Universitat Autònoma de Barcelona

EVALUACIÓN DE LA CALIDAD DE LAS MEDIDAS DE RESULTADOS DESDE LA PRESPECTIVA

SONIA LORENTE SÁNCHEZ

EVALUACIÓN DE LA CALIDAD DE LAS MEDIDAS DE RESULTADOS DESDE LA PERSPECTIVA DEL PACIENTE

SONIA LORENTE SÁNCHEZ

Tesis doctoral

Directores

Josep-Maria Losilla Vidal
Jaume Vives Brosa

Tutor
Josep-Maria Losilla Vidal

Universitat Autònoma de Barcelona

Estudios de Doctorado en Psicología de la Salud y del Deporte
Departamento de Psicobiología y de Metodología de les Ciencias de la Salud
Facultad de Psicología - Universidad Autónoma de Barcelona

2020

A mi familia,

especialmente a mi padre

ÍNDICE

1. PREÁMBULO

Las medidas aplicadas en el ámbito de la salud son, generalmente, medidas de resultados desde la perspectiva del paciente (Patient-Reported Outcomes Measures, PROMs). La aplicación de las PROMs nos permite a los profesionales de la salud evaluar el impacto de nuestras intervenciones, incrementando así nuestra calidad asistencial. En este contexto, mis compañeras de profesión y mis actuales directores, realizamos una intervención para mejorar el confort del lactante con bronquiolitis. Evaluar los resultados de esta intervención fue, sin embargo, un proceso complicado al no hallar PROMs apropiadas para medir el confort del paciente hospitalario.

Con motivo de este estudio, realizado durante el transcurso del máster de investigación, inicio la presente tesis con el objetivo principal de describir y evaluar las medidas de confort del paciente hospitalario y, como segundo objetivo, diseñar una medida de confort del lactante. Cumpliendo con el objetivo principal, realizamos una revisión psicométrica, conformando el primer artículo de este compendio. Durante la síntesis de resultados, sin embargo, detectamos que gran parte de las medidas evaluadas presentaban carencias metodológicas o bien relacionadas con su aplicabilidad. Nuestros resultados mostraron que no existían criterios unificados para evaluar la calidad de las medidas de confort del paciente hospitalario. Frente a esta situación, decidimos reformular nuestro segundo objetivo, orientándolo hacia la realización de una meta-revisión que nos aportara información relacionada con la evaluación de la calidad y de la aplicabilidad de las PROMs en el ámbito de la salud, enmarcando dicha revisión en las PROMs que evalúan la calidad de vida, constructo multidimensional relacionado con el confort del paciente. Esta meta-revisión conforma el segundo artículo de este compendio, aportando una base metodológica adecuada para evaluar y seleccionar las PROMs de mayor calidad y más apropiadas para evaluar el confort, el bienestar y la calidad de vida.

2. INTRODUCCIÓN

2.1. MEDIDAS DE RESULTADOS DESDE LA PERSPECTIVA DEL PACIENTE (PROMS)

Los resultados desde la perspectiva del paciente (Patient-Reported Outcomes, PRO) se definen como "informes del estado de salud de un paciente que provienen directamente del mismo, sin interpretación de la respuesta por parte del profesional médico o de otra persona" (Department of Health and Human Services, 2006). Este tipo de medidas se realiza mediante la aplicación de cuestionarios auto informados denominados Patient-Reported Outcomes Measures, en adelante PROMs (Weldring & Smith, 2013).

2.1.1. NIVELES DE APLICACIONES DE LAS PROMS

El hecho de que las PROMs evalúen diferentes aspectos relacionados con la salud desde la perspectiva del paciente hace posible la identificación de problemas que no podrían evaluarse con otras medidas, como la sintomatología relacionada con los efectos adversos de un tratamiento. Históricamente las PROMs se utilizaban en investigación, especialmente en ensayos clínicos que evaluaban efectos de determinados fármacos, pero en la actualidad sus aplicaciones abarcan diferentes contextos o niveles (Department of Health and Human Services, 2006; Øvretveit et al., 2017; Weldring & Smith, 2013).

De acuerdo con la revisión realizada por la Comisión Australiana de Seguridad y Calidad Sanitaria existen tres niveles de aplicaciones de las PROMs (Williams et al., 2016). Nivel micro, también denominado nivel individual o de interacción profesional-paciente, en el que las PROMs facilitarían la evaluación de resultados, la toma de decisiones diagnósticas, la elección y planificación del tratamiento y el seguimiento del paciente. Nivel meso, también denominado nivel de sistemas de salud, en el que las PROMs se aplican para comparar los resultados de tratamientos y las variaciones entre diferentes proveedores de servicios, como la diferencia de costes económicos en función de la localización geográfica de los centros asistenciales. Nivel macro, o bien organizacional, en el que las PROMs facilitan que las instituciones sanitarias puedan recoger datos considerados indicadores de calidad asistencial, como la prevalencia de infecciones o los reingresos no programados. La aplicación de las PROMs a nivel organizacional facilitaría la toma de decisiones económicas y la mejora de la calidad asistencial (Williams et al., 2016).

2.1.2. CLASIFICACIÓN DE LAS PROMS

Existen diferentes clasificaciones de las PROMs, la mayoría relacionadas con el constructo de la calidad de vida. La guía Food and Drug Administration (FDA) propone una primera clasificación de todas las PROMs, aunque limitada casi exclusivamente a su aplicación en ensayos clínicos (Department of Health and Human Services, 2006). Tal vez la clasificación más integradora, y adoptada como referente para este compendio, sería la propuesta publicada por Valderas y Alonso (2008), que clasifica las PROMs según el constructo evaluado, la población a la que se dirigen y el modelo de medida.

Así, en cuanto al constructo se refiere, las PROMs evaluarían síntomas, estado funcional, percepción de salud y calidad de vida en relación con la salud. Esta clasificación, basada en el modelo teórico de Wilson y Clearly (Wilson & Cleary, 1995), se integra en el marco que sustenta la Clasificación Internacional de Funcionamiento, Discapacidad y Salud (ICF) propuesta por la Organización Mundial de la Salud (World Health Organization, 2016). Otros constructos o conceptos que evaluarían las PROMs, al margen del modelo de Wilson y Clearly y el marco ICF, serían la satisfacción con el cuidado o la resiliencia (Valderas & Alonso, 2008).

Según la población a la que se dirigen hallamos medidas destinadas a la población general, o bien a población con características específicas relacionadas con la edad, el género, la cultura y/o la enfermedad. La mayoría de las medidas se diseñan para adultos, pero cada vez es más frecuente el desarrollo de medidas para niños o adolescentes. Algunas medidas también se centran en el género cuando la patología está vinculada al sexo, como en el caso del cáncer de próstata o el cáncer de ovario. Las PROMs de uso más frecuente, sin embargo, son aquellas aplicadas en población con patologías específicas, destinadas a capturar la carga de la enfermedad o el impacto de un tratamiento (Valderas & Alonso, 2008).

Según el modelo de medida, estos autores clasifican las PROMs de acuerdo con su métrica, dimensionalidad y adaptabilidad. El concepto de métrica se refiere al método utilizado para asignar valores numéricos a las respuestas facilitadas por los pacientes y a la elaboración de las puntuaciones, identificando así medidas que utilizan escalas sumativas y medidas que utilizan escalas comparativas, de valoración de preferencias (Viladrich & Doval, 2007). Las escalas sumativas serían, por ejemplo, las utilizadas generalmente en el ámbito psicométrico y clinimétrico, mientras que las comparativas serían más habituales en el contexto econométrico.

Cabe destacar, sin embargo, que la distinción entre medidas psicométricas y clinimétricas se considera prácticamente una cuestión de terminología y, por tanto, innecesaria, como sugieren Streiner (2003) y Viladrich y Doval (2007). Adoptando la misma posición nos referiremos, pues, a medidas psicométricas cuya tradición psicológica, como detallaremos más adelante, proporciona criterios de calidad de la medida en el ámbito de la salud y la medicina (Viladrich & Doval, 2007). La aplicación de estas medidas permitiría evaluar características psicológicas y/o aspectos subjetivos relacionados con la enfermedad o el impacto de una intervención sanitaria.

Además, hallamos las PROMs econométricas, desarrolladas a partir de métodos de la economía basados en la teoría de la decisión. Su aplicación permite obtener un valor basado en las preferencias del paciente respecto a un estado de salud combinando expectativa de vida y calidad de vida, denominado Años de Vida por Calidad, AVAC (en inglés, Quality-Adjusted Life-Year, QALY). El AVAC permitiría comparar resultados de las intervenciones sanitarias en términos de incremento de costes por cada año de salud plena conseguido (análisis de coste-utilidad) (Brazier et al., 1999, 2017; Valderas & Alonso, 2008).

Dentro del marco del modelo de medida, y con respecto a la dimensionalidad, Valderas y Alonso (2008) distinguen entre índices y perfiles. Cuando la información generada puede resumirse en un único valor se denominarían índices o medidas unidimensionales, y cuando se generan diferentes valores se denominarían perfiles o medidas multidimensionales. Esta clasificación no sería mutuamente excluyente, ya que existen medidas que pueden generar tanto índices como perfiles, como podrían ser las PROMs econométricas. Por último, la adaptabilidad se referiría a la capacidad de la medida de adecuarse a las especificidades del paciente, hallando PROMs completamente estandarizadas, parcialmente individualizadas o completamente individualizadas (Valderas & Alonso, 2008).

2.1.3. IMPLEMENTACIÓN DE LAS PROMS.

Dada la aplicabilidad y relevancia de las PROMs en el ámbito de la investigación, la práctica clínica y la toma de decisiones político-económicas, diferentes iniciativas promueven su implementación. Entre las más destacadas hallamos la guía User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice, publicada por la Sociedad Internacional de Estudios de Calidad de Vida (ISOQL) (Aaronson et al., 2011), el programa Patient-Reported Outcomes Measurement Information System (PROMIS) (Alonso et al., 2013), el programa NHS-PROM, impulsado por el Sistema Nacional de Salud de Inglaterra (NHS, 2017), las recomendaciones basadas en la evidencia publicadas por la Organización para la Cooperación y

el Desarrollo Económicos (OECD, 2017), o la reciente propuesta para la selección e implementación de las PROMs de van der Wees y colaboradores (van der Wees et al., 2019).

A pesar de las numerosas iniciativas para difundir el uso de las PROMs, las organizaciones sanitarias experimentan dificultades en su implementación (Baker, 2014). La revisión realizada por Foster y colaboradores en 2018 identifica facilitadores y barreras asociados al proceso de implementación (Foster et al., 2018). Como estrategias facilitadoras estarían favorecer la implicación de los profesionales en el proceso de selección de las PROMs más adecuadas, válidas, fiables y factibles; favorecer la implicación de los pacientes en el diseño de las medidas y garantizarles apoyo para completar los cuestionarios; invertir en nuevas tecnologías que faciliten la administración de las PROMs, disminuyendo así la carga tanto para profesional como para el paciente; adaptar la medida tanto al contexto como a las especificidades de los pacientes, y compatibilizar el proceso de implementación con los flujos de trabajo de los profesionales. Como barreras se identifican los costes económicos asociados, por ejemplo, a la obtención de la licencia de uso; la falta de apoyo organizacional (e.g. horarios, estructura del consultorio, etc.); la falta de información clara sobre los posibles beneficios de la aplicación de estas medidas, y la falta de capacitación percibida por los profesionales para la correcta administración de la medida e interpretación de los datos. Foster y colaboradores señalan que las fases iniciales de preparación y planificación, especialmente con relación a la selección de la medida más adecuada y la capacitación de los profesionales, serían fundamentales para la implementación de las PROMs (Foster et al., 2018).

2.2. EVALUACIÓN DE LA CALIDAD DE LAS PROMS

La calidad de la medida en el ámbito de la salud se ha evaluado tradicionalmente desde tres aproximaciones (Viladrich et al., 2019; Viladrich & Doval, 2007). En primer lugar, hallamos la tradición psicométrica de la psicología, cuya trayectoria se inicia en 1954, cuando la American Psychological Association (APA) elabora y publica The Technical Recommendations for Psychological Tests and Diagnostic Techniques (APA, 1954). En 1966, La American Psychological Association (APA), en colaboración con la American Educational Research Association (AERA) y el National Council on Measurement in Education (NCME) publican por primera vez The Standards for Educational and Psychological Testing, en adelante Estándares psicométricos, revisados posteriormente en 1985, 1999 y 2014. Los Estándares psicométricos proponen normas para el diseño, el desarrollo, la evaluación y la aplicación de los test y medidas, así como criterios para evaluar la validez de las interpretaciones de las puntuaciones obtenidas (APA et al., 2014).

En segundo lugar, hallamos la tradición desarrollada en el ámbito de la medicina, impulsada por The Scientific Advisory Committee of the Medical Outcomes Trust (SACMOT), la cual establece una serie de criterios mínimos para la evaluación de la calidad de la medida en el ámbito de la salud bajo la denominación Attributes and Criteria to assess Health Status and Quality of Life Instruments (Aaronson et al., 2002; Lohr et al., 1996). En el marco de esta aproximación hallamos los estándares The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink et al., 2018; Mokkink et al., 2012), diseñados especifícamente para la evaluación y selección de PROMs evaluativas, y los criterios establecidos por la Food and Drug Administration (FDA) (Department of Health and Human Services, 2006, 2009), centrados mayoritariamente a la evaluación de la calidad de PROMs aplicadas en ensayos clínicos, como se menciona anteriormente. Finalmente hallamos la tradición econométrica, impulsada por The Standing Group of Health Technology, la cual establece criterios de evaluación basados tanto en aspectos de calidad de la medida como en la obtención del índice AVAC. La versión actual (Brazier et al. 2017) es la revisión de la edición original publicada en 1999 (Brazier et al., 1999).

