

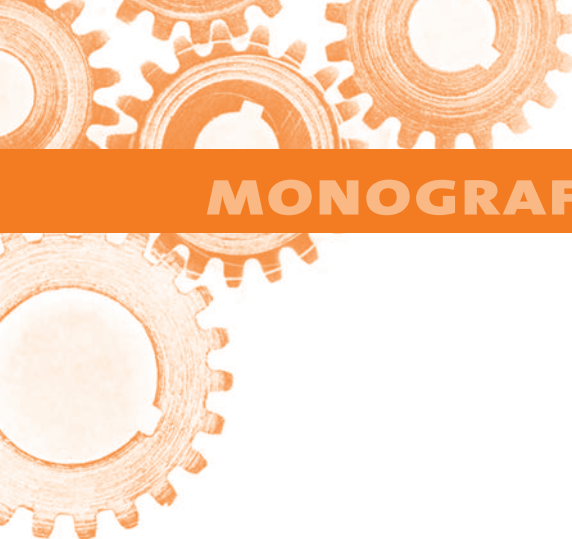


MONOGRAFÍAS DE POLÍTICA Y GESTIÓN

Resultados reportados por los pacientes (PROs)



JORDI ALONSO-CABALLERO y MONTSERRAT FERRER-FORES



MONOGRAFÍAS DE POLÍTICA Y GESTIÓN

Resultados reportados por los pacientes (PROs)



JORDI ALONSO-CABALLERO y MONTSERRAT FERRER-FORES

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos, www.cedro.org) si necesita fotocopiar o escanear algún fragmento de esta obra

© 2017 Ergon.
C/ Arboleda, 1. 28221 Majadahonda (Madrid)

ISBN: 978-84-16732-60-9
Depósito Legal: M-7693-2017
Impreso en España

Autores

JORDI ALONSO-CABALLERO

*Instituto Hospital del Mar de Investigaciones Biomédicas (IMIM).
Hospital del Mar. Barcelona.*

MONTSERRAT FERRER-FORES

*Instituto Hospital del Mar de Investigaciones Biomédicas (IMIM).
Hospital del Mar. Barcelona.*

Índice

PRÓLOGO	7
PRÓLOGO	9
I. INTRODUCCIÓN A LA MEDIDA DE LOS RESULTADOS REPORTADOS POR LOS PACIENTES (PROs)	11
II. HISTORIA DE LOS RESULTADOS REPORTADOS POR LOS PACIENTES (PROs)	13
1. Tipos de Instrumentos PRO	14
2. Utilización rutinaria de PROs.	17
– Tipos de uso rutinario de PROs.	17
– Uso rutinario en la práctica clínica: oportunidades y barreras. ...	18
– Impacto del uso rutinario de PROs en la práctica clínica	19
III. APLICACIONES DE LOS INSTRUMENTOS PRO	23
1. El programa NHS-PROM inglés: Los PROs para mejorar la calidad de la atención	23
2. Propuesta de la <i>Food and Drug Administration</i> (FDA)	28
3. PROMIS: una nueva medición dinámica de la salud percibida.	30
4. <i>Patient-Centered Outcomes Research Institute</i> (PCORI).	40
IV. BANCOS DE INSTRUMENTOS PRO	45
1. BiblioPRO: biblioteca virtual de instrumentos PRO en español	46
2. <i>Patient-Reported Outcome and Quality of Life Instruments Database</i> (PROQOLID)	48
V. EVALUACIÓN DE LOS INSTRUMENTOS PRO	51
1. <i>Evaluating Measures of Patient Reported Outcomes</i> (EMPRO)	51
2. COSMIN, un consenso internacional para la evaluación de los estudios sobre PROs	53
EPÍLOGO: CONCLUSIONES Y RECOMENDACIONES	59
BIBLIOGRAFÍA	63

Prólogo

Son recurrentes las ocasiones en las que, a quienes formamos parte del sector farmacéutico, se nos presenta la oportunidad de hablar de los pacientes: de sus dolencias, sus dudas, sus expectativas, su calidad de vida... A fin de cuentas el *leitmotiv* de nuestra actividad profesional –y hasta nuestra razón última de existir– gira en torno a ellos hasta convertirse en parte indivisible de quienes somos. Poner a los pacientes en primer lugar, lejos de ser una aspiración, encarna el valor fundacional e inspirador de todo lo que hacemos. Sencillamente, no puede ser de otro modo.

Por ello resulta tan pertinente en un sistema sanitario gobernado por la sostenibilidad y la calidad en el servicio reflexionar sobre la conveniencia de contar con un protocolo que, fundamentado en la experiencia directamente transmitida por el paciente, complemente y dimensione la evidencia científica desde una perspectiva más cercana a la vivencia en primera persona de la enfermedad. O dicho de otro modo: debatamos sobre cómo esta información proporcionada por el mismo paciente, sin pasar por tamiz interpretativo del profesional sanitario, podría revertir en una optimización del propio tratamiento al que está siendo sometido.

No son pocas las potenciales ventajas de esta suerte de bidireccionalidad metodológica que representan los instrumentos PRO (*“Patient Reported Outcomes”*, o *“Resultados reportados por pacientes”*). El seguimiento pormenorizado del tratamiento, la identificación de nuevos síntomas o dolencias, la evaluación de la calidad del cuidado recibido más allá de lo estrictamente terapéutico o la instrucción de una calidad de vida adaptada a las circunstancias particulares del paciente son argumentos de suficiente relevancia e impacto como para, cuanto menos, detenernos en ello y analizar la viabilidad de su incorporación a la rutina médica.

Confiemos en el poder del debate y el análisis como catalizador de la transformación social. Precisamente por ello, desde la Fundación AstraZeneca llevamos 13 años fomentando el diálogo entre pacientes, científicos, instituciones y profesionales sanitarios para seguir dando vida a la ciencia. O lo que es lo mismo, para seguir promoviendo iniciativas que aporten un valor real a la sociedad a través de la ciencia. Esta monografía, la cuarta que publicamos en colaboración con la Fundación Gaspar Casals, es sólo un pequeño ejemplo de este compromiso ineludible de AstraZeneca con el paciente y su entorno.

El lector de esta publicación encontrará en sus páginas la información necesaria para participar en este debate con conocimiento de causa, aportando perspectiva y lucidez a un foro del que todos formamos parte y, a la vez, del que somos últimos valedores. Aprovechemos una nueva oportunidad

de avanzar juntos hacia un sistema sanitario más eficiente, más moderno y más humano. No solo por el bien mismo del sistema sino, sobre todo, por el de quien da sentido a todo nuestro trabajo: el paciente.

Luis Cordero

*Director de Corporate Affairs
de AstraZeneca España*

Prólogo

Si buscamos mejorar la atención sanitaria, una de las estrategias más ampliamente aceptadas por todos los proveedores de servicios sanitarios plantea que debemos poner al paciente en el centro del sistema, tomando decisiones teniendo en cuenta sus preferencias y sus necesidades¹. También es reconocido en la literatura científica que el clínico toma esas decisiones basado en las mejores evidencias científicas disponibles, pero con una información pobre sobre la vivencia de enfermedad que experimentan los pacientes². Por este motivo, si queremos practicar una asistencia orientada al paciente, mejorar sus resultados en salud y su percepción o vivencia de enfermedad es necesario tomar decisiones integrando la información derivada de los ensayos clínicos sobre las intervenciones (eficacia y seguridad), con aquella que proporcionan los pacientes con una enfermedad concreta y sometidos a la intervención evaluada.

Las medidas de los resultados en salud desde la perspectiva del paciente (*o PROs, en inglés, patient-reported outcomes*) incluyen todas aquellas evaluaciones del estado de salud del paciente que condiciona su enfermedad y su tratamiento, informadas directamente por los pacientes. Estas mediciones incluyen diferentes dimensiones que expresan las percepciones del paciente en cuanto a la calidad de vida relacionada con su salud, sus síntomas, la satisfacción con los cuidados que recibe, su bienestar general y el impacto en su función³.

Hay demostración empírica de que esta información mejora la comunicación entre el profesional sanitario y el paciente, facilita la detección de problemas no reconocidos y mejora la monitorización de la respuesta a los tratamientos³.

Además de mejorar la comunicación médico-paciente, es necesario un conocimiento más profundo de la perspectiva del paciente para poder mejorar la efectividad de las intervenciones terapéuticas, definir con sus aportaciones el verdadero valor de las mismas y poder incluir esta dimensión en la variable de eficacia de los ensayos clínicos.

Por todo ello, es preciso un conocimiento de estas variables para todos los que participan en la asistencia sanitaria. Un conocimiento que debe extenderse no sólo a su definición y a sus aplicaciones, sino a las herramientas disponibles para su medición, su evaluación (validez y reproducibilidad) y, los retos que plantea su implantación desde el punto de vista de pacientes, clínicos e instituciones (Agencias de Evaluación de Tecnologías Sanitarias, Servicios Regionales de Salud, Hospitales, centro de atención primaria, etc.).

Dr. Juan E. del Llano Señarís
Director
Fundación Gaspar Casal

I. Introducción a la medida de los resultados reportados por los pacientes (PROs)

Aunque el campo de la evaluación del estado de salud desde la perspectiva de los pacientes tiene más de 50 años de historia, el término ‘Resultados Reportados por los Pacientes’ (*Patient Reported Outcomes*, PRO)⁴ ha sido acuñado recientemente, de forma específica para evitar la confusión suscitada alrededor de varios términos relacionados y muchas veces utilizados como sinónimos: “calidad de vida”, “calidad de vida relacionada con la salud”, “estado de salud” o “capacidad funcional”, entre otros. La característica común entre los distintos constructos agrupados bajo este término paraguas es que recogen información directamente del paciente sin interpretación alguna por parte de médicos u otros profesionales sanitarios⁴.

Los instrumentos PRO no deben confundirse con escalas de puntuaciones clínicas, donde el especialista completa un formulario para puntuar la gravedad de la enfermedad o los efectos del tratamiento. En este sentido, la definición de la *Food and Drug Administration*⁵ es muy explícita: “Un PRO es cualquier informe del estado de salud de un paciente, que proviene directamente del paciente, sin interpretación de la respuesta del paciente por parte de un médico o cualquier otra persona”.

Independientemente del constructo medido, los instrumentos PRO se pueden clasificar en cuatro grandes grupos de instrumentos de acuerdo a dos ejes: la población a la que se dirigen (genéricos o específicos) y el modelo de medida aplicado (psicométricos o econométricos). Los instrumentos genéricos están diseñados para ser utilizados en todo tipo de individuos, mientras que los cuestionarios específicos están diseñados para sujetos con una determinada enfermedad, síntoma o característica. Los instrumentos psicométricos son perfiles descriptivos de las dimensiones o componentes de la salud, mientras que los econométricos son índices que resumen en un único valor global el estado de salud ponderado de acuerdo a las preferencias de la población. Todos estos ejes y conceptos se describen con detalle en el apartado de tipos de instrumentos PRO.

II. Historia de los resultados reportados por los pacientes (PROs)

En 1948, el primer volumen de la revista *Cancer* describía el desarrollo de la Escala de Funcionalidad de Karnofsky⁶, y un año más tarde la revista *JAMA* publicó el primer modelo de clasificación estandarizado para la artritis reumatoide que incluía, además de medidas de signos clínicos, una escala de capacidad funcional⁷. A pesar de estos destacados precursores, hasta la década de los 70 la medición de los PROs se limitó a las actividades de la vida diaria o a las escalas de un solo ítem que medían la capacidad funcional.

El desarrollo de instrumentos genéricos se inició en la década de los 70, en parte como resultado del apoyo externo del ‘*National Center for Health Services Research*’ en Estados Unidos. El desarrollo de perfiles psicométricos comenzó con el ‘*Human Population Laboratory*’⁸ publicado en 1971, que medía la salud física, mental y social, mientras que el ‘Índice del Estado de Salud’⁹ (que más tarde se convirtió en la ‘Escala de Calidad del Bienestar’) puede considerarse el origen de las medidas econométricas basadas en preferencias. De 1970 a 1981 surgieron casi una quincena de instrumentos genéricos¹⁰. Se desarrollaron principalmente perfiles psicométricos descriptivos para su uso en la investigación de los servicios de salud, mientras que los índices econométricos fueron diseñados para planificación de servicios sanitarios y establecimiento de prioridades.

A mediados de la década de los 80 empezó la era de la reducción psicométrica. Las preocupaciones sobre los costos y la carga de la recopilación de datos dirigieron el interés hacia la eficiencia práctica para su aplicación en la clínica y hacia la eficiencia métrica para su aplicación en los ensayos clínicos y en la investigación basada en los resultados. Todo ello produjo el crecimiento del diseño de instrumentos breves. Para entender esta tendencia, vale la pena recordar el ejemplo del ‘*Sickness Impact Profile*’¹¹ publicado en 1975, que medía 12 dominios con 136 ítems. Mientras que el primer cuestionario reducido, el ‘*Medical Outcomes Study SF-20*’¹² publicado en 1988, contenía tan solo 20 ítems. Muchos instrumentos genéricos desarrollaron versiones abreviadas^{13,14} en la década de los 90.

Los avances metodológicos conseguidos en los años 70 y 80 con el auge en el desarrollo de los instrumentos genéricos y la selección de ítems para diseñar versiones reducidas se transmitieron unos años más tarde a los instrumentos PRO específicos de enfermedad. Al principio para pacientes con patologías discapacitantes en los que los resultados de calidad de vida eran especialmente relevantes, como el cáncer, la artritis, el accidente cerebrovascular o la enfermedad cardíaca, para extenderse posteriormente a prácticamente todas las especialidades médicas. La expansión prolífica se produjo en la década de los 90, debido al creciente reconocimiento de que el mantenimiento y la mejora del funcionamiento y bienestar del paciente debe ser un objetivo importante de la atención médica, particularmente para los ancianos y las personas con enfermedades crónicas¹⁵.

En España, los primeros pasos en este campo fueron dados por la escuela de psicólogos de Zaragoza quienes publicaron la adaptación transcultural del Cuestionario de Goldberg en 1979¹⁶. En 1990 se publicaron las versiones española y catalana del Nottingham Health Profile¹⁷, y entre 1994 y 1999 las versiones españolas del ‘*Sickness Impact Profile*’¹⁸, del Euroqol¹⁹, del Cuestionario de Salud SF-36²⁰ y de las viñetas COOP/WONCA²¹. De la misma manera, en la década de los 90 se inició la proliferación de adaptaciones transculturales de instrumentos específicos^{22,23}, que ha continuado creciendo exponencialmente hasta nuestros días. Sin embargo, el desarrollo de instrumentos *de novo* ha sido escaso y la mayoría de los instrumentos PRO disponibles en español son versiones derivadas²⁴ de medidas desarrolladas en otros países.

Una revisión sistemática de artículos publicados en inglés entre 1990 y 1999 sobre el desarrollo o la evaluación de instrumentos PRO²⁵ identificó 3.921 con información sobre 1.275 instrumentos: 46% específicos de patología o población, 22% genéricos, 18% específicos de dimensión, 10% índices basados en preferencias, y un 1% instrumentos individualizados. Los resultados de esta revisión permitieron cuantificar el crecimiento exponencial de nuevos instrumentos PRO, cuya producción se multiplicó por cinco en una década: desde los 144 artículos de 1990 hasta los 650 publicados en 1999.

Una revisión sistemática similar realizada en España también en 1999²⁶ identificó tan solo 61 instrumentos en total (48 desarrollados originariamente en otros países): 11 genéricos, 32 específicos de patología, 13 específicos de dimensión, y 5 de capacidad funcional. Esta discrepancia entre los más de mil instrumentos en inglés y los 61 en español refleja el retraso en la incorporación de los instrumentos PRO en nuestro país. Actualmente, sin embargo, la información más reciente procedente de BiblioPRO (biblioteca virtual de instrumentos PRO en español) indica que existen más de 1.300 instrumentos disponibles²⁴ para su uso en España.

Finalmente, la única revisión²⁷ reciente identificada sobre instrumentos PRO, realizada en la base de datos del clinicaltrials.gov, concluía que de los casi cien mil estudios localizados entre 2007 y 2013, un 27% utilizaron uno o más de estos instrumentos. No solo se sugería un incremento en la utilización de estas medidas, sino una continuidad en esta tendencia.

1. Tipos de instrumentos PRO

Desde finales de la década de 1980 se han realizado diferentes propuestas de clasificación²⁸⁻⁴², la mayoría diseñadas específicamente para la calidad de vida relacionada con la salud. Una de las primeras clasificaciones desarrolladas para los instrumentos PRO es la propuesta publicada en 2008 por Valderas y Alonso⁴³, quienes, a partir de las previas, identificaron un set mínimo común compuesto de tres ejes: el constructo, la población y el modelo de medida.

Constructo

El constructo es el rango de conceptos o dominios medidos por un instrumento, su objeto de medida. Basándose en el modelo conceptual de Wilson y Cleary⁴⁴, se diferenciaron cuatro categorías de constructo (síntomas, estado funcional, percepciones de salud y calidad de vida relacionada con la salud), y se añadió una quinta categoría que incluía otros constructos no especifi-

cados en el mencionado modelo, tales como: satisfacción con el cuidado de salud, o resiliencia. A continuación se define cada una de estas categorías y se aporta algún ejemplo de ítem.

El constructo síntoma es la percepción por parte del paciente de un estado anormal físico, emocional o cognitivo. Miden síntomas el ítem del SF-36 versión 1 “¿Tuvo dolor en alguna parte del cuerpo durante las 4 últimas semanas? ninguno, muy poco, un poco, moderado, mucho, o muchísimo”; y el ítem del Cuestionario Respiratorio St. George “Durante el último año ¿ha tenido tos casi todos los días, varios días de la semana, unos pocos días al mes, solo cuando tuvo infección en los pulmones, o nada en absoluto?”.

El estado funcional es la capacidad del individuo para realizar sus actividades. Son ejemplos de este constructo el ítem del Cuestionario Respiratorio St. George “Tardo mucho para hacer trabajos como las tareas domésticas, o tengo que parar a descansar”; y el ítem del UCLA-Dizziness Questionnaire “Cuando tengo vértigo, afecta a mis actividades diarias tales como trabajar, conducir, hacer la compra, cuidar de la familia, cuidar de mí mismo”.

Percepciones de salud son integraciones subjetivas de toda la información relacionada con un síntoma o estado funcional, como por ejemplo el ítem del SF-36 versión 1 “Creo que me pongo enfermo más fácilmente que otras personas: totalmente cierta, bastante cierta, no lo sé, bastante falsa, totalmente falsa.”

Calidad de vida relacionada con la salud se refiere a aquellos aspectos de la calidad de vida específicamente relacionados con la salud. Mide este constructo el ítem del “*The Functional Assessment of Chronic Illness Therapy* (FACIT): “Estoy satisfecho(a) con mi calidad de vida actual: nada, un poco, algo, mucho, o muchísimo”; y el del UCLA-Dizziness Questionnaire “¿Cómo influye mi problema en mi calidad global de vida? no influye para nada, influye algo, influye de una manera moderada, de forma importante, o de forma decisiva.”

Satisfacción con el cuidado de salud es el grado en que la experiencia de un paciente con el sistema sanitario es comparable a sus expectativas. Miden satisfacción el cuestionario SATFAR (Satisfacción con el servicio de dispensación farmacéutica) con el ítem “Cuando acudí a la farmacia para adquirir medicamentos el personal se mostró interesado en colaborar conmigo en el manejo de mis medicamentos: totalmente en desacuerdo, en desacuerdo, ni de acuerdo ni en desacuerdo, de acuerdo, muy de acuerdo”; y el FACIT *Treatment Satisfaction Questionnaire* para evaluar la satisfacción con el tratamiento de enfermedades crónicas con el ítem “En comparación con lo que esperaba, ¿cómo califica la eficacia del tratamiento hasta el momento: mucho peor, un poco peor, más o menos, un poco mejor, o mucho mejor?”.

Resiliencia es la habilidad para hacer frente y resistir el sufrimiento y la enfermedad, como por ejemplo el ítem de la versión para adolescentes del CHIP (*Child Health and Illness Profile*) “Imagínate que has tenido una discusión importante con un amigo íntimo y crees que él/ella no te ha entendido ni te ha escuchado: es muy improbable, improbable, probable, o muy probable que pidiera consejo a otras personas.”

Población La población de un instrumento PRO es el universo de personas para quienes es adecuado. Puede estar definido en términos de edad, género, enfermedades, síntomas o cultura. Todas estas variables del individuo son indudablemente relevantes para su caracterización desde un punto de vista clínico, epidemiológico u organizacional. Las bien conocidas diferencias en salud entre las subpoblaciones definidas de acuerdo con estas variables justifican el desarrollo de instrumentos PRO específicos.

Aunque los instrumentos PRO más numerosos son los diseñados para adultos, existe cada vez un mayor número desarrollado específicamente para niños o adolescentes. Estos últimos, generalmente disponen de diferentes versiones según las edades pediátricas. Por ejemplo, el CHIP (*Child Health and Illness Profile*) tiene versiones para niños de 6 a 11 años (*Child Report Form*, CHIP-CE/CRF), para adolescentes de 12 a 18 años (*Adolescent Edition*, CHIP-AE), y para los padres.

Los instrumentos PRO específicos de género, aunque pueden ser adecuados para cualquier tipo de estado de salud, lo más frecuente es que se desarrollen en patologías o etapas de la vida propias de un solo género. Por ejemplo, los instrumentos PRO diseñados para pacientes con cáncer de próstata son específicos del género masculino, mientras que aquellos para pacientes con cáncer de ovario, hipermenorrea, o para medir el impacto de la menopausia lo son del género femenino. Finalmente, más allá de la definición de cultura que está fuera de nuestro alcance, para la clasificación de los instrumentos PRO que nos concierne son relevantes tanto el lenguaje como el país de la población para la cual ha sido diseñado.

Modelo de medida Los dos puntos clave del modelo de medida de un instrumento PRO son el modelo teórico que sustenta su métrica y el nivel de agregación de su puntuación o dimensionalidad.

La métrica hace referencia al método utilizado para asignar valores numéricos a las respuestas dadas por el individuo y al método utilizado para la construcción de puntuaciones. Se distinguen tres grandes grupos de instrumentos PRO: psicométrico, econométrico y clinimétrico^{28,29}. Los algoritmos de puntuación de los instrumentos psicométricos y clinimétricos se basan generalmente en la suma de las respuestas de los ítems que componen la escala (con o sin pesos). La principal diferencia entre ellos reside en el método seguido para el desarrollo de la escala: los psicométricos se construyen sobre modelos teóricos aplicando análisis estadísticos sofisticados para testar dichos modelos, mientras que los clinimétricos se basan casi exclusivamente en la relevancia clínica.

