéndolo durante muchos más.

En tercer lugar, quiero mencionar aquí a mis compañeros de investigación en la Facultad y en el LASER Analytica. Además de ser muchos de ellos parte activa fundamental de las investigaciones presentadas, todos me han facilitado en gran medida la realización del trabajo que presento. De igual forma, no puedo olvidarme de los profesores de Psicología que me abrieron las puertas a la investigación hace ya más de 12 años: el Dr. D. Antonio-León Aguado Díaz y la Dra. D^a M^a Ángeles Alcedo Rodríguez. Siempre han estado disponibles desde entonces.

Finalmente, quiero expresar mi más sincero agradecimiento a todos los pacientes que a lo largo de estos años han participado de forma voluntaria en los distintos estudios presentados. Ellos son los destinatarios últimos de los trabajos presentados.

A todos ellos, gracias.

ABREVIATURAS

AP	Atención Primaria
CCAA	Comunidades Autónomas
CCI	Coefficiente de Correlación Intraclase
CONSORT	Consolidated Standards of Reporting Trials
CUDOS	Clinically Useful Depression Outcome Scale
CVRS	Calidad de Vida Relacionada con la Salud
DLQI	Dermatology Quality of Life Index
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders(4ª edición revisada)
EEUU	Estados Unidos de América
EMA	European Medicines Agency
EQ-5D	EuroQoL 5 domains Questionnaire
EUnetHTA	European Network for Health Technology Assessment
FACT-B	Functional Assessment of Cancer Therapy for patients with Breast cancer
FACT-G	Functional Assessment of Cancer Therapy - General
FDA	Food and Drug Administration
HAMD-17	Escala de Depresión de Hamilton (versión de 17 ítems)
HUI	Health Utility Index
ICG-G	Índice Clínico de Gravedad
IGP	Índice de Gravedad según los pacientes
I-QOL	<i>Incontinence Quality of Life Questionnaire</i>
IRS	Investigación de Resultados en Salud
ISOQOL	International Society for Quality of Life Research
ISPOR	International Society of Pharmacoeconomics and Outcomes Research
MADRS	Montgomery–Åsberg Depression Rating Scal
MCS	Mental Component Symmary
MIDAS	Rhode Island Methods to Improve Diagnostic Assessment and Service
NHP	Nottingham Health Profile
OMS	Organización Mundial de la Salud

PCS	Physical Component Summary
PRIME-MD	Primary Care Evaluation of Mental Disorders
PRO	Patient Reported Outcome
Rho	Coefficiente de Correlación de Spearman
SF-12	Short Form Health Survey-12 items
SF-36	Short Form Health Survey-36 items
SNS	Sistema Nacional de Salud
SOFAS	Social and Occupational Functioning Assessment Scale
TCT	Teoría Clásica de los Test
TDM	Trastorno Depresivo Mayor
TooL	Cuestionario Tolerability and quality of Life
TRI	Teoría de Respuesta al Ítem
UE	Unión Europea
UKU	Udvalg für Kliniske Undersogelser (Escala de efectos secundarios)

RESUMEN EN ESPAÑOL DEL CONTENIDO DE TESIS DOCTORAL

Desarrollo y validación de instrumentos útiles para la investigación de resultados en salud y la práctica clínica: Aplicación de la Teoría Clásica de los Test.

En el ámbito de la Investigación de Resultados en Salud, se aplican, cada vez más, instrumentos validados que incorporan la perspectiva del paciente en la evaluación de aspectos clave de las tecnologías sanitarias. La tesis doctoral aborda el proceso psicométrico de validación de tres medidas que evalúan: la satisfacción de los pacientes crónicos en tratamiento farmacológico, el impacto de los fármacos antipsicóticos en la Calidad de Vida Relacionada con la Salud (CVRS) de personas con esquizofrenia o trastorno bipolar y, por último, la gravedad del Trastorno Depresivo Mayor (TDM).

La primera investigación, centrada en el desarrollo del cuestionario SAT-Q, se llevó a cabo en 4 centros de Atención Primaria (AP). A partir de la revisión de la literatura y el trabajo en paneles con expertos y pacientes, se consiguió una versión inicial que fue sometida a validación. A tal fin, se entrevistó a 202 pacientes con enfermedad crónica y que llevaban al menos 1 año en tratamiento. Mediante análisis factorial y el índice de discriminación se obtuvo una versión de 18 ítems tipo Likert que evaluaban las áreas más importantes del modelo teórico subyacente: 2 genéricas (satisfacción con el tratamiento farmacológico y la atención sanitaria), y 4 específicas (efectividad, efectos adversos, comodidad o conveniencia y olvidos). El SAT-Q es un cuestionario de fácil cumplimentación, con adecuadas propiedades psicométricas básicas: consistencia interna (α de Cronbach) de las dimensiones aceptable ($>0,6-0,7$) y coeficientes de correlación intraclase (CCI) moderados-altos (con la excepción de eventos adversos). Se

encontró una asociación significativa (prueba de Spearman-rho-) entre sus puntuaciones (con la excepción de olvidos) y los Componentes Sumarios Físico y Mental del cuestionario genérico de CVRS SF-12.

El segundo trabajo, presentó la adaptación a España del TooL (*Tolerability and quality of Life Questionnaire*), un cuestionario de 8 ítems (Likert) que mide el impacto de los antipsicóticos en los pacientes esquizofrénicos o con trastorno bipolar. Tras la adaptación lingüística y de contenido, se entrevistó a 242 pacientes en 60 centros de salud mental. Se analizó la factibilidad, fiabilidad (consistencia interna y test-retest) y validez (constructo, convergente y de criterio). El tiempo medio necesario para cumplimentarlo fue inferior a 5 minutos. La fiabilidad fue adecuada (α de Cronbach = 0,76 y CCI = 0,90). El análisis factorial mostró una única dimensión principal y la asociación (rho) con medidas genéricas de CVRS y específicas fue significativa y moderada-alta. Finalmente, fue capaz de discriminar entre estados de gravedad clínica (U de Mann-Whitney, $p < 0,001$).

Por último, se adaptó la escala CUDOS (*Clinically Useful Depression Outcome Scale*), una medida breve (18 ítems tipo Likert) que cubre el espectro de síntomas del TDM y además recoge información de CVRS. Tras el análisis del contenido y su validación lingüística, se llevó a cabo en AP un estudio multicéntrico el que se evaluaron 338 pacientes con diagnóstico de TDM (criterios DSM-IV TR). Se aplicaron también: escala de Depresión de Hamilton (HAMD-17), SOFAS (Social and Occupational Functioning Assessment Scale), SF-36 (Sumarios Físico-PCS- y Mental-MCS-) y medidas de gravedad global de los síntomas. El tiempo medio de cumplimentación fue de 4,47 minutos. Los efectos techo/suelo se encontraron en menos del 1% de los casos y el α de Cronbach= 0,88. Los coeficientes de correlación de Pearson fueron moderados-altos

con la SOFAS, HAMD-17 y MCS; $p < 0,001$. La escala CUDOS discriminó adecuadamente entre los diferentes de gravedad ($p < 0,03$, Mann-Whitney con corrección de Bonferroni).

En conclusión, los estudios presentados mantienen un nexo común en cuanto a los principios y procedimientos psicométricos aplicados. Además, permiten poner a disposición de los profesionales sanitarios, tres nuevas medidas, útiles para la medición de constructos importantes en AP y en consultas de salud mental.

RESUMEN EN INGLÉS DEL CONTENIDO DE TESIS DOCTORAL

Development and validation of useful instruments for health outcomes research and clinical practice: Application of the Classical Test Theory

Validated instruments are being increasingly used within the framework of the Health Outcomes Research to incorporate patients' perspective into the evaluation of health technologies. This thesis presents the psychometric validation processes of three measures addressing the following constructs: the satisfaction of chronic patients undergoing a pharmacological treatment, the impact of neuroleptics on the Health-Related Quality of Life (HRQoL) of schizophrenic and bipolar patients and, finally, the severity of the Major Depressive Disorder (MDD).

The first research, focused on the development of the SAT-Q questionnaire, was carried out in 4 Primary Care (PC) centers. A preliminary version of the questionnaire was obtained from an extensive literature review and experts' and patients' panels. Chronic patients (n=202), undergoing a pharmacological treatment for at least one year, were interviewed during the validation study. Factorial analysis and the discrimination index were applied to reach a new version with 18 items (Likert type), covering the fundamental domains of the underlying theoretical model: 2 general domains (satisfaction with pharmacological treatment and the health service) and 4 specific (effectiveness, adverse events, ease of use and oversights). The SAT-Q is an easy to respond questionnaire with adequate psychometric proprieties: internal consistency (Cronbach's α) of domains $> 0.6-0.7$ and with moderate-high intraclass correlation coefficients (ICC), with the exception of adverse events domain. A statistically

significant association (Spearman rho) was found between SAT-Q dimensions (excluding oversights) and mental and physical summary components of the SF-12.

The second publication described the adaptation into Spain of the TooL (*Tolerability and quality of Life Questionnaire*) which comprises 8 Likert type items. This instrument measures the impact of neuroleptics on schizophrenic and bipolar patients. Following the linguistic and content adaptation, 242 patients were surveyed in 60 mental health centers. The psychometric proprieties tested were: feasibility, reliability (internal consistency and test-retest) and validity (construct, convergent and criterion). The mean time required to answer the TooL was lower than 5 minutes. Reliability was fairly good (Cronbach's $\alpha = 0.76$ and ICC = 0.90). A principal single domain structure was confirmed by means of factorial analysis. Moderate to high and significant associations (rho) with generic and specific instruments of HRQoL were found. Finally, the TooL questionnaire was able to capture the differences in a range of diverse clinical status according to severity (Mann-Whitney's U, $p < 0.001$).

The third manuscript addresses the adaptation of the CUDOS (*Clinically Useful Depression Outcome Scale*), a brief scale (18 Likert type items) covering the full range of MDD symptoms and generic information about HRQoL. After a review of its content and its linguistic validation, a multicenter study was carried out. A total of 338 patients diagnosed of MDD (DSM IV-TR criteria) were included. The measures applied were: Hamilton depression scale (HAMD-17), SOFAS (Social and Occupational Functioning Assessment Scale), SF-36 (Physical-PCS- and Mental-MCS- summary components) and measures of global clinical severity. The completion time mean was 4.47 minutes. Ceiling and floor effects were found in less than 1% of valid cases. Cronbach's α was 0.88. Pearson correlation coefficients with SOFAS, HAMD-17 and MCS were

moderate-high, $p < 0.001$. CUDOS scale was able to distinguish between levels of severity ($p < 0.03$, Mann-Whitney's U with Bonferroni's correction).

In conclusion, the above studies are closely related in terms of the psychometric principles considered and the procedures applied. Furthermore, from these publications, three new measures are now available for other health professionals to test relevant constructs in PC and mental health centers.

INTRODUCCIÓN

Durante el desarrollo y la evaluación de las tecnologías sanitarias, los agentes reguladores, solicitan, de forma creciente y generalizada, a las entidades promotoras (grupos de investigación, compañías farmacéuticas y comunidad científica en general), que aporten información válida sobre la opinión del paciente respecto a las consecuencias que experimenta con su consumo (Bastida et al., 2010; National Institute for Health and Clinical Excellence [NICE], 2013). Así, aspectos tales como la definición de la indicación final para su prescripción, la población candidata o el precio final de venta, se ven afectados por la valoración que realicen los pacientes sobre los beneficios o perjuicios asociados a su aplicación. Esto es así ya que se considera a los pacientes una fuente de información muy útil como destinatarios principales de la intervención (Academy of Managed Care Pharmacy [AMCP], 2005; Bastida et al., 2010; Canadian Agency for Drugs and Technologies in Health, 2006; Food and Drug Administration [FDA], 2009). Por supuesto estos aspectos son de interés una vez contrastados otros requisitos indispensables en toda tecnología sanitaria: adecuado perfil de seguridad, tolerabilidad y eficacia clínica.

Una vez aprobadas por las autoridades reguladoras, las intervenciones sanitarias están igualmente sometidas a un seguimiento continuo para analizar la seguridad y la efectividad en condiciones de práctica clínica real. En este momento, también se incluye la opinión de los pacientes sobre distintos aspectos que son de interés en la evaluación de la calidad asistencial (European Network for Health Technology Assessment [EUnetHTA], 2013). Esto es parte de lo que se conoce como Investigación de Resultados en Salud (IRS).

Investigación de Resultados en Salud (IRS)

La IRS es una disciplina del sector socio-sanitario orientada al estudio del rendimiento de las intervenciones sanitarias en condiciones de práctica clínica habitual, fuera de las condiciones más restrictivas o de laboratorio que imponen los ensayos clínicos de desarrollo inicial y los pivotaes (Badía y Lizán, 2002; Soto, 2007). Por tanto, los resultados son evaluados en términos clínicos de efectividad, pero también desde la perspectiva de la gestión sanitaria y el impacto económico. De esta forma, se persigue mejorar la calidad del Sistema Nacional de Salud (SNS) y aportar información muy útil a los profesionales implicados en el tratamiento de los pacientes, ayudando a guiar su praxis de acuerdo a la mejor evidencia disponible y conociendo la eficiencia de las distintas alternativas disponibles. Debe tenerse en cuenta que el objetivo común básico que persiguen los gobiernos, las agencias evaluadoras y los gestores sanitarios es posibilitar que los ciudadanos puedan acceder a los avances sanitarios que maximicen, no sólo el número de años vividos, sino la calidad de éstos. Eso sí, también están obligados a considerar otros condicionantes como la equidad de la prestación y el gasto sanitario, incluyendo así la evaluación económica como otro factor fundamental en la IRS (Soto, 2007).

No obstante, para que la percepción de los pacientes sobre el beneficio experimentado, o sobre otros aspectos de interés (tales como la adherencia al tratamiento o la satisfacción con las prestaciones sanitarias), sea válida y pueda ser considerada como un criterio de calidad, es necesario que se recoja aplicando una metodología definida. A tal fin, la valoración subjetiva de los pacientes se mide a través de instrumentos (cuestionarios, baterías de evaluación, índices, escalas,...) que son diseñados para la evaluación de los distintos constructos de interés y posteriormente incluidos en

protocolos de investigación. La calidad en la medición debe ser tal que garantice la posibilidad de obtener datos robustos en términos científicos y que puedan ser considerados por las agencias de evaluación (Bastida et al., 2010).

En las sociedades más industrializadas, estos procesos de evaluación están cada vez más establecidos existiendo a nivel nacional y europeo ejemplos de guías de trabajo que facilitan el intercambio de información entre las partes implicadas, a la vez que promueven la transparencia y consistencia de las decisiones tomadas (Department of Economics and Public Health Assessment, 2012; European Medicines Agency [EMA], 2014; EUnetHTA, 2013; NICE, 2013). En nuestro país, la Propuesta de Guía para la Evaluación Económica aplicada a las tecnologías sanitarias, financiada por el Ministerio de Sanidad y Consumo (Bastida et al., 2010), contempla también la necesidad de incorporar resultados informados por los pacientes, siendo necesario el empleo de instrumentos previamente validados para recoger correctamente esta información.

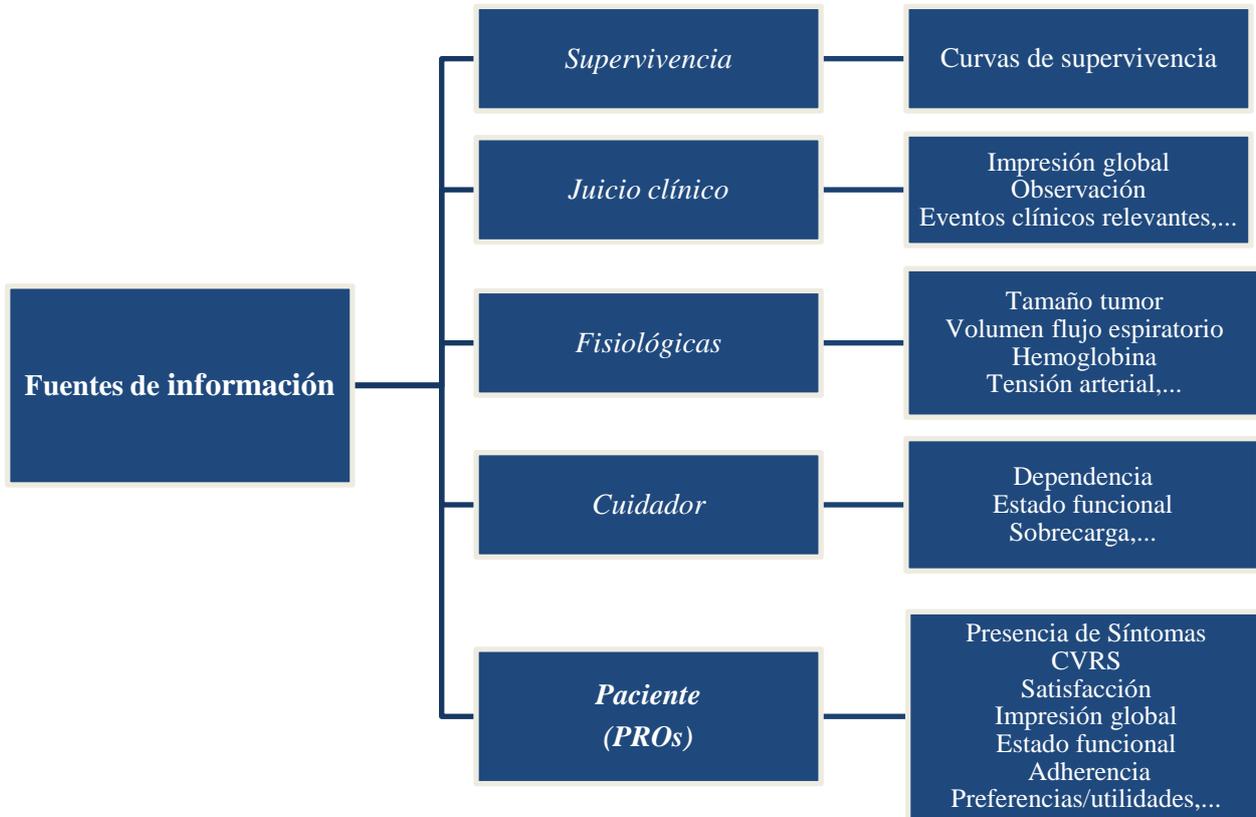
Variables de interés en la IRS

Como se ha indicado previamente, entre las variables de estudio más frecuentemente analizadas en este ámbito se encuentran las llamadas en inglés “*Patient Reported Outcomes*” (PROs) o resultados en salud informados por los pacientes. Como su nombre indica, una PRO es cualquier informe o valoración sobre la situación del paciente relacionada con su salud que viene directamente del propio paciente, sin mediar interpretación alguna de sus respuestas por parte del profesional sanitario u otra persona (FDA, 2009). Normalmente esta información se recoge con instrumentos previamente validados que son cumplimentados por los pacientes a cerca de aspectos de interés que tienen que ver principalmente con su Calidad de Vida Relacionada con la Salud (CVRS), sensaciones sobre la presencia o no de síntomas, su satisfacción con la

atención sanitaria recibida, sus preferencias por las alternativas terapéuticas existentes o el cumplimiento que hacen de las intervenciones preventivas o curativas. En este sentido, de acuerdo a la Agencia Europea de los Medicamentos (EMA, por su nombre en inglés: European Medicines Agency) y la Agencia Alimentaria y del Medicamento de los EEUU (FDA, Food and Drug Administration), el término PRO incluye todas las medidas uni o multidimensionales relacionadas con la presencia o no de síntomas y demás conceptos ya indicados. Es importante resaltar que estos agentes de referencia inciden en que estos instrumentos deben ser cumplimentados por los pacientes y reflejando su propio punto de vista (EMA, 2014; EUnetHTA, 2013; FDA, 2009).

Tal y como se refleja en la Figura 1, si bien el estudio del estado de salud de los pacientes puede ser abordado desde distintos acercamientos y perspectivas, para los organismos públicos de gestión y las agencias de evaluación, la información sobre el estado de salud del paciente debe evaluarse preguntándoles directamente a ellos ya que se consideran la mejor fuente de información (Acquadro et al., 2003; Johnston et al., 2013; Patrick et al., 2007):

Figura 1. Ejemplos de fuentes de información en IRS.



Nota: Adaptado de Acquadro (2003).

El valor de las PROs es, sin duda, el de completar la calidad de la información que tradicionalmente ha sido (y deberá seguir siéndolo) el foco de atención en el contexto sanitario por su valor básico fundamental: la supervivencia de los pacientes y la presencia de eventos adversos o riesgos asociados al consumo de medicamentos u otras tecnologías o intervenciones. La inclusión de las PROs es aún más necesaria cuando la eficacia/efectividad de la intervención no puede ser fácilmente observada de forma directa (por ejemplo en el caso de trastornos emocionales, mentales, síndromes de dolor, disfunciones sexuales, etc...). También, cuando las variables clínicas recogidas únicamente son fisiológicas o de carácter intermedio (por ejemplo, la tensión arterial o los niveles de glucosa en sangre u otros biomarcadores) o cuando los eventos clínicos

relevantes son poco frecuentes (FDA, 2009; Johnston et al., 2013; Patrick et al., 2007). Además, en patologías crónicas, cobran especial importancia por el condicionamiento que la sintomatología puede imponer sobre la capacidad del individuo para desarrollar sus actividades de la vida diaria a lo largo de su vida. Debido ente otros a los constructos abordados en las distintas PROs, su aplicación es de creciente interés actualmente entre muchos colectivos socio-sanitarios (Figura 2):

Figura 2. Agentes del ámbito de la salud con interés en las PROs.

Organismos públicos, Agencias de Evaluación y Gestores	<ul style="list-style-type: none"> • Establecimiento y evaluación continua de prioridades en las políticas sanitarias • Monitorización: seguimiento y vigilancia
Epidemiólogos	<ul style="list-style-type: none"> • Estimación de la incidencia y prevalencia-impacto de la enfermedad • Supervivencia
Farmacoeconomistas	<ul style="list-style-type: none"> • Carga de la enfermedad y necesidades no cubiertas • Evaluación económica: eficiencia (análisis coste-efectividad, coste-utilidad, coste-beneficio o minimización de costes)
Investigadores Clínicos	<ul style="list-style-type: none"> • Ensayos clínicos controlados: eficacia • Estudios naturalísticos: efectividad (real world effectiveness)
Médicos y demás profesionales sanitarios	<ul style="list-style-type: none"> • Impacto de la enfermedad y planificación de la intervención • Efectividad del arsenal terapéutico en distintas poblaciones • Seguimiento del caso individual
Pacientes y familiares	<ul style="list-style-type: none"> • Medio de intercambio de información con profesionales sanitarios • Voz en la toma de decisiones y racionalización prestaciones

Tal y como se ha indicado, existe una gran diversidad de PROs en relación a su contenido. A continuación, se presenta una breve introducción a los constructos que constituyen el objeto medido por cada uno de los instrumentos que fueron desarrollados o validados en las publicaciones que conforman el presente trabajo de tesis doctoral: la satisfacción de los pacientes crónicos en tratamiento farmacológico, el impacto de los fármacos antipsicóticos en la CVRS de personas con esquizofrenia o trastorno bipolar y, por último, la gravedad del Trastorno Depresivo Mayor (TDM).

Satisfacción del paciente

Definición y modelos teóricos

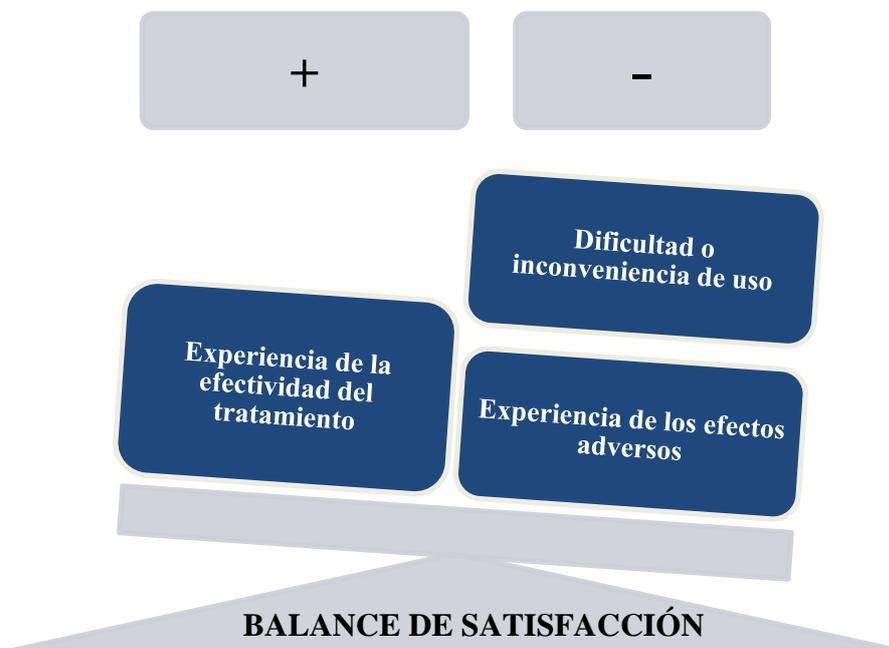
Existen múltiples definiciones de este concepto en la literatura científica. Así, la satisfacción del paciente ha sido definida de forma genérica como “*el grado en que las experiencias de un sujeto cubren sus expectativas previas*” (Asadi-Lari, Tamburini, y Gray, 2004). Con un punto de mira más cercano al ámbito sanitario, se ha definido como “la evaluación o el informe de un sujeto sobre el proceso y el resultado de su tratamiento de acuerdo a criterios predeterminados” (Weaver et al., 1997). Por su parte Hudak y Wright (2000) defienden que una medida de satisfacción es cualquier medida que incluya las evaluaciones de los pacientes o las respuestas afectivas a las distintas dimensiones que incluye la experiencia de los cuidados sanitarios (Hudak y Wright, 2000). Otro ejemplo en este caso dirigido a la evaluación de la satisfacción respecto a los tratamientos farmacológicos es la propuesta de Shikiar y Rentz (2004): “la evaluación que realiza el paciente sobre el proceso de toma de la medicación y sobre los resultados asociados con la medicación”. En todo caso parece que nuevamente se trata de un concepto multidimensional que incluye una naturaleza claramente personal y subjetiva de evaluación y que podría, incluso, reflejar el conocimiento y las expectativas del usuario en lugar de la calidad percibida por el paciente respecto al servicio y los cuidados recibidos. Así, resulta necesario para los investigadores de este campo distinguir perfectamente entre los “sentimientos” de satisfacción con el proceso de cuidados, de aquellos otros relacionados con el resultado en términos de salud.

Debido a la necesidad de establecer un marco conceptual de referencia que facilite el estudio de esta PRO, se han empleado diversos modelos teóricos. Desde la “*Social Learning Theory*” (Rotter, 1980), se podría definir la satisfacción como la

correspondencia entre la experiencia actual personal y la experiencia esperada (expectativas). En este sentido, uno de los predictores del comportamiento y de la satisfacción del paciente en una situación particular serían las “expectativas” generales y específicas. En situaciones en las que uno tiene una gran experiencia, funcionan las expectativas específicas relevantes para la situación; en las situaciones nuevas, uno aplicaría las expectativas generales que pueden haber sido desarrolladas de otras situaciones o pertenecer a la experiencia personal general. Además, estas creencias del valor esperado se formarían muy rápido cuando se ha alcanzado un cierto nivel de experiencia y serían de naturaleza más o menos estable y difícilmente modificables. En contraposición a la anterior, la “*Performance Theory*” (Oliver y De Sarbo, 1988) resta importancia a las expectativas del sujeto y defiende que la satisfacción depende básicamente del resultado clínico obtenido, siendo más alta a medida que aumenta el beneficio percibido y en todo caso es variable en función del resultado. Otros acercamientos como la “*Prospect Theory*” (Kahneman y Tversky, 1979) y la “*Cumulative Prospect Theory*” (Tversky y Kahneman, 1992) exponen cómo las personas tomarían sus decisiones bajo condiciones de incertidumbre valorando resultados pasados, ya certeros pero incluyendo los resultados de salud no experimentados todavía y que anticipan en función del bagaje previo. Además, según estos autores, el valor de un cambio en un resultado clínico no es consistente en todo el espectro de posibles cambios, siendo las pérdidas de bienestar de mayor importancia que las ganancias de la misma magnitud. Por ejemplo, si un paciente está absolutamente asintomático y con una estupenda salud, la presencia de un pequeño síntoma puede ser mucho más impactante que si uno tiene el mismo síntoma además de otros síntomas o si tiene un nivel de salud intermedio. En este sentido, una reducción en la satisfacción puede ser más importante para el paciente que una ganancia de la misma magnitud. En

suma a las anteriores, otro ejemplo teórico sería la “*Fulfillment Theory*” (Linder-Pelz, 1982) bajo la cual la satisfacción se entiende como la diferencia entre el resultado obtenido y otro resultado ideal o deseado. Una aproximación más reciente y que incluye muchos aspectos de las anteriores es la empleada en el “*Decision Balance Model*” (Atkinson et al., 2004; Atkinson, Kumar, Cappelleri, y Hass, 2005). Bajo este marco conceptual, la satisfacción global con la medicación y la adherencia resultante están determinadas por el juicio del paciente sobre si los beneficios obtenidos compensan los efectos secundarios e inconvenientes (Figura 3). En este sentido, las decisiones sobre la adherencia al tratamiento estarían influenciadas por los juicios personales sobre los atributos positivos y negativos del tratamiento y la satisfacción se constituiría como un indicador de la futura adherencia terapéutica: no es fácil que los pacientes continúen con una medicación que perciben como inefectiva, asociada con efectos secundarios o ruptura con su estilo de vida:

Figura 3. Satisfacción con el tratamiento: Decisional Balance Model (Atkinson et al., 2005).



Finalmente, uno de los modelos más difundidos es el “*Hierarchy Model*” o *Modelo Jerárquico de satisfacción* del paciente (Shikiar y Rentz, 2004), que postula la existencia de 3 niveles básicos multidimensionales que estarían interrelacionados (Figura 4):

- Satisfacción respecto a los Servicios Sanitarios (accesibilidad, relación médico-paciente, calidad de los servicios prestado, etc.);
- Satisfacción respecto al conjunto de tratamientos recibidos (sean farmacológicos o no: efectos adversos, comodidad, efectividad, costes,...) ;
- Satisfacción respecto al tratamiento farmacológico específico (donde se podrían contemplar los aspectos destacados en las teorías anteriores).

Figura 4. Niveles de satisfacción del paciente en el Modelo Jerárquico de satisfacción (Shikiar y Rentz, 2004).



Bajo este modelo comprensivo, y en armonía con la definición de Weaver et al. (1997), para la satisfacción con el tratamiento sería importante tanto el proceso de toma del medicamento como el resultado obtenido con su aplicación. Además de lo anterior, se enfatiza también la influencia que pueden tener otras variables como son las expectativas del paciente, sus experiencias pasadas, los niveles de CVRS previos y los

obtenidos y el desempeño funcional, las características de la enfermedad o los síntomas tratados.

A partir de los modelos teóricos presentados, el objetivo para cualquier investigador cuando emprende el desarrollo de una medida de satisfacción no debe ser otro que obtener un indicador robusto y fiable que permita realizar una evaluación del constructo bajo algún modelo de referencia. Por otro lado, debe considerar un rango amplio de posibles factores de influencia (sociodemográficos, características de la patología y del tratamiento, experiencias pasadas, niveles asistenciales). Así, será posible tener información válida que complemente los resultados de efectividad en la aplicación del tratamiento en la práctica clínica habitual (Shikiar y Rentz, 2004; Sitzia, 1999; Speight, 2005).

Como se verá más adelante en la publicación relativa al desarrollo del cuestionario de satisfacción del paciente crónico (SAT-Q), el Modelo Jerárquico de la Satisfacción será el marco teórico de referencia para el trabajo presentado al ser considerado el más completo y ser un referente para la International Society of Pharmacoeconomics and Outcomes Research (ISPOR: <http://www.ispor.org/>).

Importancia del estudio de la Satisfacción e Instrumentos de Medida

El estudio de la satisfacción del paciente por medio de una PRO es importante en los distintos niveles de atención planteados debido a sus implicaciones en el desarrollo de la práctica médica asistencial y en la evaluación de los recursos que la acompañan (Badía, 2005; Lenderking, 2005). Así, es destacable el valor de la satisfacción de los usuarios dentro de los programas de calidad de los servicios sanitarios para la evaluación de los mismos y la identificación de espacios de mejora (Asadi-Lari et al., 2004; Carr-Hill, 1992; Hernández, Ochando, Mora, Lorenzo, y López, 2005; Mira y Aranaz, 2000).

Junto a esto, son notables sus implicaciones clínicas en diferentes patologías. Se ha evidenciado que los pacientes satisfechos tienen mayor disposición a la cooperación con el terapeuta mostrándose más rigurosos en el seguimiento de las directrices médicas, lo cual, incide a su vez en el éxito de la terapia. Igualmente, las personas satisfechas intercambian información con su médico de forma más fiable y fluida mejorando la planificación de la intervención, la adherencia al tratamiento y así, la calidad terapéutica del mismo (Crow et al., 2002). En este sentido la satisfacción es usada cada vez más como variable final o “*end point*” en ensayos clínicos y análisis de nuevos medicamentos (Speight, 2005), y se considera relevante a la hora de analizar la adherencia terapéutica, en términos de cumplimiento o de persistencia, que resulta clave en la efectividad de los tratamientos en condiciones de práctica clínica real (Cramer et al., 2008; Crow et al., 2002; Ortega et al., 2013).

Debe hacerse constar que la evaluación de la satisfacción del paciente cobra especial interés cuando existen varias alternativas de intervención posibles, con diferente perfil de beneficios y consecuencias negativas para el paciente, sin poder establecer una preferencia clara en términos de seguridad o eficacia/efectividad. También cuando el tratamiento es farmacológico y de carácter crónico, donde la adherencia y la persistencia pueden jugar un papel central en el éxito de la intervención, y es importante prevenir y explicar posibles causas que pongan en riesgo el control de la enfermedad. Tal es la relevancia de esta PRO que, por ejemplo, en Gran Bretaña se recogen datos de satisfacción del paciente de forma sistemática a través de cuestionarios online (<http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/en/Publicationsandstatistics/PublishedSurvey/GPpatientsurvey2007/index.htm>).

En respuesta a este importante papel que puede desempeñar la satisfacción de los pacientes, son muchas las publicaciones centradas en desarrollar, desde distintas ópticas, instrumentos de medida para alguno de los aspectos incluidos en este concepto (Abetz et al., 2005; Atkinson et al., 2003; Atkinson et al., 2004; Atkinson et al., 2005; Auquier et al., 2005; Beattie, Nelson, y Lis, 2007; Campbell, Stang, y Barron, 2008; Condes et al., 2005; Meakin y Weinman, 2002; Okuda, Ohkubo, y Goto, 2004; Paddock, Veloski, Chatterton, Gevirtz, y Nash, 2000; Pleil et al., 2005; Rofail, Gray, y Gournay, 2005; Regnault, Viala-Danten, Gilet, y Berdeaux, 2010; Ruiz et al., 2005; Ruiz et al., 2008). Si bien hace unos años los cuestionarios empleados para medir la satisfacción resultaban poco fiables y válidos (Sitzia, 1999), afortunadamente, en los últimos años han sido desarrollados y adaptados a nuestro país nuevos cuestionarios psicométricamente más robustos aunque centrados principalmente a la evaluación de la satisfacción con el tratamiento (Condes et al., 2005; Ruiz et al., 2008). No obstante, ninguno permite la evaluación de áreas importantes relacionadas con la adherencia a los tratamientos como la relación con su médico, la información que el paciente recibe sobre su enfermedad o el tiempo dedicado al paciente. De ahí la pertinencia del trabajo desarrollado en el SAT-Q que pretendía crear un instrumento que permitiera una evaluación más completa de la satisfacción, incluyendo, no sólo aspectos de la satisfacción con el tratamiento, sino también con la atención recibida durante la prestación del servicio. Así, se dirige a los niveles básicos de la satisfacción del paciente en el marco teórico del Modelo Jerárquico (Shikiar y Rentz, 2004).

Calidad de Vida Relacionada con la Salud (CVRS)

Definición de CVRS

A falta de un consenso sobre la definición de este constructo, existen múltiples propuestas siendo su naturaleza multidimensional el punto de encuentro de todas ellas. Como ejemplo, la EMA destaca que es un concepto muy amplio que se puede definir como *“la percepción subjetiva de los pacientes sobre el impacto de su enfermedad y sus tratamientos en su vida diaria, en su funcionamiento físico, psicológico y social y su bienestar general”* (EMA, 2014). De forma más comprehensiva, la definición propuesta por Patrick y Erickson (1993) incluye la influencia no sólo de la enfermedad y los tratamientos sino las políticas sanitarias llevadas a cabo (Patrick y Erickson, 1993).

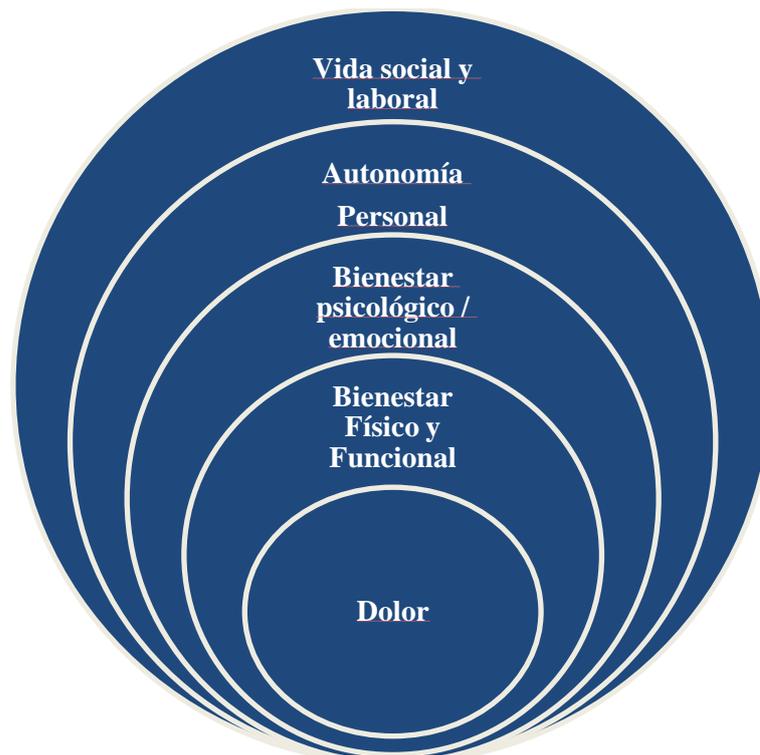
Debe diferenciarse este concepto con el genérico de Calidad de Vida que es más amplio y que se define de acuerdo a la Organización Mundial de la Salud (OMS) como *“la percepción que tiene el individuo sobre su posición en la vida dentro de su contexto cultural y los sistemas de valores bajo los que vive y en relación a sus metas, expectativas, referencias e intereses”*(WHOQOL Group, 1995; WHOQOL Group, 1998). Otra definición muy empleada en el entorno de la intervención psicosocial describe la Calidad de Vida a través de 8 dimensiones (Schalock y Verdugo, 2002; Verdugo, Schalock, Keith, y Stancliffe, 2005; Verdugo et al., 2011): bienestar emocional, relaciones personales, bienestar material (capacidad económica y aspectos materiales suficientes que le permitan una vida confortable, saludable y satisfactoria), desarrollo personal, bienestar físico, autodeterminación, inclusión social y defensa de los derechos personales (intimidad y trato recibido en su entorno).