Cabe destacar, sin embargo, que a pesar de que existen diferentes tradiciones, los criterios de calidad de la medida relacionados con la validez de la interpretación de las puntuaciones parecen derivar de los Estándares psicométricos elaborados en 1985. Así, la influencia de la tradición psicométrica se reflejaría en los criterios SACMOT (2002) (Viladrich et al., 2019; Viladrich & Doval, 2007). Su versión actual, los Estándares COSMIN, se diseñan posteriormente con el objetivo específico de mejorar el proceso de selección de las PROMs evaluativas más apropiadas para su aplicación en el campo de la investigación y la práctica clínica. Siguiendo recomendaciones de los Estándares COSMIN, la selección de las PROMs más adecuadas se basaría en la calidad de sus propiedades de medida, es decir, su validez, fiabilidad, sensibilidad, interpretabilidad, y en su factibilidad, como se detalla en la Tabla 1 (Mokkink et al., 2018; Prinsen et al., 2018).

La evaluación de la calidad de las propiedades de la medida en el ámbito de la salud, sin embargo, no puede valorarse independientemente de su uso o función. Diferentes clasificaciones permiten evaluar la calidad de la medida según este criterio. Entre las más relevantes hallamos la de Bombardier y Tugwell, que distingue entre medidas de uso diagnóstico, pronóstico y evaluativo (Bombardier & Tugwell, 1982; Tugwell & Bombardier, 1982), o la de Guyatt y colaboradores, que identifica medidas de uso evaluativo y discriminativo (Guyatt et al., 1992). La clasificación de McDowell y colaboradores, sin embargo, parece ser la más amplia hasta el momento, ya que propone hasta cuatro usos (McDowell et al., 2004): medidas de uso diagnóstico, diseñadas para diferenciar individuos, grupos o niveles de gravedad; medidas de uso predictivo o pronóstico, que

anticipan estados de salud; medidas evaluativas, utilizadas para evaluar el efecto de un tratamiento; y medidas de uso analítico, diseñadas para generar perfiles considerados de riesgo para la salud que facilitan la promoción de políticas de salud o la mejora de servicios de salud, entre otros aspectos. Dependiendo de su uso, determinadas propiedades serán más o menos relevantes para determinar la calidad de la medida de forma que, por ejemplo, las medidas evaluativas deberán demostrar elevada sensibilidad al cambio mientras que las medidas diagnósticas deberán presentar mayor capacidad de discriminación (Prinsen et al., 2018; Rosenkoetter & Tate, 2017).

La evaluación de la calidad de la medida está vinculada, además, con su factibilidad. La factibilidad o usabilidad de la medida incluye la valoración de los costes económicos derivados de su aplicación, la carga del administrador, la carga del paciente, el tiempo de respuesta, el modo de administración, etc. Estos aspectos pueden comprometer la validez y fiabilidad de los resultados si no se valoran apropiadamente. Un ejemplo en este sentido sería la carga que representa para el paciente la administración de una medida excesivamente larga o no adaptada a sus necesidades, que podría conllevar un número alto de no respuestas en los ítems y derivar, por tanto, en cuestionarios incompletos (Beattie et al., 2014; Prinsen et al., 2018; Rosenkoetter & Tate, 2017).

La evaluación de la calidad de las propiedades de las PROMs según el uso y la valoración de su factibilidad serían aspectos complementarios en el proceso de selección de la medida más adecuada. La selección de una medida válida, fiable, sensible y factible facilitaría, asimismo, su implementación en las instituciones sanitarias (Foster et al., 2018).

2.3. SÍNTESIS

La definición, aplicación, clasificación, funcionalidad, implementación y evaluación de las PROMs se sintetizan en la tabla que sigue a continuación (Tabla 1). La evaluación de la calidad de las PROMs y de su factibilidad se refleja como un proceso que recorre todas las fases de su diseño, desde su definición hasta su implementación.

Tabla 1. Síntesis. Definición, aplicaciones, clasificación, usos, evaluación de la calidad e implementación de las PROMs

DEFINICIÓN (Department of Health and Human Services, 2006)

"Las medidas de resultados reportadas por el paciente, PROMs (Patient-Reported Outcomes Measures) son cuestionarios estandarizados que evalúan la experiencia del paciente con relación a cualquier aspecto de su estado de salud"

NIVELES DE APLICACIONES (Williams et al., 2016)

Nivel Micro	Nivel Meso	Nivel Macro
Elección de diagnóstico, plan de tratamiento, seguimiento y evaluación de resultados	Efectividad de los tratamientos con relación a costes o preferencias del paciente. Diferencias entre proveedores de servicios	Rendimiento de las organizaciones, mediante la recogida de datos de indicadores de calidad como efectos adversos, prevalencia de infecciones, reingresos no programados…

CLASIFICACIÓN DESCRIPTIVA (Valderas y Alonso, 2008)

Ejes	Categorías	Categorías específicas
Constructo	Síntomas	Síntomas relacionados con enfermedades infecciosas, neoplásicas, endocrinológicas, trastornos mentales y del comportamiento, malformaciones y/o síndromes congénitos, alteraciones musculo esqueléticas…
	Estado funcional	Con relación a tareas generales, de la vida doméstica y del autocuidado, comunicación, movilidad, relaciones personales, aprendizaje y aplicación de los conocimientos
	Percepción de salud	Percepción subjetiva que integra el estado funcional y la sintomatología
	Calidad de vida	Aspectos de la calidad de vida, definida como el estado de bienestar y de confort y/o alegría percibido
	Otros conceptos	Satisfacción con el cuidado, resiliencia, entorno.
Población	Edad	Niños; adolescentes…
	Género	Todos los géneros; masculino; femenino
	Enfermedad	Categorías específicas igual que Síntomas
	Cultura	Idioma; procedencia
Modelo de medida	Métrica	Psicométricas; econométricas; otras
	Nivel de agregación	Índices o medidas unidimensionales; perfiles o medidas multidimensionales
	Adaptabilidad	Completamente estandarizada; parcialmente individualizada; completamente individualizada

FUNCIONALIDAD (USOS) (Mc Dowell et al., 2004)

Diagnóstico	Predictivo	Evaluativo	Analítico
Detectar la enfermedad Evaluar carga de la enfermedad Detectar necesidades del paciente	Anticipar estados de salud individual / poblacional	Evaluar impacto y resultados de tratamientos e intervenciones	Evaluar factores de riesgo sociales / genéticos.¡ Políticas de salud. Servcios de salud

IMPLEMENTACIÓN (Foster et al. 2018)

Facilitadores	Barreras
Medida válida, fiable, sensible y factible. Implicación de profesionales en la selección de PROMs. Adaptación de flujos de trabajo. Apoyo a los pacientes para completar cuestionarios. Nuevas tecnologías. Adaptabilidad de la medida a las especificidades de pacientes y el contexto	Costes económicos asociados a su uso y administración (e.g. licencia). Falta de apoyo organizacional. Falta de capacitación percibida por los profesionales para el correcto uso e interpretación de resultados. Falta de información sobre los beneficios de la aplicación de las PROMs

CALIDAD DE LA MEDIDA
(Mokkink et al. 2018)
(Versión adaptada de Viladrich y col. 2019)

- Validez
- Contenido
- Estructura Interna
- Validez estructural
- Consistencia Interna
- Validez cultural/
- Invariancia
- Otras propiedades
- Fiabilidad
- Error de medida
- Validez de criterio
- Hipótesis/constructo
- Sensibilidad
- Interpretabilidad
- Factibilidad
- Carga del usuario/ administrador
- Costes económicos
- Modo administración

2.4. OBJETIVOS

El objetivo general de esta tesis es proporcionar un marco metodológico adecuado para evaluar y seleccionar las medidas de salud más apropiadas, en concreto PROMs relacionadas con el confort, el bienestar y la calidad de vida. Para ello, diseñamos dos objetivos específicos:

1. Revisión psicométrica de las medidas de confort del paciente hospitalario, evaluando la calidad de las propiedades de medida, como son la validez, la fiabilidad y la sensibilidad, así como su factibilidad.
2. Meta-revisión sobre cómo evalúan los investigadores y profesionales de la salud la calidad de las PROMs, en concreto en el ámbito de la calidad de vida.

3. ARTÍCULOS QUE CONFORMAN ESTE COMPENDIO

Este compendio se conforma de dos partes. La primera parte incluye los artículos "Instruments to assess the patient comfort during hospitalization: a psychometric review protocol" (Lorente, Vives, et al. 2017), e "Instruments to assess the patient comfort during hospitalization: a psychometric review" (Lorente, Losilla, et al. 2017), correspondientes al primer objetivo, y la segunda parte incluye los artículos "Tools to assess the measurement properties of quality of life instruments: a meta-review protocol" (Lorente et al. 2018) y "Tools to assess the measurement properties of quality of life instruments: a meta-review" (Lorente et al. 2020), correspondientes al segundo objetivo de esta tesis.

Con respecto a la revisión psicométrica, seguimos recomendaciones internacionales para fomentar la transparencia, la replicabilidad y la rigurosidad metodológica de las revisiones sistemáticas (Boaz et al., 2002; CRD, 2009), registrando el protocolo en el International Prospective Register of Systematic Reviews, Center for Reviews and Dissemination (PROSPERO CRD) y publicando posteriormente su versión extendida. Como objetivos de la revisión psicométrica nos planteamos, en primer lugar, identificar las medidas que evaluaban el confort del paciente hospitalizado y, en segundo lugar, evaluar la calidad de las propiedades de las medidas y su factibilidad. Para cumplir con estos objetivos clasificamos las medidas identificadas según su modelo teórico de desarrollo; reflectivo, formativo o mixto (Coltman et al., 2008). A continuación evaluamos la calidad de las propiedades de las medidas de confort mediante la aplicación de los estándares COnsensus-based Standards for the selection of health Measurement INstruments, COSMIN (Mokkink et al., 2012) y The Quality Criteria for Measurement Properties for health questionnaires (Terwee et al., 2007). Y, por último, evaluamos la factibilidad de la medida (costes, carga del profesional y del paciente, y aceptabilidad) según la propuesta de Van der Vleuten (1996), citado en Beattie et al. (2014).

Con respecto a la meta-revisión seguimos las mismas directrices de rigurosidad metodológica que en el trabajo anterior, registrando el protocolo en PROSPERO CRD, y publicando a continuación su versión extendida. En relación con el proceso, cabe señalar que se realizó en dos fases. En una primera fase realizamos una búsqueda preliminar de revisiones en el ámbito de la calidad de vida para explorar todos los posibles descriptores que necesitaríamos para la búsqueda.

En una segunda fase elaboramos una estrategia más exhaustiva siguiendo las recomendaciones Peer Review of Electronic Search Strategies (PRESS) (McGowan et al., 2016a, 2016b), y aplicando el filtro de búsqueda de estudios de propiedades de la medida desarrollado por la Vrije University Medical Center (Terwee et al., 2009). Esta estrategia nos permitió cumplir nuestros objetivos: a) identificar las diferentes herramientas estandarizadas utilizadas para la evaluación de la calidad de la medida en el ámbito de la salud; b) decribir y comparar el contenido de las diferentes herramientas respecto a la evaluación de la validez, fiabilidad y factibilidad de la medida; y c) describir si la evaluación de la calidad de las propiedades de la medida se realiza según su uso o función. Nuestros resultados aportan una base metodológica adecuada para evaluar y seleccionar las PROMs más apropiadas en el ámbito del confort, el bienestar y la calidad de vida.

I) Artículos de este compendio

Lorente, S., Vives, J., & Losilla, J.M. (2017). Instruments to assess the patient comfort during hospitalization: a psychometric review protocol. Journal of Advanced Nursing, 73(3), 735–741. https://doi.org/10.1111/jan.13180

Lorente, S., Losilla, J.M., & Vives, J. (2017). Instruments to assess patient comfort during hospitalization: A psychometric review. Journal of Advanced Nursing, 74(5), 1001–1015. https://doi.org/10.1111/jan.13495

Lorente S., Vives J., Viladrich C. & Losilla J.M. (2018) Tools to assess the measurement properties of quality of life instruments: A meta-review protocol. BMJ Open, 8, e022829. https://doi 10.1136/bmjopen-2018-02289

Lorente, S., Viladrich, C., Vives, J. & Losilla, J.M. (2020). Tools to assess the measurement properties of quality of life instruments: a meta-review. BMJ Open, 10, e036038. http://doi.org/10.1136/bmjopen-2019-036038

Como Addenda se incluye el artículo de confort del lactante, realizado durante el Máster.

Lorente, S., Gimeno, R., Losilla, J.M., Garzón, S., & Vives, J. (2017). Benefits of the humidified low-flow oxygen therapy in infants with mild-moderate bronchiolitis Journal of Clinical Nursing, 27(5–6), 1125–1133. https://doi.org/10.1111/jocn.14140

II) La difusión de los resultados de los diferentes trabajos derivaron en la presentación de comunicaciones y pósters en congresos nacionales e internacionales.

Lorente, S., Losilla, J.M., Vives, J. (2017). Instrumentos para evaluar el confort del paciente durante su hospitalización: una revisión sistemática. II Congreso de Psicología Sanitaria y Salud Mental. Premio a la mejor comunicación científica otorgado por la Sociedad Catalana Balear de Psicología (SCBP).