Los instrumentos econométricos aparecieron para evaluar los estados de salud y asignarles un valor separadamente. Su propósito es obtener valores basados en las preferencias sobre los estados de salud de los pacientes, las poblaciones o los expertos, entre otros. Para ello, utilizan métodos que provienen del campo de la econometría basados en la teoría de la toma de decisiones. Estas preferencias son conocidas con el nombre de utilidades, que cuando se multiplican por los periodos de tiempo correspondientes dan lugar al cálculo de Años de Vida Ajustados por Calidad (AVAC).

La dimensionalidad hace referencia al número de puntuaciones generadas, distinguiéndose índices y perfiles. Estas categorías no son exclusivas, algunos instrumentos PRO producen solo una puntuación índice, otros solo perfiles, mientras que hay instrumentos que pueden generar ambos tipos de puntuaciones.

2. Utilización rutinaria de PROs

En los años 90 del siglo pasado se fueron dando una serie de fenómenos que promovieron un gran interés por incorporar los PROs en la atención clínica. Por un lado, la existencia de un número considerable de instrumentos de medida suficientemente válidos y robustos. También promovieron el uso de PROs los estudios ambiciosos de investigación centrados en los resultados de salud (como el *Medical Outcomes Study* –MOS–⁴⁵ o los *Patient Outcomes Research Teams* –PORTs–⁴⁶. Así como el énfasis en los programas de mejora de la calidad.

Tipos de uso rutinario de PROs

Van der Wees y cols.⁴⁷ diferencian tres tipos de utilización rutinaria de los PROs. Un propósito inicial sería el de guiar las decisiones clínicas relacionadas con el diagnóstico, plan de tratamiento y su evaluación. Esta es la utilización más frecuente en la mayoría de casos por parte de los proveedores de cuidados de salud y supone un uso individualizado de los PROs, que es simultáneo al proceso de atención sanitaria. Un segundo tipo de aplicación es aquél llevado a cabo a nivel de organizaciones ya sean a nivel regional o estatal y cuya finalidad no es influir en la atención clínica individual directamente sino facilitar la evaluación del rendimiento de las organizaciones proveedoras de servicios. Quizás el ejemplo más emblemático de este tipo de aplicación es el programa PROM del *National Health Service* (NHS) inglés. También hay ejemplos en los Estados Unidos (la encuesta sobre resultados de salud –HOS– en una muestra de pacientes del programa público *Medicare*) y en Holanda, en pacientes que han finalizado el tratamiento médico⁴⁷. En estos casos la recogida de datos se realiza de manera externa y desligada de la provisión del servicio.

La tercera opción es un enfoque mixto, donde las organizaciones sanitarias recogen los datos en el momento de la atención con un uso clínico y posteriormente agregan los datos para informar sobre el rendimiento a nivel regional o estatal. Así, el estado de Minnesota, en los Estados Unidos, incluye PROs de manera sistemática en la historia clínica electrónica y utiliza los datos para hacer comparaciones de la calidad de los proveedores del estado. En Holanda, en el contexto de la atención a la salud mental, la recogida sistemática de PROs se vuelca en una base de datos para desarrollar estándares de calidad a nivel nacional⁴⁷.

En los siguientes apartados revisamos con detalle las evidencias que existen hasta ahora del uso sistemático de los PROs en la práctica. En primer lugar, revisaremos los trabajos de investigación que han evaluado el impacto del uso sistemático de PROs en la atención individual del paciente. Posteriormente revisaremos el ejemplo más paradigmático del uso de los PROs en la evaluación de la calidad de los proveedores: el programa PROMS-NHS inglés después de casi 7 años de experiencia.

*Uso rutinario en
la práctica clínica:
oportunidades y
barreras*

Oportunidades

Es posible que el interés por el uso de las medidas PRO en la práctica clínica sea mayor entre los investigadores que entre los médicos en ejercicio. En todo caso, hay un cierto acuerdo en que el uso de medidas PRO en la práctica clínica puede tener diversos beneficios potenciales⁴³. En primer lugar, la información recogida mediante cuestionarios estandarizados puede facilitar la detección de problemas físicos o psicológicos que de otro modo podrían pasar inadvertidos. Las medidas PRO también se pueden aplicar como instrumentos estandarizados para el seguimiento de la progresión de la enfermedad y para evaluar el impacto del tratamiento prescrito. Otro beneficio del uso rutinario de las medidas PRO en la atención clínica diaria sería el de facilitar la comunicación médico-paciente, promoviendo así un modelo de toma de decisiones compartida. Con mucha frecuencia, los pacientes y los clínicos deben establecer prioridades comunes y compartir las expectativas sobre los resultados del tratamiento y de la enfermedad. El establecimiento de una base de comprensión conjunta puede facilitar la satisfacción de las diversas necesidades de los pacientes, y así mejorar la satisfacción de los pacientes con la atención de salud que reciben y también su adherencia al tratamiento prescrito. Finalmente, las medidas PRO en la atención clínica pueden ser utilizadas también para evaluar los resultados de la atención y/o como una estrategia para mejorar la calidad de la misma o para recompensar a los profesionales cuyos procesos y resultados sugieran una mayor calidad de la atención⁴⁸.

A esos puntos se pueden añadir otros objetivos de interés, como pueden ser su contribución a mejorar la evaluación de los verdaderos objetivos de la atención a la salud; a estandarizar las mediciones al máximo número de niveles: los pacientes con patologías diferentes, los diversos tipos de profesionales sanitarios y a través de los diferentes niveles de atención. Los PRO pueden hacernos comprender mejor la relación entre los procesos de atención y sus resultados así como averiguar si los esfuerzos y los costes de la atención a la salud tienen efectos adversos sobre los resultados de los pacientes⁴⁹.

Barreras

Quizás deberíamos empezar reconociendo que ya hace mucho tiempo que los profesionales de la salud utilizan PROs en la práctica clínica (incluso, quizás, sin saberlo). Es muy frecuente incluir en la historia clínica preguntas globales sobre la salud percibida, el dolor o la dificultad de respiración o el ahogo. En bastantes casos esas preguntas se hacen de manera completamente estandarizada (por lo que se refiere al texto de la pregunta y a las opciones de respuesta que se le ofrecen al paciente). También se utilizan con cierta frecuencia, especialmente en poblaciones más frágiles, escalas para el cribado de la discapacidad, el deterioro cognitivo o incluso para trastornos mentales comunes.

Pero también es cierto que los profesionales sanitarios se encuentran con la existencia de ciertas barreras que les hace no ser del todo optimistas respecto a la factibilidad de implementar la medición de PROs en la práctica clínica diaria. Algunos problemas son de carácter inmediato: una disponibilidad de tiempo en general insuficiente, la falta de información clara y concreta

del uso las medidas PRO, especialmente en relación a su administración y, sobre todo, el cálculo de las puntuaciones y su interpretación o el acceso a las normas poblacionales u otras más pertinentes con las que comparar las puntuaciones de sus pacientes. Otras cargas a tener en cuenta son las administrativas, que incluyen la dificultad para obtener el permiso de uso de algunos instrumentos y los costes de su uso, en particular para aquellos patentados⁴⁸. Podemos señalar también otras barreras más conceptuales. Por ejemplo, la idea que el uso de medidas PRO puede forzar al paciente a abordar problemas que no quisiera comentar con su médico. También es posible que el profesional otorgue una importancia diferente a la información escrita, en comparación con la más rica y detallada obtenida durante la realización de la historia clínica. Y una cierta desconfianza en que una información recogida con finalidades de atención individual se convierta en un indicador de calidad asistencial. Todas esas barreras causarían reticencia por parte de los profesionales para la incorporación de las medidas PRO en su práctica profesional.

Estudios cualitativos realizados en profesionales y pacientes⁵⁰ identificaron algunas barreras y dificultades para la implementación rutinaria de los PRO. Los profesionales tienden a considerar como inconvenientes importantes la falta de recursos específicos y la inadecuada preparación de los consultorios para este tipo de tarea. También consideran un problema la carga que resulta para los pacientes tener que responder esos instrumentos, así como la dificultad importante a la hora de interpretar los resultados, que en parte son redundantes con la historia clínica, una actividad que consume un tiempo muy valioso para el clínico. Y, finalmente, el desconocimiento, o incluso la desconfianza sobre la efectividad de esa actividad y de su uso para la evaluación de los propios facultativos.

Por su parte algunos de los pacientes se sentían contrariados si su médico no demostraba una reacción clara ante sus respuestas al instrumento PRO. Y también consideraban que algunas preguntas (y opciones de respuesta) resultaban dificultades de leer o de contestar. O incluso, que algunas preguntas sobre sus emociones sean demasiado intrusivas. Todos estos inconvenientes parecen muy realistas y son sin duda una barrera a superar.

Impacto del uso rutinario de PROs en la práctica clínica

La medición formal del impacto de la utilización rutinaria de PROs se ha estudiado en primer lugar, en el contexto de la atención clínica individual, especialmente en el contexto oncológico. Las bases sobre las que se asienta esta estrategia es el uso de PROs con tres finalidades distintas: como instrumentos de cribaje o diagnóstico; como un elemento más para desarrollar el plan de tratamiento; y como una medida de la evaluación del tratamiento. En los estudios más iniciales, la implantación de la administración e interpretación de los PROs era *ad hoc*, con diferencias entre los estudios. Las siguientes experiencias ya incorporaban la tecnología necesaria para que esa información se incluyera de manera automática la historia clínica y estuviera disponible en el transcurso de la visita médica.

Uno de los primeros ensayos clínicos que examinó el beneficio potencial del uso de PROs en la práctica clínica fue publicado en JAMA⁵¹. Con un diseño de ensayo clínico aleatorizado cruzado, el estudio demostró que

su uso sistemático, junto con la retroalimentación de la información a los oncólogos en las visitas de seguimiento, mejoraba la comunicación médico-paciente. En concreto, surgían más temas de interés y preocupación para los pacientes durante la visita médica. Y también se observó un alto grado de satisfacción por parte de los pacientes, sin aumentar prácticamente el tiempo de la visita. Desde entonces, diversas revisiones sistemáticas de la literatura^{43, 52-55} dan soporte a la evidencia que el uso de PRO mejora la comunicación médico-paciente, la mejoría en el registro de los síntomas de los pacientes en las historias clínicas y la influencia de la información basada en los PROs en la toma de decisiones. Sin embargo, aún son pocos los estudios que muestran un beneficio sobre los resultados en salud de los pacientes cuyos PROs se han medido e informado a los profesionales sanitarios.

Un par de revisiones sistemáticas recientes sintetizan (sin metaanalizar los resultados) 27 y 26 publicaciones, respectivamente, sobre estudios controlados en el ámbito de la atención oncológica^{3, 56}. Como no podía ser de otra forma, ambos trabajos presentan coincidencias muy notables. El estudio más reciente confirma que el uso rutinario de PROs aumenta la discusión de los resultados relevantes para el paciente en las consultas (aspectos de función social, sexual y emocional). También identifica que, en algunos estudios, pero no en todos, el uso de PROs se asoció con un mejor control de los síntomas físicos (tamaños del efecto entre pequeños y moderados). La mejoría en el control de los síntomas psicológicos fue menor. También se observó un mayor uso de medidas de soporte en el grupo experimental (los pacientes a los que se les administraba el PRO y sus médicos que recibían los resultados). La revisión de Chen y cols.³ aporta una tabla (Tabla 1) como resumen narrativo de los resultados observados. Existe evidencia de nivel muy alto en relación a la mejora de la comunicación y a la satisfacción por parte de los pacientes. El nivel es alto para la mejoría en la monitorización del tratamiento y en la detección de problemas previamente no detectados. Parece que no hay ningún impacto sobre los resultados de los pacientes. Los estudios controlados analizados no aportan información (no ya evidencia) sobre aspectos organizativos y de calidad de la atención. Sin embargo, estos aspectos se tienden a evaluar con otro tipo de diseños que con alta probabilidad no cumplirían los criterios de inclusión de estas revisiones. En todo caso, la experiencia del programa PROMS-NHS que se discute más adelante, aporta información relevante para estos aspectos.

Modelo teórico

Greenhalgh y cols.⁵⁷ comentaban, a raíz de las primeras revisiones sistemáticas que mostraban una falta de impacto del uso rutinario sobre los resultados en salud de los pacientes, que quizás el problema residía en no haber comprendido teóricamente los mecanismos a través de los cuales, usar PROs en la práctica clínica puede mejorar el resultado de la misma. La figura 1 representa las asunciones e hipótesis de los estudios de impacto realizados hasta ese momento. En él se representan los elementos y posibles mecanismos a través de los cuales la información basada en las medidas PRO en la práctica clínica podría llegar a mejorar los resultados de salud y la satisfacción con la

TABLA 1. Dirección y fortaleza de la evidencia en el uso rutinario de PROs en atención clínica en contexto oncológico.

Impacto	Dirección y fortaleza de la evidencia
Comunicación Paciente/Profesional	+++
Mejora Satisfacción del Paciente	+++
Monitorización Respuesta Tratamiento	++
Detección de problemas no identificados	++
Cambios en el manejo del paciente	+
Mejoría en los resultados	+/o
Cambios en las conductas del paciente	n/a
Mejoría de la calidad de la atención	n/a
Aumento de la transparencia/ <i>accountability</i>	n/a
Mejor funcionamiento del sistema en general	n/a

Escala de: ++++ Máximo efecto positivo a ---- máximo efecto negativo; n/a: sin datos.

Fuente: Chen J, et al. BMC Health Serv Res. 2013³.

atención recibida. El esquema muestra la complejidad de los pasos intermedios. Además, los autores señalan una serie de asunciones que tendría este modelo, que es conceptualmente simple: que la calidad de vida relacionada con la salud puede ser medida con el instrumento PRO y que realmente refleja la perspectiva del paciente; que el médico es el profesional más adecuado al cual devolverle la información sobre calidad de vida del paciente; que el proceso de toma de decisiones se produce en un solo momento en el tiempo (el de la visita médica); y, finalmente, que la educación por sí sola es capaz de lograr que los médicos utilicen los PRO en su proceso de toma de decisiones clínicas.

En su revisión, Greenhalgh y cols.⁵⁷ denuncian que los ensayos realizados hasta ese momento no habían sido óptimos para evaluar el impacto potencial del uso rutinario de PROs en la práctica clínica. Y que hay que comprender mejor los complejos procesos subyacentes, lo que contrasta con la simplicidad de la intervención que se evalúa. Hay que probar hasta qué punto los pacientes quieren hablar de su calidad de vida con los médicos. Y hasta dónde los clínicos consideran adecuado hablar de la calidad de vida de los pacientes y consideran esa información con suficiente relevancia clínica como para introducir cambios en su plan terapéutico.

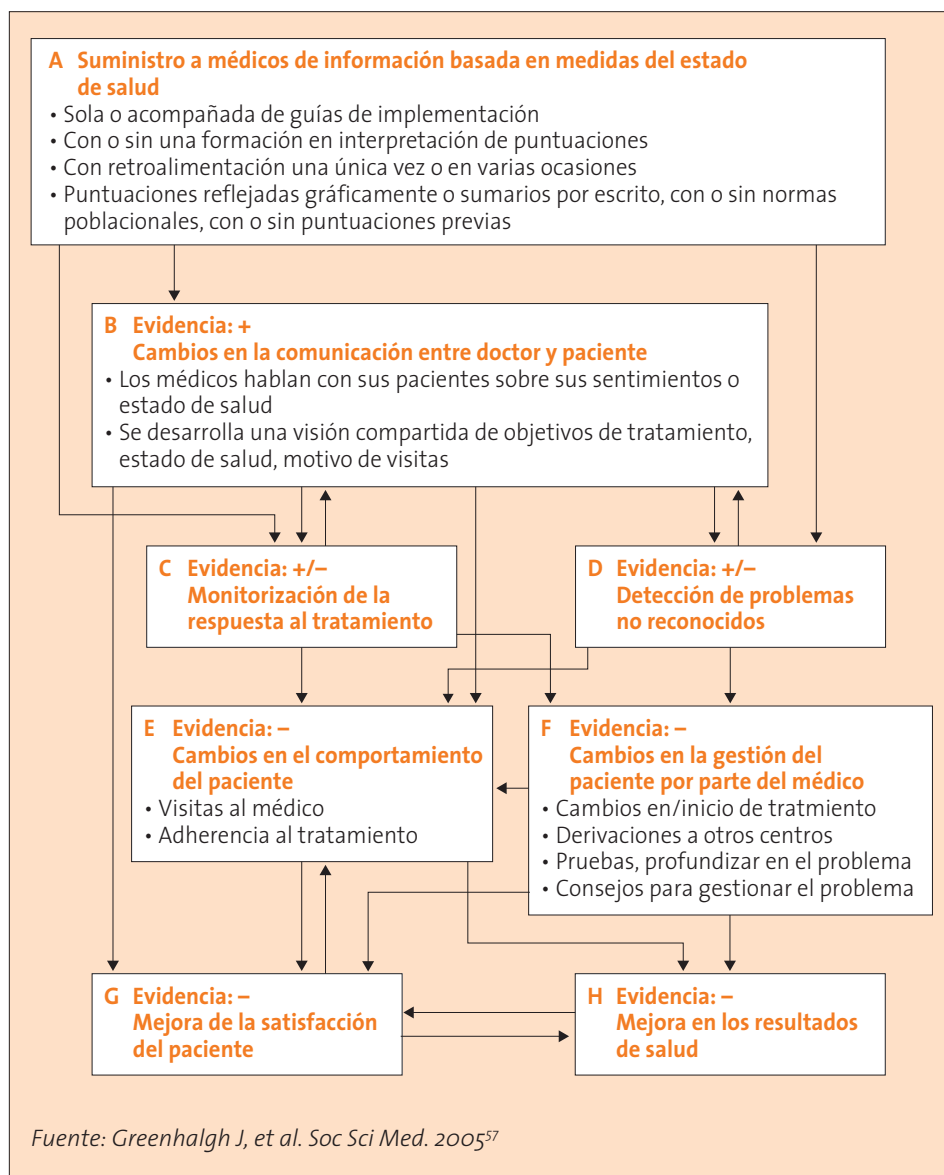


FIGURA 1. Resultados e hipótesis en las pruebas de evaluación del impacto en la toma de decisiones clínicas de medidas del estado de salud.

III. Aplicaciones de los instrumentos PRO

1. El programa NHS-PROM inglés: Los PROs para mejorar la calidad de la atención

En el contexto del Servicio Nacional de Salud (NHS) inglés, el uso rutinario de PROs en la práctica clínica ha sido impulsado para comparar los resultados entre proveedores (en general, hospitales). La idea subyacente fue que los profesionales aumentarían su productividad al evitar tratamientos innecesarios. También se pensó que se produciría una mejora de la calidad de la atención clínica al influir en el re-diseño de los servicios y en la elección del paciente⁵⁸. Este uso rutinario de PROs con finalidad de mejorar la calidad de los servicios y poder valorar el rendimiento de los proveedores es reciente, si bien se está extendiendo en diversos países, incluyendo los Estados Unidos (Darmouth y el Estado de Minnesota), Australia y Suecia (Karolinska) y su potencial ha sido reconocido en Canadá, Nueva Zelanda y otros países⁵⁹. Pero es el caso de Inglaterra, por su magnitud y nivel de evaluación, el que merece nuestra atención en esta monografía.

Desde abril del 2009, el Departamento de Salud de Inglaterra obligó a que todos los proveedores de asistencia financiada por el NHS invitaran a completar PROs a los aproximados 250.000 pacientes que se someten anualmente a una de las cuatro intervenciones quirúrgicas de carácter electivo seleccionadas para el Programa Marco NHS-PROM: prótesis de cadera, prótesis de rodilla, hernia inguinal y venas varicosas.

Como fruto de este programa, a todos los pacientes elegibles se les invita a rellenar un cuestionario preoperatorio recogiendo información sociodemográfica y de las comorbilidades del paciente, que incluye un PRO específico y un PRO genérico (el EQ-5D). Después de la intervención (a los 3 meses en el caso de las intervenciones de hernia y varices; a los 6 en el caso de las prótesis de cadera y rodilla), los pacientes reciben por correo otro cuestionario, de seguimiento, que incluye los PRO genérico y específico correspondientes y también preguntas sobre complicaciones, así como una pregunta de cambio global. Los PRO específicos son: *Oxford Hip Score* (OHS), *Oxford Knee Score* (OKS), *Aberdeen Varicose Vein Questionnaire* (AVVQ) (no se usa instrumento específico en el caso de las herniorrafias).

Estos datos se utilizan con diversas finalidades⁶⁰. Primero, como una medida de resultado del sistema de salud, como parte del Marco de Resultados del Gobierno. También lo utilizan como una manera de evaluar los resultados de las organizaciones que proveen servicios al NHS, y para incentivar la mejora de resultados incluyendo los PRO en los mecanismos de pago (por ejemplo, en las tarifas de mejores prácticas en las prótesis de cadera y rodilla, desde 2014). Organismos reguladores como la Comisión de Calidad de la Atención incorporan la información sobre PROs en sus actividades de monitorización. Asimismo, los médicos de atención primaria y los pacientes utilizan los datos accesibles sobre resultados de la cirugía para la elección de centro de referencia. También cabe destacar que, a nivel de atención primaria, se ha incluido un PRO genérico (el EQ-5D) dentro del *GP Patient Survey* (GPPS), una encuesta sobre la experiencia de los usuarios de los servicios de atención primaria que

se administra a unos 900.000 pacientes cada año, la mitad de los cuales padecen trastornos crónicos [<https://gp-patient.co.uk>]. En la actualidad se están realizando estudios piloto para valorar la posibilidad de expandir el programa PROMS a un mayor abanico de problemas de salud y procedimientos financiados por el NHS, desde trastornos mentales, atención oncológica, condiciones respiratorias y cardiovasculares crónicas⁶⁰.

Implementación y resultados del programa

El programa NHS-PROM lleva ya más de 6 años en marcha. Los primeros resultados de datos pre-operatorios fueron publicados en abril del 2010, como una estadística experimental, con actualizaciones mensuales posteriores. Los resultados de los análisis postoperatorios se publicaron en septiembre del 2010, de nuevo como una estadística experimental. A partir del 2011, los datos se presentaron separados por tipos de instrumento PRO para permitir la comparación a lo largo del tiempo. Un hito importante, en enero del 2012, fue que los resultados dejaran de considerarse una información de carácter experimental, pasando a ser un componente de las estadísticas oficiales (*Official Statistics*). En el sitio web <http://www.hscic.gov.uk/proms> aparece la información actualizada sobre este programa.