Parece claro que la CVRS es un área específica de estudio dentro de la Calidad de Vida, siempre relacionada con el impacto de la enfermedad y la atención socio-sanitaria sobre

el individuo. Por otro lado, la CVRS debe diferenciarse claramente de la presencia de los síntomas clave de una enfermedad (ej. dolor, pirosis, incontinencia...) evaluados por el propio paciente. En todo caso, la CVRS mediría el impacto de los mismos sobre la vida del paciente y no sólo la presencia o la severidad de los mismos. Tampoco se puede definir de forma simplificada como un estado de salud que presenta el paciente ya que la CVRS incluye el estado de salud y la influencia del mismo en el estado de bienestar del propio paciente (Ruiz y Pardo, 2005).

Los dominios o dimensiones normalmente incluidos en la CVRS se presentan a continuación de forma gráfica (Figura 5):

Figura 5. Ejemplos de dominios incluidos en la CVRS.



Instrumentos de medida de la CVRS

La CVRS se evalúa a través de medidas (cuestionarios, escalas, índices,...) diseñadas de forma específica y siguiendo, como ya se ha adelantado, un proceso psicométrico

estandarizado (FDA, 2009). Existen distintas clasificaciones de las medidas empleadas en la medición de la CVRS, en función de la naturaleza multidimensional del constructo a medir, las patologías que pueden afectar a la misma, la población dónde se evalúa, y las características del instrumento disponible.

En primer lugar, en relación a los aspectos de la CVRS evaluados, existen instrumentos llamados genéricos y específicos. Los primeros son aquellos que miden dominios o dimensiones compartidos por un amplio rango de poblaciones y lo hacen empleando un lenguaje familiar y comprensible que permite su aplicación en tales grupos y el estudio comparativo entre los mismos. Ejemplos de cuestionarios genéricos de uso extendido en IRS son el cuestionario *EuroQoL-5D* (EQ-5D) (Badía, Roset, Herdman, y Kind, 2001; Grupo EuroQoL, 1990), el *Nottingham Health Profile* (NHP) (Badía, Alonso, Brosa, y Lock, 1994) y el *36-Item Short Form Health Survey* (SF-36) (Rebollo, 2008; Stansfeld, Roberts, y Foot, 1997; Vilagut et al., 2008). En contraste, los instrumentos específicos, tienden a utilizar un lenguaje más adaptado a una condición o una población particular con dimensiones relevantes para las mismas, pero que pueden no ser compartidos con otros grupos. Por tanto no permiten normalmente establecer comparaciones entre distintas poblaciones. Eso sí, debido a su naturaleza específica suelen ser mucho más sensibles o capaces de detectar cambios relevantes en los grupos de interés (EUnetHTA, 2013). Ejemplos de instrumentos específicos son el *Incontinence Quality of Life Questionnaire* (I-QOL) para evaluar el impacto de los síntomas urinarios en la CVRS de los pacientes (Patrick et al., 1999; Patrick, Khalaf, Dmochowski, Kowalski, y Globe, 2013), el *Childhood Atopic Dermatitis Impact Scale* (CADIS) que mide CVRS asociada a patología dermatológica, y lo hace en una población específica como es la pediátrica (Chamlin et al., 2005; Chamlin et al., 2007), el *Dermatology Quality of Life Index* (DLQI) que evalúa también CVRS en dermatología pero en población adulta (Basra,

Fenech, Gatt, Salek, y Finlay, 2008), etcétera. Como se puede entrever, existe una gran cantidad de instrumentos específicos que se pueden ver en plataformas online como BibliPRO (<http://www.bibliopro.org/buscador/>) o PROQOLID (Patient Reported Outcome and Quality of Life Instruments Database: <http://www.proqolid.org/>), que clasifican los instrumentos, entre otros, en función de la población evaluada (infantil, adolescentes, adultos, geriátrica, hombres, mujeres,...) y en función de la patología (infecciones bacteriológicas, cardiovasculares, sistema digestivo, oftalmológicas, genitourinarias, neoplasmas, enfermedades degenerativas, depresión y ansiedad, enfermedades respiratorias,...). Normalmente se recomienda emplear en la IRS un instrumento específico y otro general para combinar las virtudes de ambos y posibilitar una evaluación exhaustiva de la CVRS de los pacientes a la vez que se permite la comparación con otros grupos (EUnetHTA, 2013).

Otra clasificación posible de las medidas de CVRS es en función de la naturaleza del mismo: el formato y métrica empleada (preguntas con respuestas tipo Likert, escalas visuales analógicas 0-100, respuestas dicotómicas, etc). Existen instrumentos que arrojan un perfil del paciente, esto es, puntuaciones para un número determinado de dimensiones como el EQ-5D, el SF-36, el NHP...) o un índice sumario de CVRS que agrega diferentes dimensiones en una única puntuación como los mencionados I-QOL, DLQI, o el FACT-G, empleado en oncología con módulos específicos de acuerdo al tipo de cáncer (<http://www.facit.org/FACITOrg/Questionnaires>), (Cella, 2000; Cella et al., 1993), (Cella, Hahn, y Dineen, 2002; King et al., 2010). Junto a estos, también hay medidas que ofrecen una combinación de índices sumarios y perfiles como el SF-36.

Finalmente, una de las distinciones de mayor peso en nuestros días entorno a los instrumentos de medida es la realizada a partir de la perspectiva de evaluación del

constructo. Así, se puede distinguir entre la perspectiva psicométrica y la econométrica. La primera, entiende la CVRS como la definida hasta ahora, la percepción de los pacientes sobre el impacto de su enfermedad, los tratamientos y las políticas sanitarias en su vida diaria, en su funcionamiento y su bienestar genera, (FDA, 2009; Patrick y Erickson, 1993). Como se ha indicado, es de especial interés para los profesionales clínicos con el fin de evaluar el estado de salud del paciente y los efectos de las intervenciones programadas. Desde la perspectiva econométrica, la CVRS asociada a un estado de salud es ponderada en función de las preferencias individuales o sociales. A pesar de la naturaleza multidimensional de la CVRS, este enfoque trata de agregar toda la información recogida sobre las distintas áreas en un único índice, normalmente medido en una escala comprendida entre: 0 (o incluso valores negativos) indicador del peor estado de salud posible o muerte hasta 1 que reflejaría el mejor estado de salud posible (NICE, 2013). La CVRS de los individuos se ajustaría pues de acuerdo a la importancia que le concede la comunidad o sociedad a esos niveles de CVRS. Se entiende que la valoración la debe realizar la sociedad como agente responsable de la financiación de los recursos sanitarios, al menos en nuestro país. Bajo este acercamiento se pretende disponer de una medida, lo más sencilla posible, pero que permita comparar de forma consistente el efecto de varias políticas sanitarias, intervenciones clínicas en distintas poblaciones o patologías (Bastida et al., 2010).

Aunque existen discrepancias sobre la pertinencia y validez de emplear estas medidas agregadas más simples (y simplificadoras) para analizar el beneficio aportado por las intervenciones sanitarias (Beresniak y Lamure, 2014; Schaeffer, 2013), la realidad es que, actualmente, las utilidades son empleadas de forma intensiva por los economistas de la salud y los gestores. La finalidad es poner en relación los años de vida ganados con cualquier intervención, con la calidad que la sociedad concede a los mismos. Son

los años de vida ajustados por calidad, que se incluyen en los análisis de coste-utilidad, tan extendidos en la evaluación económica actual (Bastida et al., 2010; EUnetHTA, 2013; NICE, 2013).

El valor de la CVRS y las utilidades descansa sobre el valor intrínseco de la recogida de información para valorar la calidad de las prestaciones o la aportación de las nuevas tecnologías frente a las ya disponibles. En suma a esto, gracias a su valor descriptivo del estado de salud poblacional general o de determinados colectivos, se constituyen como un instrumento fiable de intercambio de la información entre gestores, promotores, clínicos y pacientes. Por último pero no menos importante, su capacidad predictiva acerca de la evolución de los pacientes ha sido reconocida (Cenedese et al., 2006; Miilunpalo, Vuori, Oja, Pasanen, y Urponen, 1997; Román, Barberá, Castillo, Muñoz, y Escribano, 2013; Ruiz y Pardo, 2005; Rumsfeld et al., 1999).

Debido a la relevancia y pertinencia de la medición de la CVRS y los valores de utilidad, el uso de estas medidas está ampliamente extendida en las sociedades industrializadas existiendo agencias reguladoras en toda Europa. Como ejemplo, la European Network for Health Technology Assessment (EUnetHTA) ha publicado recientemente sus recomendaciones que son una referencia metodológica para los países de la UE (EUnetHTA, 2013). En nuestro país, existe una preferencia por el uso de instrumentos genéricos que también aporten valores de utilidad. Aun así, se reconoce la necesidad de incorporar la información extraída del empleo de medidas específicas de CVRS cuando los primeros no tienen unas propiedades psicométricas adecuadas (Bastida et al., 2010), en términos de validez de contenido y sensibilidad en la población objeto de evaluación. Estas propiedades psicométricas se explicarán en el apartado de

Propiedades psicométricas deseables en los instrumentos de medida en el entorno de la práctica de la atención socio-sanitaria.

Esquizofrenia, Trastorno Bipolar, Tratamiento Antipsicótico y CVRS

La esquizofrenia y el trastorno bipolar son síndromes mentales prevalentes en todo el mundo que afectan las funciones cognitivas y emocionales de las personas, disminuyendo su bienestar psicológico, profesional y social (Hirschfield RM, 2002; World Health Organization [WHO], 2001;). Afortunadamente, los nuevos tratamientos atípicos han demostrado ser útiles en la reducción de las recaídas y la gravedad de los síntomas (NICE, 2002; Royal Australian and New Zealand College of Psychiatrists, 2005). Sin embargo, a pesar de los avances indudables, también en términos de seguridad y tolerabilidad, los antipsicóticos tienen efectos no deseados sobre la CVRS de los pacientes: cambios en el peso corporal, la depresión y la ansiedad, las limitaciones en el desempeño de las actividades diarias, la rigidez, temblores y mareos, además de problemas de disfunción sexual. Estos efectos no deseados son acontecimientos adversos muy comunes entre los pacientes bajo tratamiento antipsicótico (Montejo, 2008; Newcomer, 2005; Novick et al., 2005) y pueden derivar en una mala adherencia a los tratamientos (Gentile, 2006; Lehman AF et al., 2004; Novick et al., 2005; Suárez y Haro, 2008). Como ya se ha indicado, la intensidad de estos síntomas puede conducir a la interrupción del tratamiento (Ascher-Svanum, Zhu, Faries, Lacro, y Dolder, 2006; Novick et al., 2005; Perkins et al., 2006; Perkins et al., 2008), lo que limita su efectividad y el aumento de los costes sanitarios (para el manejo de estos eventos adversos además de los personales y sociales). Por lo tanto, al comparar las terapias alternativas, los efectos secundarios y su impacto en la CVRS

tienen mucha importancia a fin de destacar el tratamiento antipsicótico más efectivo en cada caso.

En consecuencia, muchos esfuerzos han sido dirigidos en los últimos años a desarrollar y validar instrumentos clínicos específicos para evaluar la afectación en diferentes dominios de la CVRS de estos pacientes (Chouinard y Margolese, 2005; Gharabawi et al., 2005; Levine, Rabinowitz, Engel, Etschel, y Leucht, 2008; Lingjaerde, Ahlfors, Bech, Dencker, y Elgen, 1987; Patrick et al., 2009; Rabinowitz, Mehnert, y Eerdeken, 2006). En cuanto a los efectos secundarios, la medida más utilizada es la escala de efectos secundarios UKU (Lingjaerde et al., 1987). Este cuestionario, cumplimentado por los médicos o los pacientes consta de un total de 56 preguntas específicas (efectos cognitivos y psíquicos, síntomas extrapiramidales, síntomas asociados al sistema autónomo, hiperprolactinemia y disfunción sexual y otros). Además, también permite estudiar la interferencia de la medicación en las actividades diarias y sus consecuencias sobre el cumplimiento de la prescripción en los pacientes. Junto a esta medida, otras medidas de interés en el área, dirigidas a la presencia e intensidad de síntomas negativos y positivos de la esquizofrenia y de depresión, han sido adaptadas adecuadamente al español (Kay, Fiszbein, Vital-Herne, y Fuentes, 1990; Lobo et al., 2002; Peralta y Cuesta, 1994).

A pesar de esto, es necesario contar con un instrumento breve y sencillo, desde el punto de vista de los pacientes, para la descripción y clasificación de las consecuencias de los antipsicóticos modernos sobre la CVRS del paciente (Revicki, Matza, Flood, y Lloyd, 2005). Con este objetivo se desarrolló en Suecia el cuestionario TooL (*Tolerability and Quality of Life Questionnaire*) (Jönsson L, Lang A, y Lindström E, 2009; Lindström E, Jönsson L, y Berntsson A, 2009). Esta medida auto-aplicada refleja la percepción

subjetiva de los pacientes sobre la intensidad del beneficio y los efectos secundarios de los antipsicóticos (en pacientes con diagnóstico de esquizofrenia o trastorno bipolar). Se compone de 8 atributos y 4 niveles por dominio (escala tipo Likert). Estos dominios son el estado de ánimo, capacidades funcionales, y cualquier otro evento adverso con frecuencia asociado con fármacos antipsicóticos (fatiga, debilidad, aumento de peso, rigidez temblor, agitación, disfunción sexual, náuseas y mareos). Por lo tanto, el cuestionario se refiere tanto a cuestiones generales de CVRS (capacidades del estado de ánimo y función) extraídos de instrumentos genéricos como el EQ-5D (Badía, Roset M, Montserrat S, Herdman M, y Segura A, 1999), y a los eventos adversos frecuentemente asociados con los fármacos antipsicóticos modernos que se han descrito con anterioridad. En consecuencia, partiendo ya de un instrumento sólido a nivel conceptual, el objetivo del trabajo realizado fue realizar su adaptación cultural a España y el análisis de sus propiedades psicométricas. Así, se pretendía facilitar en práctica clínica habitual, la evaluación de los efectos secundarios en los pacientes tratados con antipsicóticos y su impacto sobre la CVRS de los pacientes.

Trastorno Depresivo Mayor y CVRS

Aunque los datos de prevalencia están sujetos a variabilidad según la metodología empleada y el ámbito poblacional y asistencial (Angst et al., 2002; Aragonés et al., 2004; Aragonés, Piñol, y Labad, 2006; Serrano-Blanco et al., 2010), el trastorno depresivo mayor (TDM) es una enfermedad mental con gran impacto en España a nivel epidemiológico, con una prevalencia-año poblacional en adultos estimada del 3,9% y una prevalencia-vida del 10,5% (Gabilondo et al., 2010; Haro et al., 2006).

El impacto del TDM sobre la CVRS de los pacientes es generalizado con gran sufrimiento personal. Esto es así debido, entre otros factores, a la morbilidad asociada

(Aragónés, Piñol, y Labad, 2009; Autonell et al., 2007; Charlson, Stapelberg, Baxter, y Whiteford, 2011; García-Campayo et al., 2008), la probabilidad de recurrencia (Romera et al., 2010; Ten Doesschate, Bockting, Koeter, y Schene, 2010; Vuorilehto, Melartin, y Isometsa, 2009), el impacto en la vida familiar y socio-laboral (Romera et al., 2010) así como el riesgo de suicidio asociado (Bernal et al., 2007). Todos estos factores contribuyen a la elevada carga socio-sanitaria y económica de la depresión (Moussavi et al., 2007; Sicras-Mainar, Blanca-Tamayo, Gutiérrez-Nicuesa, Salvatella-Pasant, y Navarro-Artieda, 2010; Ustun, Ayuso-Mateos, Chatterji, Mathers, y Murray, 2004). Como ejemplo, los costes sanitarios directos en nuestro país suponen aproximadamente el 33% del coste total, correspondiendo el 67% restante a los costes por pérdida de productividad laboral y la morbi-mortalidad asociada (Sicras-Mainar et al., 2010).

La investigación clínica en la depresión ha centrado sus esfuerzos en establecer un diagnóstico y tratamiento adecuados potenciando la efectividad de la intervención. Se persigue alcanzar la remisión completa de los síntomas (Grupo de Trabajo sobre el Manejo de la Depresión Mayor en el Adulto, 2008; National Collaborating Centre for Mental Health, 2009) debido al mejor pronóstico que implica en términos de recaídas, cronicidad y morbi-mortalidad (Romera et al., 2010; Sobocki, Ekman, Agren, Runeson, y Jonsson, 2006). Además, las estrategias de salud y las citadas guías terapéuticas, recomiendan un primer abordaje terapéutico de estos pacientes ya desde el ámbito de atención primaria y es precisamente en ese contexto asistencial donde se han constatado las mayores dificultades en la detección, el diagnóstico, el tratamiento y el seguimiento de los pacientes con depresión (Agüera, Failde, Cervilla, Díaz-Fernández, y Mico, 2010; Fernández et al., 2010). Estas carencias se habían puesto de manifiesto en nuestro país en un estudio sobre el manejo de la depresión en Atención Primaria en España, según el cual en este ámbito sólo el 34% de los pacientes con depresión llega a recibir

tratamiento con antidepresivos (Aragónés, Piñol, Labad, Folch, y Melich, 2004). Por otra parte, como se ha indicado, alcanzar la remisión es crucial para el pronóstico de la depresión y la reducción de la carga de la enfermedad (Sicras-Mainar et al., 2010; Sobocki, Jonsson, Angst, y Rehnberg, 2006; Sobocki et al., 2006). Toda esta información otorga aún mayor interés al hecho de poder identificar precozmente aquellos individuos con clínica de TDM en atención primaria y comenzar un enfoque terapéutico adecuado (Grupo de Trabajo sobre el Manejo de la Depresión Mayor en el Adulto, 2008; Lester y Howe, 2008). De hecho, ante los pacientes de riesgo, en las guías de referencia se recomienda una evaluación comprehensiva que implicaría, no sólo la relación de síntomas presentes en el sujeto, sino el análisis del deterioro funcional producido y su duración (National Collaborating Centre for Mental Health, 2009).

En armonía con estas recomendaciones, la escala CUDOS (*Clinically Useful Depression Outcome Scale*) fue diseñada en EEUU (Zimmerman, Posternak, y Chelminski, 2004; Zimmerman, Chelminski, McGlinchey, y Posternak, 2008) en el marco del proyecto MIDAS (*Rhode Island Methods to Improve Diagnostic Assessment and Service*). Este programa ha tenido como objetivo desarrollar instrumentos útiles para la práctica clínica habitual. Se trata por tanto de un instrumento breve que consta de sólo 18 preguntas con escala tipo Likert de 5 puntos para las respuestas y que se responde en menos de 3 minutos (Zimmerman y McGlinchey, 2008). Es un cuestionario fácilmente puntuable (en menos de 15 segundos) y útil para la práctica clínica habitual, ya que cubre completamente los síntomas de TDM y de trastorno distímico de la clasificación DSM-IV (American Psychiatric Association [APA], 2000) y además incluye aspectos genéricos de CVRS. Respecto a las propiedades psicométricas deseables en este tipo de instrumentos (FDA, 2009; Rey, 2007; Zimmerman et al.,

2008), la escala CUDOS ha demostrado tener unas adecuadas cualidades con un punto corte específico (punto de corte < 20 puntos) en la puntuación del cuestionario que señala la remisión del TDM, siendo por ejemplo la sensibilidad del 87,4% y la especificidad de 87,8% frente a la escala de depresión de Hamilton (Hamilton, 1960; Bobes et al., 2003) que fue el criterio considerado. Al ser el CUDOS auto-cumplimentado por el paciente, en lugar de administrado por el entrevistador, este estudio demostró un valor añadido para los médicos que buscan un criterio sencillo que les ayude a mejorar el control y seguimiento adecuados de los pacientes con TDM.

Como en el caso del cuestionario TooL, el trabajo realizado con la escala CUDOS fue llevar a cabo su adaptación cultural a España y analizar sus propiedades psicométricas con el objeto de poner a disposición de los profesionales sanitarios una herramienta útil para evaluar la CVRS de los pacientes con TDM y mejorar su detección temprana y seguimiento.

Propiedades psicométricas deseables en los instrumentos de medida en el entorno de la práctica de la atención socio-sanitaria

Debido a la naturaleza de las PROs la Psicometría, entendida como el conjunto de métodos técnicas y teorías implicadas en la medición de variables subjetivas, psicológicas, es, desde luego, un marco metodológico de referencia fundamental para el desarrollo y valoración de estas medidas (Muñiz, 2000). Desde el punto de vista de los expertos en Psicometría, los pasos a seguir para crear o adaptar a un contexto determinado una prueba están bien definidos. Aunque hay otras referencias nacionales de indudable interés (Abad, Garrido, Olea, y Ponsoda, 2006), Muñiz y Fonseca-Pedrero (2008) en su artículo sobre “La construcción de instrumentos de medida para la evaluación universitaria” detallan las principales fases que los investigadores deben

tener en cuenta a la hora de desarrollar instrumentos de medida. De forma resumida serían los siguientes (Muñiz y Fonseca-Pedrero, 2008):

- Marco general del instrumento de medida: Antes de comenzar a definir los ítems que configurarán la nueva prueba de evaluación, debe especificarse de forma clara el modelo conceptual y el contexto de aplicación. El detalle del marco conceptual facilitará que los mismos investigadores y la comunidad científica pueda tener claro el uso y la interpretación de las puntuaciones que serán clave para analizar las propiedades psicométricas de la misma.
- Definición operativa del contenido de la variable medida: el constructo a evaluar debe definirse en términos operativos mensurables para lo que resulta esencial tener bien presente el marco general y el modelo teórico subyacente. Habrá que definir todos los aspectos o dimensiones importantes de la variable (definición semántica), ponderar sus pesos y relaciones interdimensionales (definición sintáctica) para lo cual, la revisión exhaustiva de la literatura científica es fundamental.
- Especificaciones del instrumento de medida: Se debe definir el formato de presentación con las posibles adaptaciones, el modo de aplicación del mismo y los recursos materiales necesarios.
- Construcción de los ítems: En este punto, se procede a la redacción de las preguntas que configurarán el instrumento de evaluación. Los investigadores deben tener presente la necesidad de que los nuevos ítems sean de fácil lectura y comprensión, sean heterogéneos representando la máxima variabilidad y representatividad posible (pero en número equilibrado) cada dimensión considerada de interés en los pasos anteriores.

- Producción del instrumento: La nueva medida y su formato de presentación deben tener una correcta validez aparente o “*face-validity*” esto es, debe dar la impresión “de medir de manera objetiva, rigurosa, fiable y válida la variable de interés”.
- Creación de la base de datos y normas de puntuación y corrección: se debe definir cómo se interpretan y corrigen las respuestas en la medida en desarrollo a la vez que se especifican las puntuaciones de la dimensión o dimensiones evaluadas anteriormente definidas de forma diferencial.
- Desarrollo de un estudio piloto: Este punto es importante para poder testar la versión inicial del nuevo instrumento ya sea de forma cualitativa con grupos de discusión o cuantitativa (explorando además de forma preliminar propiedades psicométricas de la escala).
- Selección de instrumentos de medida convergentes: A partir de la revisión de la literatura ya comentada se podrá identificar qué otros conceptos e instrumentos de medida están relacionados con el objeto de medida en la nueva escala. El contraste de estas relaciones permitirá tener información clave sobre la validez del nuevo instrumento.
- Estudio de campo y estimación de las propiedades psicométricas: En este punto se desarrolla la investigación que permitirá poner a prueba el instrumento y contrastar las propiedades psicométricas del mismo (factibilidad, fiabilidad y validez). Un diseño correcto y un control de calidad riguroso son básicos para recoger de forma eficiente información fundamental, garantizando la representatividad y generalidad de los resultados. Esto será muy importante de cara a preparar el manual definitivo del mismo y presentar el instrumento a la comunidad científica.

Cuando se contemplan los pasos fundamentales que se han comentado para el desarrollo de una medida de evaluación, se facilita enormemente el contraste de las propiedades

psicométricas de las mismas (además de potenciarlas). Existen múltiples clasificaciones y definiciones de las mismas, pero de forma básica, siguiendo los conceptos comúnmente aplicados y testados desde la Teoría Clásica de los Test en Psicometría, serían las siguientes (American Educational Research Association, American Psychological Association, y National Council on Measurement in Education, 2004; Badía y Alonso, 2007):

- **Fiabilidad:** Con este término se apunta la necesidad de que los instrumentos de medida deben estar en lo posible, libres de errores de medida. Así, se mide la consistencia y estabilidad de las puntuaciones de la medida cuando el proceso de aplicación de la misma y las circunstancias de la medición permanecen constantes (en la medida de lo posible). Hay varios tipos de fiabilidad:
 - **Consistencia interna:** Refiere la estabilidad de las puntuaciones entre los diferentes elementos que componen el instrumento de medición.
 - **Fiabilidad test-retest:** Es la estabilidad temporal de las puntuaciones cuando se aplica el instrumento en varias ocasiones y las circunstancias de esta aplicación permanecen constantes en lo posible (no hay cambios significativos).
 - **Fiabilidad interjueces (entre observadores):** Se trata de la estabilidad de las puntuaciones obtenidas cuando el proceso de evaluación lo realizan distintos observadores capacitados para tal fin.
- **Validez:** Se refiere al grado con el que las evidencias teóricas y prácticas apoyan la interpretación de las puntuaciones del instrumento propuestas con el empleo del mismo (American Educational Research Association et al., 2004). Esta definición más que centrarse en un aspecto de validez del instrumento en sí, pone el acento sobre los niveles de evidencia que sustentan su interpretación en un aspecto o con un sentido determinado (constructo de medida, contenidos reflejados, sentido de las

respuestas, etc). No hay una única forma de examinar esta propiedad pudiéndose distinguir entre distintas fuentes o evidencias de validez:

- Evidencias basadas en el contenido del test: Se basa en el análisis de las relaciones entre el contenido de la prueba y el concepto que se quiere medir con su aplicación. Es común que se trabaje con una definición preestablecida que incluya las áreas o dimensiones que abarca, así como los límites conceptos que pudieran estar relacionados. El grado de pertinencia de cada dominio específico estará relacionado con las inferencias que se pueden establecer sobre el mismo a partir de las puntuaciones de la medida.
- Evidencias basadas en los procesos de respuesta: Se centra en las estrategias o los mecanismos que los respondedores ponen en marcha a la hora de cumplimentar el instrumento. En este sentido, el estudio de las respuestas de cada participante es útil para buscar el método de evaluación que mejor se ajusta al constructo que se pretende evaluar. Es de interés para conocer el grado en que el instrumento se puede ver afectado por otros factores relacionados con la cumplimentación del mismo y que no son relevantes en la definición del constructo que se pretende medir. Dentro de este apartado también pueden ser analizadas las valoraciones que hacen observadores o jueces sobre el concepto medido en los respondedores o examinados para comprobar en qué medida son sus respuestas se ajustan con la interpretación propuesta de las puntuaciones. La pertinencia del proceso de evaluación en relación al constructo evaluado es una variable importante en este apartado.
- Evidencias basadas en la estructura interna: En este subapartado se analiza en qué medida la relación entre los propios ítems y las dimensiones, áreas o dominios de la prueba se ajustan al constructo que se pretende evaluar y a partir

del cual se propone la interpretación de las puntuaciones de la medida. Los análisis para evaluar el grado en qué las relaciones entre los ítems soportan la asunciones sobre el objeto medido pueden variar en función del uso del instrumento.

- Evidencias basadas en las relaciones de las puntuaciones del instrumento con otras variables: Se trata de una fuente importante de validez del instrumento a desarrollar. Hay distintos tipos de información importante en este capítulo como puede ser el grado en que el instrumento nuevo permite predecir otras variables criterio supuestamente relacionadas. La fuerza de la asociación de sus puntuaciones con variables que miden constructos relacionados o divergentes también permiten evaluar la consistencia de la prueba así como el análisis de los resultados en distintos grupos que en principio deberían obtener resultados semejantes o diferentes. De esta forma, dentro de estas fuentes de validez se enmarcarían las siguientes:
 - Evidencias de tipo discriminante y convergente: Las relaciones entre las puntuaciones de la medida desarrollada y las de otras que evalúan constructos similares aportan evidencias de la validez convergente. En cambio, si se analizan las puntuaciones en relación a otros instrumentos cuyos constructos evaluados son distintos, hablaríamos de evidencias de validez discriminante.
 - Evidencias basadas en la relación con un criterio: Consiste en el estudio de la relación entre una variable externa, indicador o patrón de referencia (*gold standard*) y las puntuaciones del instrumento desarrollado. La cuestión fundamental aquí es evaluar en qué medida el nuevo instrumento permite predecir la valoración realizada por la variable criterio. Cuando la evaluación

con el nuevo instrumento es anterior a la realizada por el de referencia, se habla normalmente de evidencias de validez predictiva. En cambio, si se aplican en el mismo momento de evaluación, se entiende como una fuente de validez concurrente.

- Evidencias basadas en la capacidad de generalización de la interpretación de las puntuaciones: Son importantes cuando se trata de estimar cuál será la relación entre el nuevo instrumento y la variable criterio en una nueva situación. Este nuevo escenario que puede venir condicionado por ejemplo, por un nuevo método de evaluación del criterio, un grupo específico de respondedores, el momento de evaluación, etcétera. Normalmente, el corpus de evidencias acumulado con el progresivo empleo del instrumento desarrollado, va aportando una mayor capacidad de generalización de los resultados que se obtienen en el mismo.
- Evidencias basadas en las consecuencias de la evaluación: En este punto se refieren únicamente los resultados que se van obteniendo con la aplicación del instrumento y que pueden relacionarse con las evidencias anteriores relacionadas con la validez, haciendo que los investigadores revisen alguna de las mismas (por ejemplo, falta de sensibilidad en un grupo determinado o escasa representación de un área importante del constructo que se quiere medir).

Otra cuestión de interés en este apartado es conocer cuál es el efecto esperado con el empleo del instrumento (por ejemplo, selección de las terapias más efectivas o detección de los pacientes que requieren una rápida intervención). Cuando se define un potencial beneficio asociado al empleo de un instrumento, es interesante conocer si los resultados de su aplicación apoyan esa característica definida.

Además, de las propiedades y evidencias de validez psicométrica, cuando se pretende dedicar el instrumento a la práctica clínica y no sólo a la investigación sanitaria, es necesario considerar también su extensión, el tiempo de aplicación y la interpretación del mismo para que su uso sea factible por los profesionales especializados durante su práctica diaria. En este sentido, y a modo de ejemplo, en el ámbito de medicina clínica y la IRS, los expertos del proyecto MIDAS, (Zimmerman et al., 2008) dirigido a la generación de instrumentos útiles en práctica clínica y mejorar la evaluación de los pacientes en consulta, definen las siguientes 15 características como las más deseables para un cuestionario o instrumento de medida que sea aplicado en la práctica clínica, en este caso para la evaluación de la depresión:

1. Breve en la medida de lo posible
2. Aceptable para los pacientes
3. Cubre todos los criterios diagnósticos del DSM-IV para el trastorno depresivo mayor
4. Fiable (consistencia interna y fiabilidad test-retest)
5. Validez convergente
6. Validez discriminante
7. Indicador de la gravedad de los síntomas
8. Indicador de estado de remisión
9. La detección de casos capacidad como instrumento de cribado
10. Evalúa la función psicosocial
11. Evalúa la calidad de vida
12. Evalúa las ideas suicidas
13. Sensibilidad al cambio
14. Fácil de cumplimentar

15. Barato en su aplicación y corrección.

Como puede apreciarse, los principios y propiedades psicométricas deseables para el desarrollo de medidas propuestos desde la Teoría Clásica de los Test en Psicometría, y las recomendaciones de los expertos del grupo MIDAS no sólo están relacionadas sino que son compatibles.

Existen otras referencias genéricas con alcance internacional para el desarrollo de PROs para su aplicación en IRS y en ensayos clínicos desarrollados para obtener información sobre seguridad y eficacia de las tecnologías que serán posteriormente evaluadas. Son las propuestas por la FDA (FDA, 2009) y que han sido aceptados por las Agencias en Europa (EMA, 2014; EUnetHTA, 2013). De forma esquemática, en la siguiente Tabla 1 se presentan las propiedades psicométricas que debe tener una PRO:

Tabla 1. Propiedades Psicométricas de las PROs de acuerdo a la FDA (2009).

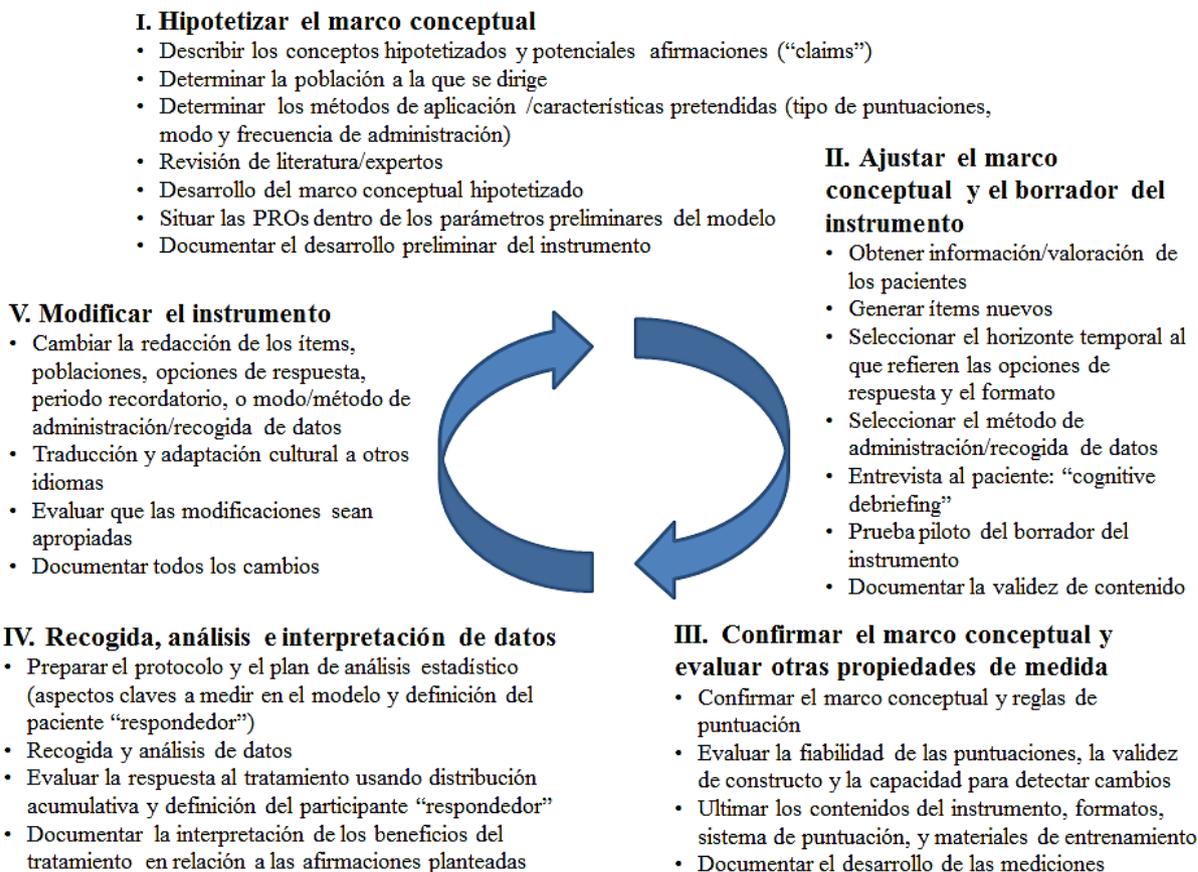
Propiedad	Tipo	¿Qué se evalúa?	Consideraciones FDA 2009
Fiabilidad	Fiabilidad test-retest o intra-observador	Estabilidad de las puntuaciones a través del tiempo cuando no se espera un cambio en el concepto de interés	<input type="checkbox"/> Coeficiente de correlación intraclase <input type="checkbox"/> Periodo temporal de evaluación (<i>recall period</i>)
	Consistencia interna	<input type="checkbox"/> Grado en el que los ítems que componen una escala miden el mismo concepto <input type="checkbox"/> Inter-correlación entre los ítems que contribuyen a una escala <input type="checkbox"/> Consistencia interna	<input type="checkbox"/> Alfa de Cronbach para el sumatorio de puntuaciones <input type="checkbox"/> Correlaciones ítem-total
	Fiabilidad inter-observador	Concordancia entre las respuestas cuando el PRO es administrado por diferentes entrevistadores	<input type="checkbox"/> Coeficiente de correlación intraclase
Validez	Validez de contenido	Evidencia de que el instrumento mide el concepto de interés, incluyendo datos provenientes de estudios cualitativos que demuestran que los ítems y dimensiones son apropiados e íntegramente relativos al concepto que pretenden medir, su uso y a la población a la que	<input type="checkbox"/> Derivación de todos los ítems <input type="checkbox"/> Calendario de entrevistas cualitativas <input type="checkbox"/> Transcripción de las entrevistas o grupos de interés <input type="checkbox"/> Ítems derivados de las transcripciones <input type="checkbox"/> Composición de los pacientes que participaron en el

Propiedad	Tipo	¿Qué se evalúa?	Consideraciones FDA 2009
		está dirigido. El probar otras propiedades de medida no reemplazará o rectificará cualquier problema con la validez de contenido	desarrollo del contenido <input type="checkbox"/> Transcripciones de las entrevistas para evaluar la comprensión del paciente
	Validez de constructo	Evidencia de que las relaciones entre ítems, dominios y conceptos se ajustan a una hipótesis <i>previa</i> acerca de las relaciones lógicas que deberían existir en las medidas de conceptos o puntuaciones relacionadas provenientes de grupos de pacientes similares o diversos	<input type="checkbox"/> Fuerza de la correlación al probar hipótesis a priori (validez discriminante y convergente) <input type="checkbox"/> Grado en el que el PRO puede distinguir entre grupos hipotéticamente a priori considerados diferentes (validez de grupos conocidos)
	Capacidad para detectar cambios relevantes	Evidencia de que un instrumento PRO puede identificar diferencias en las puntuaciones a través del tiempo en individuos o grupos (de manera similar a los usados en ensayos clínicos) que hayan experimentado cambios respecto al concepto medido.	<input type="checkbox"/> Cambios intra-sujeto a través del tiempo <input type="checkbox"/> Estadístico de tamaño del efecto

Nota: Traducción libre del autor de la tesis.