Lorente, S., Vives J., Losilla, J.M. (2017). Cuestionarios para evaluar el confort del paciente pediátrico durante su hospitalización: una revisión sistemática.
X Congreso de la Asociación Catalana de Enfermería Pediátrica. ISBN: 978-84-697-6285-1. Comunicación.

Lorente, S., Vives J., Losilla, J.M. (2017). Instrumentos de evaluación de la calidad psicométrica en el ámbito sanitario de la práctica basada en la evidencia: una meta-revisión. Simposio Meta-análisis: Avances Metodológicos, XV Congreso de Metodología de las Ciencias Sociales y de la Salud. Comunicación.

Viladrich, C., Lorente, S., Losilla, J.M., Vives, J. (2018). Tools to assess the measurement properties of quality of life instruments: a meta-review. VII European Congress of Methodology, Jena, Germany. Póster.

Viladrich, C., Lorente, S., Losilla, J.M., Vives, J. (2019). El rol del uso previsto de los cuestionarios de calidad de vida en la evaluación de las propiedades de medida: Una meta-revisión. Simposio Meta-análisis: Avances Metodológicos. XVI Congreso de Metodología de las Ciencias Sociales y de la Salud. Comunicación.

3.1. INSTRUMENTS TO ASSESS THE PATIENT COMFORT DURING HOSPITALIZATION: A PSYCHOMETRIC REVIEW PROTOCOL

INSTRUMENTS TO ASSESS THE PATIENT COMFORT DURING HOSPITALIZATION: A PSYCHOMETRIC REVIEW PROTOCOL

Sonia LORENTE, Jaume VIVES, Josep-Maria LOSILLA

Universitat Autònoma de Barcelona

Sonia LORENTE, Psychologist, PNP, MSc, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB; Paediatric nurse, General Paediatric Ward, Hospital de Terrassa, Consorci Sanitari de Terrassa; Jaume VIVES, Lecturer, PhD, and Josep-Maria LOSILLA, Lecturer, PhD, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB.

This work was supported by the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness (July 2015). No conflict of interest has been declared by the author(s). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Correspondence concerning this article should be addressed to Sonia Lorente, Psychologist, PNP, MSc, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB. Phone: +34619287300; E-mail: sonia.lorentes@e-campus.uab.cat

IMPACT STATEMENT

Measuring the patient experience is essential to improve the quality of care provided in the different health settings. Enhancing the patient comfort as a holistic experience is related to quicker discharges, fewer readmissions, increased patient satisfaction and stronger cost-benefit ratios for the institution.

ABSTRACT

Aim: The aim of this study was to analyse the psychometric properties, outcomes and utility of instruments measuring the patient comfort during hospitalization. **Background**: While there are numerous systematic reviews assessing the psychometric properties of health care instruments, none of them is devoted to evaluate the psychometric properties of instruments measuring comfort, which is considered an indicator of quality in health care and is associated to quicker discharges, increased patient satisfaction and stronger cost-benefit ratios for the institution. **Design**: Pyschometric Systematic review. **Methods**: Searches will be performed on MEDLINE, CINAHL, PsycINFO, Web of Knowledge, ProQuest Thesis&Dissertations, and grey literature, and will be focused questionnaires measuring the patient comfort as a holistic experience in any health care setting. The assessment will take into account the theoretical model upon which the instruments are built, will evaluate the psychometric properties of each study according to the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN), and will include the assessment of the quality of instruments outcomes and their cost-efficiency, acceptability and educational impact. The review protocol was registered in PROSPERO, CRD42016036290, and was supported by the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness (July 2015)**. Discussion**: The results of our psychometric review will categorise the instruments measuring the patient comfort according to their psychometric properties, methodological quality, outcomes and utility, in order to improve the quality of health care provided and the institution´s benefits. **Keywords**: comfort assessment, cost-efficiency, hospitalization, nursing care, patient comfort.

SUMMARY STATMENT: WHY THIS REVIEW IS NEEDED?

- There are no systematic reviews nor meta-analyses of health instruments measuring the patient comfort during the hospitalization; comfort is considered a direct indicator of quality in any health care.
- Assessing the psychometric properties, the outcomes and the utility of the instruments measuring the patient comfort is essential to improve the quality of care.

INTRODUCTION

Background

Measuring the patient experience is essential to improve the quality of care provided in the different health Settings (World Health Organization 2000, 2004, 2013). To date, different healthcare instruments have been developed aiming to assess the patient experience to enhance the quality of care. As reliable and valid instruments are essential in healthcare research and clinical practice, numerous systematic reviews have been developed to assess the methodological quality of these instruments. These reviews include instruments measuring health-related quality of life, pain or fatigue, to name a few (Mokkink et al. 2009). Nevertheless, there are no systematic reviews nor meta-analyses of Health instruments measuring the patient comfort during the hospitalization, being comfort a direct indicator of quality in health care (NQMC 2002; Kolcaba 2013). As enhancing the patient comfort is associated to quicker discharges, fewer readmissions, increased patient satisfaction and stronger cost-benefit ratios for the institution (Kolcaba 2001, 2013), is essential for clinicians, researchers and institutions to know how the instruments assessing comfort perform. With this purpose, their psychometric properties, their utility and their outcomes should be assessed; as the health instruments need to have high utility as well as they must be valid and reliable. Otherwise, there is a serious risk of biased results that may lead to wrong results (Terwee et al. 2007, Streiner & Norman 2008, Keszei et al. 2010).

Conceptual framework

Patient comfort is considered an individualized and holistic experience, source of patient satisfaction and well-being. The concept of comfort was framed within the Theory of Comfort, by Kolcaba and was defined theoretically as ‘the state of having met basic human needs for ease, relief and transcendence’ in four contexts (physical, psychospiritual, sociocultural and environmental) (Kolcaba & Kolcaba 1991, Kolcaba 1991, 1992, 1995, 2013). In this context, the General Comfort Questionnaire (GCQ) is likely the first instrument specifically developed to measure the patient comfort as a holistic experience, in the four contexts and registered as a multidisciplinary outcome indicator of quality in health care in the National Quality Measures Clearinghouse (Kolcaba 1992, 2013; NQMC 2002). Since then, a large range of instruments to assess the patient comfort have been developed, adapted or validated.

Aims

Due to the lack of previous systematic reviews assessing the comfort instruments performance, this study aims to examine the psychometric properties, outcomes and utility of each questionnaire measuring the patient comfort. Specific objectives:

1. Identify the health instruments measuring the patient comfort as a holistic experience during hospitalization.
2. Examine the psychometric properties of each instrument assessing their reliability and validity.
3. Examine the outcomes quality including the reproducibility, responsiveness, floor-ceiling effects and interpretability.
4. Examine the utility of each instrument assessing their cost-efficiency, acceptability and educational impact in different healthcare settings.
5. Classify the different instruments according to their psychometric properties, outcomes and utility.

METHODOLOGY

The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) will be followed (Mokkink *et al*. 2010, 2012, Terwee *et al*. 2012). This study has been registered in PROSPERO, an international database of prospectively registered systematic reviews in health and social care (protocol registration number CRD42016036290), available at: http://www.crd.york.ac.uk/ PROSPERO/display_record.asp?ID=CRD42016036290. This psychometric review will be reported using relevant elements of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Guidelines (PRISMA) (Liberati *et al*. 2009, Moher *et al*. 2009).

Search strategy

We aim to identify published instruments measuring the patient comfort during the hospitalization. We will define different combinations of keywords (using Mesh and other thesauruses, where available), in relation to the concept (e.g., comfort, theory), the setting (hospitalization or admission) and the instruments (e.g. questionnaires, scales). The following databases will be included: Medical Literature Analysis and Retrieval System (MEDLINE), by ProQuest, Cumulative Index to Nursing and Allied Health Literature (CINALH), by EBSCOhost, Psychological Information (PsycINFO), by APA PsycNET, Thesis & Dissertations, by ProQuest, and ISI Web of Knowledge (WoS, Web of Science CORE), by Thomson Reuters. To include grey literature, we will also search records in Google, and will review up to 400 links. The search will be limited by population (humans), by time (1990 to

2015) and by language (English). In addition, search alerts in CINALH & PsycINFO will be set. The Table 1 shows an example of this search strategy.

Inclusion criteria

Time frame

From 1990 **(**since the first instrument measuring the patient comfort seems to have been developed in 1992, framed within the Theory of Comfort, by Kolcaba) to 2015.

Study type

Studies developing or validating questionnaires and/or scales measuring the holistic patient comfort during the hospitalization. Protocols and guidelines, as well as those studies exclusively qualitative designed, will be excluded.

Age group

We will include the whole range of ages (newborns, toddlers, children, teenagers, young adults, middle-age adults and elderly people).

Context

We will evaluate instruments measuring the comfort in different healthcare settings where the patient is hospitalized due to acute illnesses (e.g. pneumonia, urinary tract infection), chronic pathologies (e.g. psychiatric illnesses, dementia), surgical interventions and labour issues. Therefore, the following settings will be included: general paediatric wards, general adult wards, the delivery room and the maternity ward, the operating room, the Intensive Care Unit (ICU), the Paediatric Intensive Care Unit (PICU), the Neonatal Intensive Care Unit (NICU), the postoperative area, the psychiatric ward and the hospice.

Instruments

We will include instruments developed with scales and/or subscales with close-ended answer items as Likert, Visual Analogic Scales (VAS) and/or Numerical Rating Scales (NRS) specifically designed for measuring the patient comfort as a holistic experience during the hospitalization. Instruments measuring the comfort related to screening or diagnostic tests will be excluded (e.g. colonoscopy, bronchoscopy, angiography, mammography, injections, biopsies, cystoscopy, cytology, fertility treatments), as well as those instruments assessing the comfort exclusively by physiological parameters (i.e. heart rate and/or blood pressure) or measuring the effect of a specific intervention (e.g. warm blanket vs. classical blanket or midazolam vs. fentanyl).

Data extraction

A reviewer will apply the inclusion criteria to all titles and abstracts. If no decision can be taken based solely on title and abstract alone, the full paper will be retrieved. Authors of eligible studies will be contacted to provide missing or additional data if necessary. Full-text inclusion criteria will be checked independently by two review authors. Discrepancies will be resolved through discussion (with a third author where necessary).

A pre-piloted form will be used to extract data from the included studies in order to assess the study quality and to synthesize the evidence. Extracted information of each selected instrument will include: general information (author, year, country of origin and papers); instrument detail (outcome measures, purpose/use, number of items, response categories, scale design, type of patients); utility characteristics (theoretical/conceptual framework, validity tests conducted and results, reliability tests conducted and results, response rate, sample size, setting, respondents population and demographics, ease and usefulness of interpretation, cost- efficiency, level of expertise required for scoring, acceptability, time required to completion, mode of administration, acceptability by managers and users, educational impact). Two review authors will extract data independently, and discrepancies will be identified and resolved through discussion (with a third author where necessary).

Synthesis

We will initially categorise the instruments according to the theoretical model used to design and validate the instrument: reflective vs. formative. To decide whether the model is reflective or formative we will take into account the theoretical and practical considerations detailed in the Table 2 (Coltman et al., 2008). Secondly, we will apply the COSMIN checklist in four steps (see Figure 1), and we will use the four-point scoring system (excellent, good, fair or poor) (Mokkink et al., 2010, 2012a; Terwee et al., 2011; Terwee et al., 2012). We will also apply the Quality Criteria for Measurement Properties for assessing the design, the methods and the outcomes of each instrument. These criteria consist of assessing the content validity, the internal consistency, the criterion validity, the construct validity, the reproducibility, the responsiveness, the floor and ceiling effects and the interpretability (Terwee et al., 2007).

Finally, we will assess the cost-efficiency, the acceptability and the educational impact of each instrument to discuss their utility in the real practice world, according to Van der Vleuten´s Utility Index Matrix (Van der Vleuten, 1996, cited in Beattie, Lauder, Atherton, & Murphy 2015). The extracted information using the prepilot form and the results of the whole process will be showed in tables and they will be synthesized in a narrative way.

Validity and reliability

To minimize the risk of bias during the methodological analysis of each instrument we will apply the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink et al. 2010, 2012; Terwee et al. 2012). With the same purpose, and to improve the quality of reporting of results of evaluation of health care instruments, we will apply the Quality Criteria for Measurement Properties (Terwee et al. 2007).

DISCUSSION

This psychometric review will contribute to categorise the instruments measuring the patient comfort according to their psychometric properties, their outcomes and their clinical and research utility. Concerning our results, different practical issues should be taken into account. First, assessing the psychometric properties of these instruments, as validity and reliability, would be helpful either for those researchers interested in improving the quality of these healthcare instruments, or for those clinicians interested in measuring the patient comfort during the hospitalization. Second, assessing the quality of outcomes, as reproducibility, responsiveness, floor-ceiling effects and interpretability, would be helpful for those researchers and clinicians interested in identifying and applying the best quality questionnaires measuring comfort. Third, assessing the utility of each instrument, as the cost-efficiency, the acceptability and the educational impact, would be helpful for clinicians and institutions to choose the most efficient instrument according the healthcare setting, patient and pathology. Furthermore, choosing the most suitable and efficient questionnaire may minimize the risk of getting incomplete questionnaires because of the large amount of time needed for their administering or to the important cognitive effort required from some patients with certain circumstances. Finally, we will discuss the practical and clinical applications related to the cost-efficiency as the sample needed to achieve the adequate reliability, the professional expertise required for administering the questionnaires and the institution cost-benefits resulting from the research in this field.

Limitations

The present review will only include studies published in English.

Conflict of interest

No conflict of interest has been declared by the author(s).