Participación

Las tasas de respuesta al programa NHS-PROM han sido modestas (68% en el caso de los pacientes intervenidos de prótesis de cadera y rodilla y 45% para los demás procedimientos, de menor importancia)⁵⁸. Sin embargo, este problema no parece afectar a la validez de las comparaciones de los resultados entre los hospitales, ya que los resultados se dan de manera ajustada, no cruda, teniendo en cuenta los factores conocidos asociados con la falta de respuesta (como son la edad, el nivel socioeconómico y la raza). Por otro lado, la respuesta postoperatoria es buena (85% para prótesis de cadera y rodilla; 65-75% para los procedimientos menores)⁶¹.

Comparación de proveedores

Esta comparación es uno de los principales objetivos del programa NHS-PROM. Comentaremos una aproximación muy habitual a la hora de plantear la comparación relativa de proveedores, utilizando una gráfica (Fig. 2) extraída, con permiso, de la publicación de Black⁶². La gráfica representa un “*funnel plot*” (gráfico tipo embudo) del cambio promedio en la puntuación de la escala Oxford de la Cadera (OHS) para cada uno de los centros proveedores informantes, posicionados según su valor medio de cambio en la OHS (eje vertical) y el número de procedimientos realizados en el período evaluado (eje horizontal). La línea horizontal representa el cambio promedio global (20 puntos de mejora, en este caso). Los puntos representan los proveedores (los valores son reales, pero los nombres son ficticios). Cada proveedor puede compararse con la media global y la distancia puede interpretarse, teniendo en cuenta los intervalos de confianza de las estimaciones, según los descriptores usados por el autor: mucho mejor/mucho peor que la media, cuando se sitúa por fuera del 99,8% de intervalo de confianza; mejor/peor que la media, cuando lo hace por fuera del 95%, pero no del 99,8% de ese intervalo. La forma de embudo se debe a que los intervalos de confianza son mucho más pequeños (precisos) para cuando hay un alto número de observaciones (procedimientos en este caso). Escogiendo como criterio de diferencia el más exigente (99,8%) existen

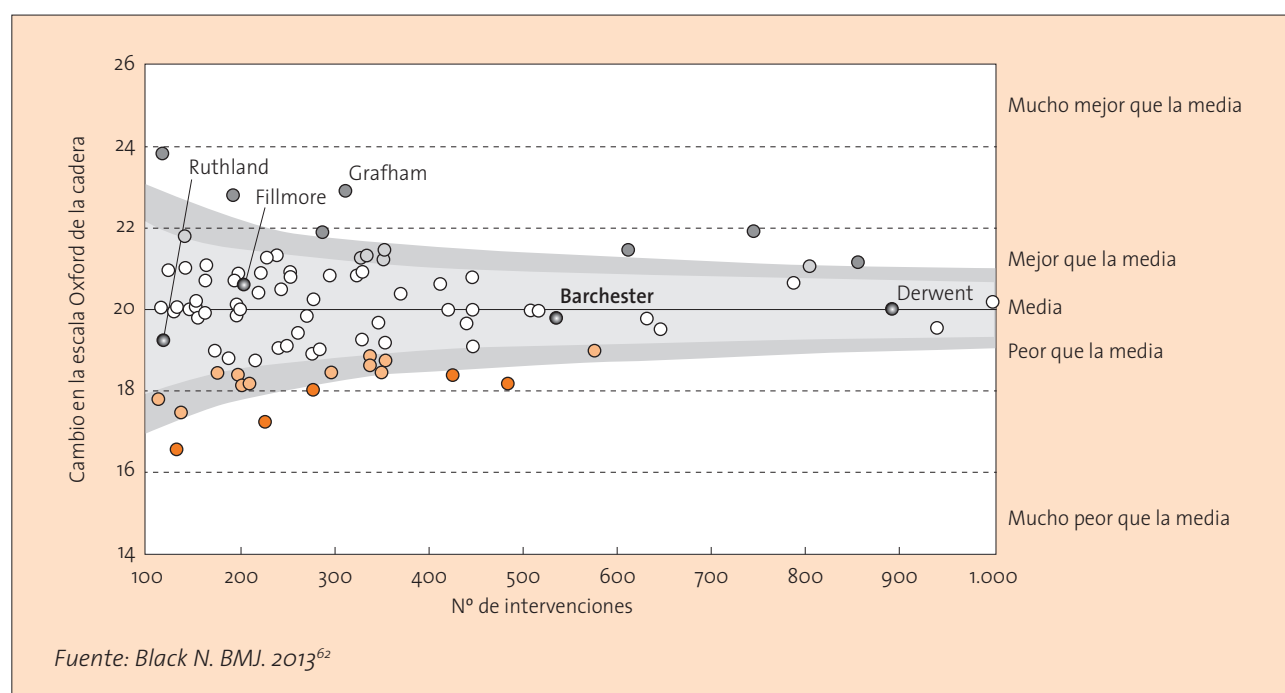


FIGURA 2. “Funnel plot” del cambio promedio en la puntuación de la escala Oxford de la Cadera para cada uno de los centros proveedores.

unos cuantos proveedores con resultados mucho mejores que la media (en gris). Estos centros se dan tanto en la parte derecha de la gráfica (alto volumen de cirugía) como en la izquierda (menor volumen).

Ajuste de los datos

Para evitar sesgos evidentes en las comparaciones entre proveedores es necesario ajustar los datos por las variables que se sabe que influyen en la casuística atendida. Estas variables incluyen, en la mayoría de casos: edad, sexo, estado de salud general, comorbilidad y el índice de privación por un lado y, la puntuación PRO preoperatoria, por el otro. La discusión de los modelos estadísticos utilizados queda fuera del ámbito de esta monografía. Una revisión interesante se puede encontrar en estos documentos: Browne y cols.⁶³, Appleby y cols.⁶⁰.

Asociación de los resultados con factores organizativos

La Figura 3 se hizo con el fin de evaluar hasta qué punto la variación en las tasas de cirugía electiva estaba influenciada por los cambios en la necesidad percibida (medida por los PRO). En concreto, y según las convenciones sobre la variabilidad en la práctica clínica, se esperaba que un aumento en las tasas de cirugía se correspondiera con una menor gravedad de los pacientes. La figura presenta, en el eje vertical, los cambios observados en los valores preoperatorios medios entre los años 2009/10 y 2011/12 en los cuestionarios específicos de las patologías (excepto para el cuarto procedimiento, herniorrafia, que corresponde a las puntuaciones del EQ-5D). En el eje horizontal, se representan los cambios (en porcentaje) en la tasa de cirugía en el mismo período. Los resultados no apoyan la hipótesis inicial. Otros análisis similares se centran en otras influencias organizativas (por ejemplo, la propiedad privada o pública de los centros, o el volumen de cirugía realizado por los mismos).

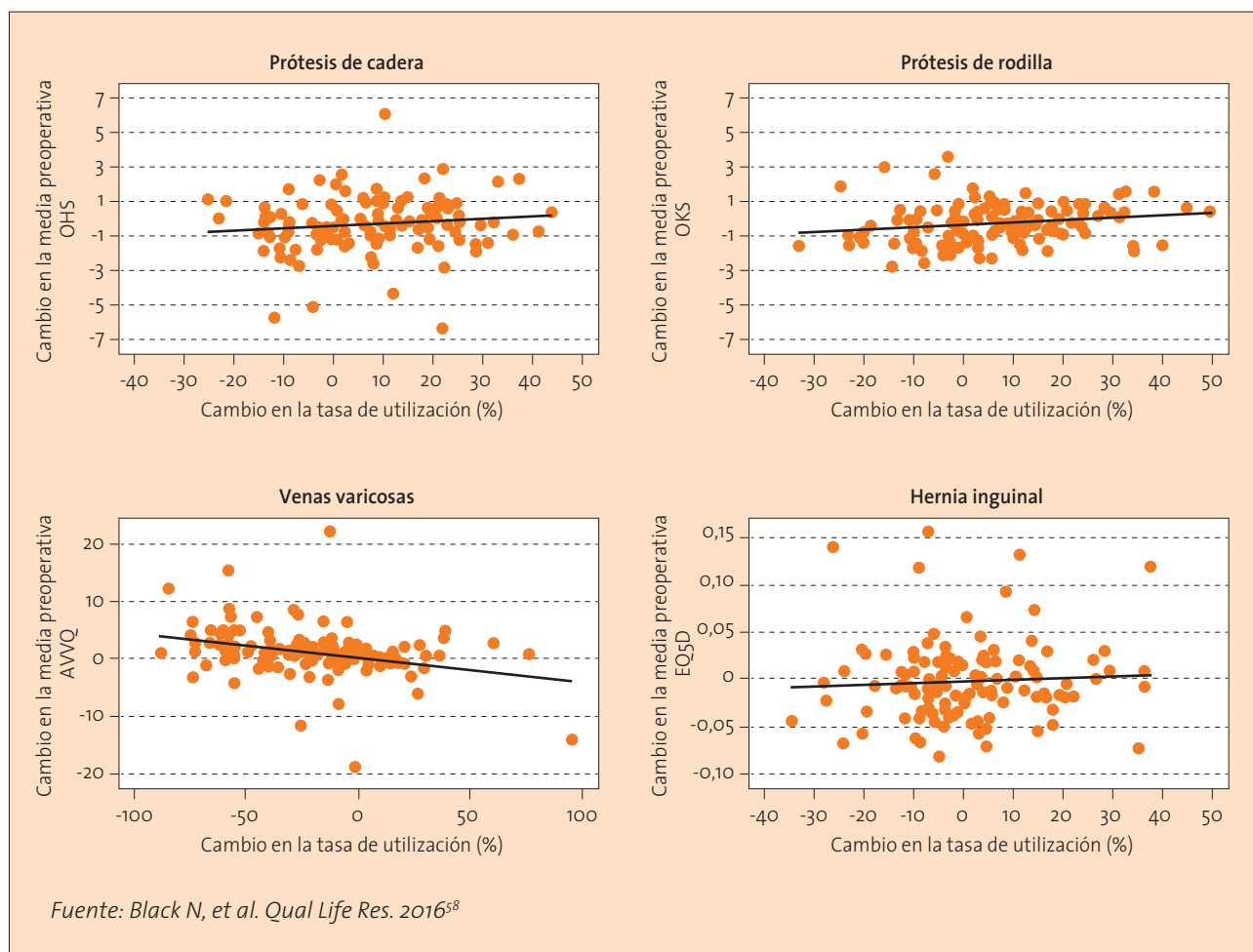


FIGURA 3. Cambios observados en los valores preoperatorios medios y cambios en las tasa de utilización o volumen quirúrgico (entre los años 2009/2010 y 2011/12, NHS Inglés).

Aspectos metodológicos

La elección de la métrica derivada de los PRO para identificar los proveedores que tienen malos resultados, tiene un impacto en la estimación de la proporción de proveedores cuya actuación es definida como “Mala”. Por ejemplo, para la prótesis de cadera, la media de riesgo de cambio ajustado en el PRO específico (Escala de Oxford), si se utiliza un indicador frecuentemente aplicado en *benchmarking*, como es una diferencia mayor a 2 desviaciones estándar de la media global, la proporción de hospitales con malos resultados es del 25%⁵⁸. Mientras que si se utiliza la proporción de pacientes que logran superar una diferencia mínimamente importante (MID), tan solo se identifica al 12% como inadecuados⁶⁴. Si el objetivo es aumentar la sensibilidad a la hora de identificar proveedores inadecuados, el primer indicador, puramente estadístico, es mejor. Ahora bien, si se prioriza evitar que se etiquete erróneamente a un proveedor como inadecuado, entonces el indicador de la MID es el óptimo.

Una preocupación previa al programa NHS-PROM para algunos cirujanos era que los pacientes pudieran confundir los resultados (la eficacia) de las intervenciones con su experiencia personal de la atención personal recibida (la calidad). Pero los datos analizados hasta ahora no apoyan esta hipótesis, puesto que se ha demostrado que solo hay una asociación muy débil ($r =$

0,2) entre las puntuaciones de los PROs y las de las medidas informadas por los pacientes de experiencia (PREMs)⁶⁵. Es decir, los PROs y PREMs miden diferentes aspectos y los pacientes pueden y deben distinguir los dominios de la salud percibida y los de la calidad de la atención recibida. Al final de la monografía volveremos a comentar esta cuestión.

Algunas lecciones sobre la práctica asistencial

La experiencia del programa NHS-PROM ha arrojado luz sobre algunos de los aspectos de la prestación de servicios de salud en Inglaterra y Gales. Como se acaba de ilustrar más arriba, se esperaba que el aumento de las tasas quirúrgicas en un área estuviera asociado con una disminución en la gravedad media preoperatoria de pacientes. Sin embargo, no se encontró una asociación significativa entre ambos indicadores⁶⁶. Esto sugiere que la decisión de operar está determinada en gran parte por los cirujanos y no reflejan realmente, o lo hacen en mucho menor grado, la necesidad percibida por los pacientes.

Es bien conocida la asociación entre el volumen hospitalario y la seguridad (por ejemplo, la mortalidad operatoria). Sin embargo, no se ha observado ninguna asociación entre el volumen quirúrgico y la eficacia medida con PROs (es decir, el cambio ajustado en las puntuaciones de los PRO no fue mejor para los proveedores con mayor volumen quirúrgico)⁶⁷. Además, también en comparación con la mortalidad, los PRO han mostrado menor variación individual entre cirujanos. Es decir, se encuentran más valores atípicos (± 3 DE) cuando la mortalidad postoperatoria se utiliza como el indicador de resultado⁶⁸.

Finalmente, los PRO tampoco dan sustento a la creencia generalizada de que los resultados son mejores cuando hay mayor competencia: no se ha observado asociación alguna entre la cantidad de competencia en una localidad y la media ajustada de cambio en las puntuaciones de PRO⁶⁹.

Conclusiones y retos de futuro

El programa NHS-PROM ha mostrado que resulta factible recoger información de PROs como una actividad rutinaria de la atención sanitaria, incluyendo también el seguimiento. Los retos logísticos y éticos son manejables. El programa ha dado lugar a una amplísima base de datos que ha sido utilizada en muchos análisis, los resultados de los cuales pueden servir sin duda para mejorar la información a la hora de tomar decisiones a múltiples niveles, desde el del ámbito clínico/paciente, al de los proveedores, los financiadores y los reguladores. La extensión de la recogida sistemática de PROs en la atención primaria y otros ámbitos puede suponer un salto cualitativo en todo este proceso.

El programa NHS-PROM no parece haber tenido impacto en una selección más adecuada de los pacientes ni en una mejora progresiva los resultados durante sus primeros años de funcionamiento⁷⁰. Esto podría explicarse por diversas razones: 1) los proveedores solo empezaron a recibir retroalimentación (información sobre los resultados de los pacientes) en el tercer año del programa; 2) la calidad de la presentación de esa información fue más bien pobre (en forma de hojas de cálculo muy poco amigables); y 3) para los cuatro procedimientos de cirugía electiva concernidos puede que exista muy poco margen de mejora, ya que todos los cirujanos y hospitales proporcionan ya una atención de buena calidad. Es muy probable que el margen de mejora

sea mucho mayor en las operaciones quirúrgicas de urgencia, por ejemplo. Parece claro que para conseguir una mejora en los procesos y en los resultados, la recogida sistemática de información PRO ha de combinarse con una mejora notable en el flujo de la información para los profesionales.

Uno de los principales retos de investigación para el futuro será clarificar las expectativas que deben depositarse en el uso rutinario de los PRO. Una síntesis realista de la literatura que está actualmente en marcha⁷¹ pretende ayudar a identificar los objetivos de los usuarios y los supuestos teóricos que subyacen a sus expectativas. Otros problemas metodológicos que deberán resolverse tienen que ver con el uso rutinario de los PRO en condiciones más exigentes como son los trastornos crónicos a largo plazo (sin un antes y después de fácil definición), los problemas de salud urgentes (sin una medición inicial posible), o la demencia (donde la disfunción cognitiva limita el grado de contribución de los propios pacientes).

Otros retos de las políticas de salud y la práctica clínica incluyen la necesidad de reunir las dos razones principales para utilizar PRO: la guía clínica y la evaluación del desempeño de los proveedores. Asegurar que el uso futuro combina ambos objetivos mejorará la calidad de los datos recogidos, y el compromiso de pacientes y clínicos, así como el beneficio potencial para la planificación sanitaria⁴⁷. Estos objetivos se verán facilitados si se adoptan nuevos métodos de recogida de los datos que reduzcan al mínimo los costes y ofrezcan una retroalimentación adecuada, tanto para la evaluación clínica como para la gestión del proveedor.

2. Propuesta de la Food and Drug Administration (FDA)

En los últimos años, se han publicado dos directrices que pretenden guiar el desarrollo de medicamentos u otras intervenciones sanitarias, en las que se revisan y plantean aspectos importantes respecto a la normativa de aplicación de los instrumentos PRO en este desarrollo. De manera que puedan incluirse resultados de PROs en la información técnica que se hace llegar a los profesionales sanitarios, así como en la comunicación con otras partes interesadas.

La cada vez mayor aplicación de los instrumentos PRO en investigación clínica incentivó a las autoridades sanitarias, en los Estados Unidos y Europa, a tenerlos en cuenta en el proceso normativo. En 2004, el Comité de Medicamentos de Uso Humano de la Agencia Europea de Medicamentos (EMA) publicó un artículo de reflexión sobre las directrices normativas para el uso de mediciones de la calidad de vida relacionada con la salud en la evaluación de medicamentos⁷². En 2006, la FDA (*Food and Drug Administration*) publicó un borrador de directrices para el sector⁷³ y en 2009 publicó la versión definitiva de la guía⁷⁴. Desde la publicación del borrador diversas sociedades científicas internacionales organizaron *workshops* para debatir el impacto de su aplicación, y se han dedicado números especiales de revistas enteramente a este tema⁷⁵.

La guía de la FDA describe cómo debe ser la investigación con instrumentos PRO para apoyar los posibles mensajes informativos en el etiquetado del producto, desde los criterios adecuados en el diseño del estudio, hasta los instrumentos PRO aplicados. Respecto al instrumento, esta guía pone un énfasis especial en la validez de contenido, aunque también requiere la valoración

de otras propiedades de medida como la fiabilidad, la validez de constructo y la sensibilidad a los cambios para asegurar que el instrumento PRO puede producir una valoración cuantitativa de los efectos de un medicamento.

Las consideraciones de la FDA respecto al diseño de los ensayos clínicos hacen referencia a la frecuencia y los periodos de evaluación, el modo de administración de los instrumentos PRO (papel, electrónico, etc.), y al tratamiento de los datos faltantes. Como en cualquier otra variable de resultado, si el objetivo es producir un mensaje de eficacia, la variable PRO debe estar pre-especificada en el plan de análisis con los ajustes apropiados en el caso de contrastes múltiples.

Para la FDA centrar el desarrollo de medicamentos en los pacientes es una prioridad clara⁷⁶, y todos los constructos medidos con instrumentos PRO, incluida la calidad de vida relacionada con la salud, pueden utilizarse como “criterios de valoración de la eficacia” en los ensayos clínicos, y en particular en el caso de fármacos que van a autorizarse para enfermedades crónicas (p. ej., cáncer, SIDA). En general, se considera que la guía establece estándares de calidad elevados para el desarrollo e implementación de instrumentos PRO en ensayos clínicos de nuevos medicamentos, y es una referencia a seguir para los patrocinadores que quieren introducir mensajes basados en PROs en las etiquetas de sus productos.

Dada la complejidad del desarrollo de los instrumentos PRO, es recomendable involucrar a la FDA en las discusiones sobre estas variables de resultado tan pronto como sea posible. También es recomendable comunicarse con el centro para la evaluación e investigación de fármacos (*Center for Drug Evaluation and Research* – CDER–) para aumentar el número de herramientas públicamente disponibles para dar soporte al etiquetado de los productos.

En 2008, el CDER de la FDA tomó el liderazgo en la formación del Consorcio PRO⁷⁷, cuya misión⁷⁸ es el desarrollo de instrumentos PRO que cumplan los requisitos de la FDA para su utilización en ensayos clínicos en los que las variables de resultado de eficacia derivadas se puedan incluir en los mensajes de las etiquetas de los productos. Los objetivos del Consorcio PRO son: permitir la colaboración precompetitiva que incluya la incorporación y el conocimiento experto de la FDA; evitar el desarrollo de múltiples instrumentos PRO para el mismo propósito; repartir los costes del desarrollo de nuevos instrumentos PRO; desarrollar instrumentos PRO que cumplan los criterios requeridos y ponerlos a disposición pública; y facilitar la revisión de los productos médicos por parte de la FDA mediante la estandarización de las variables de resultado PRO.

Un artículo publicado en 2015 presentaba la perspectiva de los diferentes interesados en este tema (industria, FDA, representantes de pacientes, y Consorcio PRO) en el área oncológica⁷⁹, y la frustración generada por los elevados estándares recogidos en la normativa de la FDA. El origen del artículo es un panel sobre “PRO y el desarrollo de medicamentos centrado en los pacientes”, formado para la sesión plenaria del *workshop* anual de la *Accelerating Anticancer Agent Development and Validation* (AAADV), para entender las razones de la baja inserción en EUA de mensajes basados en PROs en las etiquetas de los productos oncológicos. Esta baja inserción contrastaba con la mayor inserción observada en productos no oncológicos en EUA, y también

en los productos oncológicos en Europa^{80, 81}. El panel propuso como recomendaciones para solucionar las barreras detectadas⁷⁹: un rol proactivo de la FDA animando a la aplicación de variables PRO; desarrollo de planes para la medición de PRO por parte de la FDA; caracterizar y mitigar el efecto de los diseños sin doble ciego en los ensayos clínicos oncológicos; investigar en razones y patrones de datos faltantes; desarrollar recomendaciones de buenas prácticas para el análisis estadístico de datos PRO; y aumento de la formación en la metodología para instrumentos PRO por parte de la FDA y la industria farmacéutica.

3. PROMIS: una nueva medición dinámica de la salud percibida

PROMIS es el acrónimo de *Patient-Reported Outcomes Measurement Information System* (Sistema de Información de Medida de Resultados Reportados por los Pacientes). PROMIS fue uno de los programas dentro de la Hoja de Ruta de Investigación Médica para el siglo XXI que los Institutos Nacionales de Salud (NIH) de los Estados Unidos idearon en 2002 [más información en este enlace: <https://commonfund.nih.gov>] para catalizar todos los cambios necesarios para transformar el conocimiento científico en beneficios tangibles para la sociedad. Una de las 3 áreas de la Hoja de Ruta, la Reingeniería de la Investigación Clínica, se dedicó a la aceleración y el fortalecimiento de los procesos de investigación clínica a través de la adopción de una infraestructura sistemática más eficiente en el campo del descubrimiento científico. En esa área se reconoce de la importancia del concepto y la medición de los Resultados Reportados por los pacientes (PROs, por sus siglas en inglés) para la atención de salud centrada en los pacientes. Y se da un apoyo sin precedentes a una red de grupos de investigación que, en menos de una década, dará un salto cualitativo en la conceptualización e implementación de PROs.