De forma complementaria a la propuesta de Muñiz y Fonseca (2008), en la IRS, se recalca la necesidad de considerar las PROs como instrumentos “dinámicos”, en constante desarrollo dinamizado por la dialéctica científica que se establece en la evaluación de los distintos constructos de interés y las propiedades psicométricas de los instrumentos (FDA, 2009; American Educational Research Association et al., 2004). Como ejemplo de esta necesidad, el EQ-5D (Grupo EuroQoL, 1990), probablemente la medida de mayor uso en nuestros días y con más de 20 años de desarrollo en IRS, está siendo modificado tras más de veinte años desde su desarrollo inicial, aumentando los niveles de respuesta desde 3 alternativas en la versión inicial (EQ-5D 3L) a 5 (EQ-5D 5L) con el fin de mejorar su sensibilidad al cambio (Herdman et al., 2011). En la siguiente Figura 6 se presentan las fases de desarrollo de una medida PRO (FDA, 2009):

Figura 6. Fases de Desarrollo de un instrumento para evaluar PROs siguiendo el criterio de la FDA (2009).



Nota: Traducción libre del autor de la tesis.

Estos modelos son aplicables también en el apartado de la IRS dedicado a la economía de la Salud, cuando se habla de las utilidades (recordemos que se definían como la calidad de los estados de salud de los pacientes ajustados normalmente por las preferencias sociales). En este caso, también se persigue de forma inequívoca que las herramientas empleadas para su medición, tengan un desempeño psicométrico contrastado y en los mismos términos que las propiedades comentadas anteriormente (Brazier et al., 2012).

Debido a la importancia que se concede en la actualidad en el ámbito sanitario a las PROs y las propiedades psicométricas de las medidas, es de interés señalar que existen guías de calidad internacionalmente aceptadas que explicitan cómo deben ser

documentados los resultados de estas medidas incluso en el marco de los ensayos clínicos (Calvert et al., 2013; Patient Centered Outcomes research Institute, 2012; Patrick, 2013; Turner et al., 2012). Para una revisión de las más empleadas puede consultarse la página de la ISOQOL (*International Society for Quality of Life Research*) en: <http://www.isoqol.org/> o el grupo CONSORT PRO (*Consolidated Standards of Reporting Trials*) en <http://www.consort-statement.org/consort-pro-extension/>.

Para terminar este apartado, indicar que se han presentado las propiedades psicométricas fundamentales de los test desde un enfoque más clásico, que resulta más familiar para los profesionales sanitarios y la comunidad científica en el ámbito de la IRS. No obstante, otros modelos, como la Teoría de Respuesta al Ítem (TRI), están siendo cada vez más aplicados en el desarrollo y evaluación de las PROs por sus valiosas aportaciones en la medición de los constructos incluidos (Rebollo et al., 2009; Rebollo et al., 2010). Desde la TRI, se trata de estimar la variable latente no desde una relación lineal entre la puntuación verdadera del sujeto y la obtenida como lo hace el modelo clásico, sino en función del niveles del sujeto en esa variable latente y las respuestas emitidas a las preguntas individuales del test (Abad et al., 2006; Muñiz, 2010; Navas, 1994). La creciente importancia de la TRI hace que, actualmente, en IRS es común encontrar trabajos donde se complementan las evidencias obtenidas desde ambos modelos para contrastar las propiedades psicométricos de las medidas de PROs (Brazier et al., 2012). De forma más que probable, estas aproximaciones serán cada vez más extendidas debido a su utilidad en la creación de Test Adaptativos Computerizados, como el CAT-Health, recientemente validado en nuestro país por nuestro equipo de investigación en colaboración con expertos de la Universidad de Oviedo (Rebollo et al., 2010).

Este conjunto de referencias conforman la base metodológica de los trabajos que configuran el proyecto de tesis doctoral que se presenta. En el siguiente punto, se presentan los objetivos de cada uno de los tres estudios llevados a cabo para el desarrollo o la adaptación cultural y validación de cada uno de los instrumentos ya introducidos, para pasar después, a presentar ya las publicaciones incluidas en la tesis doctoral.

OBJETIVOS E HIPÓTESIS

En la tesis doctoral que se presenta, se ha perseguido una serie de objetivos generales, compartidos por las investigaciones publicadas, y otros secundarios, más específicos de cada una de las mismas. A continuación, se detallan los objetivos, desglosándolos en cada una de los trabajos incluidos:

Objetivos

Generales

- Desarrollar o adaptar culturalmente, para su uso en nuestro país, instrumentos de medida cumplimentados por los pacientes que sean fiables y válidos para la evaluación de constructos de interés en la IRS.
- Proveer a los profesionales sanitarios en España de instrumentos útiles y prácticos para su uso durante el desempeño de su práctica clínica habitual.

Secundarios

Publicación nº 1: Desarrollo y validación de un cuestionario genérico para la evaluación de la satisfacción de los pacientes crónicos: Cuestionario SAT-Q.

- Desarrollar un cuestionario genérico para la evaluación de la satisfacción de los pacientes con enfermedad crónica respecto al tratamiento farmacológico y la atención sanitaria que reciben: cuestionario SAT-Q.

Publicación 2: Validación de una medida específica para evaluar la CVRS en pacientes con esquizofrenia y desorden bipolar: cuestionario *Tolerability and quality of life* (TooL).

- Adaptar culturalmente a nuestro país y analizar las propiedades psicométricas de la versión para España del cuestionario Tool para la evaluación del impacto de los tratamientos antipsicóticos en la CVRS de los pacientes esquizofrénicos o con trastorno bipolar.

Publicación 3: Adaptación al castellano de la escala *Clinically Useful Depression Outcome Scale* (CUDOS) para la evaluación de la depresión mayor desde la perspectiva del paciente.

- Adaptar culturalmente la escala CUDOS para la detección y evaluación de la intensidad del TDM en práctica clínica habitual en el ámbito de Atención Primaria y testar sus propiedades psicométricas.

Hipótesis

Asociadas a los objetivos generales

- Si se observan los procesos psicométricos necesarios para la adaptación cultural o el desarrollo de instrumentos de medida, es posible conseguir herramientas útiles y válidas para la medición de diferentes conceptos importantes en el ámbito socio-sanitario: satisfacción de los pacientes que reciben un tratamiento farmacológico crónico, impacto en la CVRS de los tratamientos antipsicóticos en pacientes esquizofrénicos y bipolares y la severidad del trastorno depresivo mayor.
- Si los instrumentos adaptados a nuestro país o desarrollados tienen una extensión limitada a la vez que unas correctas propiedades psicométricas, será posible proveer

a los profesionales sanitarios en España de instrumentos útiles y prácticos para su uso durante el desempeño de su práctica clínica habitual.

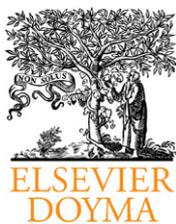
Asociadas a los objetivos secundarios

- I.** Si el cuestionario SAT-Q recoge a nivel de contenido los aspectos más importantes del modelo Jerárquico de niveles de Satisfacción y presenta unas propiedades psicométricas básicas adecuadas (factibilidad, fiabilidad y validez), entonces es posible contar con un instrumento para evaluar la satisfacción de los pacientes crónicos con el tratamiento farmacológico y con la atención sanitaria que reciben.
- II.** Si la adaptación lingüística del cuestionario TooL refleja los conceptos medidos por la versión original, es comprensible y pertinente para expertos y pacientes de nuestro país, y presenta una factibilidad fiabilidad y validez adecuadas, entonces es posible adaptar culturalmente a nuestro país el cuestionario TooL para la evaluación del impacto en la CVRS de los tratamientos antipsicóticos en pacientes esquizofrénicos o bipolares.
- III.** Si la versión traducida de la escala CUDOS incluye los contenidos de la versión original, es comprensible y pertinente para expertos y pacientes de nuestro país, y tiene una factibilidad fiabilidad y validez adecuadas, entonces es posible adaptar culturalmente a nuestro país la escala CUDOS para la evaluación del trastorno depresivo mayor en práctica clínica habitual en el ámbito de Atención Primaria.

TRABAJOS INCLUIDOS EN EL PROYECTO DE TESIS

ARTÍCULO 1 Desarrollo y validación de un cuestionario genérico para la evaluación de la satisfacción de los pacientes crónicos: Cuestionario SAT-Q.

Rebollo, P., Cuervo, J., Villa, G., Barreda, M. J., Tranche, S., Sánchez-Baragaño, M. A., y Prieto, M. A. (2010). Desarrollo y validación de un cuestionario genérico para la evaluación de la satisfacción de los pacientes crónicos: Cuestionario SAT-Q. *Aten Primaria.*, 42(12), 612-9. doi: 10.1016/j.aprim.2009.12.008



ORIGINAL

Desarrollo y validación de un cuestionario genérico para la evaluación de la satisfacción de los pacientes crónicos: Cuestionario SAT-Q

Pablo Rebollo^{a,*}, Jesús Cuervo^a, Guillermo Villa^a, María Jesús Barreda^b, Salvador Tranche^c, María Amor Sánchez-Baragaño^d y Miguel Ángel Prieto^e

^aBAP Health Outcomes Research, Oviedo, España

^bCentro de Salud Paulino Prieto, Servicio de Salud del Principado de Asturias (SESPA), Oviedo, España

^cCentro de Salud El Cristo, Servicio de Salud del Principado de Asturias (SESPA), Oviedo, España

^dCentro de Salud Teatinos-Corredoria, Servicio de Salud del Principado de Asturias (SESPA), Oviedo, España

^eCentro de Salud Vallobín-La Florida, Servicio de Salud del Principado de Asturias (SESPA), Oviedo, España

Recibido el 11 de noviembre de 2009; aceptado el 10 de diciembre de 2009

Disponible en Internet el 20 de marzo de 2010

PALABRAS CLAVE

Satisfacción;
Tratamiento
farmacológico;
Servicio sanitario;
Cuestionario

Resumen

Objetivo: Desarrollar y validar un cuestionario reducido de satisfacción del paciente crónico.

Diseño: Estudio epidemiológico, observacional, multicéntrico y transversal.

Emplazamiento: Cuatro centros de salud del Área Sanitaria IV del Principado de Asturias.

Participantes: Doscientos dos pacientes con enfermedad crónica diagnosticada y al menos un año desde inicio del tratamiento farmacológico.

Métodos: A partir de una revisión bibliográfica se construyó un banco de ítems inicial de 61 ítems relativos a satisfacción del paciente. El proceso de traducción-retrotraducción de los ítems y los paneles de expertos (n=8) y pacientes (n=30) concluyeron en una versión del cuestionario con 34 ítems (satisfacción con el tratamiento farmacológico y satisfacción con los servicios sanitarios). Finalmente, se realizó el análisis de reducción (correlación ítem total <0,2 y análisis factorial exploratorio) y validación del SAT-Q (factibilidad, fiabilidad y validez de criterio). Se aplicó el cuestionario SAT-Q y uno genérico de calidad de vida relacionada con la salud, el SF-12.

Resultados: El análisis de reducción concluyó en una versión final de 18 ítems: satisfacción general con el tratamiento (3), efectos adversos (3), olvidos (2), efectividad (3), conveniencia (4) y atención sanitaria (3). La consistencia interna de las dimensiones fue aceptable y los coeficientes de correlación intraclase moderados-altos. Se encontró una relación significativa entre las puntuaciones del SAT-Q (con la excepción de olvidos) y los Componentes Sumarios Físico y Mental del SF-12.

*Autor para correspondencia.

Correo electrónico: pablo@baphealth.com (P. Rebollo).

KEYWORDS

Satisfaction;
Pharmacological
treatment;
Health care service;
Questionnaire

Conclusiones: Se ha desarrollado y validado de forma preliminar un nuevo cuestionario integral para la evaluación de la satisfacción de los pacientes crónicos.

© 2009 Elsevier España, S.L. Todos los derechos reservados.

Development and validation of a generic questionnaire for evaluating satisfaction in patients with chronic disease: The SAT-Q Questionnaire

Abstract

Objectives: To develop and validate a brief general questionnaire to assess satisfaction in patients with chronic disease.

Design: epidemiological, observational, multicentre cross-sectional study.

Location: 4 Primary Health Care Centers from the IV Health Area in the Principality of Asturias.

Subjects: A total of 202 patients diagnosed with chronic illness who were following a pharmacological regimen for at least one year.

Methods: An extensive literature review was conducted to create an initial item pool of 61 questions regarding patient satisfaction. Then, a forward-backward translation into Spanish was carried out. Next, both the expert (n=8) and patient (n=30) panels led to a new version of 34 items (concerning satisfaction with medication and satisfaction with health service). Finally, scale item reduction (item-total correlation and exploratory factorial analysis —EFA—) and psychometrical validation (feasibility, reliability and criterion validity) of the SAT-Q— were evaluated. Both the SAT-Q and the SF-12 (to assess patient health related quality of life) were applied.

Results: Item reduction analysis resulted in 18 items: general satisfaction with medication (3), adverse-events (3), oversights (2), effectiveness (3), convenience (4) and Health services (3). Internal consistency (Cronbach α) and Intraclass Correlation Coefficients were moderate-high. Moreover, significant positive correlations between SAT-Q scores and SF-12 Physical and Mental Summary Components were found (with the exception of oversights).

Conclusions: A brief questionnaire for measuring satisfaction in chronic patients has been developed and preliminary validated.

© 2009 Elsevier España, S.L. All rights reserved.

Introducción

El estudio de la satisfacción del paciente es importante debido a sus implicaciones en el desarrollo de la práctica médica asistencial y en la evaluación de los recursos que la acompañan^{1,2}. También es destacable el valor de la satisfacción de los usuarios dentro de los programas de calidad de los servicios sanitarios³⁻⁶. En consecuencia, son muchas las publicaciones centradas en analizar este concepto de naturaleza multidimensional desde distintas ópticas. Uno de los modelos más difundidos es el modelo jerárquico de satisfacción del paciente⁷ que postula la existencia de 3 niveles básicos: satisfacción respecto a los servicios sanitarios (accesibilidad, la relación médico-paciente, la calidad de los servicios prestado, etc.); satisfacción respecto al conjunto de tratamientos recibidos (sean farmacológicos o no) y, por último, satisfacción respecto al tratamiento farmacológico específico^{7,8}. El objetivo es obtener un indicador fiable que complemente los resultados de efectividad en la aplicación del tratamiento en la práctica clínica habitual⁷⁻⁹.

En una revisión sobre los estudios realizados que evalúan «satisfacción del paciente»⁹ se concluye que, con escasas excepciones, los cuestionarios analizados son poco fiables y válidos. Afortunadamente, en los últimos años han sido desarrollados y adaptados nuevos cuestionarios limitados a

la satisfacción con el tratamiento^{10,11}. Así, se ha conseguido conocer mejor sus implicaciones clínicas en diferentes patologías: se evidencia el hecho de que los pacientes satisfechos tienen mayor disposición a la cooperación con el terapeuta mostrándose más rigurosos en el seguimiento de las directrices médicas, lo cual, incide en el éxito de la terapia. Igualmente, las personas satisfechas intercambian información con su médico de forma más fiable y fluida¹² mejorando la adherencia al tratamiento y la calidad terapéutica del mismo.

A pesar de que existen varios cuestionarios genéricos, que miden satisfacción del paciente con el tratamiento farmacológico^{13,14}, ninguno permite la evaluación de áreas importantes relacionadas con la adherencia a los tratamientos como la relación con su médico, la información que el paciente recibe sobre su enfermedad, el tiempo dedicado al paciente, etc.

Teniendo en cuenta lo anterior, en este trabajo se presenta el desarrollo y la validación preliminar de un cuestionario genérico para la evaluación de la satisfacción de los pacientes crónicos con el tratamiento farmacológico, pero también, respecto a la atención en los servicios sociosanitarios. De esta forma, será posible analizar los niveles básicos de la satisfacción del paciente en el marco teórico del modelo jerárquico de satisfacción del paciente⁷.

Sujetos y método

Fase I. Generación del banco de ítems y panel de expertos y pacientes

En primer lugar, se completó una revisión sistemática (Cochrane Library, Pubmed, EMBASE y publicaciones específicas como Value in Health o Health Technology Assessment) de todas las publicaciones de satisfacción (n=131) para la identificación de los cuestionarios genéricos y específicos. Así, se generó un listado inicial de 61 ítems. A continuación, 2 traductores bilingües tradujeron al español los ítems encontrados. Tras la lectura de los expertos, estos ítems se volvieron a traducir al inglés (retrotraducción) para asegurar su contenido equivalente. Como resultado, se obtuvo una versión modular de satisfacción del paciente (*versión 1.0*) con 61 ítems siguiendo el modelo jerárquico de satisfacción (7): satisfacción global con el tratamiento (6 ítems), efectividad del tratamiento (8 ítems), efectos secundarios del tratamiento (12 ítems), comodidad/conveniencia de administración (10 ítems), costes del tratamiento (4 ítems), olvidos de las tomas del tratamiento (5), relación médico-paciente (6 ítems), información al paciente (6 ítems) y tiempo dedicado al paciente (4). Esta versión fue revisada por 5 médicos de Atención Primaria que valoraron la importancia y la comprensibilidad de los ítems en una escala tipo Likert (0: mínimo; 4: máximo). Tras la discusión sobre las puntuaciones de cada ítem, se procedió a eliminar aquéllos que no alcanzaron una puntuación media mínima de 2,5 puntos en las valoraciones. De esta forma, se obtuvo una nueva versión del SAT-Q (*versión 1.1*) de 34 ítems (todas las preguntas referidas a los costes de las prestaciones y tratamientos se eliminaron porque no se consideraron influyentes en la satisfacción del paciente dentro del actual Sistema Nacional de Salud). Esta versión se presentó a 30 pacientes crónicos que llevaban al menos 5 años desde que fueron diagnosticados, para que evaluaran la comprensibilidad e importancia de cada uno de los ítems. No se eliminó ningún ítem (la puntuación media en cada uno de ellos fue superior a los 2,5 puntos) pero 9 fueron modificados con el fin de mejorar la comprensibilidad. Se obtuvo así, una *versión 1.2* con 34 ítems distribuidos en aspectos específicos del tratamiento farmacológico (satisfacción global con el tratamiento farmacológico, efectividad del tratamiento, efectos secundarios, conveniencia de la forma de administración/comodidad y olvidos en las tomas) y aspectos generales de satisfacción con el servicio sanitario (relación médico-paciente, tiempo dedicado al paciente e información al paciente).

Fase II. Reducción de ítems y validación de la versión 1.2 del cuestionario SAT-Q (módulo I)

Se desarrolló un estudio epidemiológico, observacional, multicéntrico y transversal. Se reclutaron pacientes que llevaban al menos un año en tratamiento de una enfermedad crónica: migraña, artritis y otras enfermedades músculo-esqueléticas, depresión, diabetes, hipertensión u otras enfermedades crónicas (principalmente cardiovasculares y neurológicas). Todos los pacientes fueron informados por sus

médicos y prestaron su consentimiento informado antes de participar en el estudio.

El tamaño muestral del estudio se estimó para poder garantizar un mínimo de 5 pacientes crónicos por cada uno de los 34 ítems presentados del cuestionario SAT-Q, de tal forma que fuera posible proceder a la reducción y validación preliminar del cuestionario. La recogida de información se realizó en 4 centros de Atención Primaria de la ciudad de Oviedo y se llevó a cabo en 2 visitas diferenciadas:

- Primera visita: se recogieron datos socio-demográficos (edad, sexo, nivel cultural, situación de convivencia e ingresos económicos) y datos clínicos (enfermedad crónica diagnosticada, tratamiento, fecha de diagnóstico e inicio de tratamiento farmacológico y comorbilidades). Además, los pacientes cumplieron la *versión 1.2* del SAT-Q y el cuestionario de salud percibida genérico SF-12.
- Segunda visita (retest): entre 2-3 semanas después de la primera visita, 50 pacientes que no habían registrado cambios en la medicación ni detectado alteración en el estado de salud (a juicio clínico) cumplieron de nuevo la *versión 1.2* del SAT-Q.

El cuestionario SAT-Q aplicado consistió en 34 preguntas de tipo Likert (5 alternativas de respuesta) relativas a la satisfacción del paciente con el tratamiento farmacológico y la evaluación de su satisfacción con la calidad del servicio.

El cuestionario de calidad de vida relacionada con la salud (CVRS) SF-12 está formado por 12 ítems del cuestionario genérico SF-36. Con su aplicación, se obtienen las puntuaciones sumarias física (PCS) y mental (MCS) de los pacientes respecto a la población general^{15,16}.

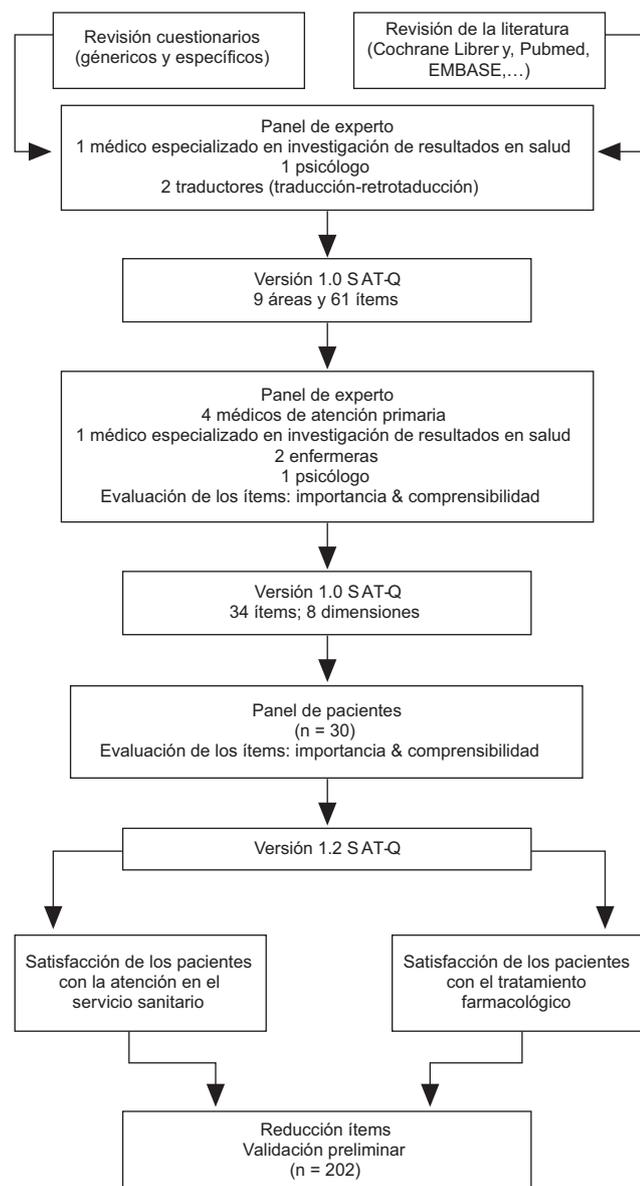
La factibilidad de la versión 1.2 del cuestionario SAT-Q se evaluó en la primera visita contabilizando el número de respuestas incompletas en comparación con el SF-12.

Para la extracción de las dimensiones del cuestionario y la reducción del número de ítems se comprobó la correlación ítem-total de los ítems de cada dimensión siguiendo un criterio de eliminación de aquellos ítems con puntuación en este índice de discriminación por debajo de 0,20. También se aplicó un análisis factorial exploratorio (AFE) de componentes principales en cada una de las dimensiones. Los ítems que presentaban un índice de comunalidad por debajo del porcentaje de varianza explicado por el primer autovalor fueron eliminados. Hecho esto, se aplicó un nuevo AFE (rotación Varimax) para comprobar la pertinencia de la extracción de una única puntuación sumaria en cada una de las dimensiones. A continuación, estas puntuaciones se transformaron linealmente en una escala 0 (peor satisfacción, más olvidos, menos conveniente el tratamiento, etc.)-100 (mejor satisfacción, menos olvidos, más conveniente el tratamiento, etc.) aplicando el algoritmo:

$$Y' = (Y_{\text{obs}} - Y_{\text{min}}) / (Y_{\text{max}} - Y_{\text{min}}) \times 100$$

Por otra parte, la fiabilidad de la información se analizó mediante el estadístico α de Cronbach (consistencia interna) y también con el coeficiente de correlación intraclase (CCI) para evaluar la estabilidad de las puntuaciones de las dimensiones (test-retest). Respecto a la validez de criterio, se analizó la relación existente entre las dimensiones que componen el SAT-Q. Se esperaba que la dimensión general

de satisfacción con el tratamiento se asociara de forma moderada con las dimensiones específicas y de forma menos intensa con la genérica de satisfacción con la atención sanitaria. En segundo lugar, se estudió también la relación entre el SAT-Q y las puntuaciones del SF-12 (PCS y MCS). Se asumió que las personas que tienen una mejor CVRS estarán más satisfechas que las personas con menor CVRS tal y como se ha reflejado en estudios anteriores tanto en satisfacción específica con el tratamiento¹⁷ como en satisfacción general con los servicios sanitarios¹⁸. En estos casos, se aplicó la prueba no paramétrica del Coeficiente de Correlación de Spearman.



Esquema general del estudio. Esquema del estudio de desarrollo y validación del cuestionario SAT-Q: estudio epidemiológico, observacional, multicéntrico y transversal.

Resultados

Participaron en el estudio 202 pacientes (70 hombres) con enfermedad crónica. El conjunto de las variables

socio-demográficas y clínicas que describen el conjunto de pacientes del estudio se recogen en la [tabla 1](#). La correlación ítem-total y los AFE permitieron encontrar una estructura con 2 dimensiones generales: satisfacción con el tratamiento farmacológico (3 ítems con el 72,35% de varianza explicada), y satisfacción con la atención sanitaria (3 ítems y 64,74% de la varianza) ([tabla 2](#)). Además, se encontraron 4 dimensiones específicas de la satisfacción con el tratamiento (efectos adversos (3 ítems; 30,22% de varianza explicada), efectividad (3 ítems; 14,96% de varianza explicada), comodidad/conveniencia (4 ítems con el 14,14% de varianza explicada) y, finalmente, olvidos (2 ítems; 10,1% de varianza explicada). [tablas 3](#).

En cuanto a la factibilidad de los cuestionarios, en el caso del SF-12, 201 pacientes (99,50%) dieron respuesta a la totalidad de las preguntas durante la primera visita. Por su parte, en el cuestionario SAT-Q el total de pacientes que lo cumplimentaron correctamente también estuvo por encima del 95% (n= 196; 97,03%). Respecto a la fiabilidad del cuestionario de satisfacción la consistencia interna fue aceptable ([tabla 4](#)). Además, los CCI entre las puntuaciones en cada una de las fueron igualmente moderadas-altas. Una vez transformadas linealmente las puntuaciones de las escalas, se observaron puntuaciones muy altas (efecto techo) en todas las dimensiones ([tabla 4](#)).

Por otro lado, respecto a la validez de criterio, la relación entre la dimensión general de satisfacción con el tratamiento farmacológico y el conjunto de las dimensiones específicas (con la excepción de olvidos de las tomas) es significativa moderada-alta¹⁷ mientras que respecto a la escala genérica de satisfacción con la atención médica, la relación es significativa pero leve ([tabla 5](#)). Además, se aprecia una relación positiva entre la dimensión general de satisfacción con el tratamiento farmacológico y el SF-12 (en especial, respecto a PCS). De igual forma, las dimensiones de efectividad, comodidad/conveniencia y efectos adversos están positivamente relacionadas con el SF-12. En el caso de la dimensión de olvidos, no se encontró ninguna asociación ([tabla 5](#)).

Discusión

En los últimos años, los agentes implicados en la investigación y gestión de tecnologías sanitarias han venido incorporando a los objetivos básicos de eficacia, seguridad y racionalización, otros importantes aspectos relacionados con la opinión del paciente como su grado de satisfacción. Se incluyen en este ámbito datos relacionados con el «grado de satisfacción con los servicios sociosanitarios», «grado de satisfacción con el tratamiento en su conjunto» y finalmente el «grado de satisfacción con el tratamiento farmacológico»^{7,8}. La medición de estos constructos se ha potenciado debido a las implicaciones que tienen en la adherencia al tratamiento y por tanto, en la efectividad de los mismos². La investigación de la que es fruto el cuestionario SAT-Q se ha desarrollado con el objeto de aportar un instrumento para la evaluación integral de los niveles de la satisfacción del paciente. La versión empleada inicialmente en el estudio (*versión 1.2*) ha sido posible gracias a los paneles de expertos y el análisis de la

Tabla 1 Características sociodemográficas y clínicas de los pacientes estudiados (n = 202)

		n (%)
<i>Sexo</i>	Varón	70 (34,65)
	Mujer	132 (65,35)
<i>Edad, media (DE)</i>	62,97 (12,93)	
<i>Años de tratamiento, media (DE)</i>	9,07 (8,21)	
<i>Grado académico; n (%)</i>	Sin estudios	49 (24,26)
	Estudios de graduado escolar	78 (38,61)
	Estudios de bachiller superior, BUP, FP	47 (23,27)
	Estudios universitarios	26 (12,87)
	No contesta	1 (0,49)
<i>Ocupación</i>	Trabajador/a fuera del hogar	41 (20,30)
	Trabajador/a en el hogar	70 (34,65)
	Jubilado/a por edad	60 (29,70)
	Jubilado/a por enfermedad	23 (11,38)
	Sin empleo	6 (2,97)
	Estudiante	1 (0,49)
	Otros	1 (0,49)
<i>Convivencia</i>	Vive solo/a	26 (12,87)
	Vive en familia o acompañado/a	174 (86,14)
	Vive en una institución	1 (0,49)
<i>Enfermedad crónica</i>	Migraña	4 (1,98)
	Artritis y otras enfermedades músculo-esqueléticas	36 (17,86)
	Depresión	31 (15,34)
	Diabetes	19 (9,41)
	Hipertensión	68 (33,67)
	Otras	44 (21,79)
<i>Comorbilidades</i>	0	38 (18,81)
	1	54 (26,74)
	2	54 (26,73)
	3	25 (12,38)
	4	24 (11,88)
	5 o más	7 (3,4)
<i>Número de tratamientos decomorbilidades</i>	0	59 (29,21)
	1	46 (22,77)
	2	31 (15,35)
	3	29 (14,35)
	4	17 (8,46)
	5 o más	20 (9,95)
<i>SF-12 MCS (Componente sumario mental):</i>	media (DE)	44,63 (13,07)
<i>SF-12 PCS (componente sumario físico):</i>	media (DE)	40,61 (11,42)

información resultante siguiendo la metodología cualitativa empleada por otros autores en trabajos anteriores^{13,14}.

El análisis aplicado en la reducción de la versión preliminar ha concluido en un cuestionario que consta de 18 ítems tipo Likert (tablas 2,3) distribuidos en 6 dimensiones: 2 genéricas (satisfacción con el tratamiento farmacológico y con la atención sanitaria), y 4 específicas (efectividad, efectos adversos, comodidad/conveniencia y por último, olvidos en la tomas). Este instrumento parece tener una aceptable consistencia interna, y un CCI en todas las dimensiones moderado-alto. Por otra parte, la correlación positiva de las puntuaciones sumarias del SF-12, tanto con la escala genérica de satisfacción con el tratamiento como con las distintas dimensiones (general, efectividad,

efectos adversos y en menor medida, conveniencia/comodidad) apoyan un rasgo de la validez de la medida (tabla 5). De esta forma, parece confirmarse el hecho de que los pacientes con un estado de salud percibido más alto tienen a estar más satisfechos con el tratamiento recibido tal y como otros autores han encontrado en el ámbito de la satisfacción con los servicios sanitarios¹⁹ o en relación a tratamientos específicos^{18,20,21}. El hecho de que en la dimensión de olvidos no se haya encontrado la misma relación con el componente PCS del SF-12 puede ser comprensible si se considera que los olvidos pueden deberse a otras causas, tales como el número de tomas o la cantidad de fármacos que el paciente debe consumir para el tratamiento de las comorbilidades. Aún así, cabe señalar

Tabla 2 Análisis factorial exploratorio de las dimensiones genéricas del cuestionario SAT-Q

<i>Varianza total explicada: ítems satisfacción general con el tratamiento farmacológico</i>			
<i>Componentes</i>	<i>Total</i>	<i>% de la varianza</i>	<i>% acumulado</i>
1	2,170	72,346	72,346
2	0,505	16,846	89,192
3	0,324	10,808	100,000
<i>Matriz de Componentes</i>			
<i>Ítems</i>	<i>Satisfacción general con el tratamiento farmacológico</i>		
<i>Satisfacción respecto al tratamiento</i>	0,891		
<i>Satisfacción con la eficacia general del tratamiento</i>	0,831		
<i>Ha obtenido los beneficios que esperaba</i>	0,828		
<i>Varianza total explicada: ítems satisfacción general con la atención sanitaria en el centro de salud</i>			
<i>Componentes</i>	<i>Total</i>	<i>% de la varianza</i>	<i>% acumulado</i>
1	1,942	64,740	64,740
2	0,652	21,726	86,467
3	0,406	13,533	100,000
<i>Matriz de Componentes</i>			
<i>Ítems</i>	<i>Satisfacción general con la atención sanitaria</i>		
<i>Actitud del médico</i>	0,864		
<i>Tiempo dedicado a su problema</i>	0,795		
<i>Información recibida</i>	0,750		

Tabla 3 Análisis factorial exploratorio de las dimensiones específicas del tratamiento farmacológico del cuestionario SAT-Q

<i>Varianza total explicada</i>				
<i>Componentes</i>	<i>Total</i>	<i>% de la varianza</i>	<i>% acumulado</i>	
<i>Efectos adversos</i>	3,627	30,224	30,224	
<i>Efectividad</i>	1,496	12,464	42,688	
<i>Conveniencia</i>	1,414	11,779	54,467	
<i>Olvidos</i>	1,065	8,874	63,341	
<i>Matriz de componentes rotados (rotación varimax)</i>				
<i>Ítems</i>	<i>Efectos adversos</i>	<i>Efectividad</i>	<i>Conveniencia</i>	<i>Olvidos</i>
<i>Interferencia del tratamiento</i>	0,817	0,083	0,144	0,113
<i>Efectos secundarios molestos</i>	0,791	0,090	0,200	0,176
<i>Limitaciones en el rendimiento personal</i>	0,715	0,235	0,143	-0,170
<i>Alivio síntomas enfermedad</i>	0,139	0,858	0,077	0,108
<i>Mejora en la calidad de vida</i>	0,352	0,753	-0,007	0,043
<i>Se siente mejor con el tratamiento</i>	-0,046	0,714	0,320	0,032
<i>Continuar el tratamiento actual</i>	-0,124	0,235	0,797	0,019
<i>Ha pensado abandonar el tratamiento</i>	0,344	0,031	0,633	0,142
<i>Administración cómoda del tratamiento</i>	0,308	0,008	0,554	0,061
<i>Ha pedido cambiar la forma de administración</i>	0,248	0,147	0,425	0,202
<i>Olvida tomar medicación</i>	0,016	-0,023	0,156	0,858
<i>Le recuerdan que tome la medicación</i>	0,089	0,171	0,084	0,854

Tabla 4 Puntuaciones del cuestionario SAT-Q y análisis de fiabilidad

Dimensiones SAT-Q	N	Mínimo	Máximo	Media	Desviación típica	α de Cronbach	CCI
General	201	0,00	100,00	55,307	21,019	0,807	0,812
Efectividad	200	0,00	100,00	64,682	21,260	0,737	0,764
Efectos adversos	200	0,00	100,00	76,864	25,286	0,753	0,632
Conveniencia	199	0,00	100,00	80,486	20,522	0,603	0,932
Olvidos	199	0,00	100,00	85,490	20,117	0,707	0,869
Atención sanitaria	194	0,00	100,00	84,536	17,518	0,724	0,835

Tabla 5 Correlación interdimensional del SAT-Q y SF-12

Correlaciones (Spearman)	General	Efectividad	Efectos Adversos	Conveniencia	Olvidos	Atención Sanitaria	PCS-12
General							
Efectividad	0,666*						
Efectos adversos	0,401*	0,300*					
Conveniencia	0,354*	0,305*	0,561*				
Olvidos	0,130	0,153**	0,198*	0,280*			
Atención sanitaria	0,223*	0,139	0,181**	0,239*	0,165**		
PCS-12	0,391*	0,300*	0,364*	0,279*	0,082	0,165**	
MCS-12	0,272*	0,181**	0,326*	0,273*	-0,037	-0,033	0,140**

PCS: componente sumario físico; MCS: componente sumario mental.

*La correlación es significativa a nivel 0,01 (bilateral).

**La correlación es significativa a nivel 0,05 (bilateral).

que en este estudio no se encontró una relación significativa entre el número de olvidos y el número de enfermedades declaradas por el paciente. Además, los AFE y la asociación moderada-alta encontrada entre la satisfacción con el tratamiento farmacológico y las distintas dimensiones específicas (con la excepción de olvidos) así como la asociación más débil respecto a la satisfacción con la atención sanitaria, permite apoyar la diferenciación de niveles de satisfacción definidos por el modelo jerárquico de la satisfacción⁷.

A pesar de lo anterior, cabe señalar que el efecto techo encontrado en las puntuaciones (tabla 4), aún siendo frecuente en encuestas de satisfacción^{10,13,22}, hace necesario realizar nuevos estudios sobre la discriminación y la sensibilidad al cambio del SAT-Q mediante registros de los cambios de tratamiento o en cambios en la salud percibida. En suma a lo anterior, sería necesario analizar la validez convergente de este instrumento frente a otros instrumentos de satisfacción ya empleados en la investigación^{11,13,14,22} y respecto a otros específicos de constructos asociados como el cumplimiento del paciente con el tratamiento farmacológico prescrito (por ejemplo, el test de Morisky-Green²³).

En conclusión, ha sido presentado un instrumento genérico breve nuevo para la evaluación de la satisfacción en pacientes con tratamiento crónico con unas correctas

propiedades psicométricas preliminares de fiabilidad y validez. De esta forma, se pretende facilitar la evaluación comprensiva de este constructo multidimensional.

Lo conocido sobre el tema:

- La satisfacción de los pacientes y el éxito de las intervenciones terapéuticas están relacionados.
- El modelo jerárquico de satisfacción respecto a los servicios sanitarios, respecto a las intervenciones terapéuticas y, respecto al tratamiento farmacológico específico.
- No existe ningún cuestionario en español que evalúe de forma comprensiva la satisfacción del paciente.

Qué aporta este estudio:

- Se ha desarrollado un instrumento de evaluación de la satisfacción de los pacientes que integra aspectos relativos al tratamiento farmacológico y también a la calidad del servicio prestado.
- Se presenta el estudio de las propiedades psicométricas básicas del nuevo cuestionario para su validación.