Funding

This work was supported by the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness (July 2015). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Ethical considerations

Non-specific ethical issues

Author contributions

All authors have agreed on the final version and meet at least one of the following criteria [recommended by the ICMJE (http://www.icmje.org/recommendations/)]:

- substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content

REFERENCES

Beattie, M., Lauder W., Atherton I. & Murphy D.J. (2015). Instruments to measure patient experience of health care quality in hospitals: a systematic review. *Systematic Reviews*, 3(1): 1-4.

Coltman, T., Deviney T. M., Midgley D.F. & Venaik S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, 61(12): 1250–1262.

Keszei, A., Novak, M. & Streiner, D.L. (2010). Introduction to health measurement scales. *Journal of Psychosomatic Research*, 68(4): 319–323.

Kolcaba K. (1991) A taxonomic structure for the concept comfort. *The Journal of Nursing Scholarship 23*(4), 237–240.

Kolcaba, K. (1992). Holistic comfort: operationalizing the construct as a nurse-sensitive outcome. *Advances in Nursing science*, 15(1): 1–10.

Kolcaba, K.(1995). The art of comfort care. *Journal of Nursing Scholarship*, 27(4): 287–289.

Kolcaba, K. (2001). Evolution of the mid range theory of comfort for outcomes research. *Nursing Outlook*, 49(2): 86–92.

Kolcaba, K. (2013) " Comfort". In *Middle Range Theories. Application to Nursing Research*, 3rd edn (Peterson, S.J. & Bredow, T.S, eds), Lippincot Williams &Wilkins, Philadelphia, pp. 193-209.

Kolcaba, K. & Kolcaba, R., (1991). An analysis of the concept of comfort. *Journal of Advanced Nursing*, 16(11): 1301–1310.

Liberati, A., Altman D.G., Tetzlaff J., Mulrow c., Ioannidis J.P., clarke M., Devereaux P.J., Kleijnen J. & Moher D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions : Explanation and Elaboration. *Annals of Internal Medicine*, 151(4): W65–W94.

Moher, D., Liberati A., Tetzlaff J. & Altman D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, 151(4): 264–269.

Mokkink L.B., Terwee C.B., Stratford P.W., Alonso J., PatrickD.L., Riphagen I., Knol D.L., Bouter L.M. & de Vet H.C.W. (2009) Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research 18*(3), 313–333.

Mokkink L.B., Terwee C.B., Knol D.L., Stratford P.W., Alonso J., Patrick D.L., Bouter L.M. & de Vet H.C.W. (2010) The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *Medical Research Methodology 10*(1), 1–8.

Mokkink L.B., Terwee C.B., Patrick D.L., Alonso J., Stratford, P.W., Knol D.L., Bouter L.M. & de Vet H.C.W. (2012) The COSMIN checklist manual. Retrieved from http://www.cosmin.nl on 28 July 2016.

National Quality Measures Clearinghouse, NQMC. (2002). In AHRQ, Agency for Healthcare Research and Quality, U.S. Department of Health & Human Services. Retrieved from http:// www.qualitymeasures.ahrq.gov/search/search.aspx?term=comfort on August 2015.

Streiner D.L. & Norman G.R. (2008) Health Measurement Scales: *A Practical Guide to their Development and Use*, 4th edn. OUP, Oxford.

Terwee C.B., Bot S.D.M., de Boer M.R., van der Windt, D.A.W.M., Knol D.L., Dekker J., Bouter L.M. & de Vet H.C.W. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology 60*(1), 34–42.

Terwee C.B., Mokkink L.B. & Patrick D.L. (2011) COSMIN checklist with 4-point scale. Retrieved from http:// www.cosmin.nl on 28 July 2016.

Terwee C.B., Mokkink L.B., Knol D.L., Ostelo R.W.J.G., Bouter, L.M. & de Vet H.C.W. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research 21*(4), 651–657.

World Health Organization (2000) Strategies for assisting health workers to modify and improve skills: developing quality health care-a process of change. Issues in Health Services. Retrieved from http://www.who.int/en/ on 28 July 2016.

World Health Organization (2004) Standards for Health Promotion in Hospitals. Retrieved from http://www.who.int/en/ on 28 July 2016.

World Health Organization (2013) Health 2020: A European Policy Framework and Strategy for the 21st Century. Retrieved from http://www.who.int/en/ on 28 July 2016.

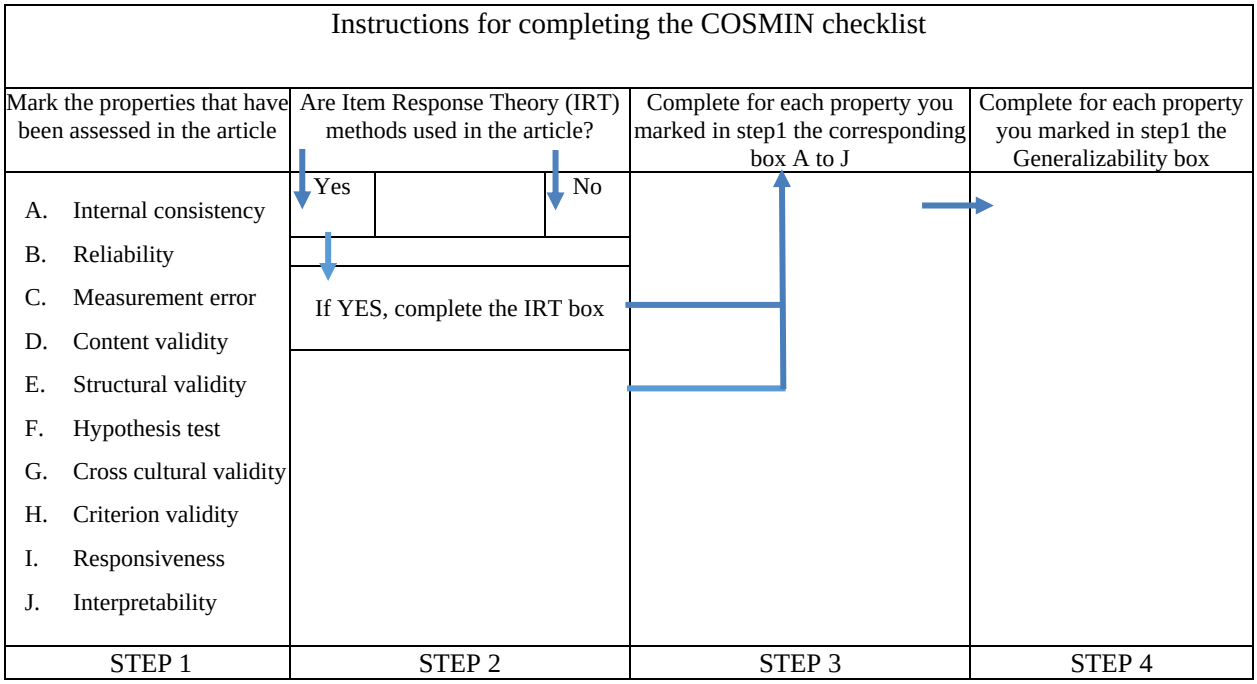
Table 1. Search strategy in CINALH & PsycINFO

CINALH (EBSCOhost)
TI (comfort* OR discomfort* OR uncomfort*) AND AB (theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission*) NOT (chair* OR ergonomic* OR thermal* OR cloth*) 1990 to 2015 Filters; Humans, English
PSYCINFO (APA PsycNET)
Title: comfort* OR discomfort* OR uncomfort* AND Abstract: theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission* NOT Any Field: chair* OR ergonomic* OR thermal* OR cloth* AND Language: English AND Population Group: Human AND Year: 1990 To 2015

Table 2. A framework for assessing reflective and formative models

Considerations	Reflective model	Formative model
Theoretical considerations		
1.Nature of construct	Latent construct exists ✓ Latent construct exists independent of the measures used	Latent construct is formed ✓ Latent constructs is a combination of its indicators
2.Direction of causality between items and latent construct	Causality from construct to items ✓ Variation in the construct causes variation in the item measures ✓ Variation in item measures does not cause variation in the construct	Causality from items to construct ✓ Variation in the construct does not cause variation in the item measures ✓ Variation in item measures causes variation in the construct
3.Characteristics of items used to measure the construct	Items are manifested by the construct ✓ Items share a common theme ✓ Items are interchangeable ✓ Adding or dropping an item does not change the conceptual domain of the construct	Items define the construct ✓ Items need not share a common theme ✓ Items are not interchangeable ✓ Adding or dropping an item may change the conceptual domain of the construct
Empirical considerations		
4.Item intercorrelation	Items should have high positive intercorrelations ✓ Empirical tests: assessing internal consistency and reliability by Cronbach alpha, average variance extracted and factors loadings (e.g. from common or confirmatory factor analysis)	Items can have any pattern of intercorrelation but should possess the same directional relationship ✓ Empirical tests: no empirical assessment of indicator reliability is possible; various preliminary analyses are useful to check directionality between items and construct
5.Item relationships with construct antecedents and consequences	Items have a similar sign and significance of relationship with the antecedents/consequences as the construct ✓ Empirical tests: establishing content validity by theoretical considerations, assessing convergent and discriminant validity empirically.	Items may not have similar significance of relationships with the antecedents/consequences as the construct ✓ Empirical tests: assessing nomological validity by using a MIMIC model, and/or structural linkage with another criterion variable
6.Measurement error and collinearity	Identifying the error in items is possible ✓ Empirical test: identifying and extracting measurement error by common factor analysis	Identifying the error is not possible if the formative measurement model is estimated in isolation ✓ Empirical test: using the vanishing tetrad test to determine if the formative items behave as predicted. Collinearity should be ruled out by standard diagnostics such as the condition index

Figure 1. The four-step procedure to complete the COSMIN checklist



Note. This figure is adapted from "The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study", by Mokkink et al., 2010, *Quality of Life Research*, 19, p. 544.

3.2. INSTRUMENTS TO ASSESS PATIENT COMFORT DURING HOSPITALIZATION: A PSYCHOMETRIC REVIEW

INSTRUMENTS TO ASSESS PATIENT COMFORT:

A PSYCHOMETRIC REVIEW

Sonia LORENTE, Josep-Maria LOSILLA, Jaume VIVES

Universitat Autònoma de Barcelona

Sonia LORENTE, Adjunct lecturer, MSc, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB; Paediatric Nurse, PNP, Hospital de Terrassa, Consorci Sanitari de Terrassa, CST; Josep-Maria LOSILLA, Associate Professor, PhD, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB; Jaume VIVES, Associate Professor, PhD, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB.

This work was supported by the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness (July, 2015). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. No conflict of interest has been declared by the author(s).

All authors have agreed on the final version, and meet at least one of the following criteria recommended by the International Committee of Medical Journal Editors, ICMJE: substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, and drafting the article or revising it critically for important intellectual content. Correspondence concerning this article should be addressed to Jaume Vives, Associate Professor, PhD, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB. Phone: +34935812331. E-mail: jaume.vives@uab.cat

ABSTRACT

Aim: To analyse the psychometric properties and the utility of instruments used to measure patient comfort, physical, social, psychospiritual and/or environmental, during hospitalization. **Background**: There are no systematic reviews nor psychometric reviews of instruments used to measure comfort, which is considered an indicator of quality in health care associated with quicker discharges, increased patient satisfaction and better cost-benefit ratios for the institution. **Design**: Psychometric review. **Data sources**: MEDLINE, CINAHL, PsycINFO, Web of Knowledge, ProQuest Thesis&Dissertations, Google. **Review methods**: We limited our search to studies published between 1990 and 2015. The psychometric analysis was performed using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN), along with the Quality Criteria for Measurement Properties. The utility of the instruments was assessed according to their cost-efficiency, acceptability and educational impact. Protocol registration in PROSPERO, CRD42016036290. **Results**: Instruments reviewed showed moderate methodological quality, and their utility was poorly reported. Thus, we cannot recommend any questionnaire without reservations, but the Comfort Scale, the General Comfort Questionnaire and their adaptations in adults and elderly patients, the Psychosocial Comfort Scale and the Incomfort des Patients de Reanimation are the most recommendable instruments to measure comfort. **Conclusions**: The methodology of the studies should be more rigorous, and authors should adequately report the utility of instruments. This review provides a strategy to select the most suitable instrument to assess patient comfort according to their psychometric properties and utility, which is crucial for nurses, clinicians, researchers and institutions.
Keywords: Psychometric Review, Validity of instruments, Patient comfort, Assessment tools, Utility of instruments, Nursing.

SUMMARY STATEMENT

Why is this review needed?

- There are no psychometric reviews of the health instruments used to assess patient comfort during hospitalization, being the comfort an indicator of quality in health care.
- Assessing the psychometric properties and the utility of the instruments measuring patient comfort is essential to improve the quality of care.

What are the key findings?

- Most instruments measuring patient comfort were valid and reliable, but no instrument can be completely recommended according to their utility, i.e., cost-efficiency, acceptability and educational impact.
- Methodologies used in studies on the measurement properties of instruments assessing patient comfort should be more accurate and authors should adequately report their utility.

How should the findings be used to influence policy/practice/research/education?

- Findings categorize the comfort instruments according to their measurement properties, allowing carers to select the most suitable comfort questionnaire in current clinical scenarios, which is essential for researchers and clinicians.
- Findings also provide a strategy to develop the most suitable comfort measurement instrument in accordance with their psychometric properties, utility and purpose, which is crucial for researchers and patients.
- Selecting the most appropriate instrument to assess comfort may improve the health care provided and patient satisfaction as well as reducing the institution costs.