El salto cualitativo en el mundo de los PROs ha consistido en desarrollar un sistema de medida eficiente (minimizando el número de ítems), flexible (permitiendo el uso intercambiable de ítems) y preciso (con un mínimo error de medida). El sistema parte de un marco conceptual de la salud previo bien conocido, el propuesto por la OMS de salud física, mental y social⁸², y aprovecha el considerable desarrollo métrico realizado hasta entonces rescatando los mejores contenidos de los muchos instrumentos PRO ya existentes (instrumentos históricos que los investigadores denominan “*legacy measures*”). Y, sobre todo, aplican y mejoran la base psicométrica que va a permitir llegar a mediciones eficientes, flexible y precisas: la teoría de respuesta al ítem (IRT de sus siglas en inglés).

La aportación fundamental de PROMIS al mundo de los PROs es su orientación por dominios o aspectos concretos de la salud (síntomas, conductas, funciones) que permite usar el mismo contenido y, sobre todo, la misma métrica en poblaciones diferentes, independientemente de los problemas de salud o trastornos crónicos de las personas que se evalúan. El avance, respecto a los instrumentos genéricos, es su adaptabilidad (los dominios se miden en la mayoría de las veces de manera independiente) y la mayor precisión, a la vez que mantienen comparabilidad y una interpretabilidad “universal”. Esta orientación, hecha realidad gracias a la aplicación de la teoría de respuesta al ítem, permite ampliar el rendimiento de los PRO que con requisitos métricos exigentes, como los que imponen las agencias reguladoras para las variables

principales en los ensayos clínicos, o los que demanda el uso individualizado de los PRO en la práctica clínica.

En este capítulo se describen y comentan el contenido, los formatos, la base psicométrica, la obtención e interpretación de las puntuaciones y los diferentes formatos de medida de los dominios de PROMIS. También se describen facilidades en línea para su administración y análisis y otros PROs estrechamente relacionados con PROMIS. Para facilitar la comunicación, al final del capítulo se incluye un cuadro con el glosario de términos utilizados.

Dominios de la salud

La primera tarea de PROMIS fue desarrollar un mapa de dominios que constituirían el sistema de información, lo que se iba a medir. Después de considerar otros modelos, el marco conceptual finalmente escogido fue el de la salud de la OMS⁸² en tres grandes áreas: física, mental y social. A través de estudios Delphi modificados se identificaron una serie de subdominios, de los que se priorizaron 5 para comenzar a trabajar: Función Física, Fatiga, Dolor, Estrés Emocional y Función Social.

Actualmente, los investigadores tienden a denominar “gran área de la salud” a cada uno de los 3 componentes principales (físico, mental y social) y a los niveles englobados en “dominios” (en el área mental: depresión, ansiedad, irritabilidad, función cognitiva, etc.), que dependen jerárquicamente de las áreas de salud. En este punto es importante hacer notar que históricamente las denominaciones usadas por los investigadores PROMIS para los diferentes niveles no han sido siempre claras ni consistentes. Inicialmente, se utilizaba la expresión “dominio” para referirse a las grandes áreas de la salud (físico, mental y social)⁸³ y por extensión a los conceptos asignados a los grupos de trabajo PROMIS (Función Física, Fatiga, Dolor, Estrés Emocional y Función Social), siendo los aspectos dentro de ellos (los actuales dominios) denominados subdominios (como Depresión y Ansiedad en el dominio del Estrés Emocional).

La figura 4 incluye los dominios completados y accesibles a todos los usuarios o en una fase muy avanzada de evaluación a mitad del 2014. Como ya se ha comentado, todos los componentes dentro de cada una de las tres áreas fundamentales se denominan dominios. Hay una serie de dominios que se consideran más generales y básicos y por tanto forman parte del Perfil de salud PROMIS (existen diversas versiones, ver más adelante).

Los dominios e ítems se han desarrollado también para la población infantil, con un proceso similar al de adultos. A finales del 2015 se habían creado y evaluado 40 dominios infantiles y más de 70 para adultos. Un listado actualizado de los dominios PROMIS disponibles puede encontrarse en este enlace: <http://www.healthmeasures.net/explore-measurement-systems/promis/obtain-administer-measures>

Aunque el énfasis de PROMIS es el enfoque de medición de dominios como unidades de interés conceptual y clínico, también se han desarrollado medidas específicas para ciertos propósitos como combinaciones de ítems que se refieren a diferentes dominios. Así, el cuestionario de Salud Global PROMIS -10 tiene 10 ítems, la mitad de ellos recogen una valoración global los 5 dominios fundamentales, más cinco preguntas sobre la percepción de la salud y de la

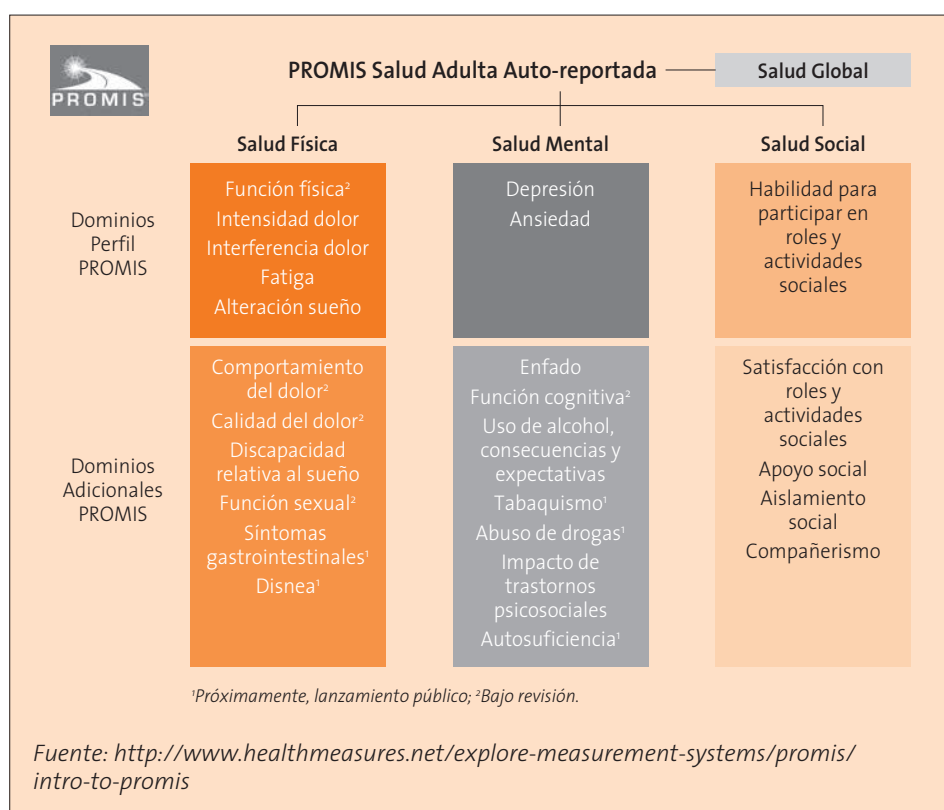


FIGURA 4. Grandes áreas de la salud percibida y dominios PROMIS desarrollados en la primera fase del proyecto (2004-2009).

calidad de vida que no pertenecen a los dominios. También se han desarrollado perfiles, es decir, combinaciones de ítems de diferentes dominios, para dar respuesta a las necesidades de medición de aspectos más globales. Los perfiles existentes se basan en la extracción de los ítems más informativos de los dominios básicos (marcados en color más intenso en la figura) y son de 29, 43 y 57 ítems, dependiendo de los intereses del usuario.

Archivo de datos existentes

Se analizaron bases de datos existentes que incluían información basada en numerosas PROs clásicas de estudios tan importantes como el estudio *Medical Outcomes*, el proyecto *Cancer-Q-Score*, el estudio *Chronic Hepatitis* y el estudio *NHLBI Cardiac Health*, entre muchos otros⁸⁴. Cada una de esas bases de datos debía incluir como mínimo mil respondientes, llegando a una base de información de más de 50.000 respondientes. Los análisis se centraron en evaluar la dimensionalidad propuesta así como en identificar los ítems de mejores propiedades dentro de cada dimensión, que podrían considerarse candidatos definitivos para la medición del dominio⁸⁴.

Librería de ítems

Su creación fue paso fundamental en las fases iniciales del proyecto⁸³. Esta librería centralizada y mantenida por el centro estadístico coordinador es una extensa base de datos relacionados de ítems provenientes de una gran cantidad de instrumentos PRO ya existentes en el momento de iniciarse el proyecto. Los investigadores de PROMIS decidieron aprovechar las ventajas y

la experiencia con ítems bien establecidos en vez de desarrollar nuevos. Así, se realizaron revisiones sistemáticas de la literatura y la consulta a expertos. En esa base de datos, cada ítem se describe según su contexto, su contenido básico o único, su período recordatorio y el instrumento PRO de origen. En la medida de lo posible, la base incluye información sobre versiones modificadas del mismo y su propiedad intelectual. Además, se analizaron bases de datos en las que existía información sobre esos ítems basados en estudios anteriores, para comprobar su comportamiento psicométrico.

La versión inicial contenía más de 10.000 ítems, de los cuales unos 7.000 estaban relacionados con los 5 dominios de interés inicial. Posteriormente se realizaron diversas modificaciones de la librería para hacerla más manejable, reduciendo el número de ítems. Pero ha ido aumentando a medida que se ha incrementado el número de dominios de PROMIS. El conjunto de todos los ítems que finalmente han sido considerados adecuados para medir un dominio se denomina **Banco de ítems**. Su composición está determinada no solo por el análisis conceptual y cualitativo de su contenido, sino por la demostración de su adecuación y robustez métricas en análisis exhaustivos tanto de los datos de estudios previos (análisis del **Archivo de datos**) como en los nuevos estudios realizados por los grupos de PROMIS.

El desarrollo de los ítems se realizó en 6 fases, algunas de las cuales ya han sido comentadas anteriormente: identificación de ítems ya existentes; clasificación y selección de ítems; revisión y modificaciones; participación de grupos focales para identificar vacíos en los dominios PROMIS; estudios cognitivos para ítems específicos y revisión final con posterior evaluación piloto. También se desarrolló un trabajo empírico para el establecimiento de las opciones de respuesta con mejor rendimiento.

Tanto el período recordatorio de los ítems como sus opciones de respuesta fueron estandarizados al máximo, intentando reducir la gran variabilidad de los formatos iniciales y, al mismo tiempo, mejorar su rendimiento. La mayor parte de los ítems PROMIS utilizan como período recordatorio la última semana, a partir de consideraciones de validez y fiabilidad. Existen algunos ítems que se refieren a las capacidades actuales sin emplear un período definido de tiempo, como los del dominio de Función Física. En cuanto a la escala de respuesta, se decidió a partir de numerosos análisis empíricos, utilizar las 5 opciones de respuesta ordinal en la mayor parte de los ítems, concretamente en los que se refieren a intensidad (1= Ninguno; 2=Leve; 3=Moderado; 4=Intenso; 5=Muy intenso), o frecuencia (1=Nunca; 2=Rara vez; 3=Algunas veces; 4=A menudo; 5=Siempre). También hay ítems de opciones preferidas (de 5 categorías) como respuesta a las categorías de duración y capacidad⁸³.

Al haber partido de una revisión y selección de los mejores ítems hasta el momento, PROMIS ha capitalizado décadas de desarrollo y mejora de la medición de la salud percibida. Además, el trabajo cualitativo detallado de expertos y grupos focales no expertos ha permitido mejorar sustantivamente inconsistencias, repeticiones e incertidumbres. En este proceso destaca la participación de pacientes y personas no profesionales que han aportado una seguridad de la relevancia de su contenido. A su vez, la necesidad de consistencia de los ítems de PROMIS ha obligado a cambiar el formato final de los ítems existentes (en muchos casos de manera mínima, en otros más), lo que limita su comparabilidad estricta.

Modelo de medida

Las escalas definitivas generadas para los 5 dominios iniciales de PROMIS fueron evaluadas en estudios empíricos para comprobar las características métricas tanto de los bancos de ítems como de las formas cortas (SF: cuestionarios cerrados creados a partir de la selección de un subconjunto de ítems). Estos estudios incluyeron población general (encuesta de un panel de algo más de 7.500 personas representativas de la población general adulta de los EEUU) así como grupos de pacientes con diversas enfermedades crónicas (cáncer, enfermedades cardíacas, artritis reumatoide, artrosis, trastornos mentales, lesión de la medula espinal y enfermedad pulmonar obstructiva crónica). Todos los estudios clínicos incluyeron un mínimo de 500 pacientes⁸⁴. Como se evaluaban más de 1.000 ítems, no todos los participantes podían responder a todos ellos y los estudios usaron un diseño de administración por bloques de tipo incompleto. Los estudios permitieron la estimación simultánea de los parámetros de los ítems a partir de modelos basados en IRT y de la diferencia media con los valores poblacionales y las razones de las desviaciones estándar.

El procedimiento de análisis incluye los siguientes pasos: 1) Análisis estadístico descriptivo tradicional a nivel de ítem (distribución de las respuestas y consistencia con los demás ítems) y de escala (media, desviación estándar, etc.); 2) Evaluación de las asunciones del modelo de IRT (unidimensionalidad, independencia local y monotonidad); 3) Modelización y evaluación del ajuste de los modelos (estimación de los parámetros, evaluación del ajuste, evaluación de las propiedades de los ítems y las de la escala); 4) Evaluar la ausencia de funcionamiento diferencial de los ítems (DIF) respecto de variables personales y clínicas relevantes; y 5) Calibración de los Bancos de ítems (estudios poblacionales, estandarización de las puntuaciones y estimaciones programación para la administración informatizada –CAT–)⁸⁴.

Puntuaciones

Una vez administrados los ítems de un dominio PROMIS (ya sea en sus formas fijas, larga –Banco de ítems– o corta, –forma corta o en presentaciones de forma dinámica, CAT–) los algoritmos de estimación basados en los modelos IRT proporcionan una estimación de la situación del individuo en el continuo medido por el dominio. Para facilitar la comprensión de la misma, esta puntuación se estandariza por el método T, dado un valor de 50 a la correspondiente a la puntuación promedio de la población adulta estadounidense, y una desviación estándar de 10 puntos. El Centro de Evaluación proporciona esa información de manera visual, tal como aparece en la gráfica (Fig. 5). La información incluye la estimación de la puntuación del individuo que ha respondido y el error estándar de la misma. La distancia entre esa puntuación y la media poblacional se marca en distancias de 5 y 10 puntos a través de líneas y de coloración. En el ejemplo de la figura, las puntuaciones para los dominios de irritabilidad, fatiga y dolor se encuentran más de 10 puntos por encima de la media poblacional (lo que corresponde a una distancia de más de una desviación típica por encima de la media).

El retorno del Centro de Evaluación incluye, además, una explicación escrita de la interpretación que debe darse a cada una de las puntuaciones. Sin duda, este tipo de informe es de utilidad para la mejor comprensión de estos instrumentos y puede facilitar su interpretación en la práctica diaria.

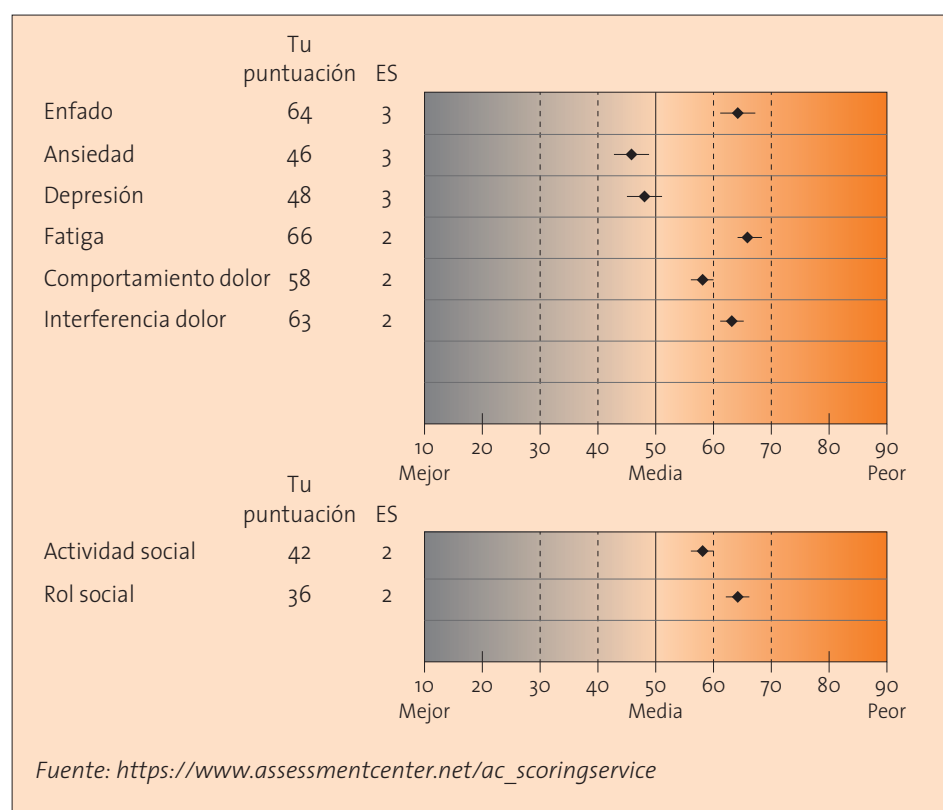


FIGURA 5. Presentación de las puntuaciones de 7 dominios, estandarizadas respecto a las medias de la población norteamericana, tal como se obtiene al responder a las escalas en Centro de Evaluación PROMIS.

Proceso de maduración y estándares científicos de PROMIS

Los investigadores de PROMIS identificaron las etapas de desarrollo (maduración) de sus instrumentos proponiendo para cada una de ellas las actividades y el nivel de rendimiento que se debían alcanzar. Ese proceso debe superarse satisfactoriamente para incluir un nuevo dominio (su banco de ítems) en el sistema PROMIS. No son una novedad, ya que se pueden encontrar en el desarrollo de otros muchos PROs. Sin embargo, el procedimiento es interesante porque incorpora aspectos poco conocidos de la metodología de la IRT respecto al desarrollo de extensiones de instrumentos, y suponen una cierta novedad en el campo de la salud.

En la versión actualizada en mayo del 2013, el modelo de maduración de un instrumento de PROMIS incluye 5 estadios: 1) Desarrollo (conceptualización y generación de los ítems); 2) Desarrollo (fase de calibración, en la que se debe haber usado análisis factorial e IRT para probar la bondad de cada ítem, además de realizar prueba de validez concurrente con instrumentos históricos y de evaluar la existencia y el impacto de un posible funcionamiento diferencial); 3) Acceso público al instrumento (cuando se han superado las valoraciones y requisitos 1 y 2); 4) Maduración (los instrumentos han superado los estándares en estudios en poblaciones diferentes, incluyendo estudios longitudinales de sensibilidad al cambio y diferencia mínima importante, y habiendo realizado análisis de DIF más extensos; en ese momento los instrumentos pueden recomendarse para el uso en la práctica clínica; y 5) Maduración completada (cuando los instrumentos han superado numero-

sas pruebas y usos exitosos en poblaciones diferentes y las puntuaciones se pueden interpretar de manera clara (tanto los cambios como los valores absolutos de la puntuación).

Los estándares de desarrollo contienen una lista detallada de características y procedimientos que hacen referencia a los siguientes conceptos: 1) Definición del concepto a medir y el modelo conceptual; 2) Generación y diseño de los ítems; 3) Construcción del *pool* de ítems; 4) Propiedades del Banco de ítems; 5) Estudios piloto y formato final del instrumento; 6) Validez; 7) Fiabilidad; 8) Interpretabilidad; y 9) Traducción a otras lenguas y Adaptación cultural.

Centro de Evaluación (Assessment Center)

Es una plataforma gratuita en línea para la recogida de datos que permite a los investigadores crear sitios web específicos para sus estudios y captura de datos de los participantes de forma segura. Los estudios pueden incluir tanto instrumentos de la biblioteca del Centro de Evaluación como instrumentos elaborados por el propio investigador. La librería de instrumentos incluye todos los tipos de instrumentos y formatos. Entre otras características importantes, el Centro de Evaluación también permite descargar instrumentos de la biblioteca para su administración en formato de papel, la personalización de ítems o de instrumentos (por ejemplo, modificando el formato, determinando el procedimiento de selección de los ítems en el CAT, o permitiendo instrucciones de salto a lo largo del cuestionario).

Una de las características más útiles del Centro es la provisión de las puntuaciones en tiempo real tanto para los CAT como para las formas cortas. La descripción de las puntuaciones PROMIS, tal como se ha comentado más arriba, se puede recibir en línea de manera inmediata a las respuestas. Esto supone un excelente material para mejorar la comunicación entre el profesional de la salud y el paciente en el caso de utilizarlos para la práctica clínica. Además, el Centro de Evaluación permite el almacenamiento de la información de salud protegida en una base de datos independiente y segura. También proporciona un sistema de información automatizado en tiempo real de la exportación de los datos acumulados. Finalmente, permite capturar la aprobación de los formularios de consentimiento informado en línea.

Los Institutos Nacionales de Salud (NIH) han estado proporcionando apoyo financiero para Centro de Evaluación en el mantenimiento y distribución de instrumentos PROMIS y en la asesoría científica de acompañamiento. A partir del 2019 se espera que estas funciones sean supervisadas y financiadas por una entidad de nueva creación llamada la Organización de la Salud PROMIS (PHO, por sus siglas en inglés), una organización sin ánimo de lucro que fue desarrollada por investigadores PROMIS. El Centro de Evaluación se puede acceder en esta dirección: <http://www.assessmentcenter.net>

La iniciativa PROMIS Internacional

Un aspecto importante de la iniciativa PROMIS ha sido el desarrollo temprano de versiones en español por la importancia creciente de este idioma en los Estados Unidos. PROMIS adoptó los métodos estándares de la adaptación cultural de instrumentos, aportando un modelo peculiar de traducción que denominan “universal”: la versión en español de los instrumentos PROMIS debería ser la misma para cualquier hispanohablante, independientemente

del país en que se hable el idioma⁸⁵. El enfoque de traducción universal implicó asegurar que los paneles de traducción, así como los estudios cualitativos sobre la comprensión y validez de los instrumentos, incluyera personas de diversos países hispanohablantes. El objetivo es maximizar las denominaciones comunes (frente a las variantes dialectales particulares), sobre todo evitando palabras que pudieran ser ofensivas o inadecuadas en alguno de los países. Para ello, las pruebas piloto, antes de llegar a la versión final del cuestionario, deben realizarse (o repetirse) en los diferentes países. La experiencia con las versiones españolas (para España) de los cuestionarios PROMIS muestra que algunas palabras adecuadas para la versión en la mayoría de países latinoamericanos requieren añadir la palabra más adecuada al contexto español (por ejemplo, “enojo [enfado]” o “piso [suelo]”). Es importante señalar que todos los dominios de PROMIS cuentan con la versión en español universal y pueden obtenerse directamente del Centro de Evaluación.