Financiación

Este trabajo es resultado de un proyecto de investigación *financiado* en parte con fondos procedentes de:

- Programa Torres Quevedo PTQ04-1-0137.
- Ayuda de Proyectos de Investigación en Empresas de la FICYT IE05-031.

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Agradecimientos

Los autores quieren mostrar su reconocimiento y mayor gratitud a la Gerencia del Área IV del Principado de Asturias así como a todos los investigadores y pacientes que, de forma desinteresada, participaron en el estudio.

Bibliografía

1. Lenderking WR. Brief reflections on treatment satisfaction. *Value in Health*. 2005;8(Supl 1).
2. Badía X. La satisfacción con el tratamiento. *Med Clin (Barc)*. 2005;125:98–9.
3. Carr-Hill RA. The measurement of patient satisfaction. *Journal of Public Health Medicine*. 1992;14:236–49.
4. Asadi-Lari M, Tamburini M, Gray D. Patients' needs, satisfaction, and health related quality of life: Towards a comprehensive model. *Health and Quality of Live Outcomes*. 2004;2.
5. Hernández ME, Ochando AG, Mora JC, Lorenzo SM, López KR. Satisfacción del paciente en una unidad de hemodiálisis: Objetivo de calidad asistencial en enfermería. *Revista de la Sociedad Española de Medicina Nefrológica*. 2005;8.
6. Mira JJ, Aranz J. La satisfacción del paciente como una medida del resultado de la atención sanitaria. *Med Clin*. 2000;114(Supl. 3):26–33.
7. Shikar R, Rentz AM. Satisfaction with medication: an overview of conceptual, methodological and regulatory issues. *Value in Health*. 2004;7:205–15.
8. Speight J. Assessing Patient satisfaction: concepts, applications, and measurement. *Value in Health*. 2005;8(Suppl. 1):S6–8.
9. Sitzia J. How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int J Qual Health Care*. 1999;11:319–28.
10. Condes E, Koldo A, Dalmau D, Estrada JM, Force L, Górgolas M, et al. Validación del cuestionario de satisfacción con el tratamiento antirretroviral: cuestionario CESTA. *Enferm Infecc Microbiol Clin*. 2005;23:586–92.
11. Ruiz M, Campillo MA, Monfort J, Pardo A, Rejas J, Soto J. Adaptación al castellano y validación del cuestionario Arthritis Treatment Satisfaction Questionnaire. *Med Clin (Barc)*. 2005;125:84–92.
12. Crowe R, Gage H, Hampson S, Hart J, Kimber A, Storey L, et al. The measurement of satisfaction with healthcare: Implications for practice from a systematic review of the literature. *Health Technology Assessment*. 2002;6:1–224.
13. Atkinson MJ, Sinha A, Hass SL, Colman SS, Kumar RN, Brod M, et al. Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *HQLO*. 2004;2:12.
14. Ruiz M, Pardo A, Rejas J, Soto J, Villasante F, Aranguren JL. Development and validation of the "Treatment Satisfaction with Medicines Questionnaire" (SATMED-Q). *Value in Health*. 2008;7:204–15.
15. Ware Jr JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34:220–33.
16. Vilagut G, Ferrer M, Rajmil M, Rebollo P, Permanyer-Miralda G, Quintana JM, et al. El cuestionario de salud SF-36 español: una década de experiencia y nuevos desarrollos. *Gac Sanit*. 2005;19:135–50.
17. Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd ed. NJ: Lawrence Erlbaum Associates; 1988.
18. Pleil AM, Coyne KS, Reese PR, Jumadilova Z, Rovner ES, Kelleher CJ. The validation of the Patient-Rated Global Assessments of Treatment Benefit, Satisfaction, and Willingness to Continue- The BSW. *Value in Health*. 2005;8(Suppl. 1): S25–34.
19. Xiao H, Barber JP. The Effect of Perceived Health Status on Patient Satisfaction. *Value in Health*. 2008;11:719–25.
20. Abetz L, Coombs JH, Keineger DL, Earle CC, Wade C, Bury-Maynard D, et al. Development of the cancer Therapy satisfaction Questionnaire: Item Generation and Content Validity Testing. *Value in Health*. 2005;8(Suppl. 1):S41–53.
21. Campbell UB, Stang P, Barron R. Survey Assessment and Continuation of a Satisfaction with Pharmacological Treatment for Urinary Incontinence. *Value in Health*. 2008;11:726–32.
22. Feurer ID, Russell RT, Pinson CW. Incorporating quality of life and patient satisfaction measures into transplant outcomes assessment program: technical and practical considerations. *Progress in Transplantation*. 2007;17:121–8.
23. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-report measure of medication adherence. *Med Care*. 1986;24:67–74.

ARTÍCULO 2. Validación de una medida específica para evaluar Calidad de Vida Relacionada con la Salud en pacientes con esquizofrenia y trastorno bipolar: versión Española del cuestionario ToolL.

Montejo, A. L., Lauffer, J. C., Cuervo, J., Rebollo, P., Cordero, L., Díez, T., y Mauriño, J. (2011). Validation of a specific measure to assess health-related quality of life in patients with schizophrenia and bipolar disorder: the 'Tolerability and quality of life' (TOOL) questionnaire. *Ann Gen Psychiatry.*, 10, 6. doi: 10.1186/1744-859X-10-6

Nº de Registro en ClinicalTrials.gov (<http://www.clinicaltrials.gov/>): NCT00692133.

PRIMARY RESEARCH

Open Access

Validation of a specific measure to assess health-related quality of life in patients with schizophrenia and bipolar disorder: the 'Tolerability and quality of life' (TOOL) questionnaire

Angel L Montejo¹, Javier Correas Lauffer², Jesús Cuervo³, Pablo Rebollo³, Luis Cordero⁴, Teresa Diez⁵, Jorge Maurino^{5*}

Abstract

Background: Perception of quality of life may differ depending on the perspective. The aim of the study was to assess the psychometric properties of the Spanish version of the 'Tolerability and quality Of Life' (TOOL) questionnaire, a specific self-rated instrument to evaluate the impact of side effects of antipsychotic drugs on health-related quality of life (HRQoL). The questionnaire consists of eight items answered on a four-point Likert scale.

Methods: A psychometric study was conducted with clinically stable outpatients with schizophrenia and bipolar disorder under antipsychotic treatment. The translation and cultural adaptation of the questionnaire was performed according to international standards. Internal consistency using the Cronbach α coefficient and test-retest reliability using the intraclass correlation coefficient (ICC) was used to assess the reliability of the instrument. Patients completed generic and specific measures of quality of life and clinical severity.

Results: A total of 238 patients were analysed, with a mean age of 42 years (SD 10.9). The mean completion time was 4.9 min (SD 4.4). Internal consistency and intraclass correlation coefficient were adequate (Cronbach α = 0.757 and ICC = 0.90). Factorial analysis showed a unidimensional structure (a single eigenvalue >1, accounting for 39.1% of variance). Significant Spearman's rank correlations between the TOOL and both generic and specific measures were found. The questionnaire was able to discriminate among the Clinical Global Impression - Severity scores (Mann-Whitney U test, $P < 0.001$).

Conclusions: The TOOL questionnaire shows appropriate feasibility, reliability, and discriminative performance as a patient-reported outcome. TOOL constitutes a valuable addition to measure the impact of adverse events of antipsychotic drugs from the patient perspective.

Background

Schizophrenia and bipolar disorder are worldwide prevalent and severe mental diseases [1,2]. Newer antipsychotic treatments have proven useful in reducing both relapses and the severity of symptoms [3]. However, weight gain, extrapyramidal symptoms, sexual dysfunction, or sedation are quite common side effects among patients under antipsychotic treatment [4]. The occurrence of these symptoms may affect patient adherence

to medication, leading to discontinuation, limiting treatment effectiveness, and increasing both personal and social costs [5,6]. Therefore, when comparing alternative therapies, side effects and their impact on patient health-related quality of life (HRQoL) could be of great importance in order to define the most efficient antipsychotic drug treatment [3,7].

The systematic assessment of patient perspective may provide valuable information that could be lost when relying only on clinical evaluation. In chronic conditions such as schizophrenia and bipolar disorder, there are advantages in using patient-reported measures to understand

* Correspondence: jorgealejandromaurino@astrazeneca.com

⁵Neuroscience Area, AstraZeneca Medical Department, Madrid, Spain
Full list of author information is available at the end of the article

complex needs and improve alliances between patients and clinicians [8]. The importance of involving patients in their own healthcare, and of patient-reported assessments, is increasingly recognised [9].

Many efforts have been made to develop or validate specific instruments to assess patient affectation within different domains [10-13]. Regarding side effects, the most widely used and specific measure is the Uvalg for Kliniske Undersøgelser (UKU) side effects scale [14]. This questionnaire, filled out by clinicians or patients, comprises 56 items that refer to psychic, neurological, autonomic and other effects. Despite this scale, more specific tools in terms of HRQoL are still needed to afford greater insight in describing and grading the impact of side effects associated to antipsychotic drugs [15,16]. A brief instrument is preferred for use in clinical practice and in investigations such as clinical management studies or clinical trials comparing the effectiveness of treatments.

A specific measure to assess HRQoL impairment related to adverse events of antipsychotic drugs has been previously developed and validated in Sweden: the 'Tolerability and quality Of Life' (TOOL) questionnaire [17]. This self-rated measure reflects the subjective perception of side effects in patients treated with antipsychotic medication. It comprises eight attributes and has four levels per domain (Likert scale: 1 = minimum to 4 = maximum impact). These domains are mood (worry-upset), function capabilities, and several adverse events frequently associated with antipsychotic treatment (fatigue-weakness, weight gain, stiffness-tremor, physical restlessness, sexual dysfunction, and dizziness-nausea). In contrast to the Drug Attitude Inventory [10] or the Subjective Well-being under Neuroleptic scale [11], the TOOL questionnaire was specifically designed to identify from the patient perspective the most common adverse events of both typical and atypical antipsychotic drugs [4,18].

The aim of the present study was to evaluate the linguistic adaptation and psychometric validation into Spanish of the TOOL questionnaire for the assessment of side effects in patients treated with antipsychotic medication and their impact on health-related quality of life.

Methods

Linguistic adaptation of the TOOL questionnaire

Forward/backward translations of the original TOOL questionnaire were completed by expert translators. Firstly, three independent Spanish experts translated the original version in Swedish into Spanish. Experts examined and compared these three different versions in order to reach a single one by consensus (intermediate version 1.0). Secondly, one Swedish expert translated this intermediate Spanish version again into Swedish (backward translation) and compared his results with

the original version to ensure conceptual equivalence (intermediate version 2.0). Finally, all the expert translators participated in the proof reading test of the intermediate version 2.0, and the final Spanish version was thus established.

Subsequently, the Spanish version was reviewed by a panel of experts (five psychiatrists and one general practitioner specialised in HRQoL). According to expert criterion, three items were modified to facilitate patient comprehension: mood, physical restlessness, and dizziness-nausea. Next, 40 clinically stable patients (20 with schizophrenia and 20 with bipolar disorder) filled out the Spanish version of the TOOL. They were also asked to review this version in terms of comprehension (C) and importance (I) using a scale ranged from 0 (lowest level of C/I) to 4 (highest level of C/I). All the items scored higher than two points (threshold) in both scales. Consequently, the Spanish version of the TOOL questionnaire was ready for validation (Additional file 1).

Psychometric validation of the TOOL questionnaire

A multicentre, non-interventional psychometric study was conducted. The study was carried out at 60 psychiatric centres throughout Spain, and was approved by the institutional review board of the University Hospital of Salamanca (NCT00692133).

Participants were outpatients treated in community healthcare centres. Eligibility criteria included being at least 18 years old, having a Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) diagnosis of schizophrenia or bipolar disorder as established by the Structured Clinical Interview for DSM-IV, being clinically stable (defined as having had no changes in severity or new treatments initiated in the last month), and taking a single oral antipsychotic medication. After complete description of the study to the participants, written informed consent was obtained.

Investigators completed the following scales: Positive and Negative Syndrome Scale (PANSS) [19] (only to patients with schizophrenia), Young Mania Rating Scale (YMRS) [20] (only to patients with bipolar disorder), Montgomery-Asberg Depression Scale (MADRS) [21], Clinical Global Impression - Severity scale (CGI-S) [22], and UKU Side Effect Rating scale (UKU) [14].

Patients completed the following instruments: (1) The EuroQol 5-Dimensions (EQ-5D) [23] and Short Form 6-Dimensions (SF-6D) [24]: taking into account that the aim of this work is not to obtain utilities but to measure patient HRQoL, in these multidimensional scales we used the unweighted scores. In both cases, the higher the scores, the better the HRQoL, and *vice versa*. (2) The Spanish version of the TOOL questionnaire. The full items of the TOOL questionnaire are shown in Figure 1.

1. Mood	5. Stiffness or tremor
<input type="checkbox"/> I do not feel anxious or depressed	<input type="checkbox"/> I have no problems with stiffness or tremor
<input type="checkbox"/> I feel a little anxious or depressed	<input type="checkbox"/> I have some problems with stiffness or tremor
<input type="checkbox"/> I feel rather much anxious or depressed	<input type="checkbox"/> I have rather much problems with stiffness or tremor
<input type="checkbox"/> I feel very anxious or depressed	<input type="checkbox"/> I have great problems with stiffness or tremor
2. Function Capabilities	6. Physical restlessness
<input type="checkbox"/> I have no difficulties with my usual activities	<input type="checkbox"/> I feel no bodily restlessness
<input type="checkbox"/> I have some difficulties with my usual activities	<input type="checkbox"/> I feel some bodily restlessness
<input type="checkbox"/> I have rather much difficulty with my usual activities	<input type="checkbox"/> I feel rather much bodily restlessness
<input type="checkbox"/> I have great difficulties with my usual activities	<input type="checkbox"/> I feel very much bodily restlessness
3. Fatigue or weakness	7. Sexual function
<input type="checkbox"/> I do not feel tired or weak	<input type="checkbox"/> I have normal sexual lust and ability
<input type="checkbox"/> I feel a little tired or weak	<input type="checkbox"/> I have somewhat reduced sexual lust and ability
<input type="checkbox"/> I feel rather much tired or weak	<input type="checkbox"/> I have very reduced sexual lust and ability
<input type="checkbox"/> I feel very tired or weak	<input type="checkbox"/> I have no sexual lust or ability
4. Body weight	8. Dizziness or nausea
<input type="checkbox"/> I am not overweight	<input type="checkbox"/> I have no problems with dizziness or nausea
<input type="checkbox"/> I am a little overweight	<input type="checkbox"/> I have some problems with dizziness or nausea
<input type="checkbox"/> I am rather much overweight	<input type="checkbox"/> I have rather much problems with dizziness or nausea
<input type="checkbox"/> I am very overweight	<input type="checkbox"/> I have great problems with dizziness or nausea

Figure 1 The 'TOlerability and quality Of Life' (TOOL) questionnaire.

Retesting was carried out 2 to 3 weeks after the first visit. In order to analyse the psychometric properties of the TOOL questionnaire, feasibility, reliability and validity were studied.

Feasibility

The item response rate of the TOOL questionnaire was described and completion time was registered as well. Finally, both floor-ceiling effects were evaluated in an exploratory analysis of the answers given to the TOOL distribution and by using percentiles 25, 50 and 75 (Tukey's Hinges).

Validity

Regarding construct validity, item-total correlation (ITC) was checked following the criterion of removing items with a score lower than 0.20 in the discrimination rate. To assess the dimensionality of the questionnaire, an exploratory factorial analysis (extraction criterion of eigenvalue >1) was also conducted. The Spearman rank correlation coefficient (r_s) between the total score of the TOOL and those of the EQ-5 D (unweighted) and the SF-6 D (unweighted) was used to test convergent validity. Correlation was considered to be high for $r_s \geq 0.5$, moderate for r_s values between 0.3 to 0.5, and low for $r_s < 0.3$. This statistic was also applied to determine the relationship between the TOOL and the MADRS, YMRS, and UKU scales.

These analyses were performed differentially according to the two groups resulting from clinical diagnosis (patients with schizophrenia or bipolar disorder). In addition, r_s was used to study the association between each of the TOOL items and SF-6 D and EQ-5 D measurements.

In order to evaluate criterion validity, patients were categorised taking into account the CGI-S to compare TOOL scores. Accordingly, the collected sample was divided into two groups: patients with no to mild involvement versus patients with moderate to severe involvement. It was hypothesised that patients with lesser involvement would obtain lower scores in the TOOL (that is, lesser side effects and a better HRQoL) than those patients with worse health status according to expert criterion. Again, a non-parametric statistic (Mann-Whitney U test) was used in these comparisons considering the skewed distribution of data.

Reliability

The internal consistency of the questionnaire was analysed using Cronbach's α . Test-retest reliability of the scores was also examined by computing the intraclass correlation coefficient (ICC).

Statistical analysis

The sample size was calculated taking into account a specific objective of calculating multiattribute utility

function reflecting patient experience of side effects of antipsychotic therapy [25]. Thus, the determination was guided by the estimation of the utility function according to which a large number of patients is needed to maximise the precision and reduce the risk of measurement error. To this end, in the present study a sample size of about 250 patients was considered. Consequently, the number of patients required for the validation process (5 to 15 patients per item) was ensured, and hypothetical comparisons between the 2 independent groups could be established (that is, schizophrenic and bipolar patients) with 80% power and a 5% significance level (2-sided tests), involving 68 patients per group.

Data tabulation, database validation and the statistical analyses were carried out using the statistical packages SPSS (version 14.0; SPSS, Chicago, IL, USA) and Stata (version 10.0; Stata, College Station, TX, USA). For all statistical tests, a level of 0.05 was considered significant.

Registration

This trial was registered under no. NCT00692133.

Results

A total of 242 patients were included in the study. For the validation analysis, 238 (121 with a diagnosis of schizophrenia and 117 with bipolar disorder) were studied. The mean age of the sample was 41.9 years (SD 10.9), and 63% were males ($n = 151$). Sociodemographic and clinical characteristics of the sample are shown in Tables 1 and 2.

The percentage of missing values in each of the items contained in the questionnaire was less than 5%. The mean completion time was 4.91 min (SD 4.48). A ceiling effect was found for the following items: stiffness-tremor and dizziness-nausea (more than 50% of all patients reported not being bothered at all by any of these symptoms).

ITC scores were above 0.2 in all items, and body weight was the only item with a discrimination index below 0.3: ITC worry-upset = 0.581; function capabilities = 0.598; fatigue-weakness = 0.633; weight = 0.284; stiffness-tremor = 0.401; physical restlessness = 0.505; sexual dysfunction = 0.360; dizziness-nausea = 0.345. Exploratory factorial analysis reflected a unidimensional structure of the TOOL questionnaire (eigenvalue = 3.331) (Table 3). This dimension explained 39.1% of the variance. The component matrix is shown in Table 2. Furthermore, the reliability of this structure in terms of internal consistency was found to be adequate (Cronbach $\alpha = 0.757$). When test-retest reliability was examined, the results showed fairly appropriate stability in the evaluations of patient involvement when no changes in clinical status were detected by the clinicians (ICC = 0.90). Since unidimensionality had been proven, it was

Table 1 Sociodemographic and clinical characteristics of the sample

Variables	Total n = 238	n (%)
Gender, male		151 (63.4)
Age, mean (SD)	41.9 (10.9)	
Median time of treatment, months (SD)	33.50 (34.88)	
Marital status, n (%):		
Single		133 (55.9)
Married		69 (29.0)
Separated		17 (7.1)
Divorced		12 (5.0)
Widowed		7 (2.9)
Living status:		
Alone		28 (11.9)
With parents		109 (46.4)
In a couple		74 (31.5)
Others		24 (10.2)
Working status:		
Working full time		44 (18.5)
Working part time		28 (11.8)
Student		5 (2.1)
Unemployed		45 (18.9)
Retired (result of disease)		89 (37.4)
Retired (age)		10 (4.2)
Others		17 (7.1)
Clinical Global Impression Severity (median score = 3.00):		
Normal		23 (9.7)
Borderline mentally ill		18 (7.6)
Mildly ill		79 (33.2)
Moderately ill		81 (34.0)
Markedly ill		32 (13.4)
Severely ill		5 (2.1)
Treatment with atypical antipsychotic drugs		227 (95.4)

possible to calculate a global score summarising the patient level in each of the eight domains. A range of scores between 8 (minimum impact) to 32 (maximum impact) was obtained. This total score was also analysed with respect to generic and specific measures to test convergent and criterion validity.

Correlations between the TOOL and generic measures of HRQoL were highly-moderately negative and significant (Table 4). These results were also observed when the TOOL scores of bipolar and schizophrenic patients were compared separately (Table 4). Negative values in r_s responded to the inverse scoring of the measures. In addition, the associations between the TOOL items and the generic measures of HRQoL (Table 4) were negative,

Table 2 Clinical characteristics of the sample

Measures	N	Mean	SD	Minimum	Maximum	Percentiles		
						25	50	75
PANSS-positive	134	13.63	6.237	7.00	36.00	9.00	12.00	17.00
PANSS-negative	134	19.91	7.938	7.00	41.00	14.75	19.00	26.00
PANSS-combined	133	-6.11	7.569	-29.00	10.00	-11.00	-6.00	-0.50
PANSS-general psychopathology	134	32.76	12.086	16.00	78.00	22.75	31.50	40.00
YMRS	118	5.61	6.777	0.00	32.00	1.00	3.00	8.25
MADRS	231	11.30	8.026	0.00	42.00	5.00	10.00	17.00
UKU-psychic	229	1.66	1.518	0.00	7.00	0.00	1.00	3.00
UKU-autonomous	224	1.87	1.974	0.00	10.00	0.00	1.00	3.00
UKU-extrapyramidal	233	1.29	1.978	0.00	10.00	0.00	1.00	2.00
UKU-total score	173	6.83	6.019	0.00	37.00	2.50	6.00	9.00
UKU-sexual	198	2.20	2.536	0.00	13.00	0.00	1.00	4.00
TOOL	233	13.462	3.430	8.00	32.00	11.00	13.00	15.00
EQ-5 D (unweighted)	234	0.743	0.248	-0.11	1.00	0.682	0.841	1.00
SF-6 D (unweighted)	236	0.785	0.125	0.39	1.00	0.708	0.793	0.893

EQ-5 D = EuroQol 5-Dimensions; MADRS = Montgomery-Asberg Depression Scale; PANSS = Positive and Negative Syndrome Scale; SF-6 D = Short Form 6-Dimensions; TOOL = 'Tolerability and quality Of Life'; UKU = Udvalg for Kliniske Undersøgelser; YMRS = Young Mania Rating Scale.

significant and mild to high. With respect to specific measures, the TOOL questionnaire correlated moderately-highly and significantly with the MADRS. Moreover, regarding CGI-S, correlation was not only moderate and significant but also higher than the one found between these scales and the HRQoL generic

measures (Table 5). Correlations between specific domains of TOOL and related UKU subscales were positive, significant and mild or moderate (Table 6).

Finally, with the aim of testing criterion validity and discriminative validity of the TOOL, patients were classified depending on their CGI-S scores, yielding two

Table 3 Spanish 'Tolerability and quality Of Life' (TOOL) questionnaire construct validity: principal component analysis and component matrix

Component	Initial eigenvalues			Extraction sums of squared loadings		
	Total	Percentage of variance	Cumulative percentage	Total	Percentage of variance	Cumulative percentage
1	3.131	39.140	39.140	3.131	39.140	39.140
2	0.955	11.934	51.074			
3	0.905	11.315	62.388			
4	0.846	10.570	72.959			
5	0.742	9.279	82.238			
6	0.563	7.032	89.269			
7	0.443	5.534	94.803			
8	0.416	5.197	100.000			

Component matrix: Spanish TOOL domains	
	Component 1
Worry-upset	0.747
Function capabilities	0.759
Fatigue-weakness	0.789
Weight gain	0.392
Stiffness-tremor	0.548
Physical restlessness	0.670
Sexual dysfunction	0.491
Dizziness-nausea	0.481

Table 4 Spearman rank correlations (r_s) between Spanish 'Tolerability and quality Of Life' (TOOL) and generic measures of health-related quality of life (HRQoL)

r_s	Function capabilities	Fatigue-weakness	Weight gain	Stiffness-tremor	Physical restlessness	Sexual dysfunction	Dizziness-nausea	EQ-5 D (unweighted)	SF-6 D (unweighted)
Worry-upset	0.501**	0.495**	0.204**	0.230**	0.416**	0.225**	0.121	-0.532**	-0.537**
Function capabilities		0.471**	0.211**	0.207**	0.375**	0.306**	0.156*	-0.591**	-0.642**
Fatigue-weakness			0.222**	0.253**	0.409**	0.291**	0.279**	-0.557**	-0.565**
Weight gain				0.213**	0.080	0.191**	0.082	-0.225**	-0.238**
Stiffness-tremor					0.243**	0.111	0.187**	-0.328**	-0.277**
Physical restlessness						0.264**	0.090	-0.422**	-0.334**
Sexual dysfunction							0.164**	-0.406**	-0.382**
Dizziness-nausea								-0.224**	-0.158*
Spanish TOOL	-	-	-	-	-	-	-	-0.720**	-0.678**
Spanish TOOL (schizophrenia)	-	-	-	-	-	-	-	-0.716**	-0.645**
Spanish TOOL (bipolar disorder)	-	-	-	-	-	-	-	-0.720**	-0.724**

*Correlation is significant at the 0.05 level (two tailed); **correlation is significant at the 0.01 level (two tailed).

groups: patients with no or only mild involvement, and patients with moderate or severe involvement. Differences between these groups of CGI-S in the TOOL and generic HRQoL scores (Mann-Whitney U test) were significant in all cases ($P < 0.001$). The results highlighted that patients with no or only mild involvement had lower TOOL scores and higher EQ-5 D and SF-6 D scores, indicating a better HRQoL (mild CGI-S: TOOL = 12.168; EQ-5 D = 0.821; SF-6 D = 0.832). In contrast, those patients with moderate-severe involvement showed higher TOOL scores and lower EQ-5 D and SF-6 D outcomes (moderate-severe CGI-S: TOOL = 14.825; EQ-5 D = 0.663; SF-6 D = 0.739).

Discussion

Agents involved in health technology research have incorporated other important aspects to the basic aims of granting new treatments efficacy and safety, such as those related to patient subjective perception [9,26].

There is published evidence on the appropriateness and accuracy of self-assessments or self-report evaluations in patients suffering from mental chronic illness [27-29]. Despite this, however, recently published studies have focused on the side effects of antipsychotic drugs and their relationship with HRQoL by following mainly clinician criterion [30]. Few studies have tried to develop specific instruments to comprehensively quantify the impact of side effects on HRQoL based on the patient perspective. This objective should be of primary importance, taking into account the high prevalence of side effects in patients undergoing antipsychotic treatments and their relationship with adherence to treatment [5].

Our results show that the TOOL questionnaire presents adequate psychometric characteristics for use in patients with schizophrenia and bipolar disorder. The completion time of the questionnaire was low (<5 min) compared with longer questionnaires such as the UKU scale. Moderate-high correlations were found in both

Table 5 Spearman rank correlations (r_s) between generic and specific measures of health-related quality of life (HRQoL)

r_s	EQ-5 D (unweighted)	SF-6 D (unweighted)	CGI-S	PANSS: general psychopathology	MADRS	YMRS
Spanish TOOL	-0.720**	-0.678**	0.399**	0.443**	0.578**	0.239**
EQ-5 D (unweighted)		0.722**	-0.306**	-0.362**	-0.483**	-0.209*
SF-6 D (unweighted)			-0.386**	-0.516**	-0.663**	-0.100
CGI-S				0.578**	0.601**	0.351**
PANSS: general psychopathology					0.758**	0.462
MADRS						0.321**

*Correlation is significant at the 0.05 level (two tailed); **correlation is significant at the 0.01 level (two tailed).

CGI-S = Clinical Global Impression - Severity; EQ-5 D = EuroQol 5-Dimensions; MADRS = Montgomery-Asberg Depression Scale; PANSS = Positive and Negative Syndrome Scale; SF-6 D = Short Form 6-Dimensions; TOOL = 'Tolerability and quality Of Life'; YMRS = Young Mania Rating Scale.

Table 6 Spearman rank correlations (r_s) between Spanish 'Tolerability and quality Of Life' (TOOL) items and UKU subscales

r_s	UKU-psychic	UKU-autonomous	UKU-extrapyramidal	UKU-sexual
Worry-upset	0.372**	0.156*	0.097	0.122
Function capabilities	0.359**	0.143*	0.218**	0.217**
Fatigue-weakness	0.403**	0.220**	0.171**	0.193**
Weight gain	0.300**	0.240**	0.239**	0.147*
Stiffness-tremor	0.192**	0.160*	0.560**	0.196**
Physical restlessness	0.378**	0.202**	0.154*	0.178*
Sexual dysfunction	0.201**	0.302**	0.167*	0.605**
Dizziness-nausea	0.153*	0.206**	0.184**	0.204**

*Correlation is significant at the 0.05 level (two tailed); **correlation is significant at the 0.01 level (two tailed).

UKU = Udvalg for Kliniske Undersøgelser.

bipolar and schizophrenic patients between the TOOL and specific clinician-rated measures as the MADRS, PANSS, and YMRS, as well as a negative relationship with generic quality of life instruments (EQ-5 D and SF-6D), thus highlighting the convergent validity of this instrument. These correlations were also similar to those found in other studies analysing the association between generic and specific measures [31]. Criterion and discriminative validity were proven by the significant differences found in the TOOL between those patients with mild involvement and patients with moderate-severe symptoms. Thus, patients suffering a worse health status according to the CGI-S obtained higher scores in the TOOL, which indicates a more severe impact on their HRQoL, as has been previously underscored [32,33].

The one-dimensional construct of the TOOL, along with its internal consistency and test-retest reliability, have been proven, thus allowing clinicians to obtain a simple and global score (ranging from 8 to 32) with the same interpretation in terms of severity as the generalised UKU scale.

Despite these results, this study presents some limitations that should be considered. A ceiling effect was observed in two items: stiffness-tremor and dizziness-nausea. This may be due to the idiosyncrasy of the population sample collected in the present study: clinically stable patients under drug treatment, based mainly on atypical antipsychotic agents. Although similar ceiling effects have been reported in other studies [31,32], these items should be tested in further investigations with schizophrenic and bipolar patients showing a worse health status in order to test discriminative power. The characteristics of the sample may limit the scope of use of this measure. Therefore, it may be arguable whether patients with greater involvement (that is, patients with severe negative symptoms or patients suffering a manic episode) could provide reliable reports when completing the questionnaire. Although there is evidence supporting

the accuracy of such measurement, as has been commented above, the questionnaire should be applied within such populations to test its psychometrical properties. Secondly, it is necessary to analyse the sensitivity of the TOOL in detecting changes in longitudinal studies.

Conclusions

The Spanish validation of the TOOL questionnaire shows appropriate feasibility, reliability, and discriminative performance as a patient-reported outcome to be used for the assessment of the impact of side effects on patient health-related quality of life. This information could be very important to improve therapeutic alliances and treatment adherence among patients with schizophrenia and bipolar disorder.

Additional material

Additional file 1: TOOL Spanish version. The Spanish version of the 'Tolerability and quality Of Life' (TOOL) questionnaire.

Acknowledgements

This study was funded by AstraZeneca Medical Department Spain. The authors would like to thank Aurelio García, Juan José Uriarte, and Fermín Mayoral for their contribution in the linguistic validation.

Author details

¹Department of Psychiatry, Hospital Universitario de Salamanca, Salamanca, Spain. ²Department of Psychiatry, Hospital del Henares, Coslada, Madrid, Spain. ³BAP Health Outcomes Research, Oviedo, Spain. ⁴Value Demonstration Unit, AstraZeneca Medical Department, Madrid, Spain. ⁵Neuroscience Area, AstraZeneca Medical Department, Madrid, Spain.

Authors' contributions

ALM, JCL, JC, PR, LC, TD, and JM conceived the study design. JC and PR performed the statistical analyses. All authors made meaningful contributions to data interpretation. ALM, JC, and JM cowrote the final draft of the manuscript. All authors read and approved the final manuscript.

Competing interests

ALM: Grants from Lilly, BMS-Otsuka, GSK, Sanofi, Astra Zeneca, Boehringer and Wyeth. Fees from Lilly, BMS-Otsuka, GSK, Sanofi, Astra Zeneca,

Boehringer and Wyeth. Served on advisory boards for Lilly, GlaxoSmithKline, Servier and AstraZeneca. LC, TD, and JM are employees of AstraZeneca. The remaining authors have no conflicts of interest.

Received: 24 December 2010 Accepted: 11 March 2011
Published: 11 March 2011

References

- McGrath J, Saha S, Chant D, Welham J: **Schizophrenia: a concise overview of incidence, prevalence, and mortality.** *Epidemiol Rev* 2008, **30**:67-76.
- El-Mallakh RS, Elmaadawi AZ, Loganathan M, Lohano K, Gao Y: **Bipolar disorder: an update.** *Postgrad Med* 2010, **122**:24-31.
- Buckley PF: **Update on the etiology and treatment of schizophrenia and bipolar disorder.** *CNS Spectr* 2008, **13**(Suppl 1):3-10.
- Edwards SJ, Smith CJ: **Tolerability of atypical antipsychotics in the treatment of adults with schizophrenia or bipolar disorder: a mixed treatment comparison of randomized controlled trials.** *Clin Ther* 2009, **31**:1345-1359.
- Perkins DO, Gu H, Weiden PJ, McEvoy JP, Hamer RM, Lieberman JA: **Predictors of treatment discontinuation and medication non-adherence in patients recovering from a first episode of schizophrenia, schizophreniform disorder, or schizoaffective disorder: a randomized, double-blind, flexible-dose, multicenter study.** *J Clin Psychiatry* 2008, **69**:106-113.
- Velligan DI, Weiden PJ, Sajatovic M, Scott J, Carpenter D, Ross R, Docherty JP: **The expert consensus guideline series: Adherence problems in patients with serious and persistent mental illness.** *J Clin Psychiatry* 2009, **70**(Suppl 4):1-46.
- Bridle C, Palmer S, Bagnall A-M, Darba J, Duffy S, Sculpher M, Riemsma R: **A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder.** *Health Technol Assess* 2004, **8**:1-187.
- Fonagy P, Mattheus R, Pilling S: **The Mental Health Outcome Measurement Initiative: Best practice guidance for local implementation adapted from the report from the Chair of the Outcomes Reference Group.** Leeds, UK: National Institute of Mental Health in England; 2005.
- Thornicroft G, Tansella M: **Growing recognition of the importance of service user involvement in mental health service planning and evaluation.** *Epidemiol Psychiatr Soc* 2005, **14**:1-3.
- Hogan T, Awad A, Eastwood R: **A self-report scale predictive of drug compliance in schizophrenics: reliability and discriminative validity.** *Psychol Med* 1983, **13**:177-183.
- Naber D, Moritz S, Lambert M, Pajonk FG, Holzbach R, Mass R, Andresen B: **Improvement of schizophrenic patients' subjective well-being under atypical antipsychotic drugs.** *Schizophr Res* 2001, **50**:79-88.
- Voruganti L, Awad A: **Personal evaluation of transitions in treatment (PETIT): a scale to measure subjective aspects of antipsychotic drug therapy in schizophrenia.** *Schizophr Res* 2002, **56**:37-42.
- Patrick DL, Burns T, Morosini P, Rothman M, Gagnon DD, Wild D, Adriaenssen I: **Reliability, validity and ability to detect change of the clinician-rated Personal and Social Performance scale in patients with acute symptoms of schizophrenia.** *Curr Med Res Opin* 2009, **25**:325-338.
- Lingjaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K: **The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients.** *Acta Psychiatr Scand* 1987, **334**:1-100.
- Revicki DA, Matza LS, Flood E, Lloyd A: **Bipolar disorder and health-related quality of life: review of burden of disease and clinical trials.** *Pharmacoeconomics* 2005, **23**:583-94.
- Alonso J, Croudace T, Brown J, Gasquet I, Knapp MR, Novick D: **Health-related quality of life (HRQL) and continuous antipsychotic treatment: 3-year results from Schizophrenia Health Outcomes (SOHO) study.** *Value Health* 2009, **12**:536-543.
- Lindström E, Jönsson L, Berntsson A: **A patient perspective on side effects of antipsychotic therapy: the TOOL instrument.** *Value Health* 2009, **12**:A361.
- Montejo AL: **Prolactin awareness: an essential consideration for physical health in schizophrenia.** *Eur Neuropsychopharmacol* 2008, **18**(Suppl 2):S108-S114.
- Kay SR, Fiszbein A, Vital-Herne M, Fuentes LS: **The Positive and Negative Syndrome Scale-Spanish adaptation.** *J Nerv Ment Dis* 1990, **178**:510-517.
- Colom F, Vieta E, Martínez-Arán A, García-García M, Reinares M, Torrent C, Goikolea JM, Banús S, Salamero M: **Spanish version of a scale for the assessment of mania: validity and reliability of the Young Mania Rating Scale [in Spanish].** *Med Clin (Barc)* 2002, **119**:366-371.
- Lobo A, Chamorro L, Luque A, Dal-Re R, Badia X, Baro E: **Validation of the Spanish versions of the Montgomery-Asberg Depression and Hamilton anxiety Rating Scales.** *Med Clin (Barc)* 2002, **118**:493-499.
- Guy W: *Early Clinical Drug Evaluation (ECDEU) Assessment Manual* Rockville, MD: U.S. National Institute of Health, Psychopharmacology Research Branch; 1976.
- EuroQol Group: **EuroQol - a new facility for the measurement of health-related quality of life.** *Health Policy* 1990, **16**:199-208.
- Brazier J, Roberts J, Deverill M: **The estimation of a preference-based measure of health from the SF-36.** *J Health Econ* 2002, **21**:271-292.
- Jönsson L, Lang A, Lindström E: **TOOL: Multi-attribute utility function reflecting patient experience of side effects to antipsychotic therapy.** *Value Health* 2009, **12**:A361.
- Willke RJ, Burke LB, Erickson P: **Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels.** *Control Clin Trials* 2009, **25**:535-552.
- Barbera M, Sanjuan J, Munarriz M, Barbeña M, Sanjuán J, Munárriz M, Novella E, Santiago C, Simó M: **Subjective experience with antipsychotics: quantitative evaluation [in Spanish].** *Actas Esp Psiquiatr* 2006, **34**:287-294.
- Bobes J, García-Portilla MP, Bascaran MT, Saiz PA, Bousoño M: **Quality of life in schizophrenic patients.** *Dialogues Clin Neurosci* 2007, **9**:215-226.
- McCabe R, Saidi M, Priebe S: **Patient-reported outcome in schizophrenia.** *Br J Psychiatry* 2007, **50**:S21-S28.
- Taksh U: **A critical review of rating scales in the assessment of movement disorders in schizophrenia.** *Curr Drug Targets* 2006, **7**:1225-1229.
- König HH, Roick C, Angermeyer MC: **Validity of the EQ-5 D in assessing and valuing health status in patients with schizophrenic, schizotypal or delusional disorders.** *Eur Psychiatry* 2007, **22**:177-187.
- Hayhurst H, Palmer S, Abbott R, Johnson T, Scott J: **Measuring health-related quality of life in bipolar disorder: relationship of the EuroQol (EQ-5D) to condition-specific measures.** *Qual Life Res* 2006, **15**:1271-1280.
- Briggs A, Wild D, Lees M, Reaney M, Dursun S, Parry D, Mukherjee J: **Impact of schizophrenia and schizophrenia treatment-related adverse events on quality of life: direct utility elicitation.** *Health Qual Life Outcomes* 2008, **6**:105.

doi:10.1186/1744-859X-10-6

Cite this article as: Montejo et al.: Validation of a specific measure to assess health-related quality of life in patients with schizophrenia and bipolar disorder: the 'Tolerability and quality of life' (TOOL) questionnaire. *Annals of General Psychiatry* 2011 **10**:6.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit



ARTÍCULO 3. Evaluación de la depresión mayor desde la perspectiva del paciente: validación al castellano de la escala CUDOS (Clinically Useful Depression Outcome Scale).