INTRODUCTION

Measuring patient experience is essential to improving the quality of care provided in different health settings (WHO 2000, 2004, 2013). To date, different health care instruments have been developed to assess patient experience during hospitalization (Bruyneel et al., 2017; Manary, Boulding, Staelin, & Glickman, 2013; Pettersen, Veenstra, Guldvog, & Kolstad, 2004). In this sense, different systematic reviews have been developed to assess the measurement properties of these instruments, which are normally used to assess patient experiences related to pain reduction, care, hospital environment, or communication with professionals, to name a few (Beattie, Lauder, Atherton, & Murphy, 2015; Ellis-Smith et al., 2016). Nevertheless, there are no systematic reviews of the health instruments used to assess patient comfort during hospitalization, with comfort considered a direct indicator of quality of health care (NQMC 2002; Kolcaba 2013). Given that improving the physical, social, psychospiritual and/or environmental patient comfort is associated with quicker discharges, fewer readmissions, increased patient satisfaction and stronger cost-benefit ratios for the institution (Kolcaba, 2001; 2013), it is essential for nurses, clinicians, researchers and institutions to know how the different instruments assessing comfort perform. With this purpose, a psychometric review to assess the validity (e.g. construct validity or content validity) and the reliability (e.g. inter-rater reliability) of instruments assessing patient comfort was conducted. In addition, since the instruments need to have high utility if they are to be used in the real-world practice, the cost-efficiency (e.g. cost of obtaining a sample), acceptability (e.g. suitability from the patient perspective), and educational impact (e.g. utility of the collected data) of each instrument measuring comfort were also assessed, according to the utility matrix proposed by Beattie et al. (2015).

Background

Patient comfort is considered an individualized and holistic experience, a source of patient satisfaction and well-being. The concept of comfort is historically associated with nursing. Different theories and definitions of comfort have been developed, but most have been restricted to physical connotations, such as pain control. However, the Theory of Comfort, by Kolcaba (1992), associated the concept of comfort with strengthening, encouragement, aid and support, and provided a theoretical significance for comfort in nursing. According to this theory, comfort is defined as the immediate experience of being strengthened by the need for relief (the experience of a patient whose specific comfort need is addressed), ease (the state of calmness or contentment), and transcendence (the state in which the patient rises above pain or problems), met in physical, psychospiritual, sociocultural, and environmental contexts; much more so than the absence of pain or other physical discomforts. When the three kinds of comfort are combined with the four contexts of experience, a twelve-cell grid is created, which is useful for

assessing the comfort needs of patients and families (e.g. privacy, pain control, information about clinical procedures, anxiety, or noisy environment), for planning interventions to address those needs, evaluating the effectiveness of those interventions to enhance the comfort, and for measuring the desired outcome in research and practice (Kolcaba, 2013). In this context, the General Comfort Questionnaire (GCQ) is probably the first instrument specifically developed to assess patient comfort as a holistic experience, registered as a multidisciplinary outcome indicator of quality in health care in the National Quality Measures Clearinghouse (Kolcaba 1992, 2013; NQMC 2002). Since then, a large range of instruments to assess patient comfort have been developed, adapted or validated, either within the Theory of Comfort or other theoretical frameworks.

To our knowledge, there are no psychometric reviews assessing the performance of comfort instruments to date. For this reason, this study aims to examine the psychometric properties and utility of each questionnaire used to measure patient comfort in order to select the most valid, reliable and useful instrument for nurses, clinicians and researchers to use in present-day health care.

THE REVIEW

Aims

1. Identify health instruments measuring patient comfort as a holistic experience during hospitalization.
2. Systematically review the measurement properties and interpretability of each comfort instrument to measure patient comfort.
3. Examine the utility of each comfort instrument according to their cost efficiency, acceptability and educational impact in different health care settings.
4. Classify the different comfort instruments according to their measurement properties, and utility.

Design

According to the protocol of this review (Lorente, Vives & Losilla, 2017), we conducted a psychometric review, applying the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) to assess the methodological quality of studies on measurement properties (Mokkink et al., 2012; Terwee et al., 2012) and the Quality Criteria for Measurement Properties to assess the quality of instruments (Terwee et al. 2007). This review was registered in PROSPERO, CRD42016036290.

Search methods

We aimed to identify published instruments to assess patient comfort during hospitalization. We defined different combinations of keywords (using Mesh and other thesauruses, where available), in relation to the concept (e.g., comfort, theory), setting (hospitalization or admission) and instruments (e.g., questionnaires, scales) (Supplementary file 1). The search was carried out between 1990 and 2015, the findings were restricted by language (English), and the following databases were included: Medical Literature Analysis and Retrieval System (MEDLINE), by ProQuest, Cumulative Index to Nursing and Allied Health Literature (CINAHL), by EBSCOhost, Psychological Information (PsycINFO), by APA PsycNET, Thesis & Dissertations, by ProQuest, and ISI Web of Knowledge (WoS, Web of Science CORE), by Thomson Reuters. To include grey literature, we also searched records in Google, and reviewed up to 400 links. In addition, search alerts in CINAHL & PsycINFO were set until December 2016.

Inclusion criteria

Time frame

From 1990, with the development of the first instrument to assess patient comfort framed within the Theory of Comfort, by Kolcaba in 1992, until 2015.

Study type

Studies developing or validating questionnaires and/or scales measuring holistic comfort during hospitalization. Protocols, guidelines, conceptual frameworks, narrative reviews, intervention studies, report opinions and qualitative studies, were excluded.

Age group

We included all age groups (new-borns, toddlers, children, teenagers, young adults, middle-aged adults and elderly people).

Context

We evaluated instruments used to measure comfort in different healthcare settings where the patient was hospitalized due to acute illnesses (e.g. pneumonia, urinary tract infection), chronic pathologies (e.g. psychiatric illnesses, dementia), surgical interventions or childbirth. Therefore, the following settings were included: general paediatric wards, general adult wards, delivery rooms and maternity wards, operating rooms, Intensive Care Units (ICU), Paediatric Intensive Care Units (PICU), Neonatal Intensive Care Units (NICU), postoperative areas, psychiatric wards and hospices.

Instruments used to assess patient comfort

We included instruments developed with scales and/or subscales with closed-ended items, as Likert, Visual Analogic Scales (VAS) and/or Numerical Rating Scales (NRS) specifically designed for measuring patient comfort as a holistic experience during hospitalization. Instruments measuring comfort during screening or diagnostic tests were excluded (e.g. colonoscopy, bronchoscopy, angiography, mammography, injections, biopsies, cystoscopy, cytology, fertility treatments), as well as those instruments assessing comfort exclusively by physiological parameters (i.e. heart rate and/or blood pressure) or those which measure the effect of a specific intervention (e.g. warm blanket vs. classical blanket or midazolam vs. fentanyl).

Search outcomes

The results of the search strategy were reported according to the PRISMA flow diagram (Figure 1). A total of 2995 references were identified through databases search, plus 20 references from Google. After removing duplicates, 2843 titles and abstracts were screened. A reviewer applied the inclusion criteria to all titles and abstracts. If no decision could be made based solely on title and abstract alone, the full paper was retrieved. The inclusion criteria were checked independently by two review authors and discrepancies were resolved through discussion (with a third author where necessary). After the assessment of 103 full-texts for eligibility, 35 articles were selected for inclusion, and 14 additional articles were identified and retained by the references search and citation alerts. A total of 49 articles were included in the psychometric review.

Quality appraisal

Quality assessment of the studies

The methodological quality of included studies was assessed by using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink et al., 2010). The COSMIN checklist consists of 12 criteria, referred to as “boxes”, to evaluate whether general requirements of the study on measurement properties are met. We used eight boxes to assess the quality of the studies on internal consistency, reliability, measurement error, structural validity, content validity (including face validity), construct validity, cross-cultural validity, and responsiveness. The criterion validity was not rated since no gold standard for instruments to measure comfort exists (Mokkink et al., 2012; Terwee et al., 2012) (See supplementary File 2 for domains and definitions). Methodological quality of every measurement property was appraised using the four-point score (poor, fair, good, excellent), according to the “worse score counts” algorithm (Terwee, Mokkink & Patrick, 2011). For

instance, if item 3 in Reliability box was scored as fair (i.e. moderate sample), and other items of the box were scored as good or excellent, the overall methodological quality of Reliability box was fair. Therefore, if any item was scored as poor (i.e. small sample), the overall methodological quality of the box was scored as poor. Two reviewers independently assessed the methodological quality of studies and discrepancies were identified and resolved through discussion (with a third author where necessary).

Quality assessment of the instruments

The quality of the results for each study was assessed on the basis of Quality Criteria for Measurement Properties proposed by Terwee et al. (2007). These criteria consist of rating the content validity, internal consistency, construct validity, reproducibility, responsiveness, floor and ceiling effects and interpretability. Criterion validity was not assessed. The quality of the results was rated as "+" (positive), "-" (negative), "?" (doubtful), or "0" (no information) (See supplementary File 3 for domains, definitions and ratings). Two reviewers independently rated the quality of results, and discrepancies were identified and resolved through discussion (with a third author where necessary).

Data abstraction

Extracted information of each selected instrument included: author, year, country, outcome measures, purpose/use, number of items, response categories, type of patients; characteristics (theoretical framework, validity tests conducted and results, reliability tests conducted and results, responsiveness tests conducted and results, response rate, sample size, setting, respondents population and demographics, level of expertise required for scoring, time required to completion, mode of administration). Once the information was extracted, we categorised the instruments according to the measurement model (reflective and/or formative) taking into account the considerations detailed in Coltman et al. (2008), and we reported different aspects related to the utility (cost-efficiency, acceptability and educational impact) according to Van der Vleuten´s Utility criteria (Van der Vleuten, 1996, cited in Beattie et al. 2015). Some authors of eligible studies were contacted to provide missing or additional data when necessary.

Synthesis

The extracted information related to the instruments characteristics was reported in a table designed *ad hoc*. Results of the methodological quality assessment were synthetized in a table, according to the COSMIN domains (Mokkink et al. 2012) and the Quality Criteria of Measurement Properties proposed by Terwee et al. (2007). The extracted information related to the utility (cost-efficiency, acceptability and educational impact) was synthetized in a table

designed according to Van der Vleuten´s Utility Index Matrix (Van der Vleuten, 1996, cited in Beattie et al. 2015).

RESULTS

Characteristics of instruments

Instruments assessing patient comfort were taken from 1991 to 2016 and were developed or adapted in the following countries: United States of America, United Kingdom, the Netherlands, Brazil, Spain, Thailand, China, Germany, Italy, Portugal, Norway, Sweden, France and Israel. Table 1 shows instruments assessing comfort, that are listed and described in the next sections: Comfort of paediatric patients, Comfort of adult patients, and Comfort in elderly patients with dementia (See supplementary File 6 for further details.

Comfort of paediatric patients

The comfort of paediatric patients was generally assessed by instruments based on the observation of specific emotions and physiological parameters. The Comfort Scale (CS) was the most common scale, with 8 items and two dimensions that account for the 84% of the variance, originally developed by Ambuel et al. (1992) to assess infant distress in the Paediatric Intensive Care Unit (PICU). The CS was based upon the concept of psychological distress, defined as a multidimensional response to internal or external aversive stimuli, that may include fear, discomfort, anxiety and pain. The measurement model used to design this instrument was reflective, with items concerned with facial expressions, muscle tension, heart rate and blood pressure.

Different authors validated the CS in PICU (Bear & Ward-Smith, 2006; Brunow de Carvalho, Lucas da Silva, Paulo, Fonseca & Belli, 1999; Courtman, Wardurgh, & Petros, 2003; Froom et al., 2008; Ista et al., 2005; Lamas et al., 2008; Nievas, Spentzas, & Bogue, 2014; Triltsch et al., 2005; Tschiedel, Müller, Schara, Felderhoff-Müser, & Dohna-Schwake, 2015). Some others authors also assessed the comfort of neonates by using the CS (Blauer, 1996; Cury, Martinez & Carlotti, 2013; Franck, Ridout, Howard, Peters, & Honour, 2011; van Dijk et al., 2000; Wielenga, de Vos, De Leeuw, & de Haan, 2004), Comfort Neo Scale (Monique van Dijk et al., 2009), Modified Comfort Scale (Gjerstad, Wagner, Henrichsen, & Storm, 2008; Lee & Young, 2005), and Adapted Comfort Scale (Caljouw et al., 2007). We also identified the Comfort Behavioural Scale (CBS) as an adaptation of the CS, with 6 items and one dimension (Amigoni et al., 2012; Andersen Bernklev, Langius-Eklof, Nakstad, & Jylli, 2015; Bai, hsu, Tang, &van Djik, 2012; Boerlage, Ista, De Jong, Tibboel, & van Djik, 2012; Boerlage et al. 2014; Da Costa Silva, Osório alves, dos Santos El Halal, dos Santos Pinheiro,& Carvalho, 2013; De Jong, Baartmans, Tuinebreijer, & van Loey, 2010; De Jong et al., 2012; Johansson & Kokinsky, 2009;

Tristão, Garcia, de Jesus, & Tomaz, 2013; Valkenburg et al. 2011). Lastly, the Paediatric Perioperative Comfort Instrument (PPCI), seven items (Moriber, 2009a), which assess the comfort of paediatric patients after surgical procedures was developed. This questionnaire was based upon the Theory of Comfort, by Kolcaba (1992), and the measurement model to design this instrument was reflective. Reflective items were related to facial expressions and verbal indicators.