A principios de 2016, se han traducido muchas de las escalas de PROMIS a muchos idiomas, especialmente las formas cortas. En la actualidad existe un proyecto de adaptación cultural para extender las traducciones a todos los idiomas en los que se realizan estudios experimentales (ensayos clínicos) con PROs.

El Comité Internacional de PROMIS

Este comité Internacional es el fruto de una colaboración de científicos y médicos interesados en la optimización y la armonización del uso de los instrumentos y los recursos relacionados PROMIS en todo el mundo, para la investigación, la atención clínica y la vigilancia poblacional. El comité es parte de la Organización de Salud PROMIS (PHO) y es instrumental para identificar, coordinar y promover las mejores prácticas para desarrollar, traducir, validar y utilizar instrumentos PROMIS en el mundo⁸⁶.

Algunas actividades científicas especialmente relevantes para ese comité incluyen el desarrollo de normas internacionales de evaluación de la validez transcultural de los instrumentos PROMIS y el desarrollo de estándares para la calibración de los instrumentos de manera consistente a nivel internacional. Finalmente, el comité incluye personas de contacto a nivel nacional y trabaja en temas como los derechos de distribución, la garantía de la privacidad de los datos, el desarrollo de herramientas específicas para cada país en la gestión de datos del Centro de Evaluación y los diversos sitios web de PROMIS [<http://www.promishealth.com/pho-international>].

Otras medidas de resultados relacionados con PROMIS

Además de la implementación del sistema PROMIS en la web de NIHPROMIS el Centro de Evaluación, algunos de los investigadores de PROMIS han participado en el desarrollo de instrumentos y plataformas de medida que pongan PROMIS al alcance de los usuarios. De las distintas herramientas y facilidades creemos oportuno destacar dos: la Caja de Herramientas NIH (NIH Toolbox) y el proyecto PROSetta Stone (la Piedra PROsetta).

La Caja de Herramientas del NIH

Es un conjunto estándar de herramientas de valoración integral desarrollado por el NIH para la evaluación de la función neurológica y del comportamiento a lo largo del tiempo. La Caja de Herramientas pretende facilitar el acceso a pruebas diagnósticas sin obligaciones de derechos de propiedad intelectual

para la actividad clínica e investigadora. Cuatro son los dominios incluidos en la caja de herramientas: Cognición, Sensación, Movimiento y Emoción. Estos dominios incluyen a su vez subdominios (por ejemplo, Función Ejecutiva o Memoria dentro de Cognición). Los instrumentos incluyen aquellos reportados por los pacientes (PROs) como los basados en el rendimiento funcional “objetivo” observados por profesionales, denominados *Performance-based Outcomes* (PFOs). Las medidas objetivas son mayoritarias en las áreas de Cognición y Movimiento (por ejemplo, el de Cognición, que incluye entre otros, el test del Control Inhibitorio de la Atención de Flankers). El área de Sensación y Emoción está compuesta mayoritariamente por PROs. Todos los instrumentos de la Caja de Herramientas han sido validados y se dispone de normas de referencia. Son aplicables a través de herramientas como el NIH iPad, un sistema de presentación bastante intuitivo y que requiere una formación mínima. La Caja de Herramientas del NIH es adecuada para usos en la población general, en personas con enfermedades crónicas y a lo largo de todo el ciclo vital (desde la edad de 3 hasta los 85 años).

La Piedra PROsetta (Stone® PROsetta)

Es un proyecto que desarrolló y aplicó métodos para vincular los instrumentos PROMIS con otros instrumentos relacionados (por ejemplo, el *SF-36*, el *Brief Pain Inventory*, el *CES-D*, el *MASQ*, o el *FACIT-Fatiga*) para expandir la gama de opciones de la evaluación de PROs dentro de una métrica común y estandarizada. Proporciona las equivalencias en puntuación entre diferentes escalas que miden el mismo resultado de salud. Las equivalencias se obtuvieron en estudios que incluían la administración de la totalidad o una parte de los ítems de ambos instrumentos previamente en los mismos individuos. Mediante la aplicación de técnicas de equiparación e igualación tres diferentes métodos de vinculación, lo que permite la creación de una tabla de equivalencias (“*Cross-Walk*”) en la que cada puntuación posible del instrumento “establecido” se hace corresponder con una puntuación T del instrumento PROMIS comparable. El proyecto PROsetta representa un paso muy importante en la clarificación de las relaciones entre los instrumentos que miden constructos similares y demuestra el potencial de modelo de medida basado en la teoría de respuesta a los ítems (IRT) [<http://www.prosetta.org/Methodology/Documents/PROSetta%20Methodology%20Report.pdf>]

Glosario para la comprensión de PROMIS

Conceptos

- **(Grandes) Áreas de Salud:** Salud Física, Mental y Social, según el modelo de la OMS.
- **Dominio:** Cada uno de los conceptos medidos en PROMIS para los que se desarrolla un Banco de Ítems y que están insertos jerárquicamente dentro de las áreas de salud. Al inicio del proyecto se denominó dominios a las (grandes) áreas de salud y subdominios a lo que actualmente se denomina dominio.

Instrumentos

- **Bancos de ítems:** conjuntos de ítems calibrados de los que se puede obtener una puntuación resumen aun utilizando solo un subconjunto de ítems (como en la administración CAT).

- **Escalas:** conjuntos de ítems calibrados para los cuales solo se obtiene una puntuación resumen si se administra un subconjunto de ítems fijo y completo del banco de ítems (por ejemplo, las Formas Cortas Fijas).
- **Formas cortas (*short forms*):** subconjuntos fijos de un Banco de ítems o preguntas para cada uno de los dominios PROMIS. Su longitud varía entre 1-11 ítems, según el dominio.
- **Perfiles:** subconjunto fijo de ítems de los dominios más generales (Ansiedad, Depresión, Fatiga, Dolor, Función Física, Alteración del Sueño y Participación en Actividades y Roles Sociales). Todos los ítems de los perfiles aparecen en las formas cortas de los dominios. Existen tres tipos de perfiles a destacar:
 - El perfil PROMIS-10 (Salud General), que incluye un ítem de cada dominio más otros cinco ítems globales de percepción de la salud y de la calidad de vida.
 - Los perfiles PROMIS-29, PROMIS-43 y PROMIS-57 para adultos, que incluyen un número de ítems creciente de cada dominio.
 - Los perfiles para la infancia (PROMIS-25, PROMIS-37 Y PROMIS-49) que pueden ser de autoadministración o para administración a los padres/tutores.
- **Repositorio (*Pool*) de ítems:** conjuntos de ítems relacionados que no se pretenden analizar con una puntuación sumario sino como medidas individuales.
- **Test Adaptativo Informatizado (CAT):** es una forma de administración flexible (no fija) en la que los ítems a aplicar al respondiente se seleccionan en el banco de ítem de manera dinámica, a partir de las respuestas anteriores del entrevistado. En promedio, las formas CAT alcanzan con 4 ítems una precisión de medida equivalente a una fiabilidad de 0,90.

Los instrumentos PROMIS disponibles se pueden consultar en el acceso al Centro de Evaluación de los NIH: <http://www.assessmentcenter.net/documents/InstrumentLibrary.pdf>

Metodología específica IRT

- **Calibración:** estimación de los parámetros de los ítems en IRT. PROMIS utiliza un modelo IRT de respuesta ordinal, donde los parámetros de los ítems son: localización (posición de la categoría de respuesta en el continuo del dominio); discriminación (representatividad del ítem del continuo). De mantenerse el supuesto IRT de invarianza, el valor estimado de estos parámetros es independiente de la muestra donde se estimen (invariantes en poblaciones diferentes).
- **Funcionamiento diferencial del ítem (DIF):** efecto por el cual los parámetros varían según algunas características de los individuos (por ejemplo, género, edad, etnia, o estado de salud). Si el DIF es importante, el ítem debe ser reemplazado y/o modificado para evitar incomparabilidad entre las medidas de diferentes poblaciones. En PROMIS se han hecho numerosos estudios de DIF entre las versiones en inglés y en español en los Estados Unidos.

Recursos en línea

- www.assessmentcenter.net
- www.promishealth.com
- www.nihpromis.org
- www.healthmeasures.net
- commonfund.nih.gov
- www.prosettastone.org

4. Patient-Centered Outcomes Research Institute (PCORI)

El Instituto de investigación de resultados centrados en los pacientes (PCORI de sus siglas en inglés) fue fundado en 2010 para financiar investigación comparativa de efectividad centrada en el paciente. En las últimas dos décadas ha ido creciendo de forma consistente la investigación basada en modelos participativos de la comunidad^{87, 88}. En EEUU, las agencias de salud federales apoyan la participación de los pacientes en investigación, la *Agency for Healthcare Research and Quality* involucra a pacientes en la identificación de temas, preguntas y resultados de investigación⁸⁹, y los *National Institutes of Health* tienen experiencia en trabajar con pacientes y organizaciones de pacientes para establecer prioridades de investigación⁹⁰. También en el Reino Unido, la participación de los pacientes en el acuerdo sobre la agenda de investigación y las actividades de investigación es cada vez más importante⁹¹. La Comisión Europea también potenció en sus Programas Marco que los foros o asociaciones de pacientes participaran como socios de los consorcios de investigación.

El amplio consenso sobre la idoneidad de la investigación centrada en el paciente⁹² ha surgido como solución a las importantes dificultades existentes actualmente en el proceso entre la publicación de la investigación y su aplicación. Mucha de la evidencia científica que se genera no llega a implementarse debido a las diferencias existentes entre la investigación y la práctica clínica, que necesitaría más información sobre cómo varía el efecto de los tratamientos en individuos y subgrupos (mujeres, niños, minorías y pacientes con comorbilidades). Con la justificación de que una mayor implicación de todos los interesados podría aumentar la relevancia de los proyectos de investigación, su transparencia y acelerar la adopción de sus resultados en la práctica.

PCORI requiere que la investigación financiada se dirija a responder preguntas que sean importantes y que midan resultados que tengan sentido para los pacientes y otros actores interesados. Para ser financiado por PCORI también se requiere que estos estén activamente implicados en la realización de la investigación. Las normas metodológicas estándar de PCORI incluyen involucrar a personas que representen la población de interés^{93, 94} y otros actores relevantes en la manera que sea adecuada y necesaria en cada contexto de investigación. Se espera que los destinatarios finales de los resultados de la investigación participen en el diseño del estudio, la selección de las medidas, la mejora del reclutamiento, la interpretación de los resultados del estudio, y / o su difusión.

Involucrar todas las partes interesadas no es un fin en sí mismo, sino una estrategia más para asegurar que la investigación, y sus resultados están centrados en los pacientes, son relevantes para los destinatarios finales, y

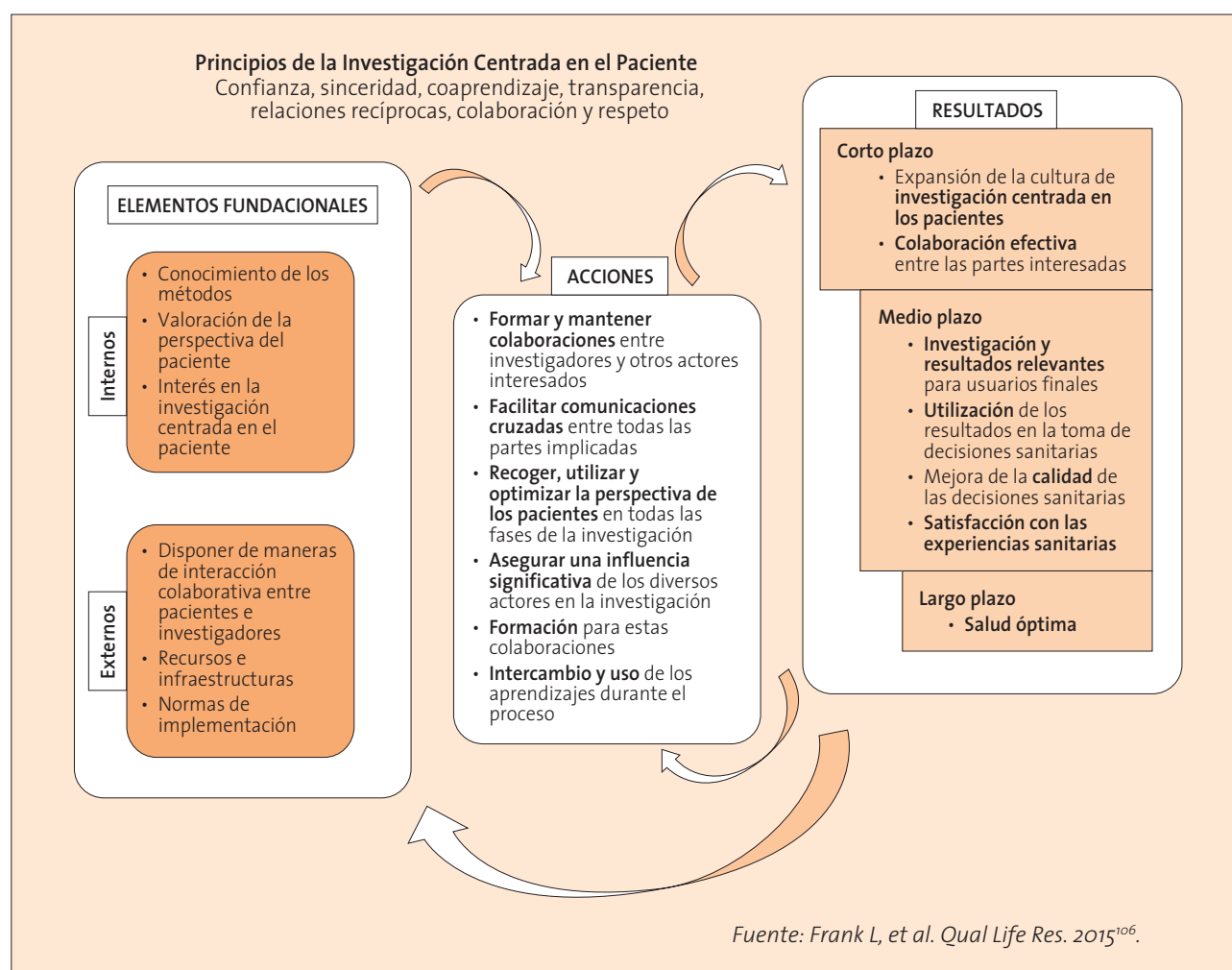


FIGURA 6. Modelo conceptual de investigación centrada en el paciente de PCORI.

se pueden difundir de manera efectiva. PCORI requiere que la implicación cumpla con el objetivo de incorporar activamente sus perspectivas, más allá de la perspectiva de los investigadores, para informar sobre las decisiones relativas a la pregunta de investigación, el diseño del estudio, las medidas utilizadas, los aspectos prácticos de implementación, particularmente los relacionados con el reclutamiento y la recolección de datos, la interpretación de los datos, y / o la difusión de los resultados.

El modelo conceptual para la investigación de resultados centrados en los pacientes del PCORI fue desarrollado a partir de la bibliografía sobre investigación comparativa de la efectividad e investigación participativa⁹⁵⁻¹⁰² y la participación de los miembros del Grupo Asesor de Participación de Pacientes de PCORI. Este modelo tiene sus raíces en la investigación participativa basada en la comunidad, en las definiciones de la investigación de resultados¹⁰³ y en los principios de la evaluación de PROs^{104, 105}.

El modelo conceptual (Fig. 6) está organizado en tres conceptos relacionados con los principios de la investigación centrada en el paciente¹⁰⁶: 1) Elementos fundacionales requeridos por este tipo de investigación; 2) Acciones para su realización; y 3) Resultados de salud obtenidos a partir de las acciones que

aplican los elementos fundacionales. Los principios de la investigación centrada en el paciente proporcionan la base ética, y deben ser demostrados. Los principios están íntimamente relacionados con los tres conceptos del modelo y pueden verse modificados por estos. Por ejemplo, la transparencia puede aumentar con el tiempo como resultado de experiencias positivas durante el proceso investigador.

Los elementos fundacionales se dividen en 3 internos (conocimiento de los métodos, valoración de la perspectiva del paciente, e interés en la investigación centrada en el paciente) y 3 externos (disponer de formas para que pacientes e investigadores actúen como compañeros, recursos e infraestructuras, y normas que permitan la implementación de los elementos fundacionales).

Las acciones representan compromiso con la investigación: 1) formación y mantenimiento de una asociación de los investigadores con las otras partes interesadas; 2) facilitar las comunicaciones cruzadas en múltiples direcciones entre todos los interesados; 3) recogida y utilización de la perspectiva de los pacientes en todas las fases de la investigación; 4) asegurar una influencia significativa de los diferentes actores en la investigación; 5) formación para formar parte de la investigación; y 6) intercambio y aplicación de los aprendizajes durante el proceso de investigación.

Se distinguen resultados a corto, medio y largo plazo. El modelo asume que los resultados retro-alimentan positiva y negativamente tanto a las acciones como a los elementos fundacionales. Los resultados a corto plazo incluyen la expansión de la cultura de la investigación centrada en los pacientes y de la asociación efectiva entre las partes interesadas. A medio plazo, que la investigación y sus resultados tengan sentido para los usuarios finales, que sean útiles para la toma de decisiones sanitarias, mejorar la calidad de dichas decisiones, y mejorar la satisfacción con las experiencias sanitarias. A largo plazo, el resultado es mejorar la salud.

Desde su creación en 2010, PCORI ha financiado cientos de estudios a través de su convocatoria regular de proyectos¹⁰⁶. El sitio web de PCORI proporciona una gran cantidad de información, en la que la organización publica cada trimestre los proyectos financiados, resultados, publicaciones y otras informaciones relacionadas. Actualmente muchos de los estudios están todavía en curso y obtendrán los resultados en los próximos años.

Un editorial del *British Medical Journal*¹⁰⁷, que comentaba un estudio sobre la efectividad de la warfarina en pacientes con accidente isquémico cerebral financiado por PCORI en 2013¹⁰⁸, defendía las potencialidades de la investigación de resultados centrados en los pacientes a partir de este ejemplo. En dicho estudio, el trabajo preliminar con pacientes había conducido a la priorización de un resultado que realmente tenía interés para ellos¹⁰⁸: el tiempo que vivían en su casa sin recurrencias y sin hospitalizaciones por complicaciones. El estudio también evaluaba eventos adversos graves y otros indicadores clínicos como resultados secundarios. Los investigadores del estudio incluyeron pacientes como co-investigadores, asegurando su implicación durante todo el proceso, desde la participación en el diseño hasta aportar ideas en el plan de análisis estadístico. El editorial enfatiza el cambio que este estudio supone en el diálogo entre pacientes y profesionales sobre las opciones de

tratamiento: los pacientes con accidente isquémico cerebral en el momento del alta serán más capaces de recordar que tomar warfarina puede ayudarles a estar en casa con buen estado de salud durante más tiempo, mientras que muchos de ellos no entenderían ni recordarían sus beneficios expresados como reducción de eventos adversos graves.

La investigación centrada en los pacientes comparte con la investigación en PROs el valor que otorga a la perspectiva del paciente. Sin embargo, a diferencia de la investigación en PROs, la investigación centrada en los pacientes requiere no solo capturar la voz de los pacientes como sujetos de investigación, sino su integración en la dirección, planificación y realización de la investigación. La investigación en PROs se dirige hacia lo que es importante para el paciente, pero la investigación centrada en el paciente añade la participación del paciente, además de en el contenido, en cómo se debe realizar la investigación: los métodos empleados para recoger la información, estrategias de reclutamiento y diseminación. El modelo de la investigación centrada en el paciente se expande más allá de la investigación en PROs.

IV. Bancos de instrumentos PRO

La gran cantidad de instrumentos existentes en la actualidad hace difícil identificar los disponibles para un uso o una población específica. La solución más rápida y más eficaz son las bases de datos de instrumentos. Existen diversas bases de datos especializadas en patologías o especialidades concretas, algunas de ellas alojadas en la página web de sus sociedades científicas. La tabla 2 recoge un listado de algunas de estas bases de datos o bibliotecas virtuales. Pero probablemente las bases de datos más importantes por la gran cantidad de instrumentos disponibles son PROQOLID y BiblioPRO que se describen con más detalle a continuación. Una consulta realizada a estas bases de datos o librerías virtuales antes de iniciar un nuevo proyecto puede ser de gran utilidad.

TABLA 2. Bibliotecas, bases de datos, o bancos de instrumentos que contienen información sobre instrumentos PRO.

Nombre (URL)	Descripción
Health and Psychosocial Instruments (HAPI) (http://bmdshapi.com/)	Es un base de datos bibliográfica disponible con suscripción que contiene un amplio registro de instrumentos y herramientas de medidas de diversas disciplinas de ciencias sociales y de la salud.
Department of Health & Human Services Measure Inventory (http://www.qualitymeasures.ahrq.gov/)	Un repositorio de medidas que se están utilizando por las agencias o subagencias del Departamento de salud y servicios sociales de USA para medir, mejorar y reportar la calidad.
Health Services and Sciences Research Resources (HSRR) (https://wwwcf.nlm.nih.gov/hsrr_search/)	Es una base de datos diseñada para localizar recursos de ciencias sociales y salud pública del área sanitaria; los usuarios pueden examinar las características de las herramientas: conjuntos de datos, software o instrumentos/índices y enlaces a recursos adicionales.
Database of Instruments for Resource Use Measurement (DIRUM) (http://www.dirum.org/instruments/search)	Una base de datos de acceso abierto de recursos y cuestionarios para utilizar en las evaluaciones económicas.
Rand Health (http://www.rand.org/health/surveys_tools.html)	Dispone de instrumentos a los que se puede acceder gratuitamente, entre ellos se encuentran algunos muy utilizados a nivel internacional y traducidos a varios idiomas como el KDQOL o PSQ.
PsycTESTS via APA PsycNET (http://www.apa.org/pubs/databases/psyc-tests/)	Esta base de datos en una amplia colección de medidas psicológicas, escalas, entrevistas o instrumentos utilizados por profesionales de las ciencias sociales y de la conducta. Contiene una base de datos electrónica con resúmenes descriptivos, texto completo y bibliografía de instrumentos.
EULAR (http://www.eular.org/)	Esta base de datos agrupa instrumentos validados con énfasis en los instrumentos PRO utilizados en reumatología.
GI-PRO Database (http://www.researchcore.org/gipro/)	Es una base de datos <i>online</i> especializada en el área de gastroenterología realizada con la colaboración del <i>National Institutes of Health</i> (NIH) sobre instrumentos PRO.