Agüera-Ortiz, L., Montón, C., Cuervo, J., Medina, E., Díaz-Cuervo, H., y Mauriño, J. (2013). Adaptación al castellano de la escala Clinically Useful Depression Outcome Scale (CUDOS) para la evaluación de la depresión mayor desde la perspectiva del paciente. *Actas Esp Psiquiatr.*, 41(5), 287-300. Recuperado de

<http://actaspsiquiatria.es/repositorio/15/85/ESP/15-85-ESP-287-300-992998.pdf> .

Nº de Registro en ClinicalTrials.gov (<http://www.clinicaltrials.gov/>): NCT01360307.

Luis Agüera-Ortiz¹
 Carmen Montón²
 Jesús Cuervo³
 Esteban Medina⁴
 Helena Díaz-Cuervo³
 Jorge Maurino⁴

Adaptación al castellano de la escala Clinically Useful Depression Outcome Scale (CUDOS) para la evaluación de la depresión mayor desde la perspectiva del paciente

¹Servicio de Psiquiatría.
 Instituto de Investigación Hospital Universitario 12 de Octubre
 Centro de investigación Biomédica en Red de Salud Mental
 (CIBERSAM)

²Centro de Salud Casablanca
 Zaragoza, España

³BAP LA-SER Outcomes, LA-SER Group
 Oviedo, España

⁴Departamento Médico, AstraZeneca
 Madrid, España

Objetivo: Adaptar al castellano la escala CUDOS (*Clinically Useful Depression Outcome Scale*) y analizar sus propiedades psicométricas en pacientes con trastorno depresivo mayor (TDM).

Método: Se realizó un estudio multicéntrico transversal de dos etapas (adaptación cultural y validación psicométrica) en el que se evaluaron pacientes en Atención Primaria con TDM diagnosticado (criterios DSM-IV TR) en los últimos 3 meses. Junto a la escala CUDOS, se aplicaron los instrumentos: PRIME-MD (*Primary Care Evaluation of Mental Disorders*), HAMD-17 (*Hamilton Depression Rating Scale*), SOFAS (*Social and Occupational Functioning Assessment Scale*), SF-36 (Componentes Físico-PCS- y Mental-MCS-) y escalas CGI-SI y PGI-SI (Escala de Impresión Clínica Global de Gravedad según clínico o paciente, respectivamente). Se analizó la factibilidad, fiabilidad y validez de la versión adaptada.

Resultados: En la fase de validación, se incluyeron 305 pacientes (69,5% mujeres) con edad media (desviación típica) de 51,75(15,53) años. El tiempo medio de cumplimentación fue de 4,47(±2,4) minutos. Los efectos techo/suelo se encontraron en menos del 1% de los casos. La consistencia interna de la escala CUDOS fue adecuada (α de Cronbach=0,88). Los coeficientes de correlación de Pearson de CUDOS con otras escalas fueron: 0,42 (SOFAS); 0,45 (HAMD-17); -0,22 (PCS); -0,65 (MCS); $p < 0,001$. La escala CUDOS discriminó adecuadamente entre los diferentes niveles clínicos de gravedad ($p < 0,03$).

Conclusiones: La versión adaptada de la escala CUDOS muestra propiedades psicométricas adecuadas como instrumento de evaluación de la depresión mayor desde la perspectiva del paciente.

Palabras clave: Trastorno depresivo mayor, Propiedades psicométricas, Atención primaria, CUDOS

Actas Esp Psiquiatr 2013;41(5):287-300

Adaptation into spanish of the Clinically Useful Depression Outcome Scale (CUDOS) for assessing major depressive disorder from the patient's perspective

Objective: To adapt the CUDOS scale (Clinically Useful Depression Outcome Scale) into Spanish and to test its psychometrical properties in a sample of patients with major depressive disorder (MDD).

Methods: A two-step cross-sectional, multicenter validation study was conducted (linguistic adaptation into Spanish and psychometric validation). The study evaluated patients attended in Primary Care with a MDD diagnosis within the last 3 months (DSM-IV TR criteria). The following scales were administered: CUDOS, PRIME-MD (Primary Care Evaluation of Mental Disorders), HAMD-17 (Hamilton Depression Rating Scale), SOFAS (Social and Occupational Functioning Assessment Scale), SF-36 (Physical -PCS- and Mental -MCS- Component Summaries), and the CGI-S & PGI-S (Clinical Global Impression for Severity of Illness scales for clinicians and patients, respectively). Feasibility, reliability, and validity of the Spanish version were assessed.

Results: In the validation study, 305 MDD patients (69.5% female) with a mean age (standard deviation-SD-) of 51.75(15.53) were included. Mean completion time was 4.47(2.4) minutes. Floor or ceiling effects were found in less than 1% of the case scores. Internal consistency was adequate (Cronbach's $\alpha = 0.88$). Pearson correlation coefficients with CUDOS were: -0.42 (SOFAS), 0.45 (HAMD-17), -0.22 (PCS), -0.65 (MCS); all $p < 0.001$. The CUDOS properly discriminated among clinical severity levels ($p < 0.03$).

Correspondencia:
 Luis Agüera-Ortiz
 Servicio de Psiquiatría
 Hospital Universitario 12 de Octubre
 Av de Córdoba km 5,400
 28041 Madrid
 Correo electrónico: luis@aguera.es

Conclusions: The adapted Spanish version of the CUDOS shows adequate psychometric properties as an evaluation instrument of major depression from the patient's perspective.

Key words: Major depressive disorder, Psychometric validation, Primary care, CUDOS

INTRODUCCIÓN

El trastorno depresivo mayor (TDM) es la enfermedad mental más frecuente en España, con una prevalencia poblacional estimada en adultos del 3,9% anual y del 10,6% a lo largo de la vida^{1,2}. Aunque estos datos están sujetos a cierta variabilidad según la metodología empleada y el ámbito poblacional o asistencial de referencia, resulta evidente su relevancia desde el punto de vista epidemiológico³⁻⁵. Es conocido el sufrimiento personal de los pacientes con TDM, la morbilidad y probabilidad de recurrencia, el impacto en la vida familiar y socio-laboral, así como el riesgo de suicidio asociado⁶⁻¹¹. Todos estos factores contribuyen a la elevada carga socio-sanitaria y económica del trastorno en nuestro país¹²⁻¹⁴.

La investigación clínica en la depresión mayor ha centrado sus esfuerzos en establecer un diagnóstico y tratamiento adecuados con objeto de maximizar la efectividad de las intervenciones terapéuticas. De esta forma se reforzarían las posibilidades de alcanzar la remisión completa de los síntomas lo que ha sido asociado con un mejor pronóstico en términos de recaídas y morbi-mortalidad^{10,13,15,16}. Con estos propósitos, las estrategias recientes de salud y las guías terapéuticas, recomiendan un primer abordaje del trastorno desde el ámbito de Atención Primaria (AP)^{13,15}, pero es precisamente en ese contexto donde se han constatado las mayores dificultades^{17,18}. Como ejemplo, se ha puesto de manifiesto que en este ámbito sólo el 34% de los pacientes llega a recibir tratamiento con antidepresivos¹⁹. Esto otorga aún mayor interés al hecho de poder identificar precozmente aquellos individuos con clínica de TDM y comenzar un enfoque terapéutico adecuado^{20,21}.

La medición de los resultados percibidos por el paciente interesa actualmente a las agencias reguladoras para valorar de forma comparada los beneficios de los fármacos^{15,22}. Los elementos subjetivos son para el paciente tan importantes o más que los biológicos por lo que es necesaria una evaluación rigurosa de los mimos²³. Instrumentos auto-aplicables fiables y sencillos son fundamentales para la recogida estructurada y sistemática de esta información y permiten además hacer al paciente más participe en el proceso terapéutico mejorando la eficiencia de la consulta médica²³⁻²⁵. En este sentido, la escala CUDOS (*Clinically Useful Depression Outcome Scale*) fue diseñada en EE.UU.²⁶⁻²⁸, bajo el proyecto MIDAS (*Rhode Island Methods to Improve Diagnostic Assessment and*

Service) que perseguía desarrollar instrumentos útiles para la práctica clínica. Es un instrumento breve que consta de 18 preguntas con respuestas tipo Likert de 5 puntos y que se puede responder rápidamente²⁹. Su corrección es sencilla y la puntuación total se calcula sumando las puntuaciones de los 16 primeros ítems (rango entre 0-mínima de sintomatología depresiva y 64 -máxima intensidad sintomática). Respecto a su validez de contenido, incluye todo el espectro de síntomas del TDM, definido por la clasificación DSM-IV y recoge además información de aspectos funcionales y de calidad de vida (ítems 17-18). En cuanto a sus propiedades psicométricas, se ha evidenciado que posee una fiabilidad y validez convergente adecuadas. Además, discrimina entre niveles clínicos de gravedad de depresión, siendo sensible a los cambios en el estado del paciente^{26,28}.

Con el objeto de aportar evidencias que permitan a los profesionales sanitarios de nuestro país valorar el empleo de esta escala en su práctica clínica, el presente trabajo presenta la adaptación cultural de este instrumento, así como el análisis inicial de la fiabilidad y validez para la evaluación clínica del TDM en consulta de AP.

METODOLOGÍA

Estudio multicéntrico, observacional, de corte transversal, implementado en condiciones de práctica clínica habitual, en 44 centros de AP de distintas CCAA de nuestro país. El protocolo del estudio fue aprobado por el Comité Ético de Investigación Clínica del Servicio Gallego de Salud (SERGAS) de la Consejería de Sanidad de la Xunta de Galicia.

Muestra

Cada investigador incluyó, de forma consecutiva según llegada a consulta, los primeros pacientes que cumplieran los criterios de selección y prestaran su consentimiento informado. Se entrevistaron 338 pacientes adultos que habían recibido el diagnóstico de TDM (criterios DSM-IV TR)³⁰ durante los últimos 3 meses. Los pacientes debían presentar síntomas de TDM o TDM en remisión parcial de acuerdo a la escala PRIME-MD^{31,32} y unas capacidades cognitivas que les permitieran responder a las preguntas del estudio. Se excluyeron pacientes que presentaban incapacidad para cumplimentar las escalas debido a enfermedad médica subyacente y aquéllos que hubieran participado en un ensayo clínico durante los 6 meses anteriores.

Tras el análisis de los criterios de selección, 33 casos tuvieron que ser excluidos, bien porque no habían recibido el diagnóstico de TDM durante los últimos 3 meses (N=14), o bien porque el diagnóstico no se ajustaba a los criterios DSM-IV TR, (N=19). En consecuencia, la muestra final válida para el estudio fue de 305 pacientes.

Debido a los propósitos del proyecto ya comentados, el estudio se realizó en dos etapas:

Etapa I. Adaptación lingüística a nuestro país de la versión original en inglés

El trabajo desarrollado en esta fase se resume en la Figura 1. La adaptación lingüística se llevó a cabo siguiendo las recomendaciones de las sociedades científicas especiali-

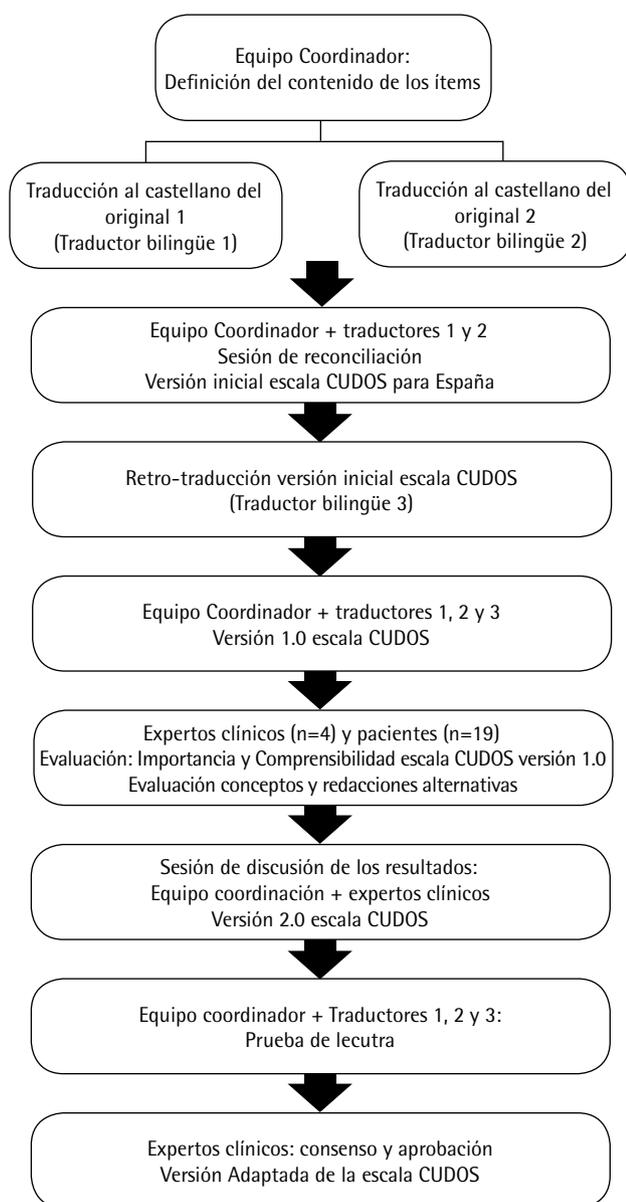


Figura 1

Etapa I: Adaptación lingüística de la escala CUDOS

zadas³³. Se contó con un equipo coordinador multidisciplinar, compuesto por expertos en el diagnóstico y tratamiento de estos pacientes y por expertos en el desarrollo de instrumentos para la evaluación de resultados en salud.

A partir de la armonización conceptual de las preguntas de la escala, se presentó la versión a un panel de médicos (n= 4) y pacientes con diagnóstico de TDM de acuerdo a los criterios DSM-IV-TR (n=19). Ambos grupos puntuaron la importancia y comprensibilidad de cada ítem en 2 escalas con rango: 0 -mínima importancia/nula comprensibilidad- hasta 4 -máxima importancia/óptima comprensibilidad-. También se les ofreció la oportunidad de presentar redacciones alternativas para abordar cada aspecto recogido. Tras el análisis de esta información (Figura 2), se consensó una versión final que fue revisada por los traductores y aprobada por todos los clínicos participantes. De esta forma, se completó la adaptación al castellano de la escala (Anexo).

Etapa II. Análisis de las propiedades psicométricas de la versión adaptada

En esta segunda etapa, los investigadores recogieron información socio-demográfica y clínica básica de cada participante y confirmaron la adecuación diagnóstica de TDM según criterios DSM-IV-TR. Durante la visita contemplada en el protocolo, aplicaron también las siguientes medidas:

Instrumentos aplicados por investigadores

- Módulo de estado de ánimo del instrumento *Primary Care Evaluation of Mental Disorders (PRIME-MD)*^{31,32}: El módulo del estado de ánimo aplicado consta de 17 preguntas dicotómicas que permiten distinguir entre la presencia de un TDM, un estado de remisión parcial o recurrencia, la distimia, depresión menor, un trastorno afectivo bipolar o una depresión de etiopatogenia medicamentosa o por trastorno físico.
- Escala de Depresión de Hamilton (*Hamilton Depression Rating Scale*), versión de 17 ítems (HAMD-17)^{34,35}: Permite evaluar la gravedad de los síntomas y valorar los cambios del paciente a lo largo del tiempo. La puntuación total presenta un rango entre 0 a 52 puntos (a mayor puntuación, mayor gravedad). Se recomienda emplear los siguientes puntos de corte: No deprimido: 0-7; Depresión ligera/menor: 8-13; Depresión moderada: 14-18; Depresión grave: 19-22; Depresión muy grave: ≥ 23 ^{35,36}.
- Escala de Impresión Clínica Global de Gravedad (CGI-SI)³⁷: Evalúa la gravedad del cuadro clínico del paciente utilizando una escala Likert de 8 niveles (desde 0-no evaluado hasta 7-extremadamente enfermo).
- Escala de Evaluación de la Actividad Social y Ocupa-

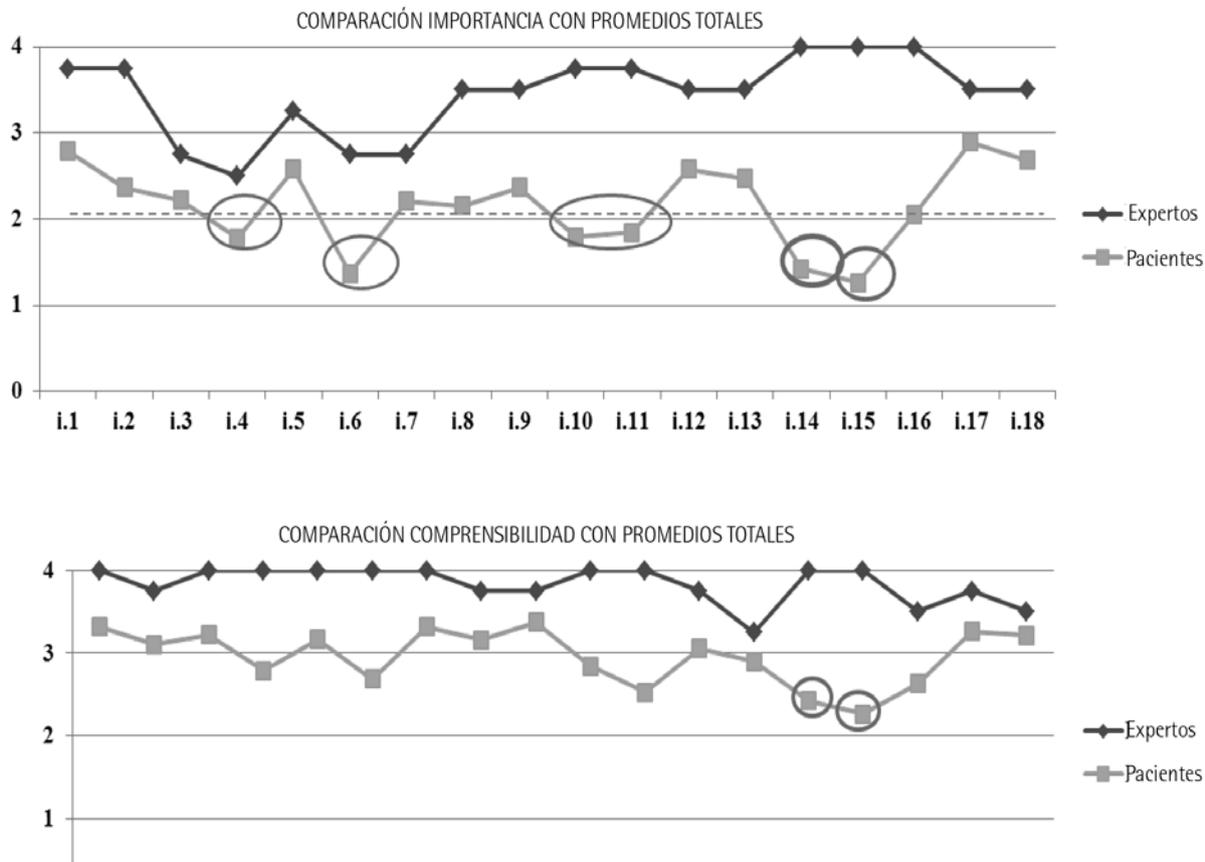


Figura 2

Importancia y comprensibilidad de la escala CUDOS según expertos clínicos (n=4) y pacientes (n=19)

cional (SOFAS)³⁸: Se trata de una escala para valorar el nivel de funcionamiento social y ocupacional del paciente sin tener en cuenta la sintomatología presente. El rango de puntuaciones incluye desde 1 punto (sujeto con un desenvolvimiento social nulo o extremadamente limitado) a 100 (el mejor funcionamiento social). Se ha establecido recientemente que una puntuación superior o igual a 80 es un punto de corte válido para definir un funcionamiento social normalizado¹⁰.

Por su parte, los pacientes informaron sobre la presencia de situaciones personales que en la literatura científica se destacan como posibles factores de riesgo (antecedentes familiares, problemas graves en personas cercanas, problemas económicos o laborales y de convivencia)^{39,40}. Además, cumplieron:

a. Cuestionario SF-36^{41,42}: Instrumento genérico de calidad de vida relacionada con la salud validado en nuestro país y de amplio uso en población general y diferentes

grupos de pacientes. Consta de 36 preguntas cuya puntuación se agrupa en dos componentes sumarios, físico (PCS) y mental (MCS) con puntuaciones estandarizadas para la población general española y normalizadas a una distribución 50 ± 10 .

b. Escala de Impresión de Gravedad Global del Paciente (PGI-SI)³⁷: Escala tipo Likert en la que el paciente valora su estado global con puntuaciones que oscilan entre 0 (no evaluado) y 7 puntos (extremadamente enfermo).

c. Versión adaptada al castellano de la escala CUDOS (Anejo)²⁶⁻²⁸: La puntuación total es el sumatorio de las respuestas a las 16 primeras preguntas. El rango teórico de las puntuaciones oscila entre 0 y 64 puntos. Mayores puntuaciones son indicativas de una mayor gravedad del cuadro depresivo. De acuerdo a los autores de la versión original, las puntuaciones globales se pueden clasificar en 5 niveles de gravedad: sin depresión (0-10 puntos), mínima (11-20), leve (21-30), moderada (31-45), grave (≥ 46).

Tabla 1		Descripción de la muestra (n= 305)	
		n	%
Sexo	Mujer	212	69,5
Formación académica	Sin estudios	46	15,1
	Bachillerato superior	46	15,1
	Estudios de graduado escolar	101	33,1
	FP superior/ universitarios	59	19,3
	Educación secundaria	53	17,4
	Convivencia	Solo	37
	Acompañado en familia	267	87,5
	En institución	1	0,3
Episodios depresivos previos, (sí)		126	41,3
Antecedentes familiares de depresión, (sí)		116	38,0
Problemas graves en personas cercanas, (sí)		93	30,5
Problemas económicos o laborales, (sí)		119	39,0
Problemas de convivencia o familiares, (sí)		90	29,5
Medicaciones para estrés o ansiedad, (sí)		201	65,9
Tratamiento antidepresivo, (sí)		267	87,5
Resultado clasificación PRIME-MD	TDM	258	84,6
	TDM en remisión parcial	47	15,4

Tamaño muestral

El número de casos a evaluar se estimó considerando el porcentaje de pacientes que respondían al tratamiento o se encontraban en remisión parcial después del mismo (33%, aproximadamente), según el trabajo original²⁷. A partir de esta información, se consideró la inclusión de al menos 176 pacientes para poder llevar a cabo el análisis de las diferencias de medias entre grupos independientes con un poder estadístico del 80%, para un tamaño del efecto de 0,4 y un nivel de confianza (1- α) del 95%. Adicionalmente se asumió que participaría en el estudio un porcentaje de pacientes (33%) para los que no habría transcurrido el tiempo mínimo requerido para poder valorar la respuesta al tratamiento (entre 6-12 semanas). Por tanto, y previendo un porcentaje de casos no válidos habitual en este tipo de estudios (10%), se estimó necesario un número mínimo de 290 pacientes.

Análisis estadístico

En primer lugar, se describieron las características clínicas y socio-demográficas del conjunto de la muestra incluida en la segunda etapa mediante estadísticos de tendencia central y de dispersión en las variables cuantitativas y por

medio de frecuencias y porcentajes en el caso de variables cualitativas. En segundo lugar, se analizaron las propiedades psicométricas de la escala:

- Distribución de las puntuaciones y factibilidad de la escala: se calcularon los Estadísticos descriptivos de cada ítem y de la puntuación total (media, desviación típica -DT-, mediana, valor mínimo y máximo y coeficiente de curtosis), el porcentaje de no respuesta y el tiempo de cumplimentación de la escala CUDOS (media recortada al 5%). Junto a esto, se comprobó el efecto techo (porcentaje de pacientes con puntuación máxima) y efecto suelo (porcentaje de pacientes con puntuación mínima).
- Funcionamiento diferencial de los ítems (DIF): se analizó de forma preliminar la existencia de sesgo en las puntuaciones, en función del género de los pacientes. A tal fin, se calculó el estadístico de Mantel-Haenszel. Este índice expresa la razón o el cociente entre la probabilidad de puntuar bajo *versus* alto en cada ítem en el grupo de varones y la probabilidad de puntuar bajo *versus* alto en el grupo de mujeres. Valores estadísticamente distintos de 1 serían indicadores de un DIF no uniforme (la probabilidad de puntuar alto o bajo en depresión sería mayor en mujeres o varones).

Tabla 2		Puntuación en las medidas de gravedad sintomática y de afectación de la calidad de vida relacionada con la salud de los pacientes				
	N	Mínimo	Máximo	Media	D.T.	
HAMD-17	305	3	43	16,75	7,32	
Escala SOFAS	305	0	90	62,53	16,77	
Componente sumario física estandarizada (PCS)	304	21,45	63,59	44,05	9,12	
Componente sumario mental estandarizada (MCS)	304	2,64	58,19	28,02	9,94	
Numero de episodios previos	126	1	30	2,80	3,70	
Índice de comorbilidad de Charlson	305	0	6	0,57	1,09	
Gravedad de la Enfermedad	CGI-SI (n=305)		PGI-SI (n=304)			
	Frecuencia	Porcentaje	Frecuencia	Porcentaje		
No evaluado	0	0,0	3	1,0		
Normal	2	0,7	14	4,6		
Dudosamente enfermo	9	3,0	21	6,9		
Levemente enfermo	75	24,6	63	20,7		
Moderadamente enfermo	137	44,9	101	33,1		
Marcadamente enfermo	74	24,3	74	24,3		
Gravemente enfermo	5	1,6	23	7,5		
Entre los pacientes más extremadamente enfermos	3	1,0	5	1,6		

- Consistencia interna, capacidad de discriminación de los ítems y validez de constructo: en términos de fiabilidad, se calculó el alfa de Cronbach como indicador de la consistencia interna y se analizó la capacidad de discriminación de los ítems que componen la escala con el cálculo de la correlación ítem-total corregida. Finalmente, se aplicó un Análisis de Componentes Principales (ACP) con el fin de aportar evidencias de la estructura interna de la escala CUDOS, descrita en el trabajo original como unidimensional²⁷.
- Validez de criterio y discriminante: se evaluaron las diferencias en las puntuaciones en la escala CUDOS en función de la gravedad, variables sociodemográficas, clasificación PRIME-MD y factores de riesgo. La gravedad sintomática se definió de acuerdo a la HAMD-17 y las escalas CGI-SI y PGI-SI y se contrastaron las diferencias con la prueba Anova (contraste de Scheffé) en caso de distribuciones paramétricas y Kruskal-Wallis y Mann-Whitney con corrección de Bonferroni en diferencias 2 a 2 como pruebas no paramétricas.
- Validez convergente: Se analizó la correlación de Pearson entre las escalas CUDOS y la HAMD-17. También se comprobó la relación de la escala CUDOS con el cuestionario SF-36 (PCS y MCS) y la escala SOFAS mediante el mismo estadístico.

Finalmente, se aplicó un análisis multivariado de regresión (método de pasos sucesivos con criterio de entrada 0,05

y salida 0,1), considerando la puntuación en la escala CUDOS como variable dependiente. Las puntuaciones en los instrumentos HAMD-17, SF-36, SOFAS junto a la edad, el índice de comorbilidad de Charlson y los factores de riesgo, fueron contrastadas como variables independientes (controlando el efecto del sexo de los pacientes como posible variable moduladora).

En todas las pruebas estadísticas se empleó un nivel de significación estadística (α) de 0,05. El software empleado en los análisis fueron el STATA v.10 y el SPSS (16.0).

RESULTADOS

Etapa I: Adaptación lingüística

Todos los ítems de la versión inicial del CUDOS fueron valorados positivamente en importancia y comprensibilidad por los expertos (Figura 2). En cambio, en el caso de los pacientes se encontraron discrepancias, referidas fundamentalmente a la importancia de algún ítem. No obstante, debe precisarse que estos ítems se mantuvieron ya que no presentaban problemas de comprensión y su importancia era alta para los expertos. Cabe señalar que cuatro enunciados fueron discutidos y ligeramente modificados durante la sesión de presentación de los resultados para mejorar su

Tabla 3	Características de los ítems de la escala CUDOS y su puntuación total										
	% E. suelo	% E. techo	% No válido	media	DT	Mediana	Min	Max	g_2	ID	^a DIF
CUDOS total (Ítems 1-16)	0,35	0,00	6,56	29,86	12,12	29,00	0,00	58,00	-0,42	-----	1,31
Ítem 1	1,64	30,82	0,00	2,69	1,09	3,00	0,00	4,00	-0,94	0,64	0,91
Ítem 2	4,93	29,93	0,33	2,64	1,19	3,00	0,00	4,00	-0,73	0,66	1,15
Ítem 3	30,13	18,21	0,98	1,76	1,48	2,00	0,00	4,00	-1,35	0,41	1,08
Ítem 4	57,33	8,33	1,64	0,91	1,30	0,00	0,00	4,00	0,32	0,21	1,41
Ítem 5	11,80	35,74	0,00	2,61	1,37	3,00	0,00	4,00	-0,84	0,47	1,25
Ítem 6	61,13	6,98	1,31	0,77	1,21	0,00	0,00	4,00	1,32	0,15	0,92
Ítem 7	11,88	16,17	0,66	2,14	1,23	2,00	0,00	4,00	-0,89	0,47	1,10
Ítem 8	8,88	21,38	0,33	2,37	1,22	2,00	0,00	4,00	-0,79	0,50	0,92
Ítem 9	5,59	29,93	0,33	2,69	1,16	3,00	0,00	4,00	-0,41	0,63	1,45
Ítem 10	25,00	21,05	0,33	1,93	1,49	2,00	0,00	4,00	-1,42	0,65	0,99
Ítem 11	30,49	17,70	0,00	1,72	1,49	2,00	0,00	4,00	-1,36	0,67	1,30
Ítem 12	6,58	18,75	0,33	2,37	1,14	2,00	0,00	4,00	-0,62	0,64	1,29
Ítem 13	9,24	14,52	0,66	2,15	1,16	2,00	0,00	4,00	-0,67	0,64	0,82
Ítem 14	61,51	8,55	0,33	0,89	1,33	0,00	0,00	4,00	0,24	0,60	1,10
Ítem 15	79,67	3,28	0,00	0,42	0,96	0,00	0,00	4,00	5,13	0,48	0,69
Ítem 16	27,63	14,80	0,33	1,64	1,40	1,00	0,00	4,00	-1,14	0,63	1,20
Ítem 17	3,72	17,57	2,95	2,49	1,09	3,00	0,00	4,00	-0,64	-----	1,12
Ítem 18	0,00	8,30	5,25	2,51	0,77	3,00	1,00	4,00	-0,36	-----	1,32

*g*₂: curtosis; ID: Índice de discriminación del ítem (correlación ítem-total corregida); DIF: funcionamiento diferencial de los ítems en función de la variable sexo (prueba de Mantel-Haenszel). ^aEn ninguno de los contrastes planteados, el DIF fue estadísticamente significativo (DIF: $p \geq 0,2$)

comprensión. De forma general, no hubo concordancia entre las puntuaciones otorgadas por los grupos participantes en este proceso (coeficientes de correlación intraclase $<0,35$; $p > 0,5$). La versión final de la escala CUDOS a validar en la segunda etapa del estudio se presenta en el Anexo.

Etapa II. Análisis de las propiedades psicométricas de la versión adaptada CUDOS

Perfil de los pacientes TDM evaluados

La muestra final incluida fue de 305 pacientes (69,5% mujeres) con una edad media (DT) de 51,75 (15,53) años. Las variables descriptivas socio-demográficas y clínicas se presentan en las Tablas 1 y 2. El 87,5% de los pacientes estaban siguiendo un tratamiento antidepresivo y el 15,4% de los pacientes presentaba TDM en remisión parcial según la clasificación PRIME-MD. De acuerdo a la clasificación de gravedad empleada en HAMD-17, el 7,9% de los pacientes no presentaban síntomas depresivos (0-7 puntos), 29,8% su-

frían depresión menor (8-13 puntos), 22,0% depresión moderada (14-18 puntos), 21% depresión grave (19-22) y un 19,3%, depresión muy grave (≥ 23 puntos).

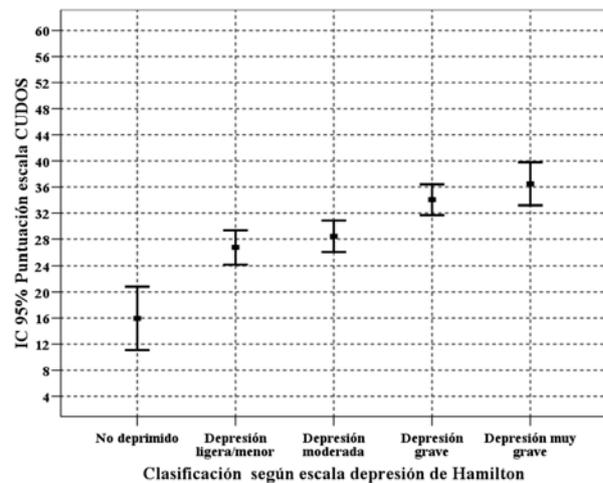
Factibilidad de la escala CUDOS y distribución de las puntuaciones

El tiempo medio de cumplimentación (recortado al 5%) fue de 4,47 ($\pm 2,40$) minutos. Por otro lado, el porcentaje de no respuesta en los ítems 1-16 fue inferior al 3% siendo posible obtener la puntuación total en el 93,44% de los pacientes (Tabla 3). En la muestra evaluada, no se encontraron pacientes con puntuación máxima y el 0,35% de los mismos presentaron una puntuación mínima. Respecto a la distribución de las puntuaciones, el coeficiente de curtosis (g_2), el estadístico de asimetría ($-0,07$) y su error típico (0,14) fueron en este caso indicadores de una distribución simétrica de las puntuaciones en la muestra del estudio, que además se ajustó a una curva normal (Z de Kolmogorov-Smirnov=0,84; $p=0,48$).

Tabla 4				
Análisis Componentes Principales (ACP): varianza explicada y matriz de componentes				
Componentes	ACP ítems 1-16		ACP eliminando ítems 4 y 6	
	Total	% de la varianza	Total	% de la varianza
1	6,11	38,19	6,03	43,10
2	1,80	11,27	1,77	12,67
3	1,5	9,81	1,04	7,41
4	1,057	6,61	0,92	6,54
Matriz de Componentes				
Ítems	Componente 1	Componente 1		
1	0,724	0,733		
2	0,745	0,750		
3	0,521	0,538		
4	0,236	-----		
5	0,548	0,545		
6	0,168	-----		
7	0,541	0,531		
8	0,571	0,574		
9	0,697	0,699		
10	0,702	0,696		
11	0,726	0,720		
12	0,722	0,724		
13	0,719	0,722		
14	0,662	0,664		
15	0,536	0,531		
16	0,698	0,689		
Correlación entre componentes 1 y 2	0,450	0,453		
α de Cronbach	0,878	0,893		

Funcionamiento diferencial de los ítems

En la Tabla 3, se presentan los valores de Mantel-Haenszel. En ningún ítem se encontró un DIF no uniforme que fuera estadísticamente significativo, indicando por tanto que los varones y las mujeres tenían una probabilidad similar de indicar puntuaciones altas o bajas en cada puntuación analizada.



Entre las categorías no deprimidos/depresión ligera-moderada/depresión grave-muy grave: $\chi^2=54,165$; $p<0,001$ (Kruskal-Wallis). Contraste Bonferroni $p<0,03$ en todas las pruebas 2a2

Figura 3

Diferencias de medias en la escala CUDOS en función de la gravedad de la depresión (escala de depresión de Hamilton)

Fiabilidad y validez de constructo

Los indicadores de consistencia interna de la escala CUDOS reflejaron unos niveles adecuados en este criterio (Tabla 4). Respecto al ID, todos los ítems presentaron valores superiores a 0,4 con 2 excepciones: "he dormido demasiado" y "he tenido mucho más apetito de lo normal", que parecen tener un poder de discriminación muy bajo, al menos en la muestra evaluada (Tabla 3).