Comfort of adult patients

The comfort of adult patients was evaluated in different health settings. The General Comfort Questionnaire (GCQ) (Kolcaba, 1992), with 48 items and twelve dimensions that accounted for a variance of 63.4%, was developed to assess the comfort of adult patients in medical and surgical wards and was based upon the Theory of Comfort (Kolcaba, 1992). The measurement model to design this instrument was mixed, reflective and formative. Reflective items were related to fear, anxiety, or information provided, while formative items were related to environmental sources of discomfort, such as light, furniture or noise. We found some adaptations of the GCQ, as the Childbirth Comfort Questionnaire (CCQ), 14 items to assess the comfort of the women during childbirth (Durnell, 2003), and Psychiatric In-Patients Comfort Scale (PICS), 38 items to assess the comfort of patients with psychiatric disorders (Alves-Apóstolo, Kolcaba, Cruz-Mendes, & Calvário-Antunes, 2007). The Psychosocial Comfort scale, 21 items, to assess the comfort of adults in medical and surgical wards (Yen, 1994). This questionnaire was also based on the Theory of Comfort, and the measurement model was reflective, with items concerned to anxiety, trust in the medical team or communication with professionals.

Different instruments were developed within other frameworks to assess the comfort of adult patients. The Pain Discomfort Scale (PDS), which consists of 10 items, was developed by Jensen, Karoly & Harris (1991) to assess the discomfort of patients with chronic pain. This questionnaire was based on the multidimensional nature of pain, and was designed as a reflective instrument, composed of items related to fear, helplessness or annoyance. The Incomforts des Patients de REAnimation (IPREA) includes 16 items to assess the comfort of the patients admitted to the Intensive Care Unit (ICU) by Kalfon et al. (2010). This tool was based on the conceptualization of stress and post-traumatic stress, and was designed as a formative questionnaire, with items concerned with environmental sources of discomfort, such as noise, light or furniture. The CS was validated to assess the comfort of adults admitted to ICU (Ashkenazy & DeKeyser-Ganz, 2011), and the Psychiatric Discomfort Scale (PDS), with 23 items, was used to assess the comfort of the patients with psychiatric disorders (Betemps, 1999).

Comfort of elderly patients with dementia

The comfort of elderly patients with dementia was assessed by diverse instruments based upon the conceptualization of the discomfort as the observation of specific emotions and body language, defined as negative and/or physical state in response to physical problems and/or environmental conditions. All of them were formative questionnaires. Formative items concerned with environmental sources of discomfort, such as noise, light or professional procedures, were included in this questionnaire. We identified the Discomfort in Alzheimer Type (DS-DAT), 9 items, (Dello Russo et al., 2008; Hurley, Volicer, Hanrahan, Houde, & Volicer, 1992); Discomfort Scale, 16 items (Morrison et al., 1998); Discomfort Behaviour Scale (DBS), 17 items (Stevenson, Brown, Dahl, Ward, & Skemp, 2006); Comfort Assessment Dying with Dementia, 14 items (CAD EOLD) (Kiely et al., 2006; Volicer, Hurley & Blasi, 2001) and Source of Discomfort Scale (SODS), 20 items (Cohen-Mansfield, Thein, Marx, Dakhee-Ali,& Jensen, 2013). In addition, we found the Hospice Comfort Questionnaire (HCQ), 49 items (Novak, Kolcaba, Steiner, & Dowd, 2001; Tanatwanit, 2011) and the End of Life Comfort Planning Questionnaire, 28 items (Oliveira, Caldeira, Amado, &Kolcaba, 2016), as adaptations from the GCQ. The former instrument measurement model was mixed (formative and reflective), and the latter was reflective.

Methodological quality

Table 2 shows the results of the methodological quality of studies on measurement properties, COSMIN, and the quality assessment of instruments, Quality Criteria of Measurement Properties. Some specific considerations are synthetized.

Comfort of paediatric patients

Content validity was only tested in two studies (Ambuel et al., 1992; Moriber, 2009), and the quality of methodology was considered as “excellent” according to the COSMIN criteria. However, the CS, by Ambuel et al. (1992), also obtained a “doubtful” score because the target population and the item selection process were not adequately reported. Reliability was tested in numerous studies, and most were rated as “fair” and “doubtful”, because authors assessed the inter-rater and/or test-retest reliability of continuous scores by using correlations (Ambuel et al. 1992; Bear & Ward-Smith, 2006; Lee & Young, 2005) or t-tests (Boerlage et al. 2012). Instrument structure, according to their factor analysis, was rated as “good” or “excellent”, although it was only tested by six authors (Ambuel et al., 1992; De Jong et al., 2012; Franck et al., 2011; Moriber, 2009; Valkenburg, 2012; Van Dijk et al., 2000). Construct validity (convergent and/or discriminative validity) was generally rated as “fair”, but “doubtful”, when the hypothesis about expected correlations between scores of instruments was not formulated. Responsiveness was also rated as “fair” and “doubtful”, in most cases when the hypothesis

about expected differences, the Smallest Detectable Change (SDC) or the Minimal Important Change (MIC), were not stated. Lastly, interpretability obtained a "doubtful" score when the comfort scores of patients were not adequately described per groups and subgroups.

Comfort of adult patients

Overall methodological quality of instruments was rated as "good", and most studies obtained a "positive" score on Quality Criteria of Measurement Properties. Nevertheless, some studies obtained a "doubtful" score in construct validity and responsiveness because the hypothesis and SDC or MIC were not stated.

Comfort of elderly patients with dementia

Content validity was tested in most studies. The methodological quality of content validity was generally rated as "excellent", and the quality of results obtained a "positive" score on Quality Criteria of Measurement Properties. Reliability was rated as "fair" and "doubtful" because authors assessed the inter-rater and/or test-retest reliability of continuous scores by using correlations (Cohen-Mansfield et al., 2013; Hurley et al. 1992; Morrison et al., 1998). The structure of the instruments, according to their factor analysis, was rated as "good", although it has only been tested by three authors (Oliveira et al., 2016; Stevenson et al., 2006; Volicer et al., 2001). Construct validity (convergent and/or discriminative validity) was rated as "fair" and "doubtful", when the hypothesis on the expected correlations between scores of instruments was not stated. Cross-cultural validity in translated questionnaires was rated as "poor" in all cases, since the multi-group factor analysis to test the construct invariance was not reported.

Utility of instruments

The utility index facilitated the evaluation of the cost-efficiency, acceptability and educational impact of each instrument. Results were showed in Table 3, and some considerations are synthetized.

Firstly, the cost-efficiency evaluated the sample size to achieve the required level of reliability according to the purpose of the instrument, administering the time, and the administrative costs of applying the questionnaire and the completion of a reliable sample (e.g. professional training or number of collaborators). Because authors did not report the sample size used to reach the required reliability, we took into account the participants of each study. When authors did not report administrative costs, we considered whether the tool was self-reported or administered by interviewers (some administrative costs) or administered by trained professionals (large administrative costs). The CS, by Ambuel et al. (1992), and associated questionnaires, reported a shorter administering time (2 minutes), but also greater administrative costs, because the questionnaire had to be applied by specifically trained professionals, either nurses or doctors, or

both. Second, the acceptability evaluated the understanding of the patients and the assessments (missing items and response rate). However, we were unable to evaluate the assessments in most cases, as information related to missing items and/or response rate was generally not reported, apart from Jensen et al. (1991), van Djik et al. (2000), Novak et al. (2001), Stevenson et al. (2006) and Boerlage et al. (2014). Lastly, the educational impact evaluated the purpose of the instrument, the scoring system and the feedback of the results. We considered that all studies reported evidence of achieved purpose, and the results were applicable in a practical context.

DISCUSSION

To our knowledge, this is the first psychometric review assessing the performance of instruments used to assess patient comfort during hospitalization. With this purpose, we analysed the measurement properties and the use of a wide variety of questionnaires assessing comfort in different settings and age groups. The CS (Ambuel et al. 1992) and associated questionnaires were the most popular instruments measuring comfort in critical paediatric units. The GCQ (Kolcaba, 1992) and associated questionnaires were focused on assessing comfort in diverse areas, as surgical, medical and psychiatric wards or childbirth. Different questionnaires were also developed to assess comfort of elderly patients with dementia (DS-DAT, DBS, CAD EOLD, SODS, Discomfort Scale, HCQ, and End of Life).

The methodological quality assessment indicated that several instruments were not as rigorously developed and validated as COSMIN and Quality Criteria of Measurement Properties recommend (Terwee et al. 2007; Mokkink et al. 2012). Poor reporting of item selection process and/or incomplete descriptions of the sample characteristics may limit the content validity (Terwee et al. 2007). The confirmatory factor analysis (CFA) to explore the dimensionality of the questionnaire and to determine the reliability coefficients of internal consistency derived from the measurement model were not always conducted and authors frequently reported results of CFA and reliability analyses from previously published research instead of analysing both in their own samples. This practice may be problematic, as a measurement model, given that factor loadings and reliability depend on the sample data, its size and missing items (De Vet, Adèr, Terwee, & Power, 2005; Floyd & Widaman 1995; Mokkink et al. 2012). Moreover, the reliability of comfort scores was sometimes analysed by correlations, so the systematic error is not taken into account and agreement is not really assessed. The evaluation of construct validity and sensitivity without testing specific hypothesis may lead to misleading conclusions, since authors may be tempted to offer alternative explanations for low correlations or little mean differences instead of concluding that the questionnaire maybe is not valid for the intended purpose (Terwee et al. 2007; Mokkink et al. 2012). Lastly, when an instrument measuring

comfort was translated and cross-culturally adapted, the multi-group factor analysis to test the construct invariance was not performed, increasing the risk of biased results in the comparison of scores (Little & Slegers 2005; Mokkink et al. 2012).

As far as the assessment of the utility of instruments is concerned, the educational impact was excellent, but cost-efficiency and acceptability were difficult to appraise because costs and assessments (missing items and response rate) were poorly reported, as pointed out in Beattie et al. (2015). For this reason, we were unable to rigorously categorize the instruments according to their utility. However, to select the right instrument for an intended purpose and a clinical scenario not only the usability should be considered, but also the attributes of comfort to be measured. In this sense, the self-reported instruments framed within the Theory of Comfort, by Kolcaba (1992), which assessed the comfort of patients in four contexts, most closely reflect the real experience of patients during hospitalization. But some instruments assessing the comfort of the paediatric or elderly patients may lead to biased conclusions, because the family experience is not usually assessed (e.g. CS or SODS). Different attributes of comfort, such as communication with professionals, information about procedures, opinions about the treatment, privacy, etc. need to be assessed even when patients are unable to refer to comfort needs. In these cases, interviewing the family may be crucial to really assess the comfort of paediatric patients or elderly patients with dementia.

Recommendations for further research

Assessing comfort patient is essential to increase patient satisfaction and institutions. Given that both patient comfort and satisfaction are considered quality indicators of the health care provided (WHO, 2000; NQMC 2002; Kolcaba 2013), instruments aiming to assess these patient experiences should be valid, reliable and useful (Keszei, Novak, & Streiner, 2010; Terwee et al. 2007). Therefore, recommendations for further research include the systematic use of methodological quality assessment checklists, as COSMIN (Mokkink et al. 2012) and/or Quality Criteria of Measurement Properties (Terwee et al. 2007), and better data reporting including subjects understanding, application costs and assessments.

Strengths and limitations

A strength of our study is the use of the COSMIN (Mokkink et al. 2012) to evaluate methodological quality and the measurement properties of included studies, and the Quality Criteria (Terwee et al. 2007) to assess the quality of instruments, along with the application of the Utility Index Matrix (Beattie et al. 2015) to assess their cost-efficiency, acceptability and educational impact and discuss their usefulness in the real cases. Concerning limitations, it should be noted that only English literature was included and, although most of the studies were

peer-reviewed published papers, there were five unpublished thesis and dissertations. Lastly, the poor reporting of some studies made it difficult to categorize the instruments according to their utility, as we had initially aimed to do.

CONCLUSIONS

Measuring patient comfort is a good practice to improve the health care provided and to increase patient satisfaction in present-day clinical scenarios. With this purpose, a number of instruments assessing comfort have been developed and validated across different settings and circumstances, most of which are valid and reliable. However, there is no instrument that can currently be wholly recommended among the questionnaires we have reviewed. The CS and the CBS were the most adequate questionnaires to assess the comfort of children in critical areas. However, the moderate methodological quality and the cost of training professionals should be taken into account when these instruments are applied. The GCQ, and their adaptations to assess both the comfort of adult patients and elderly patients with dementia, reported overall good methodological quality, and they are probably the best self-reported questionnaires, to assess the comfort of the patient in diverse scenarios, although more validations with different samples are needed. The Psychosocial Comfort scale and IPREA also reported good methodological quality, being the most recommendable instruments to assess the comfort of the patient with surgical procedures and medical diagnoses or admitted to critical areas, respectively. Nevertheless, the former showed poor reporting on costs and assessments, and the latter reported floor effects.

The findings of this psychometric review additionally provide a strategy to select and develop the most appropriate instrument to assess comfort patient in accordance to their purpose, psychometric properties, and utility. Firstly, the measurement model, which is reflective and/or formative, should be considered when a new instrument is developed, as the measurement model defines the construct of interest, either comfort or discomfort. In addition, the questionnaires devised from a mixed theoretical model, formative and reflective, allow the identification of different sources of comfort/discomfort as well as the diagnosis of the level of comfort/discomfort, respectively. So, these mixed questionnaires may be useful for nurses, clinicians and researchers to assess the comfort needs of patients and families, to plan interventions to address those needs (e.g. pain reduction, hand massage, coaching or guided imagery to reduce the anxiety), and to evaluate the effectiveness of those interventions to enhance comfort and well-being. Secondly, questionnaires assessing comfort must be adequately validated and their reliability should be correctly established to ensure their methodological quality. In this sense, the correct validation of translated and cross-culturally adapted questionnaires is also required to ensure the invariance of the construct.