TABLA 3. Los objetivos específicos de BiblioPRO	
	• Identificar los instrumentos PRO actualmente disponibles en español.
	• Poner el máximo número de instrumentos PRO a disposición del público en una biblioteca virtual de acceso libre y gratuito, permitiendo una selección informada.
	• Promover el respeto de los derechos de propiedad intelectual y otros requerimientos legales para el acceso apropiado y uso de los instrumentos PRO en castellano.

Fuente: BiblioPRO.

**1. BiblioPRO:
Biblioteca virtual
de instrumentos
pro en español**

BiblioPRO (<http://www.bibliopro.org>) es una biblioteca virtual de instrumentos PRO en español. Es una iniciativa del Grupo de investigación en Servicios Sanitarios del IMIM-Instituto Hospital del Mar de Investigaciones Médicas en colaboración con investigadores españoles¹⁰⁹, diseñada en el contexto de la Red IRYSS (2003-2006) y desarrollada después en CIBERESP (Consorcio de Investigación Biomédica En Red de Epidemiología y Salud Pública). El comité científico de BiblioPRO está compuesto por investigadores del CIBERESP, CIBER de Salud Mental (CIBERSAM), CIBER de Enfermedades Neurodegenerativas (CIBERNED), Red de Investigación en Servicios de Salud en Enfermedades Crónicas (REDISSEC) y de empresas de investigación contratadas (IMS Health).

BiblioPRO nació para diseminar y divulgar los resultados de investigación en el ámbito de los PRO, facilitando a investigadores, clínicos, gestores socio-sanitarios y profesionales de la industria de nuestro país el acceso a dichos instrumentos. Tiene como principal objetivo dar respuesta a la necesidad detectada de una orientación basada en la evidencia para identificar los instrumentos PRO disponibles o escoger el instrumento más adecuado para un uso concreto. El proyecto fue desarrollado, además, con la voluntad de promover el reconocimiento de la propiedad intelectual y el respeto a las condiciones de uso de cada instrumento. (Tabla 3)

En una década, BiblioPRO se ha convertido en la librería virtual con mayor número de instrumentos PRO tanto en el ámbito nacional como internacional, con más de 1.300. Además, es totalmente de acceso libre, mientras que otras iniciativas similares proporcionan acceso gratuito a una parte limitada, pero para obtener información detallada requieren algún tipo de pago por suscripción o consulta. Para acceder a BiblioPRO no es necesario registrarse, tan solo es necesario hacerlo para solicitar sublicencias y para descargar algunos materiales adicionales. Registrarse es gratuito, su finalidad es que quede almacenada toda la información sobre las sublicencias solicitadas, y los usuarios registrados tienen acceso a una funcionalidad adicional, “Mi BiblioPRO”, desde la que pueden obtener su listado histórico de los permisos de uso obtenidos.

**Identificación
de instrumentos
candidatos**

La identificación de instrumentos PRO se realiza mediante revisiones sistemáticas periódicas de literatura utilizando un algoritmo con alta sensibilidad aplicado a Pubmed. Para ello se diseñó un protocolo de búsqueda específico para la identificación de artículos con información sobre el desarrollo, adaptación, características métricas o uso de instrumentos PRO en España¹¹⁰. La

TABLA 4. Revisiones sistemáticas BiblioPRO

	2005-2007	2008-2010	2011-2012	2013-2014
Nº de títulos revisados	2.097 (699/año)	4.830 (1610/año)	3.928 (1964/año)	4.218 (2109/año)
Nº de resúmenes revisados	431	1.977	1.495	1.497
Nº de artículos revisados	330	681	893	1.107
Nº de nuevos instrumentos PRO	223 (74,3/año)	403 (134,3/año)	492 (246,0/año)	— En proceso

Fuente: BiblioPRO

revisión de títulos y resúmenes se realiza mediante pares de revisores y un tercer experto en caso de desacuerdo. La extracción de datos de los artículos seleccionados se realiza mediante un formulario específico diseñado para obtener la información necesaria para completar las fichas técnicas de la biblioteca que disponen de información sobre: autores, artículo de publicación del instrumento, dirección o correo electrónico, y características del instrumento como constructo, población, modelo de medida, número de ítems, escalas, área, o palabras clave.

Las primeras revisiones sistemáticas se realizaron por periodos de 3 años (2005-7 y 2008-10), para irse reduciendo, primero a 2 años (2011-12 y 2013-14), hasta hacer revisiones anuales a partir de 2015. Las revisiones anuales garantizan la actualización de la información contenida en la web. La tabla 4 muestra el aumento de publicaciones y de instrumentos PRO observado desde 2005. El número de títulos detectados con la búsqueda ha pasado de 699/año en el periodo 2005-2007 a 2109/año en el periodo 2013-2014. De forma similar, el número de nuevos instrumentos PRO identificados ha pasado de 74,3 anuales en 2005-2007 a 246 anuales en 2011-2012.

Clasificación de cuestionarios con una nueva taxonomía

Dado su carácter científico, BiblioPRO abordó la discusión sobre las bases conceptuales de los PRO y las taxonomías de clasificación. Como resultado de ello, los investigadores del comité científico desarrollaron un nuevo y exhaustivo sistema de clasificación⁴³ basado en los dos modelos conceptuales más ampliamente aceptados en este campo, el de Wilson & Cleary para la calidad de vida⁴⁴ y el de la Clasificación Internacional de Funcionamiento, Discapacidad y Salud de la OMS¹¹. La taxonomía BiblioPRO considera tres dominios principales: el concepto o constructo medido, la población diana a la que va dirigido, y el modelo de medición. Este sistema de clasificación es relativamente simple, así como robusto y útil. BiblioPRO cuenta con un “buscador inteligente” que permite identificar los instrumentos para un uso concreto, combinando las palabras clave y los elementos de esta taxonomía, que se describe más detalladamente en el apartado de “Tipos de instrumentos PRO”.

Contenido y actividad

BiblioPro cuenta con 7.300 usuarios registrados, y anualmente recibe cerca de 45.000 visitas y gestiona más de 400 sublicencias. La página web proporciona toda la información disponible sobre cada uno de los 1.350 instrumentos PRO identificados para su uso en España, desde Información básica sobre los

autores hasta información detallada sobre los ítems y dominios medidos. El concepto más frecuentemente medido por los instrumentos identificados son síntomas, seguidos de la calidad de vida relacionada con la salud; estos instrumentos son en su mayoría específicos de patología especialmente del área de salud mental.

BiblioPRO intenta localizar a los autores o propietarios de los derechos de autor de los instrumentos PRO en español identificados para solicitar la inclusión en la biblioteca. Cuando estos lo permiten, la web ofrece acceso directo a los cuestionarios, y a otros materiales necesarios como manuales de administración, puntuación o interpretación. La página web de BiblioPRO está preparada para poder distribuir los instrumentos y sus licencias de uso según las preferencias acordadas con el propietario intelectual (libremente, con registro de las sublicencias otorgadas gratuitamente o con cobro de sublicencias por todos o por una parte de los solicitantes).

Cuando no es posible descargar directamente el instrumento desde BiblioPRO pero es posible hacerlo desde otra página web se redirecciona a través de un link. Cuando no existe ninguna de estas posibilidades se recomienda contactar con el autor de la versión española o de la versión original (en caso de ser adaptado). El efectivo respeto de los derechos de propiedad intelectual es central para BiblioPRO. El concepto de *copyright* es legalmente complejo y es entendido de manera diferente según los países. Es relativamente fácil acceder a copias no formales de los instrumentos sin seguir el proceso legal correspondiente.

2. Patient-Reported Outcome and Quality of Life Instruments Database (PROQOLID)

PROQOLID (*Patient-Reported Outcome and Quality of Life Instruments Database*) es una iniciativa de Mapi Research Trust (<http://www.proqolid.org>). Actualmente este banco de instrumentos está centrado en los instrumentos de investigación para evaluación de resultados clínicos (COAs, siglas en inglés de '*clinical outcome assessments*'). Según el glosario de la FDA¹¹² los COA serían 'cualquier evaluación que pueda estar influenciada por la elección, juicio o motivación de las personas, y pueda apoyar evidencias tanto directas como indirectas de los beneficios de los tratamientos'. A diferencia de los biomarcadores que dependen completamente de procesos o algoritmos automáticos, los COAs dependen de la implementación, interpretación y declaración de un paciente, un clínico, o un observador. Los cuatro tipos de COAs son los resultados reportados por pacientes (PRO), los resultados reportados por clínicos (ClinRO), los resultados reportados por observadores (ObsRO), y los resultados basados en pruebas de rendimiento (PerfO).

PROQOLID con su nombre inicial, QOLID (*Quality of Life Instruments Database*), fue diseñado en 2002 en colaboración con el Dr. Marcello Tamburini (*National Cancer Institute*, Milan), que previamente había creado la página web QLMed.org con el propósito de proveer a los investigadores o clínicos implicados en la evaluación de los resultados reportados por los pacientes con una biblioteca comprensiva de instrumentos disponibles *online*.

En la primera etapa se incluyeron 470 instrumentos PRO ordenados en varias categorías. Cada año se ha ido añadiendo información, llegando recientemente a los 900 instrumentos. La web recibe aproximadamente unas 800 visitas diarias. Recientemente esta base de datos se ha englobado dentro de

la plataforma de e-comercio ePROVIDE que agrupa además de la información de PROQOLID otros buscadores de estudios o guías que incluyan tanto instrumentos para medir PROs, como otros COA. Aunque una proporción elevada de los contenidos de esta librería son los instrumentos PRO, y especialmente aquellos que miden calidad de vida relacionada con la salud.

Se puede acceder a una parte de la información contenida en PROQOLID de manera gratuita. Los no suscriptores pueden disponer de información básica de los instrumentos: nombre de los autores, *copyright*, objetivo, área e indicación terapéutica, tipo de COA, idioma original en el que fue desarrollado, y referencias bibliográficas. Los suscriptores pueden disponer además de: información de contacto y condiciones de uso, copia o muestra del instrumento, lista de los idiomas en los que está disponible y sus referencias bibliográficas, versiones electrónicas, características (concepto medido, número de ítems, población diana, periodo recordatorio, tiempo y modo de administración), dominios, puntuaciones, documentación sobre la validez de contenido, y características métricas evaluadas en el estudio de validación original.

Para que un instrumento sea elegido para incluirse en este banco, debe cumplir los siguientes criterios: disponer de una publicación que describa el desarrollo o validación; un propietario intelectual claramente identificado; y una versión master del cuestionario en inglés provista por los propietarios intelectuales, autores o distribuidores disponible para ser enviada a los usuarios potenciales bajo petición. Es por ello que, a diferencia de otros buscadores, permite acceder directamente al instrumento en la mayoría de casos. PROQOLID permite realizar búsquedas por dominios medidos, por áreas terapéuticas o por palabras clave.

V. Evaluación de los instrumentos PRO

Los instrumentos PRO deben cumplir unos mínimos estándares de calidad para ser útiles tanto en investigación como en la práctica clínica. Aunque ha habido diferentes iniciativas para evaluar los instrumentos PRO, actualmente existen dos herramientas estandarizadas bien establecidas: *Consensus-based Standards for the selection of health status Measurement INstruments* (COSMIN) y *Evaluating the Measurement of Patient-Reported Outcomes* (EMPRO). COSMIN y EMPRO son, de hecho, herramientas complementarias ya que la primera valora la calidad metodológica de los estudios que evalúan las propiedades de uno o varios instrumentos PRO, y la segunda la calidad de los instrumentos (es decir, hasta qué punto sus propiedades métricas se ajustan a los criterios de calidad establecidos).

1. Evaluating Measures of Patient Reported Outcomes (EMPRO)

EMPRO permite una evaluación estandarizada, mediante consenso de expertos, de cada instrumento a través de la evidencia científica disponible. Los desarrolladores de EMPRO definieron la calidad de un instrumento PRO como “el grado de confianza en que todo posible sesgo ha sido minimizado y que la información sobre el proceso que llevó a su desarrollo y evaluación sea clara y accesible”⁵⁵. EMPRO combina 3 aspectos fundamentales: 1) atributos de medición bien descritos y establecidos, 2) revisores expertos para llevar a cabo las evaluaciones, y 3) puntuaciones que permiten una comparación directa entre instrumentos PRO.

EMPRO ha demostrado excelente fiabilidad⁵⁵ en términos de consistencia interna (alfa de Cronbach= 0,95) y concordancia inter-evaluador (coeficientes de correlación intraclase 0,87-0,94), así como su utilidad para comparar el funcionamiento de instrumentos genéricos⁵⁵ y específicos de alguna enfermedad, como la insuficiencia cardíaca¹¹³, cáncer de prostata¹¹⁴ y patologías del hombro¹¹⁵.

EMPRO está basado en una exhaustiva lista de 39 criterios acerca de los 8 atributos ideales que deberían cumplir los instrumentos PRO según las directrices del Comité Científico del *Medical Outcomes Trust*¹¹⁶:

1. “Modelo conceptual y de medición” (7 criterios), que evalúa cómo se han obtenido los conceptos a medir y hasta qué punto ha participado la población a la que va dirigido el instrumento;
2. “Fiabilidad” (8 criterios), que considera hasta qué punto el instrumento está libre de error aleatorio, tanto si se ha evaluado la consistencia interna como la reproducibilidad test-retest;
3. “Validez” (6 criterios), que considera hasta qué punto el instrumento mide aquello que pretende medir incluyendo la validez de contenido, de constructo y de criterio.
4. “Sensibilidad al cambio” (3 criterios), que evalúa la capacidad de detectar cambios en el tiempo;
5. “Estrategias para la interpretación” (3 criterios), que evalúa las estrategias propuestas para asignar sentido a las puntuaciones del instrumento;

6. “Carga de administración” (7 criterios), que considera el tiempo, esfuerzo y otras demandas requeridas tanto para administrar como para responder el instrumento;
7. “Modos de administración alternativos” (2 criterios), que evalúa la equivalencia entre los modos de administración propuestos para el instrumento, tales como auto-administrado, administrado por entrevista cara a cara, entrevista telefónica o asistida por ordenador;
8. “Adaptaciones culturales o lingüísticas” (3 criterios), que evalúa el proceso de adaptación seguido para conseguir la equivalencia de las versiones adaptadas.

Todos los atributos EMPRO se acompañan de una breve descripción para facilitar la comprensión del sentido original y garantizar una aplicación estandarizada durante el proceso de evaluación. La conformidad con cada ítem se puede responder en una escala de 4 puntos Likert, desde 4 (totalmente de acuerdo) a 1 (en total desacuerdo). El recuadro de “sin información” se puede marcar en caso de información insuficiente. Cinco ítems permiten responder “no aplicable”. Se recomienda proporcionar comentarios detallados para justificar cada puntuación EMPRO. Estos comentarios ayudan en la interpretación de las puntuaciones EMPRO.

La evaluación EMPRO consiste en dos rondas consecutivas. En la primera ronda, cada experto evalúa independientemente el instrumento asignado mediante la revisión del texto completo de artículos identificados a través del proceso de revisión sistemática aplicando la herramienta EMPRO. En la segunda ronda, cada experto recibe las puntuaciones del otro/s experto/s que evalúa/n el mismo instrumento. En caso de discrepancias, primero, se les invita a resolverlas mediante la discusión hasta llegar a un consenso y, si es necesario, se pueden resolver mediante un tercer revisor.

Se calculan puntuaciones específicas de cada atributo y una puntuación global. La información detallada y los algoritmos para obtener las puntuaciones EMPRO se encuentran disponibles *online* (<http://www.bibliopro.org/buscador/987/evaluating-the-measurement-of-patient-reported-outcomes-version-espanola>). Primero se calcula la media de los ítems aplicables para cada atributo (cuando al menos 50% han sido puntuados), teniendo en cuenta que a aquellos ítems donde se señala la opción de respuesta “sin información” se les asigna la puntuación 1 (puntuación más baja posible); y, segundo, esta media cruda se transforma linealmente a un rango que va de 0 (peor puntuación posible) a 100 (mejor puntuación posible). Para los atributos de “fiabilidad” y “carga” se calculan subpuntuaciones independientes ya que cada uno consta de dos componentes: “consistencia interna” y “reproducibilidad” para la fiabilidad, así como “de respuesta” y “administración” para carga. Para la fiabilidad, se escoge la subpuntuación más alta para sus dos componentes para representar al atributo.

Además de las puntuaciones específicas de los atributos, se obtiene una puntuación global calculando la media de cinco atributos relacionados con las propiedades de medición: “modelo conceptual y de medida”, “fiabilidad”, “validez”, “respuesta al cambio”, e “interpretabilidad”. Cuando el objetivo es evaluar una versión adaptada, en el cálculo de la puntuación global debería incluirse también la puntuación obtenida en el atributo de “adaptaciones lin-

güísticas o culturales”. La puntuación global solo se calcula cuando al menos tres de estos cinco atributos tienen puntuación. Las puntuaciones EMPRO se consideran razonablemente aceptables si llegan como mínimo a los 50 puntos (de los 100 puntos máximos teóricos). El umbral se escogió en base a las recomendaciones globales de los revisores en los dos primeros estudios EMPRO¹¹³. La curva *Receiver Operating Characteristic* (ROC) se calculó para evaluar la coincidencia entre las puntuaciones de los atributos EMPRO y las recomendaciones globales de los revisores. El área bajo la curva ROC fue de 0,87, y el punto de corte sugerido fue 51.

La herramienta EMPRO proporciona una puntuación para cada uno de los 8 atributos y una puntuación global que oscila de 0 a 100 puntos, así como un grado de recomendación general del instrumento.

2. COSMIN, un consenso internacional para la evaluación de los estudios sobre PROs

COSMIN es el acrónimo de la expresión: *CO*nsensus-based *S*tandards for the *S*election of *h*ealth status *M*easurement *I*nstruments (Estándares para la selección de instrumentos de medida del estado de salud). Esta iniciativa fue concebida en el año 2005 en el contexto de desarrollos metodológicos para el análisis de revisiones sistemáticas de la literatura en el seno de la Colaboración Cochrane¹¹⁷. En un principio, se centró en el desarrollo de una lista de comprobación (*checklist*) para evaluar la calidad científica de los artículos que reportaban las propiedades métricas de los instrumentos de medida de la salud. Es decir, una escala de calidad de las evidencias disponibles como las que se usan para poder cualificar el grado de validez interna y externa en las revisiones sistemáticas. Una característica importante de esa iniciativa fue su deliberado carácter internacional, así como el uso de metodologías cualitativas formales en todo el proceso. COSMIN ha ido evolucionando y su pretensión es que esa misma lista de comprobación y otros recursos sean útiles para comparar instrumentos PRO y facilitar su selección. En la actualidad, la iniciativa ha desarrollado las siguientes herramientas COSMIN: 1) Taxonomía y definiciones de las propiedades métricas; 2) Lista de comprobación (*checklist*) para evaluar la calidad de los estudios; 3) Filtro de búsqueda de estudios; 4) Protocolo para revisiones sistemáticas de instrumentos PRO; 5) Base de datos de revisiones sistemáticas de PROs; y 6) Guía para la selección de PROs en un Conjunto Básico de Resultados (*Core Outcome Set*). Además de su uso en las revisiones sistemáticas de la evidencia sobre los instrumentos PRO, COSMIN puede tenerse en cuenta a la hora de elaborar o evaluar manuscritos científicos sobre PROs o en el diseño de los estudios para evaluar sus propiedades métricas. En este capítulo se revisan las aportaciones más destacadas.

Taxonomía y lista de comprobación

Se organizó un estudio Delphi (de 4 rondas) en el que participaron 51 expertos internacionales de diferentes disciplinas (psicología, epidemiología, estadística y medicina)^{118,119}. Los expertos fueron interrogados en primer lugar sobre cuáles son las propiedades de medida que deberían ser incluidas en la lista de comprobación. Como es habitual en este tipo de metodología, solo se incluyeron en la definitiva aquellas propiedades para las cuales se había alcanzado un nivel alto de acuerdo (un consenso mínimo del 67% de los expertos debían estar de acuerdo o muy de acuerdo en su inclusión).

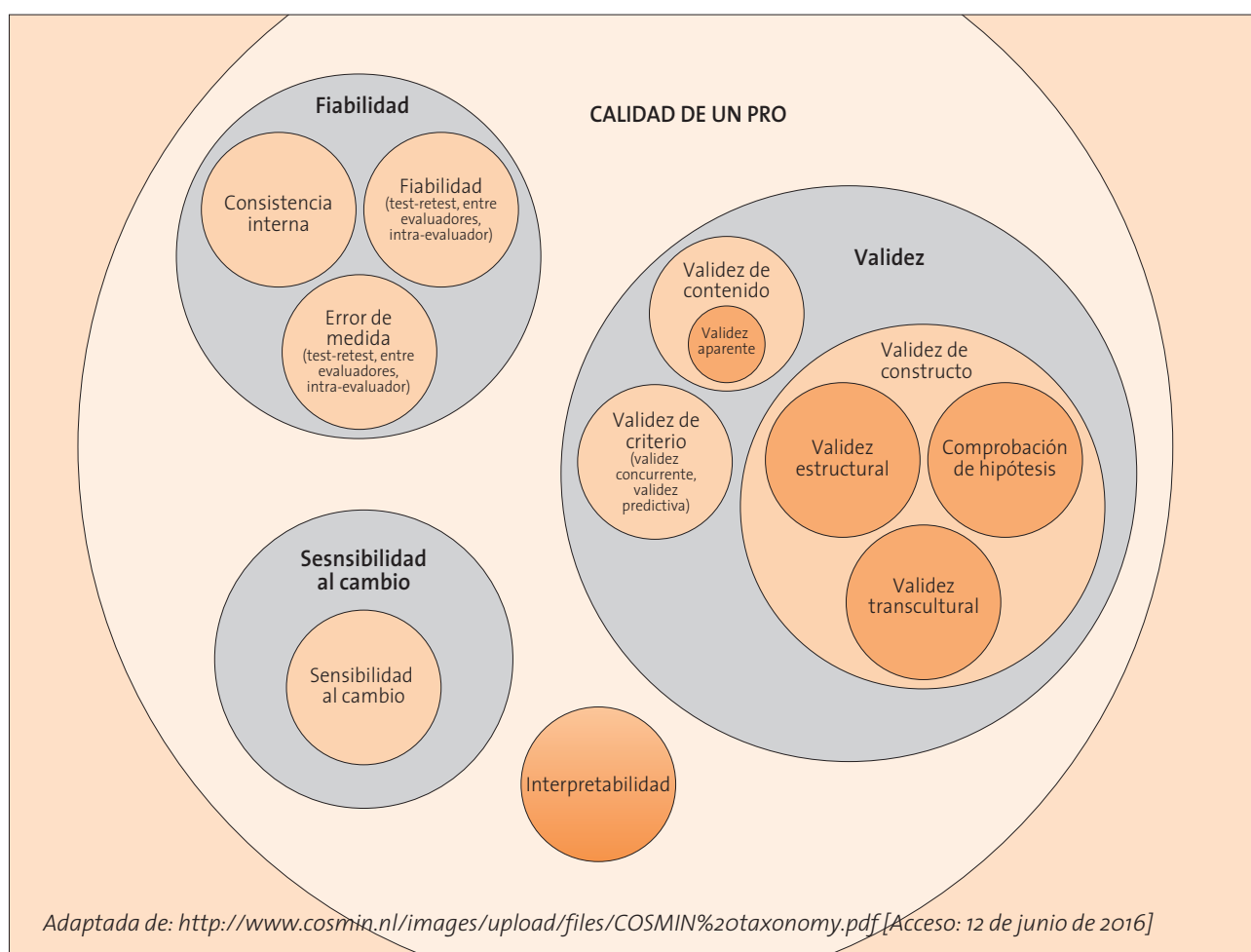


FIGURA 7. Representación gráfica de la Taxonomía COSMIN.