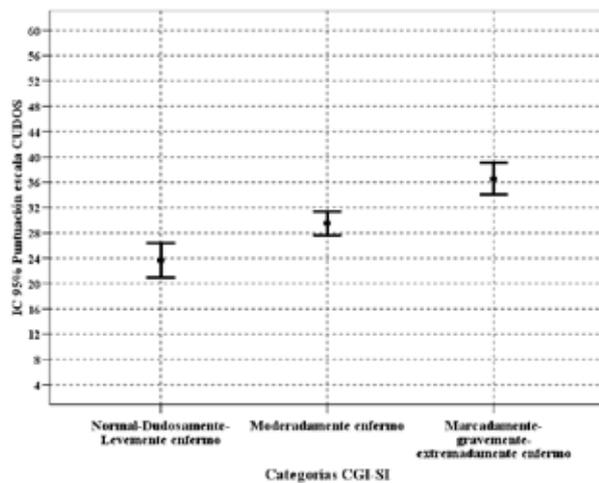
En el ACP llevado a cabo con los 16 ítems de la puntuación sumaria (Tabla 4), se contrastó, en primer lugar, la adecuación muestral para el análisis (Kaiser-Meyer-Olkin=0,86, $p<0,001$). A continuación, en el gráfico de sedimentación (figura no mostrada) se reflejó la existencia de un primer componente dominante (autovalor=6,11) que explicaba el 38,2% de la varianza total frente al 11,3%, 9,8% y 6,6% del resto de componentes con autovalores mayores que 1. Tras aplicar la rotación (Normalización Oblimin con Kaiser), en la matriz de configuración se pudo comprobar que 6 ítems presentaban pesos significativos ($>0,3$) en múltiples componentes. Esos ítems referían además a contenidos diferenciados (ítems 1, 3, 6-7, 10-11) y la solución matricial era difícilmente interpretable. En consecuencia, se forzó el análisis para la extracción de los 2 componentes principales (que explicarían en torno al 50% de la varianza) con rotación ortogonal pero la correlación entre estos 2 componentes fue importante (0,45) y 7 preguntas presentaban pesos significativos en ambos factores (ítems: 5, 7, 10-13,

Tabla 5		Validez discriminante escala CUDOS				
Variables		Puntuación escala CUDOS				
		N	Media	DT	ETM	*p
Sexo	Varón	87	29,03	11,87	1,27	0,482
	Mujer	198	30,22	12,24	0,87	
Convivencia	Solo	34	29,74	13,52	2,32	0,837
	Acompañado en familia	250	29,91	11,96	0,76	
Edad	<40 años	67	29,58	12,62	1,54	0,150
	40-65 años	157	30,92	12,24	0,98	
	>65 años	61	27,43	11,04	1,41	
Formación	Ninguna/básica	135	29,76	12,44	1,07	0,519
	Superior	56	27,84	11,95	1,60	
Resultado en escala PRIME-MD	Síntomas de TDM	243	30,79	11,80	0,76	0,003
	Remisión parcial	42	24,48	12,70	1,96	
Episodios depresivos previos	Sí	121	31,01	12,23	1,11	0,228
	No	164	29,01	12,01	0,94	
Antecedentes familiares depresión	Sí	110	31,85	10,88	1,04	0,022
	No	152	28,29	13,21	1,07	
Problemas graves en personas cercanas	Sí	85	33,61	11,19	1,21	0,001
	No	191	28,26	12,40	0,90	
Problemas económicos o laborales	Sí	113	32,50	11,68	1,10	0,002
	No	172	28,12	12,13	0,92	
Problemas de convivencia/familiares	Sí	87	35,09	11,06	1,19	0,001
	No	196	27,55	11,91	0,85	

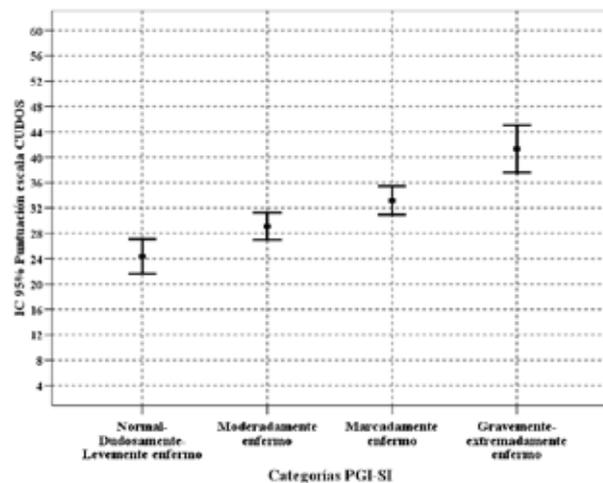
*Prueba de Mann-Whitney

Tabla 6		Correlación entre las puntuaciones de la escala CUDOS, SF-36 y escala SOFAS de funcionamiento social del paciente			
	SF-36: PCS	SF-36: MCS	Escala CUDOS	Escala SOFAS	
SF-36: MCS	-0,074				
Escala CUDOS	-0,224(**)	-0,650(**)			
Escala SOFAS	0,351(**)	0,368(**)	-0,423(**)		
Ítem 17 CUDOS	-0,219(**)	-0,588(**)	0,645(**)	-0,456(**)	
Ítem 18 CUDOS	-0,145(*)	-0,559(**)	0,604(**)	-0,464(**)	

** La correlación es significativa al nivel 0,01 (bilateral)



$F_{2,284}=25,447$; $p<0,001$. Contraste Scheffé $p<0,002$ en todas las pruebas 2a2



$\chi^2=47,208$; $p<0,001$ (Kruskal-Wallis). Contraste Bonferroni $p<0,05$ en todas las pruebas 2a2

Figura 4

Puntuación de la escala CUDOS en función de las escalas de Impresión Clínica de Gravedad (CGI-SI) e Impresión de Gravedad de los Pacientes (PGI-SI)

16), generando igualmente problemas de interpretación. A partir de estos datos, se realizó la extracción de un único componente observándose que todos los ítems de la escala correlacionaban de forma moderada-alta con el mismo (con la excepción de las preguntas 4 y 6). Debido a que estos ítems de la escala con baja capacidad de discriminación deberían ser propuestos para su eliminación, se recalcularon los índices de la consistencia interna y el ACP (Tabla 4). Debe señalarse que cuando se analizaron los residuos de la matriz de correlaciones (entre observadas y reproducidas), en ningún caso una solución multidimensional fue claramente superior en el porcentaje de residuales no redundantes frente a la unidimensional.

Validez de criterio y discriminante

Para la evaluación de estas propiedades, se contrastaron las diferencias de las puntuaciones en la escala CUDOS en función de distintos criterios. En relación a la clasificación de gravedad de la HAMD-17, la escala CUDOS discriminó entre tres categorías: no deprimidos / depresión ligera-moderada / depresión grave-muy grave ($p<0,03$ con corrección de Bonferroni; Figura 3). También se encontraron diferencias significativas en las puntuaciones de la escala CUDOS en relación a las categorías definidas en las escalas CGI-SI y PGI-SI ($p<0,002$ y $p<0,05$; Figura 4). En suma a estas evidencias, en la Tabla 5 se presentan contrastes adicionales considerando variables socio-demográficas y relativas a situaciones personales que pudieran constituir un factor de riesgo para la aparición o el mantenimiento de trastornos depresivos.

Validez convergente

El coeficiente de correlación de Pearson para la CUDOS y la escala HAMD-17 fue moderado, positivo y significativo ($r=0,451$; $p<0,001$). Junto a esto, la puntuación total en la escala CUDOS (ítems 1-16) así como las preguntas relativas al grado de interferencia en la vida del paciente y la CVRS general (ítems 17 y 18, respectivamente) se asociaron de forma moderada y significativa respecto la puntuación del componente mental del SF-36 (MCS) y la escala SOFAS (Tabla 6).

Factores asociados a la puntuación CUDOS

Las variables independientes incluidas en el modelo multivariado explicaron el 50% de las diferencias en las puntuaciones en la escala CUDOS ($R^2=0,495$. $F_{3,252}=82,34$; $p<0,001$). Los pesos relativos de cada variable incluida en el modelo final fueron los siguientes (coeficientes estandarizados): MCS ($\beta=-0,526$; $p<0,001$), HAMD-17 ($\beta=0,260$; $p<0,001$) y problemas de convivencia ($\beta=-0,121$; $p=0,01$). Por último, precisar que ninguna de estas variables en el modelo final presentó problema alguno de colinealidad, tras analizar los índices de condición y las proporciones de varianza.

DISCUSIÓN

El objetivo del presente estudio fue la adaptación cultural y el análisis de las propiedades psicométricas de la escala

CUDOS, un instrumento auto-aplicado breve cuya versión original demostró ser de utilidad en la práctica clínica para la evaluación eficiente de los pacientes con depresión^{26-29,43}.

Si bien las escalas de depresión de Hamilton y de Montgomery-Åsberg son las más frecuentemente utilizadas para la evaluación clínica de los pacientes con TDM^{35,44}, están diseñadas para recoger información estrictamente desde el punto de vista del clínico. Como ya se ha indicado, la perspectiva de los pacientes es un complemento esencial para identificar síntomas relevantes y evaluar cómo afectan a su vida diaria^{13,15,22,25}. Entre otros, ejemplos de instrumentos autoaplicados disponibles en nuestro país son el ampliamente utilizado Inventario de Depresión de Beck (BDI)⁴⁵, el PHQ-15⁴⁶ (centrado en síntomas somáticos asociados a depresión y ansiedad) y el método basado en escalas visuales analógicas, empleado con éxito en pacientes ambulatorios evaluados en un centro de salud mental⁴⁷. Frente a estos últimos más recientes, la escala CUDOS ofrece una descripción más completa de los síntomas asociados al TDM de acuerdo al DSM-IV y permite además evaluar otros aspectos importantes como la capacidad funcional del paciente y su afectación en la calidad de vida relacionada con la salud.

Respecto a factibilidad de uso de la escala, el tiempo medio de cumplimentación de la versión adaptada (4,5 minutos, DT=2,4) fue mayor al indicado por los autores de la versión original (3 minutos)²⁹. Aún así, resultó un tiempo razonable en el contexto de AP.

En relación a su estructura interna, los resultados del ACP presentados junto a los índices de consistencia interna y discriminación permiten asumir la unidimensional de la misma. No obstante, hay limitaciones que deben ser consideradas. Entre otras, aunque la escala CUDOS (ítems 1-16) presenta alta consistencia interna con valores similares a los publicados previamente en la versión original²⁷ (α de Cronbach de 0,88 y 0,9, respectivamente), los ítems 4 y 6 deberían ser analizados de nuevo en trabajos posteriores con el objeto de valorar su posible eliminación, al menos en el cálculo de la puntuación de la escala CUDOS.

A pesar de lo anterior, la validez de criterio y discriminante de la escala CUDOS se confirmó, en primer lugar, mediante las diferencias de medias encontradas entre tres categorías en función de la HAMD-17: no deprimidos / depresión ligera-moderada / depresión grave-muy grave. También se contrastaron como estadísticamente significativas las diferencias de acuerdo a diversas medidas de severidad de afectación o síntomas (CGI-SI y PRIME-MD). En suma a esto, la validez convergente de la versión que presentamos parece razonable: sus puntuaciones se asociaron de forma significativa y moderada con la escala de Hamilton (0,45 en la versión española y 0,69 en la original; ambas $p < 0,001$). En función de estas diferencias en la HAMD-17, las puntuaciones en la escala CUDOS para pacientes TDM podrían agruparse en torno a 3 clústeres: 0-24-no deprimidos

/24-32-depresión leve o moderada />32: depresión grave o muy grave. Esta primera aproximación, distinta de la facilitada por los autores originales, deberá ser analizada en detalle en próximos estudios con diversas muestras y grupos con el fin de precisar mejor los baremos. Entendemos que lo mismo es aplicable al funcionamiento diferencial de los ítems y la escala. Aunque se ha podido contrastar que no se producía sesgo por DIF en relación al sexo, es necesario realizar estos análisis (u otros más completos, aunque menos intuitivos como los Logit o análisis de Rasch) respecto a nuevos contextos y variables.

Por último, en el modelo multivariado de regresión, las variables independientes significativamente asociadas con la puntuación CUDOS fueron el componente mental del SF-36 (MCS), la HAMD-17 y los problemas de convivencia de los pacientes. Estos resultados están en armonía con los publicados recientemente en nuestro país también en AP⁴⁰.

Dentro de las limitaciones de este estudio, debemos señalar que la información presentada procede de una única visita de evaluación, sin seguimiento, por lo que no se ha podido presentar información sobre la sensibilidad para detectar cambios clínicamente significativos en la situación de los pacientes⁴⁸. También debe indicarse que la población analizada corresponde a pacientes de AP y las puntuaciones podrían ser diferentes en pacientes más graves y crónicos como los asistidos en atención especializada.

En conclusión, la versión española de la escala CUDOS, presenta buenas capacidades psicométricas, con una fiabilidad y un rendimiento descriptivo y discriminativo adecuados. Adicionalmente es fácilmente cumplimentada por los pacientes. Estos hechos, junto a la posibilidad de obtener una puntuación sumaria de forma muy rápida, facilitan su uso en la evaluación del TDM durante práctica clínica en Atención Primaria.

CONFLICTO DE INTERESES

El estudio ha sido financiado por el Departamento Médico de AstraZeneca España. E.M y J.M son empleados de AstraZeneca España. J.C y H.D-C son empleados de BAP LA-SER Outcomes España. L.A y C.M no tienen conflictos de intereses que declarar.

BIBLIOGRAFÍA

1. Haro JM, Palacin 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.
2. Gabilondo A, Rojas-Farreras S, Vilagut G, Haro JM, Fernández A, Pinto-Meza A, et al. Epidemiology of major depressive episode in a southern European country: results from the ESEMeD-Spain project. *J Affect Disord*. 2010;120(1-3):76-85.
3. Aragones E, Pinol JL, Labad A, Masdeu RM, Pino M, Cervera J.

- Prevalence and determinants of depressive disorders in primary care practice in Spain. *Int J Psychiatry Med.* 2004;34(1):21-35.
4. Angst J, Gamma A, Gastpar M, Lepine JP, Mendlewicz J, Tylee A. Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci.* 2002;252(5):201-9.
 5. Serrano-Blanco A, Palao DJ, Luciano JV, Pinto-Meza A, Lujan L, Fernandez 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(2):201-10.
 6. Autonell J, Vila F, Pinto-Meza A, Vilagut G, Codony M, Almansa J, et al. One year prevalence of mental disorders comorbidity and associated socio-demographic risk factors in the general population of Spain. Results of the ESEMeD-Spain study. *Actas Esp Psiquiatr.* 2007;35(Suppl 2):4-11.
 7. Charlson FJ, Stapelberg NJ, Baxter AJ, Whiteford HA. Should global burden of disease estimates include depression as a risk factor for coronary heart disease? *BMC Med.* 2011;9:47.
 8. Ten Doesschate MC, Bockting CL, Koeter MW, Schene AH. Prediction of recurrence in recurrent depression: a 5.5-year prospective study. *J Clin Psychiatry.* 2010;71(8):984-91.
 9. Vuorilehto MS, Melartin TK, Isometsa ET. Course and outcome of depressive disorders in primary care: a prospective 18-month study. *Psychol Med.* 2009;39(10):1697-707.
 10. Romera I, Perez V, Menchon JM, Delgado-Cohen H, Polavieja P, Gilaberte I. Social and occupational functioning impairment in patients in partial versus complete remission of a major depressive disorder episode. A six-month prospective epidemiological study. *Eur Psychiatry.* 2010;25(1):58-65.
 11. Bernal M, Haro JM, Bernert S, Brugha T, de GR, Bruffaerts R, et al. Risk factors for suicidality in Europe: results from the ESEMED study. *J Affect Disord.* 2007;101(1-3):27-34.
 12. Sicras-Mainar A, Blanca-Tamayo M, Gutierrez-Nicuesa L, Salvatella-Pasant J, Navarro-Artieda R. Impact of morbidity, resource use and costs on maintenance of remission of major depression in Spain: a longitudinal study in a population setting. *Gac Sanit.* 2010;24(1):13-9.
 13. Grupo de Trabajo sobre el Manejo de la Depresión Mayor en el Adulto. Guía de Práctica Clínica sobre el Manejo de la Depresión Mayor en el Adulto. Madrid, 2008.
 14. Valladares A, Dilla T, Sacristan JA. La depresión: una hipoteca social. Últimos avances en el conocimiento del coste de la enfermedad. *Actas Esp Psiquiatr.* 2009;37(1):49-53.
 15. National Collaborating Centre for Mental Health. Depression. The treatment and management of depression in adults. London: National Institute for Health and Clinical Excellence, 2009.
 16. Sobocki P, Ekman M, Agren H, Runeson B, Jonsson B. The mission is remission: health economic consequences of achieving full remission with antidepressant treatment for depression. *Int J Clin Pract.* 2006;60(7):791-8.
 17. Agüera L, Failde I, Cervilla JA, Diaz-Fernandez P, Mico JA. Medically unexplained pain complaints are associated with underlying unrecognized mood disorders in primary care. *BMC Fam Pract.* 2010;11:17.
 18. Fernandez A, Pinto-Meza A, Bellon JA, Roura-Poch P, Haro JM, Autonell J, et al. Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. *Gen Hosp Psychiatry.* 2010;32(2):201-9.
 19. Aragonés E, Pinol JL, Labad A, Folch S, Melich N. Detection and management of depressive disorders in primary care in Spain. *Int J Psychiatry Med.* 2004;34(4):331-43.
 20. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006;163(11):1905-17.
 21. Lester H, Howe A. Depression in primary care: three key challenges. *Postgrad Med J.* 2008;84(996):545-8.
 22. Willkie R, Burke L, Rickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Control Clin Trials.* 2009;25:535-52.
 23. Rey E. Using patient report outcomes in the office: a challenge for the physician. *An Med Interna.* 2007;24(9):413-4.
 24. Montejo AL, Lauffer JC, Cuervo J, Rebollo P, Cordero L, Diez T, et al. Validation of a specific measure to assess health-related quality of life in patients with schizophrenia and bipolar disorder: the 'Tolerability and quality of life' (TOOL) questionnaire. *Ann Gen Psychiatry.* 2011;10:6.
 25. Maurino J, Sanjuan J, Haro JM, Diez T, Ballesteros J. Impact of depressive symptoms on subjective well-being: the importance of patient-reported outcomes in schizophrenia. *Patient Prefer Adherence.* 2011;5:471-4.
 26. Zimmerman M, Posternak MA, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry.* 2004;161(10):1911-3.
 27. Zimmerman M, Chelminski I, McGlinchey JB, Posternak MA. A clinically useful depression outcome scale. *Compr Psychiatry.* 2008;49(2):131-40.
 28. Zimmerman M, Martinez J, Attiullah N, Friedman M, Toba C, Boerescu DA, et al. Determining remission from depression on two self-report symptom scales: a comparison of the Quick Inventory of Depressive Symptomatology and the Clinically Useful Depression Outcome Scale. *Compr Psychiatry.* 2012;53(7):1034-8.
 29. Zimmerman M, McGlinchey JB. Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Ann Clin Psychiatry.* 2008;20(3):125-9.
 30. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). Washington D.C.: American Psychiatric Publishing, 2000.
 31. Spitzer RL, Williams JB, Kroenke K, Linzer M, deGruy FV, III, Hahn SR, et al. Utility of a new procedure for diagnosing mental disorders in primary care. The PRIME-MD 1000 study. *JAMA.* 1994;272(22):1749-56.
 32. Baca E, Saiz J, Agüera L, Caballero L, Fernandez-Liria A, Ramos J, et al. Validación de la versión española del PRIME-MD: un procedimiento para el diagnóstico de trastornos mentales en atención primaria. *Actas Esp Psiquiatr.* 1999;27(6):375-83.
 33. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. *Value Health.* 2009;12(4):430-40.
 34. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry* 1960;23:56-62.
 35. Bobes J, Bulbena A, Luque A, Dal-Re R, Ballesteros J, Ibarra N. A comparative psychometric study of the Spanish versions with 6, 17, and 21 items of the Hamilton Depression Rating Scale. *Med Clin (Barc).* 2003;120(18):693-700.
 36. Romera I, Perez V, Menchon JM, Polavieja P, Gilaberte I. Optimal cutoff point of the Hamilton Rating Scale for Depression according to normal levels of social and occupational functioning. *Psychiatry Res.* 2011;186(1):133-7.
 37. Guy W. ECDEU Assessment Manual for Psychopharmacology (Revised). Rockville: National Institute of Mental Health, Dept of

- Health, Education and Welfare publication (ADM), 1976.
38. Spitzer R, Gibbon M, Endicott J. Global Assessment Scale (GAS), Global Assessment of Functioning (GAF), Social and Occupational Functioning Scale (SOFAS). Mental Health Status, Functioning and Disability Measures. In: First M, editor. Handbook of psychiatric measures (1st ed.). Washington DC: American Psychiatric Association, 2000; p. 96-100.
 39. King M, Walker C, Levy G, Bottomley C, Royston P, Weich S, et al. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. *Arch Gen Psychiatry*. 2008;65(12):1368-76.
 40. Bellon JA, de Dios LJ, King M, Moreno-Kustner B, Nazareth I, Monton-Franco C, et al. Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain. *Psychol Med*. 2011;41(10):2075-88.
 41. Vilagut G, Ferrer M, Rajmil L, Rebollo P, Permanyer-Miralda G, Quintana JM, et al. The Spanish version of the Short Form 36 Health Survey: a decade of experience and new developments. *Gac Sanit*. 2005;19(2):135-50.
 42. Rebollo P. Use of health-related quality of life "SF Group" questionnaires (SF-36, SF-12 and SF-6D) in the framework of clinical trials and routine clinical practice. *Med Clin (Barc)*. 2008;130(19):740-1.
 43. Zimmerman M, McGlinchey JB, Posternak MA, Friedman M, Boerescu D, Attiullah N. Differences between minimally depressed patients who do and do not consider themselves to be in remission. *J Clin Psychiatry*. 2005;66(9):1134-8.
 44. Lobo A, Chamorro L, Luque A, Dal-Re R, Badia X, Baro E. Validation of the Spanish versions of the Montgomery-Asberg depression and Hamilton anxiety rating scales. *Med Clin (Barc)*. 2002;118(13):493-9.
 45. Sanz J, Vázquez C. Fiabilidad, validez y datos normativos del Inventario para la Depresión de Beck. *Psicothema*. 1998;10(2):303-18.
 46. Ros MS, Comas VA, Garcia-Garcia M. Validación de la Versión Española del Cuestionario PHQ-15 para la evaluación de síntomas físicos en pacientes con trastornos de depresión y/o ansiedad: estudio DEPRE-SOMA. *Actas Esp Psiquiatr*. 2010;38(6):345-57.
 47. Cordero-Villafafila A, Ramos-Brieva JA, Gutierrez LR, Arranz ML. A brief and rapid procedure to measure the intensity of depressive symptoms in Primary Care. *Actas Esp Psiquiatr*. 2011;39(1):45-8.
 48. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-9.

ANEXO I		Versión en castellano de la escala CUDOS (Clinically Useful Depression Outcome Scale)				
Durante la SEMANA PASADA, INCLUYENDO HOY.....		No es cierto en absoluto (0 días)	Rara vez cierto (1-2 días)	A veces cierto (3-4 días)	A menudo cierto (5-6 días)	Casi siempre cierto (cada día)
1.	Me he sentido triste o deprimido	0	1	2	3	4
2.	He perdido interés en mis actividades cotidianas	0	1	2	3	4
3.	He tenido poco apetito y no me apetecía comer	0	1	2	3	4
4.	He tenido mucho más apetito de lo habitual	0	1	2	3	4
5.	He tenido dificultades para dormir	0	1	2	3	4
6.	He dormido demasiado	0	1	2	3	4
7.	Me he sentido muy nervioso, resultándome difícil permanecer quieto	0	1	2	3	4
8.	Me he sentido físicamente más lento o pesado	0	1	2	3	4
9.	He tenido poca energía	0	1	2	3	4
10.	Me he sentido culpable	0	1	2	3	4
11.	He pensado que era un fracasado	0	1	2	3	4
12.	He tenido problemas de concentración	0	1	2	3	4
13.	He tenido más problemas para tomar decisiones que habitualmente	0	1	2	3	4
14.	He deseado estar muerto	0	1	2	3	4
15.	He pensado en suicidarme	0	1	2	3	4
16.	He pensado que no había esperanza en el futuro	0	1	2	3	4
17.	En general, ¿en qué medida los síntomas de depresión han interferido o le han causado problemas en su vida diaria durante la semana pasada?					
	0: Nada	1: Un poco	2: Moderadamente	3: Bastante	4: Mucho	
18.	Durante la semana pasada, ¿cómo calificaría su calidad de vida en general?					
	0: Muy buena, mi vida no podría ser mejor	1: Bastante buena, la mayoría de mis cosas están yendo bien	2: Las partes buenas y malas están más o menos igualadas	3: Bastante mala, la mayoría de mis cosas están yendo mal	4: Muy mala, mi vida no podría ser peor	

RESULTADOS

Tal y como se ha podido revisar en el apartado anterior, en el que se han presentado las 3 publicaciones incluidas en el proyecto de tesis doctoral, cada una de las mismas se centró en una medida PRO cuyo interés había sido introducido previamente. De este modo, en primer lugar, se trabajó en el desarrollo de un nuevo instrumento para evaluar la satisfacción de los pacientes crónicos con tratamiento respecto a los niveles propuestos en el modelo Jerárquico de satisfacción y en el ámbito de Atención Primaria (AP). En segundo lugar, se llevó a cabo en distintos centros especializados de Salud Mental una adaptación cultural completa de un cuestionario muy breve, ya existente fuera de nuestras fronteras y válido para la evaluación del impacto de los tratamientos antipsicóticos y los efectos secundarios, extrapiramidales, en la CVRS de los pacientes esquizofrénicos y con trastorno bipolar. En tercer lugar, de nuevo en centros de AP, se adaptó culturalmente (incluyendo también el estudio de sus propiedades psicométricas básicas), una escala específica, también para la evaluación de la CVRS, pero en pacientes con TDM. Esta medida, escala CUDOS, además de permitir una completa evaluación de este constructo, permitía detectar a los pacientes cuyo tratamiento antidepresivo no estaba siendo efectivo lo cual debería redundar en un mejor manejo del paciente.

A continuación, se presentan los principales hallazgos de cada una de las investigaciones prestando atención a los objetivos e hipótesis planteados en cada una de las mismas.

ARTÍCULO 1: Desarrollo y validación cuestionario SAT-Q.

Respecto al estudio de desarrollo del cuestionario de satisfacción de los pacientes crónicos en tratamiento, SAT-Q, se partió de una revisión sistemática de la literatura con un examen de las publicaciones por 2 revisores expertos en IRS. Junto a éstos, el trabajo de los expertos traductores independientes permitió la generación de un banco de ítems inicial (n=61 ítems) que abordaban aspectos apuntando a los niveles descritos en el modelo de referencia, haciendo factible el desarrollo del instrumento, al menos de forma inicial. Este pool primario fue analizado a nivel de contenido y discutido por los 2 expertos en IRS y 5 médicos de Atención Primaria de reconocido prestigio (panel de expertos) en el ámbito clínico y de la investigación. Se valoró especialmente la importancia y la comprensibilidad de los ítems y dimensiones en una escala tipo Likert (0: mínima importancia/comprensibilidad - 4: máxima importancia/comprensibilidad). Tras una revisión de los valores medios en cada aspecto comentado y el posterior consenso de los expertos, se obtuvo una nueva versión del SAT-Q de 34 ítems que se ocuparían de medir el nivel de satisfacción del paciente con el servicio sanitario a distintos niveles y el tratamiento farmacológico ya que las entendían como las relevantes en nuestro contexto. La satisfacción respecto al conjunto de tratamientos y otras indicaciones (dietas, ejercicio, hábitos, etc...) los expertos clínicos la entendían como diseminada en las preguntas de satisfacción con la atención sanitaria en general y en la satisfacción general con el tratamiento recibido. Con el objeto de completar el estudio del contenido de los ítems y de paso, realizar un estudio piloto con la versión inicial del instrumento, se presentó la versión de 34 ítems a 30 pacientes crónicos, para que evaluaran la comprensibilidad e importancia de cada uno de los ítems y respondieran al mismo. Además, se les pedía que aportaran nuevas cuestiones que creyeran relevantes para estar satisfechos con la atención sanitaria y su tratamiento.

Como resultado, todos los ítems les resultaron relevantes y no propusieron nuevas alternativas de contenido que no estuvieran ya reflejadas. No obstante, 9 preguntas fueron modificadas con el fin de mejorar su grado de comprensibilidad. Se obtuvo así, la versión del SAT-Q que, posteriormente, se sometería al estudio de validación. A tal fin, se diseñó un trabajo de campo para implementarlo en un total de 4 centros de AP del Principado de Asturias recogiendo información válida de 202 pacientes crónicos (65,35% mujeres), con una edad media de casi 63 años que llevaban en torno a 9 años con tratamiento farmacológico para cualquier enfermedad crónica (principalmente artritis, depresión, diabetes e hipertensión).

En los resultados de los análisis estadísticos aplicados, se comprobó la factibilidad del cuestionario para ser empleado en el entorno de la AP: sólo el 3% de los participantes no pudieron completarlo en su totalidad. Por otra parte, la correlación ítem-total y el análisis factorial reflejaron la posibilidad de obtener una estructura básica más reducida conformada por 18 ítems distribuidos en varias dimensiones. Se encontraron 2 dimensiones generales: satisfacción con el tratamiento farmacológico (72,35% de varianza explicada -3 ítems-), y satisfacción con la atención sanitaria (64,74% de la varianza-3 ítems-). También se encontraron 4 dimensiones específicas de la satisfacción con el tratamiento: efectos adversos del tratamiento (30,22% -3 ítems-); efectividad del tratamiento (14,96% -3 ítems-), comodidad/conveniencia de la administración del tratamiento (14,14% -4 ítems-) y olvidos en las tomas (10,1% -2 ítems-). Estas puntuaciones se transformaron linealmente en una escala con límite inferior en 0 (peor satisfacción, más olvidos, menos conveniente el tratamiento...) hasta el límite superior situado en 100 (mejor satisfacción, menos olvidos, más conveniente el tratamiento,...) aplicando el algoritmo:

$$Y' = (Y_{\text{obs}} - Y_{\text{min}}) / (Y_{\text{max}} - Y_{\text{min}}) \times 100$$

Donde Y_{obs} es el valor obtenido por el paciente, Y_{max} es el valor máximo de la escala en bruto –sin transformar-, e Y_{min} es el valor mínimo de la escala en bruto.

La consistencia interna de las dimensiones fue aceptable (rango α de Cronbach: 0,6-0,8) así como la estabilidad de las puntuaciones (rango coeficiente correlación intraclassa –CCI-:0,63-0,93. La puntuación más baja en el CCI se obtuvo en los efectos adversos sufridos por los pacientes). Se observó una distribución de las puntuaciones asimétrica en las subescalas con un marcado efecto techo (los pacientes tendían a mostrar unos valores de satisfacción muy altos). Finalmente, se encontró una asociación positiva de intensidad leve-moderada y estadísticamente significativa entre las puntuaciones de satisfacción general con el tratamiento, efectividad del mismo y su perfil de efectos adversos con las puntuaciones sumarias del SF-12 (física y mental), tal y como se preveía.

ARTÍCULO 2: Adaptación cultural y validación del cuestionario Tool.

En el caso del cuestionario Tool, al tratarse de una medida existente y en funcionamiento en el país de origen (Suecia), el primer paso consistió en realizar una adaptación de contenido a nuestro contexto. Tras confirmar el equipo coordinador del estudio, autores de la publicación, que el contenido de los 8 ítems de la versión original era pertinente y apuntaban hacia todos los aspectos fundamentales a considerar en la evaluación del impacto de los tratamientos antipsicóticos sobre la CVRS de los pacientes en nuestro país, se procedió a la adaptación lingüística con 3 expertos traductores independientes. El trabajo del equipo de traducción, el comité coordinador y varios profesionales sanitarios (panel de expertos) permitió tener la versión del Tool

adaptada culturalmente para validar (durante este proceso 3 ítems fueron modificados en su redacción: estado de ánimo, temblores y mareos-náuseas). Esta versión fue revisada de forma preliminar en una fase piloto por 40 pacientes sintomáticamente estables (20 esquizofrénicos y 20 con trastorno bipolar) y se pudo comprobar que no había ninguna dificultad para la comprensión y cumplimentación de la medida. En el posterior trabajo de campo para la validación psicométrica participaron 60 centros de Salud Mental distribuidos por las distintas CCAA y se entrevistaron 242 pacientes (238 finalmente incluidos en el análisis: 121 con diagnóstico de esquizofrenia y 117 con trastorno bipolar), el 95% bajo tratamiento antipsicótico. El estudio psicométrico realizado para la evaluación de las propiedades del cuestionario TOOL se centró en el análisis de la factibilidad, la validez (constructo, convergente y de criterio) y la fiabilidad (consistencia interna y test-retest).

Se obtuvo un índice de cumplimentación del total de la escala superior al 95%, con un tiempo de cumplimentación inferior a 5 minutos, lo cual, hace que sea fácilmente aplicable en práctica clínica. El efecto techo-suelo en las puntuaciones se observó sólo en molestias asociadas a los temblores y mareos-nauseas.

En suma a lo anterior, la fiabilidad del cuestionario TOOL, en términos de consistencia interna fue aceptable (α de Cronbach = 0,75). Sólo un ítem presentó un índice de discriminación inferior a 0,3 (“peso corporal”). A partir de este dato, se procedió a la realización del análisis factorial exploratorio excluyendo e incluyendo dicho ítem. Contando con este ítem, se encontró una estructura unidimensional que explicaba aproximadamente el 40% de la varianza total y con la que todos los ítems presentaban una asociación moderada-alta (rango: 0,39-0,79). Considerando que el ítem “peso corporal” tenía un interés claro desde el punto de vista clínico (por las implicaciones

que puede tener por ejemplo en términos de adherencia al tratamiento), su índice de discriminación era superior al mínimo requerido y presentaba una mínima asociación con el factor general subyacente (0,39), el análisis estadístico posterior se incluyó en los análisis posteriores.

En la evaluación de la estabilidad de estas puntuaciones, se calculó el CCI entre los pacientes que no habían experimentado cambios entre la visita basal y el retest, según criterio clínico, y se obtuvieron valores muy altos: 0,90 (IC 95%= 0,73-0,96).

En el análisis de la validez convergente, los coeficientes de correlación de Spearman (ρ) entre las puntuaciones del TOOL con las medidas EQ5D y SF6D fueron negativas, moderadas-altas y significativas (ρ : -0,72, $p < 0,001$ y -0,68, $p < 0,001$ escala SF6D). Estos resultados fueron muy semejantes a los obtenidos cuando se comparó la relación entre el cuestionario TOOL y las escalas de utilidad de forma separada por cada uno de los grupos diagnósticos (esquizofrenia y trastorno bipolar). Además, la relación entre las puntuaciones del TOOL y el instrumento específico de depresión, MADRS, fue igualmente moderada y significativa (ρ : 0,58, $p < 0,01$). También, los ítems del TOOL se asociaron de forma diferencial con las subescalas de la UKU de efectos secundarios como cabía esperar.

Por último, dentro del análisis de la validez de criterio se recodificó la variable “gravedad de los síntomas”, según criterio clínico, comparando las puntuaciones en el cuestionario TOOL y las escalas de utilidad (EQ-5D y SF-6D) de los pacientes que se encontraban en estado normal o levemente enfermo con los que se encontraban en un estado moderadamente grave o muy grave. Como resultado, se encontraron diferencias significativas en las medidas analizadas. Así, los pacientes que tenían un diagnóstico indicativo de mayor gravedad, presentaban puntuaciones más altas (indicativas de peor

estado en el TOOL) respecto a aquellos otros que tenían mejor estado y puntuaciones más bajas en las medidas de utilidad (indicativas de un peor estado de salud percibido).

ARTÍCULO 3. Adaptación cultural y validación de la escala CUDOS.

La última de las publicaciones describe, como se ha indicado, la adaptación cultural de un instrumento específico de depresión (TDM de acuerdo a los criterios DSM-IV-TR) creado por el grupo MIDAS en EEUU para la evaluación de la CVRS de los pacientes con TDM y la rápida detección de los casos con recaída sintomática o respuesta subóptima al tratamiento.

Debido a los objetivos de este estudio, los pasos aplicados en el caso del trabajo anterior fueron los seguidos también en el caso de la adaptación de la escala CUDOS. Así, en una primera etapa se analizó el contenido de sus 18 ítems por el equipo coordinador compuesto por psiquiatras, médicos de AP, psicólogos (uno experto en análisis cualitativo de la información y otro en métodos psicométricos) y expertos en IRS y se realizó la adaptación lingüística con 3 traductores independientes con los que se trabajó para consensuar una versión inicial que reflejara fielmente el contenido de la original (armonización conceptual). Esta primera forma de la escala fue revisada por un comité de expertos (n= 4) y de forma preliminar por pacientes con diagnóstico de TDM de acuerdo a los criterios DSM-IV-TR (n=19), lográndose consensuar una versión final aprobada por todos los expertos y que sería la aplicada en el trabajo decampo. El rango teórico de las puntuaciones se calcula con las 16 primeras preguntas (estructura básicamente unidimensional) en la versión original y oscila entre 0 (sin depresión) y 64 puntos (mayor gravedad posible del TDM).

Para el análisis de las propiedades psicométricas de la versión adaptada de la escala CUDOS se llevó a cabo un estudio multicéntrico observacional transversal en 44

centros de AP de distintas CCAA. Se entrevistaron un total de 338 pacientes adultos con TDM (criterios DSM-IV-TR) diagnosticado en los 3 meses anteriores a la entrevista (un total de 305 fueron considerados válidos).

El tiempo medio requerido para la cumplimentación del total de la escala (inferior a 5 minutos), el porcentaje de no respuesta a los 16 ítems que configuran la puntuación de la misma (inferior al 3%) y el análisis de la distribución de las puntuaciones en los ítems confirmaron la factibilidad de uso del instrumento en el entorno de AP. Tampoco se evidenció un funcionamiento diferencial de los ítems (prueba de Mantel-Haenszel) en función del sexo de los pacientes.

A partir del análisis de consistencia interna y el índice de discriminación se encontró que dos de los ítems de la escala (“he dormido demasiado” y “he tenido mucho más apetito de lo normal”) no parecían estar alineados con el conjunto (correlación ítem total: 0,21 y 0,15, respectivamente) y se analizó por tanto la validez de constructo de la escala CUDOS con y sin estos ítems.

El análisis de componentes principales, si bien no reflejó una estructura subyacente rotundamente unidimensional (38-44% varianza explicada por el primer factor), el análisis de las soluciones matriciales rotadas y el análisis de los residuos de las matrices de correlaciones (observadas y reproducidas) no permitieron concluir sobre la pertinencia o superioridad de una estructura distinta a la planteada por los autores de la versión original. En función de estos datos, el equipo coordinador decidió mantener la propuesta de la versión original a la espera de que futuras aplicaciones de la escala, permitan contrastar con nuevas datos estos hallazgos.

Por el contrario, los datos encontrados en el análisis de la validez de criterio y discriminante sí que fueron más robustos. La escala CUDOS, permitió distinguir entre

los niveles de gravedad sintomática definidos por la prueba de referencia (HAMD-17): no deprimidos frente a depresión ligera o moderada y éstas 2 frente depresión grave o muy grave ($p < 0,03$ con corrección de Bonferroni). También se encontraron diferencias significativas en las puntuaciones de la escala CUDOS en función del criterio de gravedad de los clínicos y los propios pacientes. En suma a esto, se hallaron diferencias estadísticamente significativas en las puntuaciones según la clasificación de la gravedad sintomática, realizada a partir de otras escalas relevantes en AP como la PRIME-MD (“síntomas de TDM” frente a “remisión parcial”), y otras cuestiones de interés como problemas de relación con personas del entorno de los pacientes, problemas económicos o laborales y problemas de convivencia.

Por último, se observó que la escala CUDOS se asociaba de forma significativa, negativa y moderada-alta con el componente sumario mental del cuestionario genérico SF-36 y la escala SOFAS de desenvolvimiento social. También se encontró una asociación moderada pero positiva con los ítems genéricos de CVRS incluidos en la escala (ítems 17 y 18). Así la validez convergente de la versión adaptada a nuestro país de la escala CUDOS parecía también confirmada por los estadísticos aplicados. Así, la escala CUDOS presentaba un contenido relevante y una capacidad discriminativa tal, que permitían pensar en futuros trabajos de investigación para profundizar en su validación.