Otherwise, the results may lead to biased conclusions. Lastly, instruments measuring patient comfort should report cost-efficiency, acceptability and educational impact, which are crucial in the daily practice, both for professionals and patients.

Funding

This work was supported by the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness (July 2015). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflict of interest

No conflict of interest has been declared by the author(s). All authors have agreed on the final version, and meet at least one of the following criteria recommended by the International Committee of Medical Journal Editors, ICMJE: substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, and drafting the article or revising it critically for important intellectual content.

Author contributions

All authors have agreed on the final version and meet at least one of the following criteria [recommended by the ICMJE (http://www.icmje.org/recommendations/)]:

- substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data;
- drafting the article or revising it critically for important intellectual content

REFERENCES

Alves-Apóstolo, J. L., Kolcaba, K., Cruz-Mendes, A., & Calvário-Antunes, M. T. (2007). Development and psychometric evaluation of the Psychiatric In-patients Comfort Scale (PICS). *Enfermeria Clinica*, *17*(1), 17–23. http://doi.org/10.1016/S1130-8621(07)71760-6

Ambuel, B., Hamlett, K. W., Marx, C. M., & Blumer, J. L. (1992). Assessing distress in pediatric intensive care environments: the COMFORT scale. *Journal of Pediatric Psychology*, *17*(1), 95–109. http://doi.org/10.1093/jpepsy/17.1.95

Amigoni, A., Mozzo, E., Brugnaro, L., Gentilomo, C., Stritoni, V., Michelin, E., & Pettenazzo, A. (2012). Assessing sedation in a pediatric intensive care unit using Comfort Behavioural Scale and Bispectral Index: these tools are different. *Minerva Anestesiologica*, *78*(3), 322–329.

Andersen, R. D., Bernklev, T., Langius-Eklof, A., Nakstad, B., & Jylli, L. (2015). The COMFORT behavioural scale provides a useful assessment of sedation, pain and distress in toddlers undergoing minor elective surgery. *Acta Paediatrica*, *104*(9), 904–909. http://doi.org/10.1111/apa.13045

Ashkenazy, S., & DeKeyser-Ganz, F. (2011). Assessment of the reliability and validity of the Comfort Scale for adult intensive care patients. *Heart & Lung*, *40*(3), e44-51. http://doi.org/10.1016/j.hrtlng.2009.12.011

Bai, J., Hsu, L., Tang, Y., & Van Dijk, M. (2012). Validation of the COMFORT Behavior Scale and the FLACC Scale for Pain Assessment in Chinese Children after Cardiac Surgery. *Pain Management Nursing*, *13*(1), 18–26. http://doi.org/10.1016/j.pmn.2010.07.002

Bear, L. A., & Ward-Smith, P. (2006). Interrater reliability of the COMFORT scale. *Pediatric Nursing*, *32*(5), 427.

Beattie, M., Lauder, W., Atherton, I., & Murphy, D. J. (2015). Instruments to measure patient experience of health care quality in hospitals: a systematic review. *Systematic Reviews*, *3*(1), 4. http://doi.org/10.1186/2046-4053-3-4

Betemps, E. J. (1999). A Self-Administered Instrument to Measure Psychiatric Discomfort of Persons With Serious Mental Illness. *Psychiatric Services*, *50*(1), 107–108. http://doi.org/10.1176/ps.50.1.107

Blauer, T. (1996). *A simultaneous comparison of three neonatal pain scales during common newborn intensive care unit procedures*. University of Utah.

Boerlage, A. A., Ista, E., de Jong, M., Tibboel, D., & van Dijk, M. (2012). The COMFORT behavior scale: is a shorter observation period feasible?. *Pediatric Critical Care Medicine*, *13*(2), e124–e125. http://doi.org/10.1097/PCC.0b013e3182192d92

Boerlage, A. A., Ista, E., Duivenvoorden, H. J., de Wildt, S. N., Tibboel, D., & van Dijk, M. (2014). The COMFORT behaviour scale detects clinically meaningful effects of analgesic and sedative treatment. *European Journal of Pain*, *19*(4), 473–479. http://doi.org/10.1002/ejp.569

Brunow de Carvalho, W., Lucas da Silva, P. S., Paulo, C. S., Fonseca, M. M., & Belli, L. A. (1999). Comparison between the Comfort and Hartwig sedation scales in pediatric patients undergoing mechanical lung ventilation. *São Paulo Medical Journale*, *117*(5), 192–196.

Bruyneel, L., Coeckelberghs, E., Buyse, G., Casteels, K., Lommers, B., Vandersmissen, J., … Vanhaecht, K. (2017). Validation of the Child HCAHPS survey to measure pediatric inpatient experience of care in Flanders. *European Journal of Pediatrics*. http://doi.org/10.1007/s00431-017-2919-7

Caljouw, M. A. A., Kloos, M. A. C., Olivier, M. Y., Heemskerk, I. W., Pison, W. C. R., Stigter, G. D., & Verhoef, A.-M. J. H. (2007). Measurement of pain in premature infants with a gestational age between 28 to 37 weeks : Validation of the adapted COMFORT scale. *Journal of Neonatal Nursing*, *13*(1), 13–18. http://doi.org/10.1016/j.jnn.2006.11.007

Cohen-Mansfield, J., Thein, K., Marx, M., Dakheel-Ali, M., & Jensen, B. (2013). Sources of Discomfort in Persons With Dementia. *JAMA Internal Medicine*, *173*(14), 1378–1379. http://doi.org/10.1001/jamainternmed.2013.6480.Author

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research*, *61*(12), 1250–1262. http://doi.org/10.1016/j.jbusres.2008.01.013

Courtman, S. P., Wardurgh, A., & Petros, A. J. (2003). Comparison of the bispectral index monitor with the Comfort score in assessing level of sedation of critically ill children. *Intensive Care Medicine*, *29*(12), 2239–2246.

Cury, M. R. J., Martinez, F. E., & Carlotti, A. P. C. P. (2013). Pain assessment in neonates and infants in the post-operative period following cardiac surgery. *Postgraduate Medical Journal*, *89*(1048), 63–7. http://doi.org/10.1136/postgradmedj-2012-130808

da Costa Silva, C., Osório Alves, M. M., dos Santos El Halal, M. G., dos Santos Pinheiro, S., & Carvalho, P. R. A. (2013). A comparison of gradual sedation levels using the Comfort-B scale and bispectral index in children on mechanical ventilation in the pediatric intensive care unit. *Revista Brasileira de Terapia Intensiva*, *25*(4), 306–311. http://doi.org/10.5935/0103-507X.20130052

de Jong, A., Baartmans, M., Tuinebreijer, W., & van Loey, N. E. E. (2010). Reliability , validity and clinical utility of three types of pain behavioural observation scales for young children with burns aged 0 – 5 years. *Pain*, *150*(3), 561–7. http://doi.org/10.1016/j.pain.2010.06.016

de Jong, A., Tuinebreijer, W., Bremer, M., van Komen, R., Middelkoop, E., & van Loey, N. E. E. (2012). Construct validity of two pain behaviour observation measurement instruments for young children with burns by Rasch analysis. *Pain*, *153*(11), 2260–6. http://doi.org/10.1016/j.pain.2012.07.021

de Vet, H. C. W., Adèr, H. J., Terwee, C. B., & Pouwer, F. (2005). Are factor analytical techniques used appropriately in the validation of health status questionnaires? A systematic review on the quality of factor analysis of the SF-36. *Quality of Life Research*, *14*(5), 1203–1218. http://doi.org/10.1007/s11136-004-5742-3

Dello Russo, C., Di Giulio, P., Brunelli, C., Dimonte, V., Villani, D., Renga, G., & Toscani, F. (2008). Validation of the Italian version of the Discomfort Scale - Dementia of Alzheimer Type. *Journal of Advanced Nursing*, *64*(3), 298–304. http://doi.org/10.1111/j.1365-2648.2008.04810.x

Durnell, K. (2003). *Exploring the Presence of Comfort within the Context of Childbirth. Thesis*. University of Michigan.

Ellis-Smith, C., Evans, C., Bone, A., Henson, L., Dzingina, M., Kane, P., & Higginson, I. (2016). Measures to assess commonly experienced symptoms for people with dementia in long-term care settings: A systematic review. *BMC Medicine*, *14*, 38–49. http://doi.org/10.1186/s12916-016-0582-x

Floyd, F. J., & Widaman, K. F. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment*, *7*(3), 286–299. http://doi.org/10.1037/1040-3590.7.3.286

Franck, L. S., Ridout, D., Howard, R., Peters, J., & Honour, J. W. (2011). A comparison of pain measures in newborn infants after cardiac surgery. *Pain*, *152*(8), 1758–1765. http://doi.org/10.1016/j.pain.2011.03.017

Froom, S. R., Malan, C. A., Mecklenburgh, J. S., Price, M., Chawathe, M. S., Hall, J. E., & Goodwin, N. (2008). Bispectral Index asymmetry and COMFORT score in paediatric intensive care patients. *British Journal of Anaesthesia*, *100*(5), 690–696. http://doi.org/http://dx.doi.org/10.1093/bja/aen035

Gjerstad, A. C., Wagner, K., Henrichsen, T., & Storm, H. (2008). Skin conductance versus the modified COMFORT sedation score as a measure of discomfort in artificially ventilated children. *Pediatrics*, *122*(4), e848-53. http://doi.org/10.1542/peds.2007-2545

Hurley, A. C., Volicer, B. J., Hanrahan, P. A., Houde, S., & Volicer, L. (1992). Assessment of discomfort in advanced Alzheimer patients. *Research in Nursing & Health*, *15*(5), 369–77. http://doi.org/10.1002/nur.4770150506

Ista, E., M, van D., Tibboel, D., M, de H., Aneja, R., van Dijk, M., … Aneja, R. (2005). Assessment of sedation levels in pediatric intensive care patients can be improved by using the COMFORT "behavior" scale. *Pediatric Critical Care Medicine*, *6*(1), 58–92 8p. http://doi.org/10.1097/01.PCC.0000149318.40279.1A

Jensen, M. P., Karoly, P., & Harris, P. (1991). Assessing the affective component of chronic pain: Development of the pain discomfort scale. *Journal of Psychosomatic Research*, *35*(2–3), 149–154. http://doi.org/10.1016/0022-3999(91)90069-Z

Johansson, M., & Kokinsky, E. (2009). The COMFORT behavioural scale and the modified FLACC scale in paediatric intensive care. *Nursing in Critical Care*, *14*(3), 122–130. http://doi.org/10.1111/j.1478-5153.2009.00323.x

Kalfon, P., et al. (2010). Development and validation of a questionnaire for quantitative assessment of perceived discomforts in critically ill patients. Intensive care medicine, 36 (10), 1751 - 1758. https://doi.org/10.1007/s00134-010-1902-9

Keszei, A., Novak, M., & Streiner, D. L. (2010). Introduction to health measurement scales. *Journal of Psychosomatic Research*, *68*(4), 319–323. http://doi.org/10.1016/j.jpsychores.2010.01.006

Kiely, D. K., Volicer, L., Teno, J., Jones, R. N., Prigerson, H. G., & Mitchell, S. L. (2006). The Validity and Reliability of Scales for the Evaluation of End-of- Life Care in Advanced Dementia. *Alzheimer Disease and Associated Disorders*, *20*(3), 176–181.https://doi.org/10.1097/00002093-200607000-00009

Kolcaba, K. (1992). Holistic comfort: operationalizing the construct as a nurse-sensitive outcome. *Advances in Nursing Science*, *15*(1), 1–10. https://doi.org/10.1097/00012272-199209000-00003

Kolcaba, K. (2001). Evolution of the mid range theory of comfort for outcomes research. *Nursing Outlook*, *49*(2), 86–92. https://doi.org/10.1067/mno.2001.110268

Kolcaba, K. (2013). Comfort. In Lippincott Williams & Wilkins (Ed.), *Middle Range Theories. Application to nursing research.* (Third edit, p. 372). Philadelphia.

Lamas, A., López-Herce, J., Sancho, L., Mencía, S., Carrillo, A., Santiago, M. J., & Martínez, V. (2008). Assessing sedation in critically ill children by bispectral index, auditory-evoked potentials and clinical scales. *Intensive Care Medicine*, *34*(11), 2092–2099. http://doi.org/10.1007/s00134-008-1198-1

Lee, W. K., & Young, B. W. Y. (2005). Measuring the sedation level of mechanically ventilated infants by a modified COMFORT scale. *Hong Kong Journal of Paediatrics*, *10*(3), 189–195.

Little, T. D., & Slegers, D. W. (2005). Factor Analysis: Multiple Groups. *Behavioral Science*, *2*, 617–623. http://doi.org/10.1002/0470013192.bsa221

Lorente S. , Vives J. & Losilla J. M. (2017) Instruments to assess the patient comfort during hospitalization: a psychometric review protocol. *Journal of Advanced Nursing, 73* (3), 735-741. doi: 10.1111/jan.13180

Manary, M. P., Boulding, W., Staelin, R., & Glickman, S. W. (2013). The Patient Experience and Health Outcomes. *New England Journal of Medicine*, *368*(3), 201–203. http://doi.org/10.1056/NEJMp1211775

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D. L., … de Vet, H. C. W. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, *19*(4), 539–549. http://doi.org/10.1007/s11136-010-9606-8

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. W. (2012). *The COSMIN checklist manual*. Retrieved from http://www.cosmin.nl/images/upload/files/COSMIN checklist manual v9.pdf

Moriber, N. (2009). *Evaluating the Reliability and Validity of the Pediatric Perioperative Comfort Instrument: A Pschometric Study. Thesis*. Rush University.