Las propiedades se clasificaron en: A) Consistencia Interna; B) Fiabilidad; C) Error de Medida; D) Validez de Contenido; E) Validez Estructural; F) Test de Hipótesis; G) Validez Transcultural; H) Validez de Criterio; I) Sensibilidad al Cambio; y J) Interpretabilidad. Posteriormente se desarrolló la taxonomía COSMIN, en las que estas propiedades se resumen en 4 agrupaciones: Fiabilidad (Consistencia interna, Fiabilidad, Error de medida); Validez (Constructo –estructural, hipótesis, transcultural–; Contenidos y Criterio); Sensibilidad al Cambio e Interpretabilidad.

La lista de comprobación está compuesta por 12 cajas (de la A a la J), una caja para cada una de las propiedades métricas. Además, hay una caja para los requisitos específicos de los instrumentos basados en la teoría de respuesta al ítem (IRT) y otra para la evaluación de la generalización de los resultados. Cada una de las cajas incluye entre 5 y 13 ítems o preguntas que evalúan el diseño y los análisis estadísticos explicitando los estándares que deben cumplir los estudios sobre instrumentos de salud. Un ejemplo de estándar de diseño para la propiedad de Consistencia Interna (A) es el siguiente: “¿Se reportó el número de ítems sin respuesta?”. Inicialmente la adecuación de este estándar se valoraba con un Sí, un No, aunque posteriormente se desarrollaron 4 categorías de valoración¹²⁰. En el caso de la

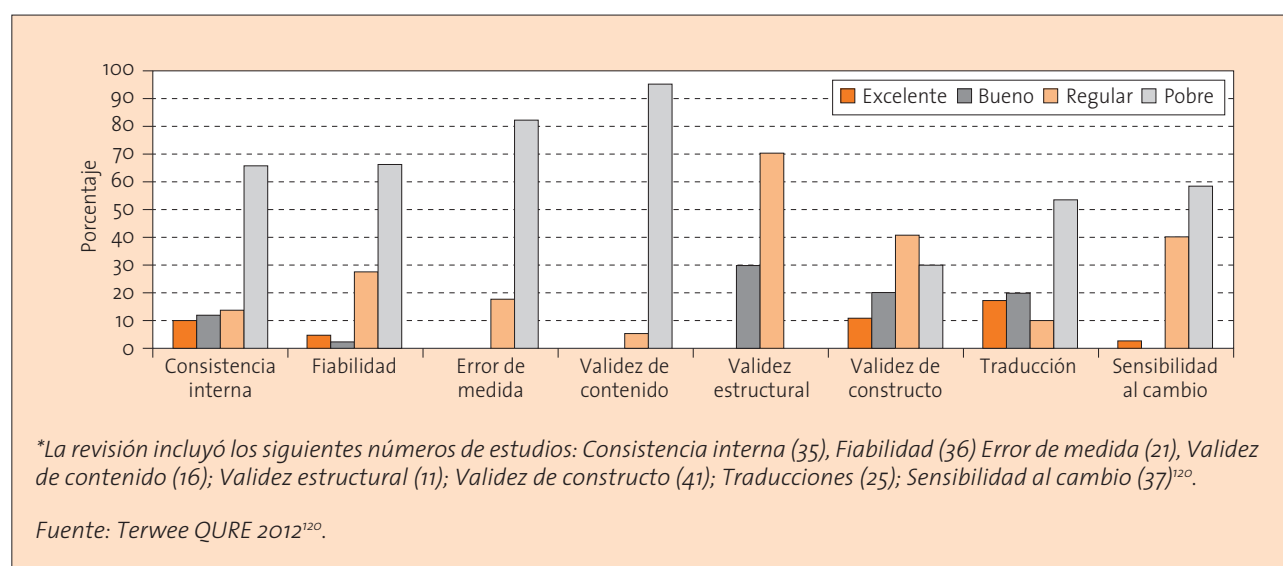


FIGURA 8. Resultados de la revisión de la calidad de estudios sobre PROs realizados con la guía COSMIN^{120*}

propiedad Consistencia Interna hay 7 estándares que se refieren al diseño. Un ejemplo de estándar de análisis estadístico para esta misma propiedad es: (para estudios siguiendo la Teoría clásica del Test-CTT): “¿Se calculó el coeficiente alfa de Cronbach o el KR-20?”. Para la Consistencia Interna hay 3 estándares de análisis estadístico¹¹⁸.

La compleción de la guía de comprobación se realiza en 4 etapas: 1) Identificar las propiedades de medida evaluadas en el artículo y seleccionar las cajas (A-J) que se van a completar; 2) Determinar si el estudio está basado en teoría clásica del test (CTT, por sus siglas en inglés) o en la teoría moderna del ítem (IRT, por sus siglas en inglés); 3) Determinar el grado en el que el estudio cumple con los estándares de buena calidad metodológica para las propiedades métricas evaluadas; y 4) Determinar la generalización de los resultados del estudio. En total hay 89 estándares de los cuales, 52 corresponden a requisitos de diseño, 18 a requisitos de análisis estadísticos y 19 son de carácter general.

En su versión más reciente, cada uno de estos estándares se debe puntuar con una escala categórica de 4 niveles: “Excelente / Bueno / Regular / Pobre”, lo que permite calcular una puntuación de la calidad metodológica para cada una de las 8 propiedades de medida incluidas en la guía COSMIN¹²⁰. Así, por ejemplo, esta puntuación se calculó para 8 cuestionarios de medida de la discapacidad de cuello, a partir de la información obtenida en 46 artículos. 41 de esos artículos evaluaron la validez de constructo y tan solo en 13 de ellos (30%) los resultados fueron calificados de “Excelente” o “Buenos”. Lo que supone que para la mayoría, los estudios que presentaron datos sobre validez de constructo eran de calidad “Regular” (40%) o “Pobre” (30%). La figura 8 presenta los resultados de esa revisión, lo que muestra una evidencia global de la calidad de los estudios. La calidad metodológica resultó muy baja para las propiedades de Validez de contenido y todas las de Fiabilidad (Error de medida, Consistencia interna y Fiabilidad). Pero también para la Sensibilidad al cambio y las Adaptaciones transculturales.

Protocolo para la revisión sistemática de instrumentos PRO

Las revisiones sistemáticas de instrumentos PRO son importantes para una selección basada en la evidencia a la hora de usar estos instrumentos de medida. Para poder llevar a cabo esa revisión el grupo COSMIN desarrolló un procedimiento de 10 pasos, que puede ser consultado en la web de la iniciativa [accesible en este enlace: <http://database.cosmin.nl/>]. Este procedimiento incluye el desarrollo de filtros de búsqueda así como la manera de aplicar la guía de comprobación COSMIN para valorar la calidad de los estudios. Además, COSMIN propone un método de síntesis de la mejor evidencia a partir del número de estudios, su calidad, determinada a partir de la lista de comprobación, y la consistencia de los resultados que permitiría determinar la fortaleza de la evidencia que existe para cada una de las propiedades de medida. Como notan los investigadores COSMIN, este procedimiento es similar al que se propone el grupo de trabajo GRADE [accesible en este enlace: <http://www.gradeworkinggroup.org>]. Así, por ejemplo, se propone considerar una evidencia como fuerte, cuando existen resultados consistentes con valores superiores a 0,70 para el coeficiente de correlación intraclase (ICC) o para el índice Kappa en 2 o más estudios calidad buena o en un solo estudio de calidad excelente. La tabla 5 incluye un listado más exhaustivo de los criterios de calidad de las propiedades métricas de un instrumento PRO que propone COSMIN para valorar cada una de las propiedades métricas de los instrumentos PRO^{121, 122}.

Base de datos de revisiones sistemáticas de instrumentos PRO

Como parte de la iniciativa COSMIN, el grupo mantiene una base de datos continua de todas las revisiones sistemáticas de los instrumentos PRO. Las revisiones sistemáticas son herramientas importantes para informar mejor la selección de PROs tanto en la investigación como en la práctica clínica. Y también lo son para identificar vacíos de conocimiento sobre la calidad de las propiedades métricas de los instrumentos PRO. Según se informa en la web COSMIN, esta búsqueda se actualizó por última vez el 19 de junio de 2014 y la base de datos contiene actualmente 569 revisiones sistemáticas únicas de PROs, accesible en este enlace: <http://database.cosmin.nl>

Parte de la actividad del grupo COSMIN se basa, precisamente, en realizar búsquedas en las bases de datos PubMed y EMBASE para identificar las revisiones sistemáticas sobre PROs publicadas anualmente. Para ello los investigadores de la iniciativa desarrollaron y validaron un filtro de búsqueda de estudios sobre propiedades de medida en PubMed¹²³. La búsqueda inicial mostró una sensibilidad del 97,4%, con un valor predictivo positivo del 4,4%. Ese filtro ha sido posteriormente adaptado para las bases de datos EMBASE, CINAHL y MEDLINE (con OVID) y se encuentra accesible en el sitio web de COSMIN. [El filtro para EMBASE puede encontrarse en este enlace: <http://www.cosmin.nl/images/upload/files/Translated%20search%20filter%20for%20measurement%20properties%20for%20EMBASE.pdf>]

Para que una revisión sistemática se incluya en la base de datos COSMIN se deben cumplir las siguientes condiciones:

- El estudio debe ser una revisión sistemática (es decir, de haber realizado al menos una búsqueda sistemática en al menos una base de datos electrónica).
- El objetivo de la revisión debe ser el de identificar todos los instrumentos PRO de interés y debe haber sido realizada con la intención de evaluar o resumir las evidencias sobre sus propiedades de medida.

TABLA 5. Criterios y estándares de calidad usados en COSMIN.

Propiedad	Valoración	de los Estándares de Calidad
Fiabilidad		
Consistencia interna	+	(Sub)escala unidimensional y alfa de Cronbach $\geq 0,70$
	?	Dimensionalidad desconocida o alfa de Cronbach indeterminada
	–	(Sub)escala no unidimensional o alfa de Cronbach $< 0,70$
Error de medida	+	MID $>$ SDC o MID fuera de LOA
	?	MID sin definir
	–	MID $\leq$ SDC o MID equivale o dentro de LOA
Fiabilidad	+	ICC/Kappa ponderada $\geq 0,70$ o la r de Pearson $< 0,80$
	?	NI ICC/Kappa ponderada, NI la r de Pearson determinadas
	–	ICC/ Kappa ponderada $< 0,70$ o la r de Pearson $< 0,80$
Validez		
Validez de contenido	+	La población diana considera que todos los ítems del cuestionario son relevantes y considera que el cuestionario está completo
	?	La población no ha estado involucrada
	–	La población diana considera algunos ítems del cuestionario irrelevantes o considera que el cuestionario está incompleto
Validez de constructo		
Validez estructural	+	Los factores explican como mínimo 50% de la varianza
	?	No se menciona la varianza explicada
	–	Los factores explican $< 50\%$ de la varianza
Comprobación hipótesis	+	(Correlación con un instrumento que mida el mismo constructo $\geq 0,50$ o como mínimo 75% de los resultados concuerdan con las hipótesis) y correlación mayor con constructos relacionados que con constructos no relacionados
	?	Únicamente se indican correlaciones con constructos no relacionados
	–	Correlación con un instrumento que mida el mismo constructo $< 0,50$ o $< 75\%$ de los resultados concuerdan con las hipótesis o correlación menor con constructos relacionados que con constructos no relacionados
Sensibilidad al cambio		
Sensibilidad al cambio	+	(Correlación con un instrumento que mida el mismo constructo $\geq 0,50$ o como mínimo 75% de los resultados concuerdan con las hipótesis o AUC $\geq 0,70$) y correlación mayor con constructos relacionados que con constructos no relacionados
	?	Únicamente se determinan correlaciones con constructos no relacionados
	–	Correlación con un instrumento que mida el mismo constructo $< 0,50$ o $< 75\%$ de los resultados concuerdan con las hipótesis o AUC $\geq 0,70$ o correlación menor con constructos relacionados que con constructos no relacionados

[...] número de referencia. MID: Mínima Diferencia Importante; SDC: Cambio Mínimo Detectable; LOA: Límites de Acuerdo; ICC: Coeficiente de Correlación Intraclass; AUC: Área Bajo la Curva ROC.

+: valoración positiva; ?: valoración indeterminada; -: valoración negativa.

Fuente: Schellingerhout JM, QURE 2012121.

- El constructo o concepto de interés para la revisión debe ser el estado de salud (o algún aspecto concreto dentro del mismo), que se define como: (a) los procesos biológicos y fisiológicos, o (b) síntomas, o (c) el funcionamiento físico, o (d) el funcionamiento social y el psicológico, o (e) la percepción de salud general, o (f) la calidad de vida relacionada con la salud de la vida (tal como se define en el modelo de Wilson y Cleary⁴⁴).
- La población de estudio debe contener seres humanos (pacientes o población en general).
- Los instrumentos de interés deben ser PROs, que se definen como instrumentos aplicables a estudios longitudinales para monitorear los cambios en la salud a través del tiempo (la medida de resultado es la variable dependiente).
- El estudio deberá evaluar e informar sobre al menos una o más propiedades de medida de los PROs incluidos.

Nuevos desarrollos del grupo COSMIN

COSMIN está realizando nuevas consultas a expertos con la finalidad de actualizar la guía de comprobación, pues con su uso se han identificado algunas limitaciones¹¹⁷. En primer lugar, la necesidad de mejorar los estándares sobre las evidencias de validez de contenido, a la vista de las exigentes propuestas sobre el desarrollo y obtención del contenido de los PRO que han propuesto agencias reguladoras como la FDA (véase la sección correspondiente en esta monografía). En segundo lugar, el uso cada vez mayor de la Teoría de Respuesta al Ítem para la evaluación de las propiedades de medida de los PROs, requiere una extensión de los estándares existentes en la guía de COSMIN para los estudios basados en esa aproximación. Finalmente, aunque COSMIN se desarrolló para evaluar la calidad de los estudios de las medidas PRO, su metodología debería ser adecuada para otros instrumentos relevantes en la evaluación de la atención centrada en los pacientes, como son las medidas reportadas por los clínicos (como puede ser la valoración de un test de máxima distancia alcanzada en 6 minutos, *6-Minute Walking Test*, u otros).

Sin embargo, el cambio más importante que viene experimentando la iniciativa COSMIN es el nuevo énfasis del uso de la guía de comprobación para la evaluación directa de los instrumentos PRO. La robustez de la metodología COSMIN se está dirigiendo no solo a valorar la calidad de los estudios que informan sobre las propiedades métricas de los PRO, sino sobre todo a sintetizar el rendimiento de esas propiedades en cada uno de los PRO. Y así poder convertirse en un método de apoyo a la selección de los mejores PRO. Si bien esta ha sido siempre la finalidad última de COSMIN, esta orientación supone un cambio interesante no exento de dificultades. Además de las limitaciones mencionadas en el párrafo anterior, el principal problema es la complejidad (y hasta cierto punto la redundancia) de los estándares propuestos. Se necesitará información empírica para hacer de la guía de comprobación COSMIN una herramienta eficiente para la selección de instrumentos PRO de acuerdo a su nivel de rendimiento de medición.

Epílogo: Conclusiones y recomendaciones

La lectura de esta monografía deja claro que los instrumentos PRO han seguido una evolución ascendente llegando a ser una opción fundamental, no solo en la investigación de la efectividad comparativa, sino por su gran potencial para influir la realidad diaria en los servicios sanitarios, añadiendo valor a la atención clínica individual, mejorando la calidad de los procesos y los incentivos profesionales, el *benchmarking* y la comparación de proveedores, y los sistemas de información de salud poblacional.

Los instrumentos PRO tienen un lugar de importancia creciente en la medicina y la investigación centrada en el paciente. Pero, como también queda patente en la presente obra, para alcanzar su máximo impacto potencial se debe avanzar en muchos aspectos que implican tanto a los investigadores como a los profesionales sanitarios y los usuarios, así como a los gestores, reguladores y los responsables políticos de la salud.

A continuación hacemos algunas recomendaciones que nos parecen adecuadas para la consecución de un mayor impacto potencial de los instrumentos PRO en nuestro entorno.

Educación de los distintos implicados. Los profesionales sanitarios deben estar convencidos de que la naturaleza subjetiva de los PROs tiene un valor propio e importante y que no es una debilidad de estas variables. Elementos centrales de esa educación serían: el valor intrínseco de su contenido (conceptos de alta complejidad difícilmente observables con otros instrumentos); la producción de valores sintéticos, válidos y precisos, predictivos de necesidades y costes sanitarios; y el hecho de ser apropiados para diversos usos y contextos, que incluyen desde la atención clínica individual hasta perspectivas más generales del sistema sanitario. La conexión de los datos individuales clínicos y poblacionales está mejorando sustancialmente los sistemas de información de la salud, permite una monitorización más adecuada de los problemas de salud y contribuye a la priorización de problemas a distintos niveles. El proceso de educación debe incluir una mejor identificación de las necesidades informativas a cubrir por los PROs y debe ayudar a moldear las expectativas de los implicados en su uso.

Implementación tecnológica más rápida (de la tecnología existente y de la futura): la captura de datos y su inmediata evaluación y retorno en el ámbito de la atención individualizada es claramente factible. En este sentido, las formas flexibles de los instrumentos (Test Adaptativos Informatizados o CAT) son clave. En otros ámbitos (por ejemplo, a nivel institucional), la inmediatez no es necesaria, pero sí se necesita acortar el tiempo entre la producción y la evaluación de la información. La incorporación de instrumentos PRO en la HC electrónica, los registros de enfermedades y otros sistemas de información de salud los puede convertir en parte importante de la provisión y la gestión de servicios, así como para la evaluación y comparación de los proveedores de los servicios y como un elemento de valor específico en los sistemas de vigilancia de la salud poblacional. Parte de la tecnología debe ser aplicada a mejorar la

comparabilidad de la información obtenida con diferentes instrumentos PRO para poder interpretar su información de manera homogénea independiente del tipo de instrumento que la haya generado. Los desarrollos recientes de mapeo y co-calibración de instrumentos (como el proyecto PROsetta Stone, por ejemplo) ya lo permiten.

Evaluación coste-efectividad a través de instrumentos PRO que incorporan utilidades o preferencias de poblaciones por los estados de salud (como el EQ-5D, por ejemplo). Para muchas enfermedades cuyos tratamientos se dirigen fundamentalmente a mejorar la salud percibida (sin afectar probablemente a la supervivencia), incluir los costes en un indicador de resultado se convierte en una herramienta que puede mejorar los procesos para la priorización de los servicios de salud. Este es también el objetivo de la atención basada en el valor que promueve el Consorcio Internacional para los Sets Estandarizados de Medidas de Resultados de Salud (*International Consortium for Health Outcomes Measurement's Standardized Outcome Sets* – ICHOM–) ¹²⁴, y que entre 2013-2015 ha completado los sets para las 21 patologías que representan el 45% de la carga de enfermedad global.

Otros conceptos reportados por los pacientes, como los PREMs (*Patient-Reported Experience Measures*), que miden aspectos relevantes de la calidad percibida de los procesos de atención sanitaria y complementan el valor de los PROs. Los PREMs típicamente recogen la experiencia sobre la atención, los tratamientos y el soporte recibido, aspectos todos ellos relevantes para la mejora de la calidad de los servicios. Y sus puntuaciones son en buen parte independientes de la salud percibida. Aunque su desarrollo métrico actual es más limitado que el de los PROs, los PREMs representan otro elemento imprescindible en la medicina y la investigación centrada en los pacientes.

Una Hoja de Ruta para los instrumentos PRO

Atención clínica

Deben llegar a comprender hasta qué punto los instrumentos PRO son un elemento de ayuda para la atención centrada en el paciente; que su información rigurosa y sintética tiene un carácter complementario, y no sustitutivo, del arsenal de información clínica y parámetros biológicos. Estos últimos son en su mayoría resultados “intermedios”, si se comparan con los resultados finales (recuperación funcional, alivio sintomático, mejora del bienestar, etc.). Deben exigir las condiciones adecuadas para que la administración y uso de los instrumentos PRO puedan ser uno de los elementos de la atención habitual sin sacrificar ninguna de las muchas otras actividades imprescindibles en su crucial tarea.

Gestión sanitaria

Los gestores deben seguir redirigiendo la atención para centrarla más en los pacientes y en sus resultados, no solo en los procesos. Sin embargo, todavía no hay suficiente evidencia como para ser prescriptivos en la interpretación de resultados a nivel de proveedores (por ejemplo, en la comparación entre proveedores de los cambios observados con los servicios prestados o del estado de salud de la población atendida). Se requiere pues investigación evaluativa sobre la aportación de los instrumentos PRO a la comparación

de proveedores. En todo caso, estos instrumentos pueden aportar un retorno de alto valor para los propios centros, para su autoconocimiento y para la monitorización de la evolución de los resultados de su atención y de la población a la que atienden.