Para concluir este apartado, simplemente indicar que los trabajos publicados presentan datos que permiten poner a disposición tanto de investigadores en IRS como a los profesionales sanitarios 3 nuevas PROs en nuestro país que era el objetivo fundamental perseguido en este proyecto de tesis doctoral. Como se verá en el apartado de Discusión está claro que desde el punto de vista de la investigación en IRS y la Psicometría se

pueden proponer nuevas investigaciones y análisis que aporten información adicional que permita avanzar y hacer más preciso el estudio de sus propiedades, los resultados aquí presentados representan un primer paso culminado sobre el que seguir avanzando.

DISCUSIÓN

En la actualidad, el interés de contar con instrumentos psicométricamente robustos para la evaluación de los resultados obtenidos en salud desde la perspectiva de los pacientes, es ampliamente reconocido en nuestro país y en la UE (Bastida et al., 2010; EMA, 2014; EUnetHTA, 2013; NICE, 2013).

Los distintos planes de intervención sanitaria, vigentes para el Gobierno de España y en distintas CCAA, hacen hincapié en la necesidad de abordar la salud de la población desde un modelo biopsicosocial para maximizar el alcance de las acciones propuestas y optimizar los recursos disponibles (Generalitat de Catalunya, 2012; Ministerio de Sanidad, 2012; Ollero, Sanz, Padilla, y (coords), 2012; Velasco, 2010). Como se reconoce en la justificación de los citados planes sanitarios e intervenciones, la creciente prevalencia de la cronicidad de las patologías que afectan a la población precisan de mantener un enfoque integrador donde los pacientes y la sociedad en su conjunto reciben la atención sanitaria de forma activa, como agentes fundamentales en el éxito de la intervención y también en la evaluación de la misma (Grupo de Trabajo de la Sociedad Española de Medicina Interna (SEMI) y la Sociedad Española de Medicina Familiar y Comunitaria (semFYC), 2011). Así, muchos indicadores de la calidad de los servicios pasan a estar mediados por la percepción del paciente y la sociedad. Sirvan como ejemplo las recomendaciones recogidas por el Ministerio de Sanidad de nuestro país en su Estrategia para el Abordaje de la Cronicidad en el Sistema Nacional de Salud (Ministerio de Sanidad, 2012) en la que se recomienda la incorporación de una estrategia evaluativa con una triple visión basada, entre otras, en la IRS. De esta forma, se incluye la mortalidad y morbilidad asociada a la enfermedad pero también la calidad de vida de los pacientes, los años de vida ajustados por calidad, la esperanza de vida en

buena salud y la satisfacción de los pacientes. Además, se incide en la necesidad de promover la innovación tecnológica y su aplicación desde una racionalización de los recursos basados en la evidencia para lo cual resulta fundamental disponer de instrumentos de evaluación fiables y válidos.

En esta tesis doctoral, se incluyen tres publicaciones científicas que diseminan los resultados de otros tantos proyectos de investigación cuyo objetivo compartido fue el de poner a disposición de los profesionales sanitarios y estudiosos de los resultados en salud, nuevas herramientas válidas para evaluación de constructos de interés no sólo en la IRS sino para la práctica clínica diaria. A continuación, se discutirán las principales aportaciones y limitaciones de cada una esos trabajos. Junto a esto, se introducirán brevemente nuevos proyectos de estudios psicométricos de estas medidas PROs aquí publicadas, que se han venido desarrollando desde entonces. Se debe hacer constar en este punto que se comparte la necesidad de considerar los instrumentos de medición como objetos sujetos a constante evolución dialéctica con el objeto de contar cada vez con instrumentos más refinados y mejores (American Educational Research Association et al., 2004; FDA, 2009). En consecuencia, como se verá en la discusión de cada publicación, el estudio de sus propiedades no se ha limitado a lo presentado en la publicación, sino que ha ido progresando con la vista puesta en la continua revisión y mejora de los mismos.

ARTÍCULO 1: Desarrollo y validación cuestionario SAT-Q.

La investigación para el desarrollo del cuestionario SAT-Q ha permitido tener una medida que posibilita la evaluación de los niveles de la satisfacción del paciente de mayor interés en un mismo instrumento. Se trata un cuestionario breve, de fácil cumplimentación y corrección con tan sólo 18 ítems tipo Likert que cubren las

dimensiones del modelo teórico subyacente de mayor interés (Shikiar y Rentz, 2004): 2 genéricas (satisfacción con el tratamiento farmacológico y con la atención sanitaria), y 4 específicas (efectividad, efectos adversos, comodidad/conveniencia y por último, olvidos en la tomas). Las propiedades psicométricas básicas del cuestionario SAT-Q son al menos aceptables, en términos de consistencia interna y estabilidad de las puntuaciones con la excepción del dominio de eventos adversos que probablemente debido a la presencia de alguna molestia en el intervalo temporal entre la visita basal y el retest haya podido influir y por tanto deberá ser testado nuevamente en nuevos trabajos. Los resultados del análisis factorial y la asociación moderada-alta encontrada entre la satisfacción con el tratamiento farmacológico y las distintas dimensiones específicas (con la excepción de olvidos) así como la asociación más débil respecto a la satisfacción con la atención sanitaria, permite apoyar la diferenciación de niveles de satisfacción definidos por el Modelo Jerárquico de la satisfacción (Shikiar y Rentz, 2004). Por otra parte, la correlación positiva de las puntuaciones sumarias del SF-12, parecen confirmar el hecho de que los pacientes con un estado de salud percibido más alto tienden a estar más satisfechos con los servicios sanitarios (Xiao y Barber, 2008) o en relación a tratamientos específicos (Abetz et al., 2005; Campbell et al., 2008; Pleil et al., 2005).

Entre las limitaciones más importantes, cabe señalar el efecto techo encontrado en las puntuaciones. Aunque parece ser una constante en los estudios que evalúan este concepto (Condes et al., 2005; Feurer, Russell, y Pinson, 2007; Ruiz et al., 2008), hace necesaria la puesta en marcha de nuevos estudios que evalúen su validez convergente frente a otros instrumentos de satisfacción. También interesa comprobar su asociación y capacidad de discriminación respecto a constructos relacionados como el cumplimiento del paciente, medido, por ejemplo, con el test de Morisky-Green (Morisky, Green, y

Levine, 1986) o la versión actual de 8 ítems validada en distintos contextos (Kim et al., 2014; Reynolds et al., 2014; Yang et al., 2014).

Trabajos posteriores de análisis de las propiedades psicométricas del cuestionario SAT-Q

En respuesta a las limitaciones comentadas, en trabajos completados desde la publicación (Prieto et al., 2010a; Prieto et al., 2010b), se analizó, en primer lugar, la asociación de las dimensiones genéricas del SAT-Q con el control de la enfermedad y la adherencia al tratamiento en pacientes con hipertensión arterial de riesgo cardiovascular alto o muy alto (n=1.517), atendidos en centros de AP distribuidos por todo el país. A partir del análisis estadístico, se encontraron valores significativamente más bajos, indicativos de peor satisfacción, entre aquellos pacientes cuya presión arterial no estaba controlada de acuerdo a los criterios clínicos y entre los que no cumplían adecuadamente con la prescripción médica. Si bien no se pudo establecer una relación causal por el diseño del estudio, sí se confirmó la asociación entre respecto a esas variables, sin duda, relevantes desde el punto de vista de los profesionales sanitarios.

Además de lo anterior, se llevó a cabo un estudio multicéntrico nacional en el que se estudió, a través del SAT-Q, la satisfacción en pacientes portadores de un trasplante renal (n=206) que seguían un régimen inmunosupresor, de acuerdo a la práctica clínica de cada centro especializado participante (hospitales de referencia). En este estudio, recientemente publicado, aunque también se observó un considerable efecto techo en las puntuaciones del cuestionario, se encontró una relación positiva significativa y leve-moderada, entre la satisfacción global con el tratamiento inmunosupresor y la edad y el estado de salud de los pacientes (Ortega et al., 2013). Además, los pacientes que no habían presentado valores anómalos del tratamiento en sangre (indicativo de una mayor

adherencia al tratamiento) estaban más satisfechos que los que sí habían registrado valores anómalos y tenían una evaluación clínica más favorable. También, los pacientes con 1-2 tomas del medicamento tenían unos niveles de satisfacción más altos que los que tomaban 3 ó más. El nivel de funcionamiento del injerto (muy bueno-bueno vs regular-malo) se asoció igualmente con la satisfacción. En los modelos de regresión multivariados se confirmó que el estado de salud de los pacientes, el número de tomas del medicamento y la evaluación clínica del injerto eran variables asociadas a las puntuaciones del SAT-Q.

Trabajos más recientes con el SAT-Q (Campillo-Álvarez, Rodríguez-Aguilella, Cuervo, Castejón, y Rebollo, 2013) se han centrado en el análisis de la estructura interna y la reducción del mismo desde nuevas aproximaciones psicométricas, como el modelo de Crédito Parcial de Rasch (Masters, 1982). Como resultado, se obtuvo una medida más breve, 10 ítems, claramente unidimensional y centrada únicamente en la evaluación de la satisfacción del paciente crónico con el tratamiento farmacológico. Finalmente, desde la Teoría de Respuesta al Ítem, se ha abordado también el estudio del comportamiento diferencial de los ítems del pool de ítems del cuestionario en función de la patología crónica de los pacientes. A partir de este estudio, se identificaron diferencias entre pacientes con diabetes mellitus y pacientes con hipertensión arterial en ítems relacionados con eventos adversos, olvidos en la medicación y conveniencia de la administración (Campillo-Álvarez, Rodríguez-Aguilella, Castejón, Cuervo, y Rebollo, 2013). Estos análisis permiten continuar avanzando en la mejora del conjunto de ítems del SAT-Q y el ajuste del mismo de acuerdo a las patologías a analizar.

ARTÍCULO 2: Adaptación cultural y validación del cuestionario Tool.

Respecto al cuestionario Tool, además de completar la adaptación lingüística y de contenido, en el análisis de sus propiedades de medición se ha conseguido comprobar que la versión española es, en primer lugar, aplicable o factible desde el punto de vista de la práctica clínica habitual de profesionales en salud mental, debido a su reducido número de ítems, su alta tasa de respuestas completas y el tiempo de cumplimentación reducido. Junto a lo anterior, el cuestionario TOOL tiene una estructura unidimensional, que hace referencia a la intensidad del efecto de los tratamientos antipsicóticos y sus posibles efectos adversos sobre la CVRS de los pacientes, una variable de interés para las agencias de evaluación de medicamentos (NICE, 2002) y también para los clínicos por sus implicaciones en la adherencia a los tratamientos (Perkins et al., 2008). La consistencia interna y su estabilidad en las puntuaciones indican que se trata de un cuestionario fiable.

En términos de otras evidencias que apoyan la validez del instrumento, las puntuaciones del cuestionario TOOL se relacionaron de forma moderada-alta y significativa, con las escalas de utilidad más empleadas en la actualidad (EQ-5D y SF-6D), al igual que se había comprobado previamente con otras medidas específicas (Hayhurst, Palmer, Abbott, Johnson, y Scott, 2006; Konig, Roick, y Angermeyer, 2007). De igual forma, mostró una relación moderada significativa y positiva frente a cuestionarios específicos como el MADRS. Por último, la validez de criterio del cuestionario TOOL se contrastó mediante la capacidad del cuestionario para distinguir según el grado de severidad del trastorno según juicio clínico y con la relación con un cuestionario específico como el UKU de efectos secundarios (Briggs et al., 2008). Así, el instrumento es válido para la

medición de la afectación de la CVRS en los pacientes con tratamiento farmacológico antipsicótico.

A pesar de la muestra incluida en el estudio (un total de 238 pacientes esquizofrénicos o con trastorno bipolar) de 60 centros de Salud Mental, el estudio presenta una serie de limitaciones. Una que merece ser destacada es debida, precisamente, a las características de la muestra incluida. Se trata de pacientes estables en la sintomatología y por tanto, sería legítimo cuestionarse si a través de esta medida se podría obtener información fiable sobre la afectación de los tratamientos en aquellos pacientes en una fase más aguda, florida o con una afectación mayor. Por tanto, nuevos análisis en esa población deberían ser llevados a cabo para poder considerarlo aplicable al conjunto de pacientes con cuadros psiquiátricos más intensos. Por otro lado, debido a la naturaleza transversal del diseño del estudio, no se pudo contrastar la capacidad del cuestionario Tool para detectar cambios clínicos relevantes en la situación de los pacientes (sensibilidad al cambio), ni hacer una estimación de las diferencias mínimas clínicamente relevantes en sus puntuaciones que resultaría de interés (FDA, 2009; Revicki, Hays, Cella, y Sloan, 2008).

Trabajos posteriores con el cuestionario Tool

Como se ha descrito en la publicación, el cuestionario Tool hace referencia a dominios de la CVRS que se pueden ver afectados por efectos adversos frecuentemente asociados al uso de antipsicóticos: estado emocional, limitaciones en actividades cotidianas, cansancio o debilidad, rigidez o temblores, inquietud, mareos o náuseas, disfunción sexual y el sobrepeso. A partir de la validación de esta medida específica, se llevó a cabo un trabajo de estimación de las utilidades o preferencias sociales asociadas a los estados de salud definidos en el Tool. A tal fin, se aplicó una aproximación estadística,

la función de multi-atributo de utilidad, que ya había sido contrastada con éxito por los creadores de una de las medidas de utilidad más importantes en la actualidad en EEUU y Canadá, el Health Utility Index (HUI) (Feeny et al., 2002; Torrance, Furlong, Feeny, y Boyle, 1995; Torrance et al., 1996). Como resultado, se publicó el algoritmo de estimación de los valores de utilidad en España para el TooL (Montejo et al., 2011) y se analizaron a su vez, las propiedades psicométricas de esta medida cuyo rango teórico de puntuaciones oscila entre 0 (peor estado de salud posible) y 1 (mejor estado de salud posible). Este trabajo representó el primer estudio realizado en nuestro país para la estimación de utilidades directamente con pacientes esquizofrénicos y con trastorno bipolar.

Además de lo anterior, la estimación de las utilidades en estos pacientes brindó la posibilidad de comparar la concordancia en la valoración de los estados de salud que se obtiene con la aplicación de medidas genéricas (con valores de utilidad obtenidos de la población general) y específicas de utilidad (en este caso con la valoración directa de los pacientes). Se aplicó para ello el método de Bland-Altman y se calculó el CCI entre las medidas comparadas. Los resultados de esta comparación fueron claros (Cuervo, Rebollo, Mauriño, Cordero, y Castejón, 2011): apenas se producía concordancia entre las distintas medidas de utilidad, incluso entre las propias medidas genéricas. Además, los valores de utilidad obtenidos en población general, empleando instrumentos genéricos, subestimaban de forma importante el impacto de los síntomas descritos sobre los pacientes. La magnitud de las diferencias era muy importante teniendo en cuenta el rango teórico de las escalas. Las implicaciones de nuestros hallazgos se suman a las múltiples voces críticas con el empleo de instrumentos genéricos para el cálculo de los años de vida ajustados por calidad. Como alternativa, se está proponiendo establecer nuevas aproximaciones que pretenden ser más robustas para la medición de este

parámetro tan relevante en la evaluación económica de las intervenciones sanitarias o incluso un cambio en el parámetro de efectividad (Beresniak y Lamure, 2014; Schaeffer, 2013).

ARTÍCULO 3. Adaptación cultural y validación de la escala CUDOS.

Como se ha comentado al inicio de este apartado, la demanda de herramientas útiles es mayor aun si cabe entre los profesionales de AP ya que, en los planes de intervención sanitaria recientes, desempeñan un rol especial por su situación estratégica como puerta de entrada en el circuito sanitario y como los médicos más cercanos al paciente. Debido a esto, tienen mucha importancia en la prevención de muchos problemas en salud y cuando esto no es posible, en su rápida detención y tratamiento de las complicaciones asociadas. Las implicaciones positivas de su labor son evidentes en términos de un mejor pronóstico para los pacientes y un menor consumo de recursos sanitarios para los gestores. De ahí que sea fundamental que cuenten con instrumentos válidos en la evaluación de las patologías que, a nivel epidemiológico, clínico y social, son más importantes, como es el caso del TDM (Grupo de Trabajo sobre el Manejo de la Depresión Mayor en el Adulto, 2008; National Collaborating Centre for Mental Health, 2009).

El análisis psicométrico al que se sometió la versión adaptada al castellano de la escala CUDOS mostró que el cuestionario presenta una buena factibilidad, en términos de efectos suelo/techo y porcentaje de respuesta. El tiempo de cumplimentación obtenido fue ligeramente mayor al publicado en la versión original (4,5 minutos frente los 3 minutos), pero posibilita su uso en práctica clínica (Zimmerman et al., 2008; Zimmerman y McGlinchey, 2008). Por otro lado, se obtuvo una consistencia interna similar a la publicada en EEUU (Zimmerman et al., 2008). En cuanto a la validez de

criterio, las puntuaciones de la escala CUDOS correlacionaron moderadamente con la escala de Hamilton, aunque la correlación fue menor que la publicada anteriormente (Zimmerman et al., 2008). Además, permitió distinguir entre tres categorías agregadas: no deprimidos/depresión ligera-moderada/depresión grave-muy grave. En el mismo sentido, discriminaba entre los niveles de gravedad definidos por los médicos y los pacientes (con una mayor capacidad para distinguir datos reportados por el paciente). Respecto a la validez convergente, la escala se asoció, como cabía esperar, con todas las puntuaciones sumarias del cuestionario SF-36, siendo más intensa la correlación con la componente mental y con la escala de salud mental. Finalmente, se observó correlación con la escala de funcionalidad SOFAS, obteniéndose puntuaciones significativamente menores en la escala CUDOS para el grupo con buena funcionalidad.

Así, parece legítimo proponer la escala CUDOS para la evaluación del impacto en la CVRS y la intensidad sintomática en pacientes TDM en AP para recoger información desde el punto de vista de los pacientes que complementa la obtenida por los clínicos a través de las medidas como la HAMD-17, normalmente empleada por psiquiatras y psicólogos pero que son más extrañas a los médicos de AP (Bobes et al., 2003; Hamilton, 1960; Lobo et al., 2002; Ramos-Brieva y Cordero-Villafafila, 1988; Romera, Pérez, Menchón, Polavieja, y Gilaberte, 2011).

Trabajos posteriores con la escala CUDOS

Aunque se trata del trabajo más reciente, se ha podido llevar a cabo un análisis centrado en el estudio de la capacidad diagnóstica y predictiva de la escala CUDOS frente al criterio de referencia para el diagnóstico de TDM, como es la escala HAMD-17 ya mencionada (Agüera et al., 2012). El objetivo fue aportar información útil a los médicos para la interpretación rápida de la puntuación de la escala y de paso, comparar esta

información con la publicada por los autores de la versión original (Zimmerman et al., 2004; Zimmerman et al., 2008; Zimmerman et al., 2012a; Zimmerman et al., 2012b). Como resultado, la capacidad de la escala CUDOS para predecir la remisión (HAMD-17) fue adecuada (área bajo la curva= 0,837; IC95%= 0,740-0,934; $p < 0,001$). Con los datos en España (Agüera et al., 2012), una puntuación de 24 puntos en la escala CUDOS se consideró un punto de corte útil en práctica clínica para detectar casos con respuesta subóptima al tratamiento con síntomas de TDM clínicamente significativos (sensibilidad: 81% y especificidad: 77%). Esta puntuación fue más alta que la publicada en la versión original donde el punto de corte establecido era de 20 puntos (Zimmerman et al., 2004). Posibles explicaciones a estas diferencias pueden situarse en las características de la muestra: los pacientes TDM incluidos en EEUU tenían unas puntuaciones medias en la HAMD-17 significativamente inferiores a los valores medios de la muestra en nuestro país (tamaño del efecto de la diferencia de 0.67 puntos). Así, según el baremo aceptado de la HAMD-17, en EEUU la mayoría de los pacientes presentaban una afectación leve mientras que en la muestra nacional, el 43% presentaban síntomas depresión moderados o graves y el 19,3%, síntomas muy graves. No obstante, estos resultados deberán ser analizados en futuras investigaciones que incluya un seguimiento prospectivo de los pacientes TDM con distinta gravedad con el fin de precisar mejor los baremos.

CONCLUSIONES

En esta tesis doctoral se han presentado los proyectos de investigación diseñados e implementados con el objetivo común de hacer posible la medición de variables de interés en el ámbito de la investigación socio-sanitaria y la práctica clínica de los profesionales de la salud. Se podrían destacar las siguientes conclusiones:

- I.** Desde el modelo biopsicosocial de la salud cada vez más implantado en nuestro país, las opiniones de los pacientes sobre la atención recibida y los resultados en salud son variables de interés en la evaluación de la calidad de los servicios y la eficiencia comparada de las intervenciones específicas.
- II.** La información relativa a la experiencia subjetiva en salud de los pacientes puede ser recogida de forma sistemática y eficiente a través de instrumentos previamente validados, siguiendo los principios y métodos que han sido desarrollados en la Psicometría.
- III.** La aplicación de instrumentos válidos para la medición de la CVRS y el impacto de la enfermedad o las intervenciones es también de utilidad para los profesionales sanitarios en su práctica clínica habitual debido al valor predictivo y discriminativo de estos instrumentos. Estos instrumentos aportan información, normalmente no recogida en la historia clínica, pero que realmente son el objetivo de la práctica clínica: la restauración de niveles de CVRS aceptables para el sujeto paciente.
- IV.** El cuestionario SAT-Q, desarrollado para la evaluación de la satisfacción de los pacientes con enfermedad crónica, tiene unas propiedades psicométricas básicas adecuadas permitiendo obtener información válida en AP sobre dos dominios de interés: satisfacción general de los pacientes con la atención recibida y satisfacción

con la efectividad del tratamiento, los efectos adversos, la comodidad y conveniencia del mismo. Estos aspectos se han asociado además a la adherencia al tratamiento farmacológico y la CVRS de los pacientes.

- V.** El cuestionario breve TooL ha sido adaptado con éxito a nuestro país y sus propiedades psicométricas permiten hablar de una estructura unidimensional que evalúa el impacto de los tratamientos antipsicóticos sobre la CVRS de los pacientes. Además distingue entre distintos niveles de gravedad clínica y se asocia a otras medidas de referencia en el estudio de efectos secundarios de los tratamientos farmacológicos y genéricas de CVRS de los pacientes.
- VI.** La escala CUDOS de depresión, es una medida de utilidad clínica en AP para la evaluación de la CVRS de los pacientes con TDM, que distingue entre los pacientes que se encuentran en remisión sintomática o con afectación leve frente a los que presentan una situación moderada o grave.
- VII.** Las características de los instrumentos presentados posibilitan que sean aplicados en condiciones de práctica clínica real en AP (SAT-Q y CUDOS) y en salud mental (TooL), estando a disposición de los profesionales sanitarios de nuestro país.
- VIII.** Aunque los tres instrumentos difieren en las áreas evaluadas y la población objetivo, mantienen de forma inequívoca un nexo común en cuanto a los principios y procedimientos estadísticos implicados en su desarrollo o su adaptación cultural y validación.
- IX.** Todos los instrumentos presentados han seguido un continuo proceso de estudio que está dando sus frutos en términos de nuevos avances en el conocimiento de sus cualidades psicométricas y su refinamiento, siempre con la vista puesta en obtener cada vez mejores herramientas de medición de las PROs.

INFORME FACTOR IMPACTO DE LAS PUBLICACIONES PRESENTADAS EN EL PROYECTO DE TESIS DOCTORAL

Las revistas donde se publicaron las investigaciones que constituyen el contenido de la tesis doctoral tienen factor de impacto científico de acuerdo a las clasificaciones de las revistas científicas realizadas por:

- Journal Citation Report realizada por la editorial Thomson Reuters (portal **Web of Knowledge**): http://wokinfo.com/products_tools/analytical/jcr/las
- SCImago Journal y Country Rank (SJR) desarrollado por SCIMAGO Lab (portal **SCOPUS**): <http://www.scimagojr.com/index.php>

Indicadores considerados:

- Factor Impacto de la Revista (JFI^{TM}) de la editorial Thomson Reuters (año de referencia 2012).
- *SCImago Journal Rank (SJR)*: indicador de la influencia científica media de un artículo en la revista a través de la importancia global de la discusión generada a nivel científico global (años de referencia presentados: 2005-2012).
- *Total cites*: refleja la evolución del número medio de citas recibidas por la revista en los 3 años anteriores (años de referencia presentados: 2005-2012).
- *Cites per doc*: mide el impacto científico de la revista en los 2 últimos años y empleando la misma fórmula que en el JFI^{TM} de Thomson Reuters (años de referencia presentados: 2005-2012).

Atención Primaria: ISSN 0212-6567

- *JFITM* en 2012: **0,96**
- Indexada en: Index Medicus/Medline, Excerpta Medica/EMBASE, IBECS, IME, SCOPUS, Medes, Science Citation Index Expanded.

Tabla 2. Factor Impacto Publicación en Atención Primaria.

Indicadores	2005	2006	2007	2008	2009	2010	2011	2012
SJR	0,196	0,220	0,239	0,229	0,241	0,207	0,246	0,186
Total Cites (3years)	391	469	335	331	242	204	243	226
Cites por Doc. (2years)	0,71	0,68	0,52	0,58	0,51	0,68	0,85	0,96

Annals of General Psychiatry: ISSN 1744859X

- *JFITM* en 2012: **2,12**
- Indexada en: Cases Database, Citebase, DOAJ, Embase, Google Scholar, Index Copernicus, OAIster, PsycINFO, PubMed, PubMed Central, Science Citation Index Expanded, SCImago, Scirus, Scopus, SOCOLAR, Zetoc.

Tabla 3. Factor de Impacto Publicación en Annals of General Psychiatry.

Indicadores	2005	2006	2007	2008	2009	2010	2011	2012
SJR	0,234	0,358	0,491	0,652	0,696	0,639	0,595	0,594
Total Cites (3years)	27	80	87	152	180	190	206	203
Cites / Doc. (2years)	0,75	1,81	1,71	2,20	2,07	2,15	2,23	2,12

Actas Españolas de Psiquiatría: ISSN 1139-9287

- *JFITM* en 2012: **0,79**
- Indexada en: Index Medicus/Medline, EMBASE, SCOPUS, Medes, Science Citation Index Expanded

Tabla 4. Factor de Impacto Publicación Actas Españolas de Psiquiatría.

Indicadores	2005	2006	2007	2008	2009	2010	2011	2012
SJR	0,210	0,165	0,224	0,218	0,224	0,239	0,321	0,251
Total Cites (3years)	88	72	93	108	111	140	142	134
Cites per doc. (2years)	0,46	0,34	0,47	0,60	0,54	0,80	0,79	0,79

REFERENCIAS BIBLIOGRÁFICAS

Abad, F., Garrido, J., Olea, J., y Ponsoda, V. (2006). *Introducción a la Psicometría. Teoría Clásica de los Tests y Teoría de la Respuesta al Ítem*. Madrid: Universidad Autónoma de Madrid.

Abetz, L., Coombs, J. H., Keininger, D. L., Earle, C. C., Wade, C., Bury-Maynard, D. , Copley-Merriman, K., y Hsu, M. A. (2005). Development of the cancer therapy satisfaction questionnaire: item generation and content validity testing. *Value.Health*, 8(1), S41-S53.

Academy of Managed Care Pharmacy (AMCP). (2005). *The AMCP Format for Formulary Submissions, version 2.1 A Format for Submissions of Clinical and Economic Data in Support of Formulary Consideration by Health Care Systems in United States*. Estados Unidos: Academy of Managed Care Pharmacy (AMCP).

Acquadro, C., Berzon, R., Dubois, D., Leidy, N. K., Marquis, P., Revicki, D., y Rothman, M. (2003). Incorporating the patient's perspective into drug development and communication: an ad hoc task force report of the Patient-Reported Outcomes (PRO) Harmonization Group meeting at the Food and Drug Administration, February 16, 2001. *Value Health.*, 6(5), 522-531.

Agüera, L., Montón, C., Medina, E., Mauriño, J., Díaz, H., y Cuervo, J. (2012, Noviembre). *Assessing Depressive Symptoms in Primary Care: Psychometrical Properties of the Spanish Version of the Clinically Useful Depression Outcome*

Scale. Póster presentado en el 15th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Berlín.

Agütera, L., Failde, I., Cervilla, J. A., Díaz-Fernández, P., y Mico, J. A. (2010). Medically unexplained pain complaints are associated with underlying unrecognized mood disorders in primary care. *BMC.Fam.Pract.*, *11*(17), 17.

American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. (2004). *Standards for educational and psychological testing* (2nd ed.). Washington, D.C.: American Educational Research Association.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. Washington D.C.: American Psychiatric Association.

Angst, J., Gamma, A., Gastpar, M., Lepine, J. P., Mendlewicz, J., y Tylee, A. (2002). Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci.*, *252*(5), 201-209.

Aragonés, E., Piñol, J. L., y Labad, A. (2006). The overdiagnosis of depression in non-depressed patients in primary care. *Fam.Pract.*, *23*(3), 363-368.

Aragonés, E., Piñol, J. L., y Labad, A. (2009). Comorbilidad de la depresión mayor con otros trastornos mentales comunes en pacientes en Atención Primaria. [Comorbidity of major depression with other common mental disorders in primary care patients]. *Aten Primaria.*, *41*(10), 545-551.

- Aragonés, E., Piñol, J. L., Labad, A., Folch, S., y Melich, N. (2004). Detection and management of depressive disorders in primary care in Spain. *Int.J Psychiatry Med.*, 34(4), 331-343.
- Aragonés, E., Piñol, J. L., Labad, A., Masdeu, R. M., Pino, M., y Cervera, J. (2004). Prevalence and determinants of depressive disorders in primary care practice in Spain. *Int.J Psychiatry Med.*, 34(1), 21-35.
- Asadi-Lari, M., Tamburini, M., y Gray, D. (2004). Patients' needs, satisfaction, and health related quality of life: towards a comprehensive model. *Health Qual.Life Outcomes*, 2, 32. Recuperado de PM:15225377
- Ascher-Svanum, H., Zhu, B., Faries, D., Lacro, J. P., y Dolder, C. R. (2006). A prospective study of risk factors for nonadherence with antipsychotic medication in the treatment of schizophrenia. *J Clin Psychiatry.*, 67(7), 1114-1123.
- Atkinson, M. J., Kumar, R., Cappelleri, J. C., y Hass, S. L. (2005). Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. *Value.Health*, 8 Suppl 1, S9-S24. Recuperado de PM:16336491
- Atkinson, M. J., Sinha, A., Hass, S. L., Colman, S. S., Kumar, R. N., Brod, M. y Rowland, C. R. (2004). Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. *Health Qual.Life Outcomes.*, 2(1), 12.

- Atkinson, M. J., Stewart, W. C., Fain, J. M., Stewart, J. A., Dhawan, R., Mozaffari, E., y Lohs, J. (2003). A new measure of patient satisfaction with ocular hypotensive medications: The Treatment Satisfaction Survey for Intraocular Pressure (TSS-IOP). *Health Qual.Life Outcomes.*, 1(1), 67. Recuperado de PM:14617372
- Auquier, P., Pernoud, N., Bruder, N., Simeoni, M. C., Auffray, J. P., Colavolpe, C., Francois, G., Gouin, F., Manelli, J. C., Martín, C., Sapin, C., y Blache, J. L. (2005). Development and validation of a perioperative satisfaction questionnaire. *Anesthesiology*, 102(6), 1116-1123.
- Autonell, J., Vila, F., Pinto-Meza, A., Vilagut, G., Codony, M., Almansa, J., Muñoz P. E., Torres, J.V., Alonso, J. y Haro, J. M. (2007). One year prevalence of mental disorders comorbidity and associated socio-demographic risk factors in the general population of Spain. Results of the ESEMeD-Spain study. *Actas Esp Psiquiatr.*, 35 Suppl 2:4-11., 4-11.
- Badía, X., y Alonso, J. (2007). *La medida de la Salud: Guía de escalas de medición en español*. Colección Medicina Eficiente (4ª Edición ed.). Barcelona: Fundación Lilly, Editec.
- Badía, X., y Lizán, L. (2002). Reflexiones sobre la investigación de resultados en salud. *Aten Primaria*, 30(6), 388-391.
- Badía, X. (2005). La satisfacción con el tratamiento. *Med Clin.(Barc.)*, 125(3), 98-99.
- Badía, X., Alonso, J., Brosa, M., y Lock, P. (1994). Reliability of the Spanish version of the Nottingham Health Profile in patients with stable end-stage renal disease. *Soc Sci.Med.*, 38(1), 153-158.

- Badía, X., Roset M, Montserrat S, Herdman M, y Segura A. (1999). La versión española del EuroQol: descripción y aplicaciones. *Medicina Clínica (Barc)*, 112(1), 79-86.
- Badía, X., Roset, M., Herdman, M., y Kind, P. (2001). A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis.Making.*, 21(1), 7-16.
- Basra, M. K., Fenech, R., Gatt, R. M., Salek, M. S., y Finlay, A. Y. (2008). The Dermatology Life Quality Index 1994-2007: a comprehensive review of validation data and clinical results. *Br.J.Dermatol.*, 159(5), 997-1035.
Recuperado de PM:18795920
- Bastida, J., Oliva, J., Antoñanzas, F., García-Altés, A., Gisbert, R., Mar, J., y Puig-Junoy, J. (2010). Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias. *Gaceta Sanitaria*, 24(2), 154-170.
- Beattie, P. F., Nelson, R. M., y Lis, A. (2007). Spanish-language version of the MedRisk Instrument for Measuring Patient Satisfaction With Physical Therapy Care (MRPS): preliminary validation. *Phys.Ther.*, 87(6), 793-800.
- Beresniak, A., y Lamure, M. (2014). ECHOUTCOME: European Consortium in Healthcare Outcomes and Cost-Benefit research. Lyon. ECHOUTCOME.
Recuperado el 14 de Abril, 2014, de: <http://www.echoutcome.eu/>.
- Bernal, M., Haro, J. M., Bernert, S., Brugha, T., de, G. R., Bruffaerts, R. Lepine, J. R., de Girolamo, G., Vilagut, G., Gasquet, I, Torres, J. V., Kovess, V., Heider, D.,

- Neeleman, J., Kessler, R y Alonso, J. (2007). Risk factors for suicidality in Europe: results from the ESEMED study. *J Affect.Disord.*, 101(1-3), 27-34.
- Bobes, J., Bulbena, A., Luque, A., Dal-Re, R., Ballesteros, J., y Ibarra, N. (2003). 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 [A comparative psychometric study of the Spanish versions with 6, 17, and 21 items of the Hamilton Depression Rating Scale]. *Med Clin (Barc.)*, 120(18), 693-700.
- Brazier, J., Rowen, D., Mavranzouli, I., Tsuchiya, A., Young, T., Yang, Y., Barkham, M. e Ibbotson, R. (2012). Developing and testing methods for deriving preference-based measures of health from condition-specific measures (and other patient-based measures of outcome). *Health Technol.Assess.*, 16(32), 1-114.
- Briggs, A., Wild, D., Lees, M., Reaney, M., Dursun, S., Parry, D., y Mukherjee, J. (2008). Impact of schizophrenia and schizophrenia treatment-related adverse events on quality of life: direct utility elicitation. *Health Qual Life Outcomes.*, 6(105), 105.
- Calvert, M., Blazeby, J., Altman, D. G., Revicki, D. A., Moher, D., y Brundage, M. D. (2013). Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA.*, 309(8), 814-822.
- Campbell, U. B., Stang, P., y Barron, R. (2008). Survey assessment of continuation of and satisfaction with pharmacological treatment for urinary incontinence. *Value.Health.*, 11(4), 726-732.

- Campillo-Álvarez, A., Rodríguez-Aguilella, A., Castejón, N., Cuervo, J., y Rebollo, P. (2013, Noviembre). *Differential Item Functioning in a Satisfaction Item Bank*. Póster presentado en el 16th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Dublín.
- Campillo-Álvarez, A., Rodríguez-Aguilella, A., Cuervo, J., Castejón, N., y Rebollo, P. (2013, Mayo). *SAT-Q Questionnaire: Application of Rasch measurement for selecting items and improving the assessment of patient satisfaction*. Póster presentado en el 18th Annual International Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Nueva Orleans.
- Canadian Agency for Drugs and Technologies in Health. (2006). *Guidelines for the economic evaluation of health technologies: Canada* (3rd ed.). Ottawa: Canadian Agency for Drugs and Technologies in Health.
- Carr-Hill, R. A. (1992). The measurement of patient satisfaction. *J Public Health Med.*, *14*(3), 236-249.
- Cella, D. (2000). Assessment methods for quality of life in cancer patients: the FACIT measurement system. *International Journal for Pharmaceutical Medicine*, *14*, 78-81.
- Cella, D., Hahn, E. A., y Dineen, K. (2002). Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res.*, *11*(3), 207-221.

- Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., Silverman, M., Yellen, S. B., Winicour, P., y Brannon, J. (1993). The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol.*, *11*(3), 570-579.
- Cenedese, E., Speich, R., Dorschner, L., Ulrich, S., Maggiorini, M., Jenni, R., y Fischler, M. (2006). Measurement of quality of life in pulmonary hypertension and its significance. *Eur Respir.J.*, *28*(4), 808-815. doi:09031936.06.00130405 [pii];10.1183/09031936.06.00130405 [doi]. Recuperado de PM:16707511
- Chamlin, S. L., Cella, D., Frieden, I. J., Williams, M. L., Mancini, A. J., Lai, J. S., y Chren, M. M. (2005). Development of the Childhood Atopic Dermatitis Impact Scale: initial validation of a quality-of-life measure for young children with atopic dermatitis and their families. *J Invest Dermatol.*, *125*(6), 1106-1111.
- Chamlin, S. L., Lai, J. S., Cella, D., Frieden, I. J., Williams, M. L., Mancini, A. J., y Chren, M. M. (2007). Childhood Atopic Dermatitis Impact Scale: reliability, discriminative and concurrent validity, and responsiveness. *Arch Dermatol.*, *143*(6), 768-772.
- Charlson, F. J., Stapelberg, N. J., Baxter, A. J., y Whiteford, H. A. (2011). Should global burden of disease estimates include depression as a risk factor for coronary heart disease? *BMC.Med.*, *3*(9), 47. doi: 10.1186/1741-7015-9-47.
- Chouinard, G., y Margolese, H. C. (2005). Manual for the Extrapiramidal Symptom Rating Scale (ESRS). *Schizophr.Res.*, *76*(2-3), 247-265.