Regulación

El papel de las agencias gubernamentales en otros países (como la FDA americana, entre muchas otras) ha sido clave para aumentar la aplicación de los instrumentos PRO y de la medicina centrada en el paciente. En nuestro país faltan guías claras del Ministerio de Sanidad y de las Comunidades Autónomas para la incorporación de los instrumentos PRO en el arsenal de indicadores sobre: las variables finales en los ensayos clínicos para la evaluación de nuevos fármacos e intervenciones terapéuticas; la calidad de la atención de los servicios y de los proveedores; la salud de la población, para definir los planes de salud; y las desigualdades en salud.

Investigación

Las entidades financiadoras de investigación, siguiendo la tendencia global hacia la investigación participativa, deberían establecer mecanismos y herramientas para favorecer la comunicación entre investigadores y otros actores implicados, especialmente los pacientes, en el diseño y la diseminación de proyectos de investigación para garantizar resultados relevantes y aplicables a la práctica. Los investigadores deben aplicar los estándares de calidad disponibles en la actualidad para seleccionar y/o desarrollar instrumentos PRO adecuados (EMPRO, COSMIN, o guía FDA según el objetivo); así como diseñar y describir los proyectos de investigación siguiendo las recomendaciones de la extensión de la guía CONSORT para los PROs¹²⁵. Tanto esta última como la guía de la FDA coinciden en poner el énfasis en la evidencia sobre las características métricas de los instrumentos PRO utilizados y en el balance entre los PROs y otros resultados relevantes.

Bibliografía

1. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. *N Engl J Med*. 2012; 366: 780-1.
2. Calkins DR, Rubenstein LV, Cleary PD, Davies AR, Jette AM, Fink A, et al. Failure of physicians to recognize functional disability in ambulatory patients. *Ann Intern Med*. 1991; 114: 451-4.
3. Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res*. 2013; 13: 211.
4. Doward LC, McKenna SP. Defining patient-reported outcomes. *Value Health*. 2004; 7 Suppl 1: S4-8.
5. FDA. Guidance for Industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Food and Drug Administration (FDA). Disponible en: <http://www.fda.gov/downloads/Drugs/Guidances/UCM193282.pdf> [Acceso: 20 febrero, 2013].
6. Karnofsky DA, Abelmann WH, Craver LF, Burchenal JH. The use of the nitrogen mustards in the palliative treatment of carcinoma - With particular reference to bronchogenic carcinoma. *Cancer*. 1948; 634-56.
7. Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. *JAMA*. 1949; 140: 659-62.
8. Belloc NB, Breslow L, Hochstim JR. Measurement of physical health in a general population survey. *Am J Epidemiol*. 1971; 93: 328-36.
9. Fanshel S, Bush JW. A health-status index and its application to health-services outcomes. *Operations Res*. 1970; 18: 1021-66.
10. McHorney CA. Health status assessment methods for adults: past accomplishments and future challenges. *Annu Rev Public Health*. 1999; 20: 309-35.
11. Gilson BS, Gilson JS, Bergner M, Bobbitt RA, Kressel S, Pollard WE, et al. The Sickness Impact Profile. Development of an outcome measure of health care. *Am J Public Health*. 1975; 65: 1304-10.
12. Stewart AL, Hays RD, Ware JE. The MOS Short-form General Health Survey. Reliability and validity in a patient population. *Med Care*. 1988; 26: 724-35.
13. Ware J Jr., 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.
14. de Bruin AF, Diederiks JP, de Witte LP, Stevens FC, Philipsen H. The development of a short generic version of the Sickness Impact Profile. *J Clin Epidemiol*. 1994; 47: 407-18.
15. Tarlov AR. Shattuck lecture--the increasing supply of physicians, the changing structure of the health-services system, and the future practice of medicine. *N Engl J Med*. 1983; 308: 1235-44.
16. Muñoz PE, Vázquez JL, Rodríguez Insausti F, Pastrana E, Varo J. Adaptación española del General Health Questionnaire (G.H.Q.) de D.P. Goldberg (Un método de identificación de casos psiquiátricos en la comunidad). *Arch Neurobiol*. 1979; 42: 139-58.
17. Alonso J, Anto JM, Moreno C. Spanish version of the Nottingham Health Profile: translation and preliminary validity. *Am J Public Health*. 1990; 80: 704-8.
18. Badia X, Alonso J. Adaptación de una medida de la disfunción relacionada con la enfermedad: la versión española del Sickness Impact Profile. *Med Clin (Barc)*. 1994; 102: 90-5.

19. Badia X, Fernandez E, Segura A. Influence of socio-demographic and health status variables on evaluation of health states in a Spanish population. *Eur J Public Health*. 1995; 5 : 87-93.
20. Alonso J, Prieto L, Anto JM. La versión española del "SF-36 Health Survey" (Cuestionario de Salud SF-36): Un instrumento para la medida de los resultados clínicos. *Med Clin (Barc)*. 1995; 104: 771-6.
21. Lizan Tudela L, Reig Ferrer A. Adaptacion transcultural de una medida de la calidad de vida relacionada con la salud: la version española de las vinetas COOP/WONCA. *Aten Primaria*. 1999; 24: 75-82.
22. Rodriguez-Fornells A, Rodriguez-Martinez A, Jarne-Esparcia A, Soler-Pujol R, Miarons-Tuneu R, Grau-Fernandez A. Estudio factorial y adaptación de la escala de calidad de vida en la esquizofrenia (QLS). *Rev de Psicol Gral y Aplic*. 1995; 48: 353-64.
23. Alvarez-Ude F, Galan P, Vicente E. Adaptación transcultural y validación preliminar de la versión española del Kidney Disease Questionnaire (Cuestionario de la Enfermedad Renal). *Nefrologia*. 1997; 17 : 486-96.
24. Pardo Y, Castellvi P, Ferrer M, Garin O, Alonso J; Comité Científico BiblioPRO. Nuevos instrumentos en español de resultados percibidos por los pacientes: revisión sistemática BiblioPRO 2011-2012. II Congreso Iberoamericano de Epidemiología y Salud Pública, XXXIII Reunión Científica de la SEE, XVI Congresos SESPAS, X Congreso APE. 2-4 septiembre de 2015, Santiago de Compostela. *Gaceta Sanitaria*. 2015; 29 Supl: 393.
25. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ*. 2002; 324: 1417-21.
26. Badia X, Baro E. Cuestionarios de salud en España y su uso en atención primaria. *Aten Primaria*. 2001; 28: 349-56.
27. Vodicka E, Kim K, Devine EB, Gnanasakthy A, Scoggins JF, Patrick DL. Inclusion of patient-reported outcome measures in registered clinical trials: Evidence from ClinicalTrials.gov (2007-2013). *Contemp Clin Trials*. 2015; 43: 1-9.
28. Patrick DL, Erickson P. Health Status and Health Policy. Allocating resources to Health Care. New York: Oxford University Press; 1993.
29. Tully MP, Cantrill JA. Subjective outcome measurement--a primer. *Pharm World Sci*. 1999; 21: 101-9.
30. FDA. Draft guidance for industry on patient-reported outcome measures: Use in medicinal product development no support labeling claims. *Federal Register*. 2006; 71: 5862-3.
31. Schmidt S, Bullinger M. Current issues in cross-cultural quality of life instrument development. *Arch Phys Med Rehabil*. 2003; 84: S29-S34.
32. Emery MP, Perrier LL, Acquadro C. Patient-reported outcome and quality of life instruments database (PROQOLID): frequently asked questions. *Health Qual Life Outcomes*. 2005; 3: 12.
33. McDowell I, Newell C. Measureing Health. A guide to rating scales and questionnaires. New York: Oxford University Press; 1996.
34. Osoba D. A taxonomy of the uses of health-related quality-of-life instruments in cancer care and the clinical meaningfulness of the results. *Med Care*. 2002; 40: III31-8.
35. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985; 38: 27-36.
36. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care*. 1989; 27: S217-32.
37. Guyatt GH, Veldhuyzen van Zante SJO, Feeny DH, Patrick DL. Measuring quality of life in clinical trials: a taxonomy and review. *Can Med Assoc J*. 1989; 140: 1441-8.
38. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med*. 1993; 118: 622-9.

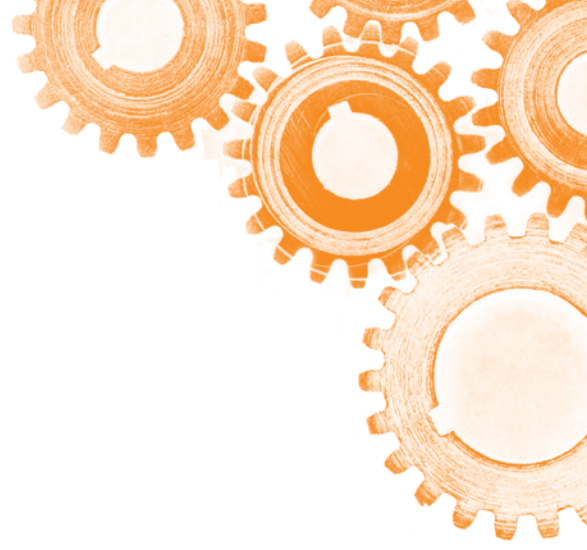
39. MacKeigan LD, Pathak DS. Overview of health-related quality-of-life measures. *Am J Hosp Pharm.* 1992; 49: 2236-45.
40. Patrick DL, Chiang YP. Measurement of health outcomes in treatment effectiveness evaluations: conceptual and methodological challenges. *Med Care.* 2000; 38: 1114-25.
41. Bowling A. *Measuring health: A review of quality of life measurement scales.* Buckingham: Open University Press; 1991.
42. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care.* 2007; 45: S3-11.
43. Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. *Qual Life Res.* 2008; 17: 1125-35.
44. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA.* 1995; 273: 59-65.
45. Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *JAMA.* 1989; 262: 925-30.
46. Salive ME, Mayfield JA, Weissman NW. Patient outcomes research teams and the Agency for Health Care Policy and Research. *Health Serv Res.* 1990; 25: 697-708.
47. van der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q.* 2014; 92: 754-75.
48. Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: challenges and opportunities. *Qual Life Res.* 2009; 18: 99-107.
49. Greenfield S, Nelson EC. Recent developments and future issues in the use of health status assessment measures in clinical settings. *Med Care.* 1992; 30: MS23-MS41.
50. McHorney CA, Earl Bricker DJr. A qualitative study of patients' and physicians' views about practice-based functional health assessment. *Med Care.* 2002; 40: 1113-25.
51. Detmar SB, Muller MJ, Schornagel JH, Wever LD, Aaronson NK. Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA.* 2002; 288: 3027-34.
52. Valderas JM. Measurement of health and disability. *Lancet.* 2007; 370: 483-4.
53. Valderas JM, Alonso J, Guyatt GH. Measuring patient-reported outcomes: moving from clinical trials into clinical practice. *Med J Aust.* 2008; 189: 93-4.
54. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res.* 2008; 17: 179-93.
55. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value Health.* 2008; 11: 700-8.
56. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di DD, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. *J Clin Oncol.* 2014; 32: 1480-501.
57. Greenhalgh J, Long AF, Flynn R. The use of patient reported outcome measures in routine clinical practice: lack of impact or lack of theory? *Soc Sci Med.* 2005; 60: 833-43.
58. Black N, Burke L, Forrest CB, Ravens Sieberer UH, Ahmed S, Valderas JM, et al. Patient-reported outcomes: pathways to better health, better services, and better societies. *Qual Life Res.* 2016; 25: 1103-12.

59. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf.* 2014; 23: 508-18.
60. Appleby J, Devlin N, Parkin D. Using patient reported outcomes to improve Health Care. West Sussex,UK: 2016.
61. Hutchings A, Neuburger J, Grosse FK, Black N, van der Meulen J. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. *Health Qual Life Outcomes.* 2012; 10: 34.
62. Black N. Patient reported outcome measures could help transform healthcare. *BMJ.* 2013; 346: f167.
63. Browne J, Jamieson L, Lewsey J, van der Meulen J, Black N, Cairns J et al. Patient Reported Outcome Measures (PROMs) in elective surgery report to the Department of Health. London: Health Services Research Unit, London School of Hygiene & Tropical Medicine and Clinical Effectiveness Unit, Royal College of Surgeons of England.
64. Neuburger J, Hutchings A, van der Meulen J, Black N. Using patient-reported outcomes (PROs) to compare the providers of surgery: does the choice of measure matter? *Med Care.* 2013; 51: 517-23.
65. Black N, Varaganum M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Qual Saf.* 2014; 23: 534-42.
66. Black N, Varaganum M, Hutchings A. Influence of surgical rate on patients' reported clinical need and outcomes in English NHS. *J Public Health (Oxf).* 2014; 36: 497-503.
67. Varaganum M, Hutchings A, Black N. Relationship between patient-reported outcomes of elective surgery and hospital and consultant volume. *Med Care.* 2015; 53: 310-6.
68. Feng Y, Pistollato M, Charlesworth A, Devlin N, Propper C, Sussex J. Association between market concentration of hospitals and patient health gain following hip replacement surgery. *J Health Serv Res Policy.* 2015; 20: 11-7.
69. Varaganum M, Hutchings A, Black N. Do patient-reported outcomes offer a more sensitive method for comparing the outcomes of consultants than mortality? A multilevel analysis of routine data. *BMJ Qual Saf.* 2015; 24: 195-202.
70. Varaganum M, Hutchings A, Neuburger J, Black N. Impact on hospital performance of introducing routine patient reported outcome measures in surgery. *J Health Serv Res Policy.* 2014; 19: 77-84.
71. Greenhalgh J, Pawson R, Wright J, Black N, Valderas JM, Meads D, et al. Functionality and feedback: a protocol for a realist synthesis of the collation, interpretation and utilisation of PROMs data to improve patient care. *BMJ Open.* 2014; 4: e005601.
72. EMEA. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. EMEA/CHMP/EWP/139391/204. London, 2004.
73. US Department of Health and Human Services. Draft guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. February 2006. Disponible en: <http://www.fda.gov/ohrms/dockets/98fr/o6d-0044-gd10001.pdf> [Acceso: 14 enero,2011].
74. US Department of Health and Human Services. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. December 2009. Disponible en: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf> [Acceso: 14 enero,2011].
75. Sloan JA, Halyard MY, Frost MH, Dueck AC, Teschendorf B, Rothman ML. The Mayo Clinic manuscript series relative to the discussion, dissemination, and operationalization of the Food and Drug Administration guidance on patient-reported outcomes. *Value Health.* 2007; 10 Suppl 2: S59-63.

76. Food and Drug Administration Safety and Innovation Act, S 3187, 112th Cong (2012). Disponible en: <http://www.gpo.gov/fdsys/pkg/PLAW-112publ144/pdf/PLAW-112publ144.pdf> [Acceso: 17 junio, 2014].
77. Coons SJ, Kothari S, Monz BU, Burke LB. The patient-reported outcome (PRO) consortium: filling measurement gaps for PRO end points to support labeling claims. *Clin Pharmacol Ther.* 2011; 90: 743-8.
78. US Food and Drug Administration. Guidance for Industry and FDA staff: qualification process for drug development tools. January 2014. Disponible en: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf> [Acceso: 15 marzo, 2015].
79. Basch E, Geoghegan C, Coons SJ, Gnanasakthy A, Slagle AF, Papadopoulos EJ, et al. Patient-reported outcomes in cancer drug development and US regulatory review: Perspectives from industry, the Food and Drug Administration, and the patient. *JAMA Oncol.* 2015; 1: 375-9.
80. Gnanasakthy A, Mordin M, Clark M, DeMuro C, Fehnel S, Copley-Merriman C. A review of patient-reported outcome labels in the United States: 2006 to 2010. *Value Health.* 2012; 15: 437-42.
81. Clark MJ, Harris N, Griebisch I, Kaschinski D, Copley-Merriman C. Patient-reported outcome labeling claims and measurement approach for metastatic castration-resistant prostate cancer treatments in the United States and European Union. *Health Qual Life Outcomes.* 2014; 12: 104.
82. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol.* 2010; 63: 1179-94.
83. DeWalt DA, Rothrock N, Yount S, Stone AA. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care.* 2007; 45: S12-S21.
84. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007; 45: S22-31.
85. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof.* 2005; 28: 212-32.
86. Alonso J, Bartlett SJ, Rose M, Aaronson NK, Chaplin JE, Efficace F, et al. The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. *Health Qual Life Outcomes.* 2013; 11: 210.
87. Viswanathan M, Ammerman A, Eng E, Garlehner G, Lohr KN, Griffith D, et al. Community-based participatory research: Assessing the evidence. Evidence report/technology assessment (Summary), N° 99. Prepared by RTI-University of North Carolina Evidence-based Practice Center under Contract N° 290-02-0016). Agency for Healthcare Research and Quality. AHRQ Publication 04-E022-2.
88. Cashman SB, Adeky S, Allen AJ, III, Corburn J, Israel BA, Montano J, et al. The power and the promise: working with communities to analyze data, interpret findings, and get to outcomes. *Am J Public Health.* 2008; 98: 1407-17.
89. Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Floyd N. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. *J Clin Epidemiol.* 2010; 63: 491-501.
90. National Institutes of Health. From public advocacy to research priorities: NHLBI listens and responds. National Heart, Lung, and Blood Institute. Disponible en: <http://www.nhlbi.nih.gov.ezproxy.welch.jhmi.edu/public/nhlbi-listens.pdf> [Acceso: 23 abril, 2014].
91. Staley K. Exploring impact: Public involvement in NHS, public health and social care research. INVOLVE.

92. Concannon TW, Meissner P, Grunbaum JA, McElwee N, Guise JM, Santa J, et al. A new taxonomy for stakeholder engagement in patient-centered outcomes research. *J Gen Intern Med*. 2012; 27: 985-91.
93. Forsythe LP, Szydowski V, Murad MH, Wang Z, Elraiyah TA, Hickam DH. A systematic review of approaches for engaging patients in research on rare diseases. *J Gen Intern Med*. 2014; 3: 788-800.
94. Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. *JAMA*. 2014; 312: 1513-4.
95. Shippee ND, Domecq-Garces JP, Prutsky-Lopez GJ, Wang Z, Elraiyah TA, Nabhan Met al. Patient and service user engagement in research: A systematic review and synthesized framework. *Health Expect*. 2015; 18: 1151-66.
96. Curtis P, Slaughter-Mason S, Thielke A, Gordon C, Pettinari C, Ryan K. PCORI Expert Interviews Project: Final report. Portland,OR: Center for Evidence-based Policy, Oregon Health & Science University. Patient-Centered Outcomes Research Institute.
97. Deverka PA, Lavalley DC, Desai PJ, Esmail LC, Ramsey SD, Veenstra DL, et al. Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *J Comp Eff Res*. 2012; 1: 181-94.
98. Mullins CD, Abdulhalim AM, Lavalley DC. Continuous patient engagement in comparative effectiveness research. *JAMA*. 2012; 307: 1587-8.
99. Workman T, Maurer M, Carman K. Unresolved tensions in consumer engagement in CER: a US research perspective. *J Comp Eff Res*. 2013; 2: 127-34.
100. Concannon TW, Fuster M, Saunders T, Patel K, Wong JB, Leslie LK, et al. A systematic review of stakeholder engagement in comparative effectiveness and patient-centered outcomes research. *J Gen Intern Med*. 2014; 29: 1692-701.
101. Witt CM, Chesney M, Gliklich R, Green L, Lewith G, Luce B, et al. Building a strategic framework for comparative effectiveness research in complementary and integrative medicine. *Evid Based Complement Alternat Med*. 2012; 2012: 531096.
102. Hoffman A, Montgomery R, Aubry W, Tunis SR. How best to engage patients, doctors, and other stakeholders in designing comparative effectiveness studies. *Health Affairs*. 2010; 29: 1823-41.
103. Clancy CM, Eisenberg JM. Outcomes research: measuring the end results of health care. *Science*. 1998; 282: 245-6.
104. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013.
105. Brundage M, Blazeby J, Revicki D, Bass B, de VH, Duffy H, et al. Patient-reported outcomes in randomized clinical trials: development of ISOQOL reporting standards. *Qual Life Res*. 2013; 22: 1161-75.
106. Frank L, Forsythe L, Ellis L, Schrandt S, Sheridan S, Gerson J, et al. Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. *Qual Life Res*. 2015; 24: 1033-41.
107. Concannon TW. Can patient centered outcomes research improve healthcare? *BMJ*. 2015; 351: h3859.
108. Xian Y, Wu J, O'Brien EC, Fonarow GC, Olson DM, Schwamm LH, et al. Real world effectiveness of warfarin among ischemic stroke patients with atrial fibrillation: observational analysis from Patient-Centered Research into Outcomes Stroke Patients Prefer and Effectiveness Research (PROSPER) study. *BMJ*. 2015; 351: h3786.
109. Garcia-Duran M, Castellvi P, Ferrer M, Alonso J, on behalf of the BiblioPRO Scientific Committee. BiblioPRO: Online Library of Patient-Reported Outcomes in Spanish. *PRO Newsletter*. 2012; 48: 7-10.
110. Valderas JM, Mendivil J, Parada A, Losada-Yanez M, Alonso J. Development of a geographic filter for PubMed to identify studies performed in Spain. *Rev Esp Cardiol*. 2006; 59: 1244-51.

111. Cieza A, Stucki G. The International Classification of Functioning Disability and Health: its development process and content validity. *Eur J Phys Rehabil Med*. 2008; 44: 303-13.
112. FDA. U.S. Food and Drug Administration. Clinical Outcome Assessment (COA): Glossary of Terms. Disponible en: <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/ucm370262.htm> [Acceso: 15 enero, 2015].
113. Garin O, Herdman M, Vilagut G, Ferrer M, Ribera A, Rajmil L, et al. Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures. *Heart Fail Rev*. 2014; 19: 359-67.
114. Schmidt S, Garin O, Pardo Y, Valderas JM, Alonso J, Rebollo P, et al. Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments. *Qual Life Res*. 2014; 23: 2169-81.
115. Schmidt S, Ferrer M, Gonzalez M, Gonzalez N, Valderas JM, Alonso J, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. *J Shoulder Elbow Surg*. 2014; 23: 434-44.
116. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res*. 2002; 11: 193-205.
117. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Braz J Phys Ther*. 2016; 20: 105-13.
118. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010; 19: 539-49.
119. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol*. 2010; 10: 22-9.
120. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012; 21: 651-7.
121. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res*. 2012; 21: 659-70.
122. Terwee CB, Gerding MN, Dekker FW, Prummel MF, van der Pol JP, Wiersinga WM. Test-retest reliability of the GO-QOL: a disease-specific quality of life questionnaire for patients with Graves' ophthalmopathy. *J Clin Epidemiol*. 1999; 52: 875-84.
123. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009; 18: 1115-23.
124. Porter M. A strategy for health care reform—toward a value-based system. *N Engl J Med* 2009; 361: 109-12.
125. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA*. 2013; 309: 814-22.



ESINFU0048 Febrero 2017