- Condes, E., Aguirrebengoa, K., Dalmau, D., Estrada, J. M., Force, L., Gorgolas, M. Badía, X., y Podzamczar, D. (2005). Validación del cuestionario de satisfacción con el tratamiento antirretroviral: Cuestionario CESTA. *Enferm.Infecc.Microbiol.Clin.*, 23(10), 586-592.
- Cramer, J. A., Roy, A., Burrell, A., Fairchild, C. J., Fuldeore, M. J., Ollendorf, D. A., y Wong, P. K. (2008). Medication compliance and persistence: terminology and definitions. *Value Health.*, 11(1), 44-47.
- Crow, R., Gage, H., Hampson, S., Hart, J., Kimber, A., Storey, L. y Thomas, H. (2002). The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature. *Health Technol.Assess.*, 6(32), 1-244.
- Cuervo, J., Rebollo, P., Mauriño, J., Cordero, L., y Castejón, I. (2011, Noviembre). *Estimating Utilities in Schizophrenic and Bipolar Patients from Disease-Specific or Generic Instruments Assessing Patients' Health States: Where Are the Differences?* Póster presentado en el 14th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Madrid.
- Department of Economics and Public Health Assessment. (2012). *Choices in Methods for Economic Evaluation: A methodological guide*. Saint-Denis La Plaine: Haute Autorité de Santé.
- European Medicines Agency. *Health-technology-assessment bodies*. Recuperado el 4 de Febrero, 2014, de:
http://www.ema.europa.eu/ema/index.jsp?curl=pages/partners_and_networks/general/general_content_000476.jspymid=WC0b01ac0580236a57.

European Network for Health Technology Assessment. (2013). *Guideline. Endpoints used for relative effectiveness assessment of pharmaceuticals: Health-related quality of life and utility measures*. Bélgica: EUnetHTA. Recuperado el 14 de Abril, 2014, de:

<https://5026.fedimbo.belgium.be/sites/5026.fedimbo.belgium.be/files/Health-related%20quality%20of%20life.pdf>

Feeny, D., Furlong, W., Torrance, G. W., Goldsmith, C. H., Zhu, Z., DePauw, S., Denton, M. y Boyle, M. (2002). Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care.*, 40(2), 113-128.

Fernández, A., Pinto-Meza, A., Bellón, J. A., Roura-Poch, P., Haro, J. M., Autonell, J., Palao, D. J., Peñarrubia, M. T., y Fernández, R. (2010). Is major depression adequately diagnosed and treated by general practitioners? Results from an epidemiological study. *Gen.Hosp.Psychiatry.*, 32(2), 201-209.

Feurer, I. D., Russell, R. T., y Pinson, C. W. (2007). Incorporating quality of life and patient satisfaction measures into a transplant outcomes assessment program: technical and practical considerations. *Prog.Transplant.*, 17(2), 121-128.

Food and Drug Administration, F. (2009). *Guidance for Industry. Patient-Reported Outcomes Measures: Use in Medical Product Development to Support Labeling Claims*. Silver Spring, MD: U.S. Department of Health and Human Services. Recuperado el 12 de Marzo, 2014, de:

<http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf>

- Gabilondo, A., Rojas-Farreras, S., Vilagut, G., Haro, J. M., Fernández, A., Pinto-Meza, A., y Alonso, J. (2010). Epidemiology of major depressive episode in a southern European country: results from the ESEMeD-Spain project. *J Affect.Disord.*, *120*(1-3), 76-85.
- García-Campayo, J., Ayuso-Mateos, J. L., Caballero, L., Romera, I., Aragonés, E., Rodríguez-Artalejo, F, Quail, D. y Gilaberte, I. (2008). Relationship of Somatic Symptoms With Depression Severity, Quality of Life, and Health Resources Utilization in Patients With Major Depressive Disorder Seeking Primary Health Care in Spain. *Prim.Care Companion.J Clin Psychiatry.*, *10*(5), 355-362.
- Generalitat de Catalunya, D. d. S. (2012). *Plan de Salud de Cataluña 2011-2015* (Departamento de Salud ed.). Barcelona: Departamento de Salud.
- Gentile, S. (2006). Long-term treatment with atypical antipsychotics and the risk of weight gain : a literature analysis. *Drug Saf.*, *29*(4), 303-319.
- Gharabawi, G. M., Bossie, C. A., Lasser, R. A., Turkoz, I., Rodriguez, S., y Chouinard, G. (2005). Abnormal Involuntary Movement Scale (AIMS) and Extrapyramidal Symptom Rating Scale (ESRS): cross-scale comparison in assessing tardive dyskinesia. *Schizophr.Res.*, *77*(2-3), 119-128.
- Grupo de Trabajo de la Sociedad Española de Medicina Interna (SEMI) y la Sociedad Española de Medicina Familiar y Comunitaria (semFYC). (2011). *Documento de Consenso Atención al Paciente con Enfermedades Crónicas*. Sevilla: MERGABLUM. Edición y Comunicación, SL. Recuperado el 15 de Abril, 2014, de:<http://www.fesemi.org/documentos/1370366301/publicaciones/consensos/conferencia-nacional-atencion-paciente-cronico.pdf>

Grupo de Trabajo sobre el Manejo de la Depresión Mayor en el Adulto. (2008). *Guía de Práctica Clínica sobre el Manejo de la Depresión Mayor en el Adulto*. Madrid: Plan nacional para el SNS del MSC. Avalia-t N^a 2006/06. Recuperado el 26 de Agosto de 2013, de:

http://www.guiasalud.es/GPC/GPC_424_Dep_Adult_compl.pdf

Grupo EuroQoL. (1990). EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy.*, 16(3), 199-208.

Hamilton, M. (1960). A rating scale for depression. *J Neurol.Neurosurg.Psychiatry.*, 23(1), 56-62.

Haro, J. M., Palacín, C., Vilagut, G., Martínez, M., Bernal, M., Luque, I., Codony, M., Dolz, M. y Alonso, J. (2006). Prevalencia de los trastornos menatales y factores asociados: resultados del estudio ESEMeD-España [Prevalence of mental disorders and associated factors: results from the ESEMeD-Spain study]. *Med Clin.(Barc.)*, 126(12), 445-451.

Hayhurst, H., Palmer, S., Abbott, R., Johnson, T., y Scott, J. (2006). Measuring health-related quality of life in bipolar disorder: relationship of the EuroQol (EQ-5D) to condition-specific measures. *Qual Life Res.*, 15(7), 1271-1280.

Herdman, M., Gudex, C., Lloyd, A., Janssen, M., Kind, P., Parkin, D., Bonsel, G. y Badía, X. (2011). Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.*, 20(10), 1727-1736.

Hernández, M., Ochando, A., Mora, J., Lorenzo, S., y López, K. (2005). Satisfacción del paciente en una unidad de hemodiálisis: Objetivo de calidad asistencial en enfermería. *Rev Soc Enferm Nefrol*, 8(2), 90-96.

Hirschfield RM. (2002). *Guideline Watch: Practice Guideline for the Treatment of Patients with Bipolar Disorder, 2nd Edition*. Arlington, VA: American Psychiatric Association. Recuperado el 2 de Septiembre, 2013, de: http://www.psych.org/psych_pract/treatg/pg/prac_guide.cfm.

Hudak, P. L., y Wright, J. G. (2000). The characteristics of patient satisfaction measures. *Spine*, 25(24), 3167-3177.

Johnston, B. C., Patrick, D. L., Busse, J. W., Schunemann, H. J., Agarwal, A., y Guyatt, G. H. (2013). Patient-reported outcomes in meta-analyses-Part 1: assessing risk of bias and combining outcomes. *Health Qual Life Outcomes.*, 11(1), 109-111.

Jönsson L, Lang A, y Lindström E. (2009). TooL: Multi-Attribute Utility Function Reflecting Patient Experience of Side Effects to Antipsychotic Therapy. *Value in Health*, 12(7), A361.

Kahneman, D. y Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263-292.

Kay, S. R., Fiszbein, A., Vital-Herne, M., y Fuentes, L. S. (1990). The Positive and Negative Syndrome Scale-Spanish adaptation. *J Nerv.Ment.Dis.*, 178(8), 510-517.

Kim, J. H., Lee, W. Y., Hong, Y. P., Ryu, W. S., Lee, K. J., Lee, W. S., y Morisky, D. E. (2014). Psychometric properties of a short self-reported measure of

- medication adherence among patients with hypertension treated in a busy clinical setting in Korea. *J Epidemiol.*, 24(2), 132-140.
- King, M. T., Cella, D., Osoba, D., Stockler, M., Eton, D., Thompson, J., y Eisenstein, A. (2010). Meta-analysis provides evidence-based interpretation guidelines for the clinical significance of mean differences for the FACT-G, a cancer-specific quality of life questionnaire. *Patient.Relat Outcome.Meas.*, 1, 119-26. doi: 10.2147/PROM.S10621.
- Konig, H. H., Roick, C., y Angermeyer, M. C. (2007). Validity of the EQ-5D in assessing and valuing health status in patients with schizophrenic, schizotypal or delusional disorders. *Eur.Psychiatry.*, 22(3), 177-187.
- Lehman AF, Lieberman JA, Dixon LB, McGlashan TH, Miller AL, Perkins DO y Kreyenbuhl J. (2004). Practice Guideline for the Treatment of Patients with Schizophrenia (Second Edition). *Am. J. Psychiatry.*, 161(2), 1-56.
- Lenderking, W. R. (2005). Brief reflections on treatment satisfaction. *Value.Health*, 8 *Suppl 1*, S2-S5. Recuperado de PM:16336485
- Lester, H., y Howe, A. (2008). Depression in primary care: three key challenges. *Postgrad.Med J.*, 84(996), 545-548.
- Levine, S. Z., Rabinowitz, J., Engel, R., Etschel, E. y Leucht, S. (2008). Extrapolation between measures of symptom severity and change: an examination of the PANSS and CGI. *Schizophr.Res.*, 98(1-3), 318-322.
- Linder-Pelz, S. U. (1982). Toward a theory of patient satisfaction. *Soc Sci.Med.*, 16(5), 577-582.

- Lindström E, Jönsson L, y Berntsson A. (2009). A Patient Perspective on Side Effects of Antipsychotic Therapy: The Tool Instrument. *Value in Health*, 12(7), A361.
- Lingjaerde, O., Ahlfors, U. G., Bech, P., Dencker, S. J., y Elgen, K. (1987). The UKU side effect rating scale. A new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. *Acta Psychiatr.Scand.Suppl.*, 334:1-100., 1-100.
- Lobo, A., Chamorro, L., Luque, A., Dal-Re, R., Badía, X., y Baro, E. (2002). Validación de las versiones en español de Montgomery-Asberg Depression Rating Scale y la Hamilton Anxiety Rating Scale para la evaluación de la depresión y de la ansiedad [Validation of the Spanish versions of the Montgomery-Asberg depression and Hamilton anxiety rating scales]. *Med Clin (Barc.)*, 118(13), 493-499.
- Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Meakin, R., y Weinman, J. (2002). The 'Medical Interview Satisfaction Scale' (MISS-21) adapted for British general practice. *Fam.Pract.*, 19(3), 257-263.
Recuperado de PM:11978716
- Miilunpalo, S., Vuori, I., Oja, P., Pasanen, M., y Urponen, H. (1997). Self-rated health status as a health measure: the predictive value of self-reported health status on the use of physician services and on mortality in the working-age population. *J Clin Epidemiol.*, 50(5), 517-528.

- Ministerio de Sanidad, Servicios Sociales e Igualdad. (2012). *Estrategia para el Abordaje de la Cronicidad en el Sistema Nacional de Salud*. Madrid: Centro de Publicaciones. Ministerio de Sanidad, Servicios Sociales e Igualdad.
- Mira, J. J., y Aranaz, J. (2000). La satisfacción del paciente como una medida de resultado de la atención sanitaria [Patient satisfaction as an outcome measure in health care]. *Med Clin.(Barc.)*, *114 Suppl 3:26-33.*, 26-33. Recuperado de: http://calite-revista.umh.es/indep/web/satisf_paciente.pdf
- Montejo, A. L. (2008). Prolactin awareness: an essential consideration for physical health in schizophrenia. *Eur.Neuropsychopharmacol.*, *18 Suppl 2:S108-14.*, S108-S114.
- Montejo, A. L., Correas-Lauffer, J., Mauriño, J., Villa, G., Rebollo, P., Díez, T. y Cordero, L. (2011). Estimation of a multiattribute utility function for the Spanish version of the TooL questionnaire. *Value.Health.*, *14(4)*, 564-570.
- Morisky, D. E., Green, L. W., y Levine, D. M. (1986). Concurrent and predictive validity of a self-reported measure of medication adherence. *Med Care.*, *24(1)*, 67-74.
- Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., y Ustun, B. (2007). Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet.*, *370(9590)*, 851-858.
- Muñiz, J. (2000). *Teoría Clásica de los Tests*. Madrid: Pirámide.
- Muñiz, J. (2010). Las Teorías de los Tests: Teoría Clásica y Teoría de Respuesta a los Ítems. *Papeles del Psicólogo*, *3(1)*, 57-66.

- Muñiz, J., y Fonseca-Pedrero, E. (2008). Construcción de Instrumentos de Medida para la Evaluación Universitaria. *Revista de Investigación en Educación*, 5, 13-25.
- National Collaborating Centre for Mental Health. (2009). *Depression. The treatment and management of depression in adults*. London: National Institute for Health and Clinical Excellence.
- National Institute for Clinical Excellence (NICE). (2002). *Guidance on the use of newer (atypical) antipsychotic drugs for the treatment of schizophrenia. Technology Appraisal N° 43*. London: National Institute for Health and Clinical Excellence.
- National Institute for Health and Clinical Excellence, N. (2013). *Guide to the methods of technology appraisal 2013. Process and methods guides*. London: National Institute for Health and Clinical Excellence.
- Navas, M. (1994). Teoría Clásica de los Test versus Teoría de Respuesta al Ítem. *Psicológica*, 15, 175-208.
- Newcomer, J. W. (2005). Second-generation (atypical) antipsychotics and metabolic effects: a comprehensive literature review. *CNS Drugs*, 19 Suppl 1:1-93., 1-93.
- Novick, D., Bousoño, M., Suárez, D., Olivares, J. M., Montejo, A. L., Haro, J. M., Edgell, E. T., y Ratcliffe, M. (2005). Use of concomitant medication with antipsychotic treatment in outpatients with schizophrenia: results from the European Schizophrenia Outpatients Health Outcomes (SOHO) study. *Prog.Neuropsychopharmacol.Biol.Psychiatry*., 29(6), 972-982.
- Okuda, M., Ohkubo, K., y Goto, M. (2004). Development of patient satisfaction questionnaire in allergic rhinitis. *Alerugi*, 53(12), 1195-1202.

Oliver, R., y De Sarbo, W. (1988). Response determinants in satisfaction judgements. *Journal of Consumer Research*, 14, 495-507.

Ollero, M., Sanz, R. y Padilla, C. (coords). (2012). *Plan Andaluz de Atención Integrada a Pacientes con Enfermedades Crónicas 2012/2016*. Sevilla: Consejería de Salud. Junta de Andalucía. Recuperado el 14 de Abril, 2014, de:

http://www.juntadeandalucia.es/salud/export/sites/csalud/galerias/documentos/p_2_p_2_planes_integrales/PIEC.pdf

Ortega, F., Otero, A., Crespo, J. F., Delgado, J. F., Borro, J. M., y Cuervo, J. (2013). Satisfaction and adherence with immunosuppressant treatment in renal transplant patients living with a working graft. *J Nephrol.*, 26(2), 297-305.

Paddock, L. E., Veloski, J., Chatterton, M. L., Gevirtz, F. O., y Nash, D. B. (2000). Development and validation of a questionnaire to evaluate patient satisfaction with diabetes disease management. *Diabetes Care*, 23(7), 951-956. Recuperado de PM:10895845

Patient Centered Outcomes Research Institute. (2012). Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA.*, 307(15), 1636-1640.

Patrick, D. (2013). Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *Value Health.*, 16(4), 455-456.

Patrick, D. L., Burke, L. B., Powers, J. H., Scott, J. A., Rock, E. P., Dawisha, S., O'Neill, R. y Kennedy, D. L. (2007). Patient-reported outcomes to support

medical product labeling claims: FDA perspective. *Value Health., 10 Suppl 2:S125-37., S125-S137.*

Patrick, D. L., Burns, T., Morosini, P., Rothman, M., Gagnon, D. D., Wild, D. y Adriaenssen, I. (2009). Reliability, validity and ability to detect change of the clinician-rated Personal and Social Performance scale in patients with acute symptoms of schizophrenia. *Curr.Med Res Opin., 25(2), 325-338.*

Patrick, D. L., y Erickson, P. (1993). *Health Status and Health Policy. Quality of Life in Health Care Evaluation and Resource Allocation.* New York: Oxford University Press.

Patrick, D. L., Khalaf, K. M., Dmochowski, R., Kowalski, J. W. y Globe, D. R. (2013). Psychometric performance of the incontinence quality-of-life questionnaire among patients with overactive bladder and urinary incontinence. *Clin Ther., 35(6), 836-845.*

Patrick, D. L., Martin, M. L., Bushnell, D. M., Yalcin, I., Wagner, T. H. y Buesching, D. P. (1999). Quality of life of women with urinary incontinence: further development of the incontinence quality of life instrument (I-QOL). *Urology., 53(1), 71-76.*

Peralta, V., y Cuesta, M. J. (1994). Psychometric properties of the positive and negative syndrome scale (PANSS) in schizophrenia. *Psychiatry Res., 53(1), 31-40.*

Perkins, D. O., Gu, H., Weiden, P. J., McEvoy, J. P., Hamer, R. M. y Lieberman, J. A. (2008). Predictors of treatment discontinuation and medication nonadherence in patients recovering from a first episode of schizophrenia, schizophreniform

disorder, or schizoaffective disorder: a randomized, double-blind, flexible-dose, multicenter study. *J Clin Psychiatry*, 69(1), 106-113.

Perkins, D. O., Johnson, J. L., Hamer, R. M., Zipursky, R. B., Keefe, R. S., Centorrino, F., Green, A. I., Glick I. B., Kahn, R. S. Sharma, T., Tohen, M., McEvoy, J. P., Weiden, P. J., y Lieberman, J. A. (2006). Predictors of antipsychotic medication adherence in patients recovering from a first psychotic episode. *Schizophr.Res.*, 83(1), 53-63.

Pleil, A. M., Coyne, K. S., Reese, P. R., Jumadilova, Z., Rovner, E. S., y Kelleher, C. J. (2005). The validation of patient-rated global assessments of treatment benefit, satisfaction, and willingness to continue-the BSW. *Value.Health.*, 8 Suppl 1:S25-34., S25-S34.

Prieto, M., Cuervo, J., Divison, J., Abellán, J., Llisterri, J., Pérez-Monteoliva, R., Carrasco, J. L., Quijano, J. J., Villa, G., y Rebollo, P. (2010^a, Marzo). *Evaluación de la satisfacción del paciente con hipertensión arterial de riesgo alto o muy alto con tratamiento farmacológico combinado*. Póster presentado en la 15^a Reunión Nacional de la SEH-LELHA organizada por la Sociedad Española de Hipertensión y la Liga Española para la Lucha contra la Hipertensión Arterial, Zaragoza.

Prieto, M., Divison, J., Abellán, J., Llisterri, J., Pérez-Monteoliva, R., Carrasco, J., Quijano, J. J., Cuervo, J., y Rebollo, P. (2010b). *High-risk hypertensive patients: Assessment of blood pressure control. Adherence to combination treatments and satisfaction*. Póster presentado en el 20th European Meeting on Hypertension organizado por la European Society of Hypertension, Oslo.

Rabinowitz, J., Mehnert, A., y Eerdeken, M. (2006). To what extent do the PANSS and CGI-S overlap? *J Clin Psychopharmacol.*, 26(3), 303-307.

Ramos-Brieva, J. A., y Cordero-Villafáfila, A. (1988). A new validation of the Hamilton Rating Scale for Depression. *J Psychiatr.Res.*, 22(1), 21-28.

Rebollo, P. (2008). Utilización del "grupo SF" de los cuestionario de calidad de vida (SF-36, SF-12 y SF-6D) en el marco de ensayos clínicos y en la práctica clínica habitual [Use of health-related quality of life "SF Group" questionnaires (SF-36, SF-12 and SF-6D) in the framework of clinical trials and routine clinical practice]. *Med Clin.(Barc.)*, 130(19), 740-741.

Rebollo, P., Castejón, I., Cuervo, J., Villa, G., García-Cueto, E., Díaz-Cuervo, H., Zardain, P. C., Muñiz, J., y Alonso J. (2010). Validation of a computer-adaptive test to evaluate generic health-related quality of life. *Health Qual Life Outcomes.*, 8(147), 147-148. doi: 10.1186/1477-7525-8-147.

Rebollo, P., García, C. E., Zardain, P. C., Cuervo, J., Martínez, I., Alonso, J., Ferrer, M., y Muñiz J. (2009). Desarrollo del CAT-Health, primer test adaptativo informatizado para la evaluación de la calidad de vida relacionada con la salud en España [Development of the CAT-Health, the first adaptive computerized test (CAT) for the evaluation health-related quality of life in Spain]. *Med Clin (Barc.)*, 133(7), 241-251.

Regnault, A., Viala-Danten, M., Gilet, H., y Berdeaux, G. (2010). Scoring and psychometric properties of the Eye-Drop Satisfaction Questionnaire (EDSQ), an instrument to assess satisfaction and compliance with glaucoma treatment. *BMC.Ophthalmol.*, 10(1), 1-10. doi: 10.1186/1471-2415-10-1.

- Revicki, D., Hays, R. D., Cella, D. y Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.*, 61(2), 102-109.
- Revicki, D. A., Matza, L. S. Flood, E., y Lloyd, A. (2005). Bipolar disorder and health-related quality of life: review of burden of disease and clinical trials. *Pharmacoeconomics.*, 23(6), 583-594.
- Rey, E. (2007). Trasladar las medidas de salud autopercebida a la consulta: un reto para el clínico [Using patient report outcomes in the office: a challenge for the physician]. *An.Med Interna.*, 24(9), 413-414. Recuperado de: <http://scielo.isciii.es/pdf/ami/v24n9/editorial.pdf>
- Reynolds, K., Viswanathan, H. N., Muntner, P., Harrison, T. N., Cheetham, T. C., Hsu, J. W., Glod, D. T., Silverman, S., Grauer, A., Morisky, D. E., y O'Malley, C. D. (2014). Validation of the Osteoporosis-Specific Morisky Medication Adherence Scale in long-term users of bisphosphonates. *Qual Life Res.*, 7, 1-12. Doi: 10.1007/s11136-014-0662-3.
- Rofail, D., Gray, R., y Gournay, K. (2005). The development and internal consistency of the satisfaction with Antipsychotic Medication Scale. *Psychol.Med.*, 35(7), 1063-1072.
- Román, A., Barbera, J. A., Castillo, M. J., Muñoz, R., y Escribano, P. (2013). Calidad de vida relacionada con la salud en una cohorte nacional de pacientes con hipertensión arterial pulmonar o hipertensión pulmonar tromboembólica crónica [Health-related quality of life in a national cohort of patients with pulmonary

- arterial hypertension or chronic thromboembolic pulmonary hypertension]. *Arch.Bronconeumol.*, 49(5), 181-188.
- Romera, I., Pérez, V., Menchón, J. M., Delgado-Cohen, H., Polavieja, P., y Gilaberte, I. (2010). Social and occupational functioning impairment in patients in partial versus complete remission of a major depressive disorder episode. A six-month prospective epidemiological study. *Eur.Psychiatry.*, 25(1), 58-65.
- Romera, I., Pérez, V., Menchón, J. M., Polavieja, P., y Gilaberte, I. (2011). Optimal cutoff point of the Hamilton Rating Scale for Depression according to normal levels of social and occupational functioning. *Psychiatry Res.*, 186(1), 133-137.
- Rotter, J. (1980). *Social Learning and Clinical Psychology*. New York: Johnson Reprint Company (publicado originalmente en 1954).
- Royal Australian and New Zealand College of Psychiatrists. (2005). Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the treatment of schizophrenia and related disorders. *Aust.N.Z.J Psychiatry.*, 39(1-2), 1-30.
- Ruiz, M., Campillo, M. A., Monfort, J., Pardob, A., Rejas, J., y Soto, J. (2005). Adaptación al castellano y validación del cuestionario Arthritis Treatment Satisfaction Questionnaire [Linguistic adaptation into Spanish and validation of the Arthritis Treatment Satisfaction Questionnaire]. *Med.Clin.(Barc.)*, 125(3), 84-92.

- Ruiz, M. A., Pardo, A., Rejas, J., Soto, J., Villasante, F., y Aranguren, J. L. (2008). Development and validation of the "Treatment Satisfaction with Medicines Questionnaire" (SATMED-Q). *Value Health.*, 11(5), 913-926.
- Ruiz, M., y Pardo, A. (2005). Calidad de vida relacionada con la salud: definición y utilización en la práctica médica. *Pharmacoeconomics*, 2(1), 31-43.
- Rumsfeld, J. S., MaWhinney, S., McCarthy, M., Jr., Shroyer, A. L., VillaNueva, C. B., O'Brien, M., Moritz, T. E. Henderson, W. G., Grover, F. L., Sethi, G. K., y Hammermeister, K. E. (1999). Health-related quality of life as a predictor of mortality following coronary artery bypass graft surgery. Participants of the Department of Veterans Affairs Cooperative Study Group on Processes, Structures, and Outcomes of Care in Cardiac Surgery. *JAMA.*, 281(14), 1298-1303.
- Schaeffer, S. (2013). ECHOUTCOME: Challenging QALYs. BioCentury. The Bernstein Report on BioBusiness., 9(2), 1.
- Schalock, R. L., y Verdugo, M. A. (2002). *Handbook of quality of life for human service practitioners*. Washington DC: American Association of Mental Retardation.
- Serrano-Blanco, A., Palao, D. J., Luciano, J. V., Pinto-Meza, A., Luján, L., Fernández, A., Roura, P., Bertsch, J., Mercader, M., y Haro, J.M. (2010). 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.*, 45(2), 201-210.

- Shikiar, R. y Rentz, A. M. (2004). Satisfaction with medication: an overview of conceptual, methodologic, and regulatory issues. *Value.Health.*, 7(2), 204-215.
- Sicras-Mainar, A., Blanca-Tamayo, M., Gutiérrez-Nicuesa, L., Salvatella-Pasant, J. y Navarro-Artieda, R. (2010). Impacto de la morbilidad, uso de recursos y costes en el mantenimiento de la remisión de la depresión mayor en España: estudio longitudinal de ámbito poblacional [Impact of morbidity, resource use and costs on maintenance of remission of major depression in Spain: a longitudinal study in a population setting]. *Gac.Sanit.*, 24(1), 13-19. doi:10.1016/j.gaceta.2009.05.012
- Sitzia, J. (1999). How valid and reliable are patient satisfaction data? An analysis of 195 studies. *Int.J.Qual.Health Care*, 11(4), 319-328.
- Sobocki, P., Ekman, M., Agren, H., Runeson, B. y Jonsson, B. (2006). The mission is remission: health economic consequences of achieving full remission with antidepressant treatment for depression. *Int.J Clin.Pract.*, 60(7), 791-798.
- Sobocki, P., Jonsson, B., Angst, J. y Rehnberg, C. (2006). Cost of depression in Europe. *J Ment.Health Policy Econ.*, 9(2), 87-98.
- Soto, J. (2007). Implicación de la investigación en resultados en salud en la mejora continua de la calidad asistencial del Sistema Nacional de Salud. *An Med Interna*, 24, 517-519.
- Speight, J. (2005). Assessing patient satisfaction: concepts, applications, and measurement. *Value.Health*, 8 Suppl 1, S6-S8. Recuperado de PM:16336490

Stansfeld, S. A., Roberts, R., y Foot, S. P. (1997). Assessing the validity of the SF-36 General Health Survey. *Qual Life Res.*, 6(3), 217-224.

Suárez, D., y Haro, J. M. (2008). Overview of the findings from the European SOHO study. *Expert.Rev.Neurother.*, 8(6), 873-880.

TenDoesschate, M. C., Bockting, C. L., Koeter, M. W., y Schene, A. H. (2010). Prediction of recurrence in recurrent depression: a 5.5-year prospective study. *J Clin Psychiatry.*, 71(8), 984-991.

Torrance, G. W., Feeny, D. H., Furlong, W. J., Barr, R. D., Zhang, Y., y Wang, Q. (1996). Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care.*, 34(7), 702-722.

Torrance, G. W., Furlong, W., Feeny, D., y Boyle, M. (1995). Multi-attribute preference functions. Health Utilities Index. *Pharmacoeconomics.*, 7(6), 503-520.

Turner, L., Shamseer, L., Altman, D. G., Weeks, L., Peters, J., Kober, T., Dias, S, Schulz, K. F., Plint, A. C., y Moher, D. (2012). Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane.Database.Syst.Rev.*, 14(11), MR000030.

doi: 10.1002/14651858.MR000030.pub2., MR000030.

Tversky, A., y Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, 5, 297-323.

Ustun, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., y Murray, C. J. (2004). Global burden of depressive disorders in the year 2000. *Br.J Psychiatry.*, 184, 386-392.

Velasco, S. (2010). *Programa de atención biopsicosocial al malestar en atención primaria. Area VII-Mieres-2010. Principado de Asturias. Proyecto de Evaluación de Proceso, Resultados e Impactos.* Principado de Asturias: AsturSalud, Portal de Salud del Principado de Asturias. Recuperado el 16 de Abril, 2014, de:

http://www.asturias.es/Astursalud/Ficheros/AS_Salud%20Publica/AS_Promocion%20de%20la%20Salud/Salud%20de%20las%20mujeres%20asturianas/evaluacion%202%C2%AA%20etapa%20malestar_MIERES_140111.pdf

Verdugo, M. A., Schalock, R. L., Keith, K. D., y Stancliffe, R. J. (2005). Quality of life and its measurement: important principles and guidelines. *J Intellect.Disabil.Res.*, 49(Pt 10), 707-717.

Verdugo, M., Canal, R., Fernández, S., Bermejo, L., Alonso, S., Tamarit, J., López, M. A., Castillo, T., Leturia, F. J., y Planella, J. (2011). *Modelo de calidad de vida aplicado a la atención residencial de personas con necesidades complejas de apoyo. La toma de decisiones para el desarrollo de su proyecto de vida. Las personas con necesidades complejas de apoyo, con graves discapacidades.* Madrid: Ministerio de Sanidad, Política Social e Igualdad. Instituto de Mayores y Servicios Sociales (IMSERSO).

Vilagut, G., Valderas, J. M., Ferrer, M., Garín, O., López-García, E., y Alonso, J. (2008). Interpretación de los cuestionarios de salud SF-36 y SF-12 en España:

componentes físico y mental [Interpretation of SF-36 and SF-12 questionnaires in Spain: physical and mental components]. *Med Clin (Barc)*, 130(19), 726-735. doi: 10.1157/13121076.

Vuorilehto, M. S., Melartin, T. K., y Isometsa, E. T. (2009). Course and outcome of depressive disorders in primary care: a prospective 18-month study. *Psychol.Med.*, 39(10), 1697-1707.

Weaver, M., Patrick, D. L., Markson, L. E., Martin, D., Frederic, I., y Berger, M. (1997). Issues in the measurement of satisfaction with treatment. *Am.J.Manag.Care*, 3(4), 579-594. Recuperado de PM:10169526

WHOQOL Group. (1995). The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci.Med.*, 41(10), 1403-1409.

WHOQOL Group. (1998). The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. *Soc Sci.Med.*, 46(12), 1569-1585.

World Health Organization. (2001). *The World Health report 2001. Mental Health: New understanding, New Hope*. Ginebra: World Health Organization. Recuperado de: <http://www.who.int/whr/2001/en/>

Xiao, H., y Barber, J. P. (2008). The effect of perceived health status on patient satisfaction. *Value.Health.*, 11(4), 719-725.

- Yang, A., Wang, B., Zhu, G., Jiao, Z., Fang, Y., Tang, F., Ma, C., Zhao, Y., Cheng C., y Zhong, M. (2014). Validation of Chinese version of the Morisky Medication Adherence Scale in patients with epilepsy. *Seizure.*, 23(4), 295-299.
- Zimmerman, M., Chelminski, I., McGlinchey, J. B., y Posternak, M. A. (2008). A clinically useful depression outcome scale. *Compr.Psychiatry.*, 49(2), 131-140.
- Zimmerman, M., Martinez, J., Attiullah, N., Friedman, M., Toba, C., y Boerescu, D. A. (2012a). How should residual symptoms be defined in depressed patients who have remitted? *Compr.Psychiatry.*, 54(2), 91-6. doi: 10.1016/j.comppsy.2012.06.010.
- Zimmerman, M., Martinez, J., Attiullah, N., Friedman, M., Toba, C., Boerescu, D. A., y Rahgeb, M. (2012b). Determining remission from depression on two self-report symptom scales: a comparison of the Quick Inventory of Depressive Symptomatology and the Clinically Useful Depression Outcome Scale. *Compr.Psychiatry.*, 53(7), 1034-1038.
- Zimmerman, M., y McGlinchey, J. B. (2008). Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Ann.Clin.Psychiatry.*, 20(3), 125-129.
- Zimmerman, M., Posternak, M. A., y Chelminski, I. (2004). Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry.*, 161(10), 1911-1913.

ANEXO:

COMUNICACIONES EN CONGRESOS Y PUBLICACIONES CON LOS INSTRUMENTOS DESARROLLADOS.

Dentro del equipo de IRS en LASER Analytica, hemos continuado investigando y trabajando, entre otras, con las medidas de PROs incluidas en las publicaciones siguiendo la meta ya comentada de profundizar en sus propiedades psicométricas y optimizar la medición de los constructos definidos.

ARTÍCULO 1: Desarrollo y validación cuestionario SAT-Q.

- Cuervo, J., Rebollo, P., Zardain, P., Prieto, M. A., Tranche, S., Barreda, M. J., y Sánchez-Baragaño, M. A. (2008, Noviembre). *Development and Validation of the Satisfaction Questionnaire (SAT-Q) in Patients with Chronic Disease.* Comunicación Oral presentada en el en el ISPOR 11th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Atenas.
- Prieto, M. A., Cuervo, J., Divison, J. A., Abellán, J., Llisterri, J. L., Pérez-Monteoliva, R., Carrasco, J. L., Quijano, J. J., Villa, G., y Rebollo, P. (2010, Marzo). *Evaluación de la satisfacción del paciente con hipertensión arterial de riesgo alto o muy alto con tratamiento farmacológico combinado.* Póster presentado en la 15ª Reunión Nacional de la SEH-LELHA organizado por la Sociedad Española de Hipertensión y la Liga Española para la Lucha contra la Hipertensión Arterial, Madrid.

- Prieto, M. A., Divison, J. A., Abellán, J., Llisterri, J. L., Robles, N. R., Pérez-Monteoliva, R., Carrasco, J. L., Quijano, J. J., Cuervo, J., y Rebollo, P. (2010, Junio). *High-risk hypertensive patients: Assessment of blood pressure control. Adherence to combination treatments and satisfaction*. Póster presentado en el 20th European Meeting on Hypertension organizado por la European Society of Hypertension, Oslo.
- Ortega, F., Rebollo, P., Cuervo, J., en nombre del grupo de Investigación del estudio SATISFY. (2010, Octubre). *Determinantes de la Satisfacción con el tratamiento inmunosupresor en pacientes portadores de un injerto renal funcionante*. Póster presentado en el XL Congreso de la SEN organizado por la Sociedad Española de Nefrología, Granada.
- Cuervo, J., Rebollo, P., Prieto, M. A., Tranche, S., Barreda, M. J., y Sánchez-Baragaño, M. A. (2013, Febrero). *Desarrollo y Validación de un Cuestionario Genérico para la Evaluación de la Satisfacción de los Pacientes Crónicos: Cuestionario SAT-Q*. Comunicación Oral presentada en la I Jornada Científica organizada por BiBlioPro, Barcelona.
- Campillo-Álvarez, A., Rodríguez-Aguilella, A., Cuervo, J., Castejón, N., y Rebollo, P. (2013, Mayo). *SAT-Q Questionnaire: Application of Rasch measurement for selecting items and improving the assessment of patient satisfaction*. Póster presentado en el ISPOR 18th Annual International Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Nueva Orleans.
- Campillo-Álvarez, A., Rodríguez-Aguilella, A., Castejón, N., Cuervo J., y Rebollo, P. (2013, Noviembre). *Differential Item Functioning in a Satisfaction Item Bank*.

Póster presentado en el ISPOR 16th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Dublín.

- Ortega, F., Otero, A., Crespo, J. F., Delgado, J. F., Borro, J. M., y Cuervo, J. (2013) Satisfaction and adherence with immunosuppressant treatment in renal transplant patients living with a working graft. *Journal of Nephrology*; 26(2):297-305.

ARTÍCULO 2: Adaptación cultural y validación del cuestionario Tool.

- Mauriño, J., Cordero, L., Montejo, A. L., Rebollo, P., Cuervo, J., Díez, T., Tafalla, M., y Hernández, R. (2008, Noviembre). *The Spanish Version of the Tool Questionnaire: A Useful Measure for Evaluating the HRQoL and Utilities from Schizophrenic and Bipolar Patients*. Póster presentado en el en el ISPOR 11th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Atenas.
- Cuervo, J., Rebollo, P., Mauriño, J., Cordero, L., y Castejón, I. (2011, Noviembre). *Estimating Utilities in Schizophrenic and Bipolar Patients from Disease-Specific or Generic Instruments Assessing Patients' Health States: Where Are the Differences?(I)*. Póster presentado en el ISPOR 14th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Madrid.
- Cuervo, J., Rebollo, P., Mauriño, J., Cordero, L., Castejón, I. y Díaz-Cuervo H. (2012, Mayo). *Estimación de utilidades en pacientes con trastorno bipolar y esquizofrenia a partir de instrumentos específicos o genéricos; ¿Cuáles son las diferencias?(II)*. Póster presentado en las XXXII Jornadas de Economía de la Salud organizadas por la Asociación Española de Economía de la Salud (AES), Bilbao.

ARTÍCULO 3. Adaptación cultural y validación de la escala CUDOS.

- Agüera, L., Montón, C., Medina, E., Cuervo, J., Rodríguez-Aguilella, A., y Mauriño, J. (2011, Noviembre). *The Spanish Version of the Clinically Useful Depression Outcome Scale. A Valid Instrument to Evaluate Depressive Symptoms in Primary Care.* Póster presentado en el ISPOR 14th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Madrid.
- Agüera, L., Montón, C., Medina, E., Mauriño, J., Díaz, H., y Cuervo, J. (2012, Noviembre). *Assessing Depressive Symptoms in Primary Care: Psychometrical Properties of the Spanish Version of the Clinically Useful Depression Outcome Scale.* Póster presentado en el ISPOR 15th Annual European Congress organizado por la International Society of Pharmacoeconomics and Outcomes Research, Berlín.