Morrison, R. S., Ahronheim, J. C., Morrison, G. R., Darling, E., Baskin, S. A., Morris, J., … Meier, D. E. (1998). Pain and discomfort associated with common hospital procedures and experiences. *Journal of Pain and Symptom Management*. Netherlands: Elsevier Science. http://doi.org/10.1016/S0885-3924(97)00261-3

National Quality Measures Clearinghouse, NQMC. (2002). In AHRQ, Agency for Healthcare Research and Quality, U.S. Department of Health & Human Services. Retrieved from http://www.qualitymeasures.ahrq.gov/search/search.aspx?term=comfort. on August 2015.

Nievas, I. F. F., Spentzas, T., & Bogue, C. W. (2014). SNAP II index: an alternative to the COMFORT scale in assessing the level of sedation in mechanically ventilated pediatric patients. *Journal of Intensive Care Medicine*, *29*(4), 225–228. http://doi.org/http://dx.doi.org/10.1177/0885066613475422

Novak, B., Kolcaba, K., Steiner, R., & Dowd, T. (2001). Measuring comfort in caregivers and patients during late end-of-life care. *American Journal of Hospice & Palliative Care*, *18*(3), 170–80 3p 1p. http://doi.org/10.1177/104990910101800308

Oliveira, S. M., Caldeira, S. M. A., Amado, J. C., & Kolcaba, K. (2016). Cultural adaptation and validation of the Portuguese End of Life Spiritual Comfort Questionnaire in Palliative Care patients. *Porto Biomedical Journal*, *In press*, 6–11. http://doi.org/10.1016/j.pbj.2016.08.003

Pettersen, K. I., Veenstra, M., Guldvog, B., & Kolstad, A. (2004). The Patient Experiences Questionnaire: Development, validity and reliability. *International Journal for Quality in Health Care*, *16*(6), 453–463. https://doi.org/10.1093/intqhc/mzh074

Stevenson, K. M., Brown, R. L., Dahl, J. L., Ward, S. E., & Skemp, M. (2006). The Discomfort Behaviour Scale: A Measure of Discomfort in the Cognitively Impaired Based on the Minimum Data Set 2.0. *Research in Nursing & Health*, *29*, 576–587. http://doi.org/10.1002/nur.20168

Tanatwanit, Y. (2011). *Comfort as experienced by Thai older patients with advanced cancer*. *Thesis*. Retrieved from: http://cuislandora.wrlc.org/islandora/object/etd% 3A105/datastream/PDF/view

Terwee, C., Bot, S., de Boer, M., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., … de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. http://doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C., Mokkink, L., Knol, D., Ostelo, R. W. J. G., Bouter, L. M., & de Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651–657. http://doi.org/10.1007/s11136-011-9960-1

Terwee, C., Mokkink, L., & Patrick, D. (2011). COSMIN checklist with 4-point scale. Retrieved from http://www.cosmin.nl

Triltsch, A. E., Nestmann, G., Orawa, H., Moshirzadeh, M., Sander, M., Grosse, J., … Spies, C. D. (2005). Bispectral index versus COMFORT score to determine the level of sedation in paediatric intensive care unit patients: a prospective study. *Critical Care (London, England)*, *9*(1), R9-17. http://doi.org/10.1186/cc2977

Tristão, R. M., Garcia, N. V. M., de Jesus, J. A. L., & Tomaz, C. (2013). COMFORT behaviour scale and skin conductance activity: What are they really measuring? *Acta Paediatrica*, *102*(9), 402–406. http://doi.org/10.1111/apa.12325

Tschiedel, E., Müller, O., Schara, U., Felderhoff-Müser, U., & Dohna-Schwake, C. (2015). Sedation monitoring during open muscle biopsy in children by Comfort Score and Bispectral Index - A prospective analysis. *Paediatric Anaesthesia*, *25*(3), 265–271. http://doi.org/10.1111/pan.12547

Valkenburg, A. J. (2012). *The COMFORT-behavior scale is useful to assess pain and distress in 0-to 3-year-old children with Down syndrome. Thesis*.

van Dijk, M., de Boer, J. B., Koot, H. M., Tibboel, D., Passchier, J., & Duivenvoorden, H. J. (2000). The reliability and validity of the COMFORT scale as a postoperative pain instrument in 0 to 3-year-old infants. *Pain*, *84*(2–3), 367–377. http://doi.org/10.1016/S0304-3959(99)00239-0

van Dijk, M., Roofthooft, D. W., Anand, K. J., Guldemond, F., de Graaf, J., Simons, S., … Tibboel, D. (2009). Taking up the challenge of measuring prolonged pain in (premature) neonates. *Clinical Journal of Pain*, *25*(7), 607–616. http://doi.org/10.1097/AJP.0b013e3181a5b52a

Volicer, L., Hurley, a C., & Blasi, Z. V. (2001). Scales for evaluation of End-of-Life Care in Dementia. *Alzheimer Disease and Associated Disorders*, *15*(4), 194–200. http://doi.org/10.1097/00002093-200110000-00005

Wielenga, J. M., de Vos, R., de Leeuw, R., & de Haan, R. J. (2004). COMFORT Scale: a reliable and valid method to measure the amount of stress of ventilated preterm infants. *Neonatal Network*, *23*(2), 39–44. http://doi.org/10.1891/0730-0832.23.2.39

World Health Organization. (2000). Strategies for assisting health workers to modify and improve skills: developing quality health care-a process of change. *Issues in Health Services*. Retrieved from http://apps.who.int/iris/bitstream/10665/66285/1/WHO_EIP_OSD_00.1.pdf

World Health Organization. (2004). Standards for Health Promotion in Hospitals. Retrieved from http://apps.who.int/iris/bitstream/10665/107549/1/e82490.pdf

World Health Organization. (2013). Health 2020: A European Policy Framework and Strategy for the 21st Century. Retrieved from http://www.euro.who.int/__data/assets/pdf_file/0011/199532/Health2020-Long.pdf?ua=1

Yen, M. (1994). *Patient comfort and its relation to selected process factors: Scaling and model testing. Dissertation*. Ann Arbor, MI: University of Minnesota.

Figure 1. PRISMA Flow Diagram

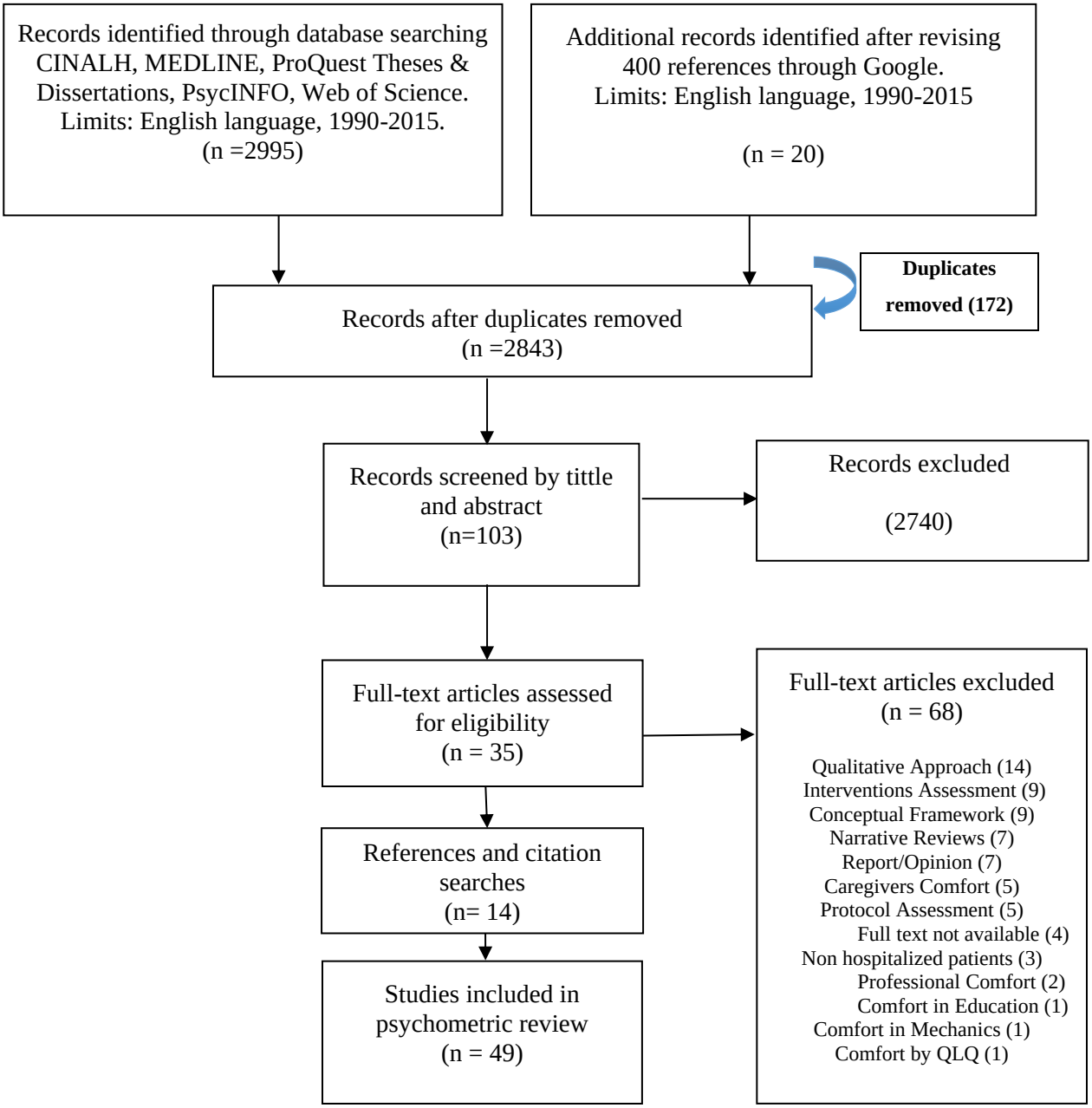


Table 1. Studies overview

Authors	Year	Questionnaire
Comfort of paediatric patients		
Ambuel et al.	1992	Comfort Scale (CS)
Blauer	1996	Comfort Scale (CS)
Bruno de Carvalho et al.	1999	Comfort Scale (CS)
Van Djik et al.	2000	Comfort Scale (CS)
Courtman et al.	2003	Comfort Scale (CS)
Wielenga et al.	2004	Comfort Scale (CS)
Ista et al.	2005	Comfort Scale (CS)
Triltsch et al.	2005	Comfort Scale (CS)
Bear & Ward-Smith	2006	Comfort Scale (CS)
Froom et al.	2008	Comfort Scale (CS)
Lamas et al.	2008	Comfort Scale (CS)
Franck et al.	2011	Comfort Scale (CS)
Cury et al.	2013	Comfort Scale (CS)
Nievas et al.	2014	Comfort Scale (CS)
Tschiedel et al.	2015	Comfort Scale (CS)
Lee & Young	2005	Modified Comfort Scale
Caljouw et al.	2007	Adapted Comfort Scale
Gjerstad et al.	2008	Modified Comfort Scale
Van Djik et al.	2009	Confort Scale Neo
Johansson & Kokinsky	2009	Confort Behavioural Scale (CBS)
de Jong et al.	2010	Confort Behavioural Scale (CBS)
Amigoni et al.	2012	Confort Behavioural Scale (CBS)
Bai et al.	2012	Confort Behavioural Scale (CBS)
Boerlage et al.	2012	Confort Behavioural Scale (CBS)
de Jong et al.	2012	Confort Behavioural Scale (CBS)
Valkenburg	2012	Confort Behavioural Scale (CBS)
da Costa Silva et al.	2013	Confort Behavioural Scale (CBS)
Tristão et al.	2013	Confort Behavioural Scale (CBS)
Boerlage et al.	2014	Confort Behavioural Scale (CBS)
Andersen et al.	2015	Confort Behavioural Scale (CBS)
Moriber	2009	Pediatric PerioperativeComfort Instrument (PPCI)
Comfort of adult patients		
Jensen et al.	1991	Pain Discomfort Scale (PDS)
Kolcaba	1992	General Comfort Questionnaire (GCQ)
Yen	1994	Psychosocial Comfort
Betemps	1999	Psychiatric Discomfort Scale (PDS)
Durnell et al.	2003	Childbirth Comfort Questionnaire (CCQ)
Alves-Apóstolo et al.	2007	Psychiatric In-Patients Comfort Scale (PICS)
Kalfon et al.	2010	Pediatric Perioperative Comfort Instrument (IPREA)
Ashkenazy& DeKeyser-Ganz	2011	Comfort Scale (CS)
Comfort of elderly patients with dementia		
Hurley et al.	1992	Discomfort Scale-Dementia of Alzheimer Type (DS-DAT)
Morrison et al.	1998	Discomfort Scale (DS)
Novak et al.	2001	Hospice Comfort Questionnaire (HCQ)
Volicer et al.	2001	Comfort Assessment Dying with Dementia (CAD-EOLD)
Kiely et al.	2006	Comfort Assessment Dying with Dementia (CAD-EOLD)
Stevenson et al.	2006	Discomfort Behaviour Scale (DBS)
Dello Russo et al.	2008	Discomfort Scale-Dementia of Alzheimer Type in Italian (DSDAT)
Tanatwanit et al.	2011	Hospice Comfort Questionnaire in Thai (HQC)
Cohen- Mansfield et al.	2013	Source of Discomfort Scale (SODS)
Oliveira et al.	2016	End of Life Comfort Planning Questionnaire

Note: For further details and instruments characteristics, see the Supplementary File 6.

Table 2. Methodological quality

Author	Year	Questionn.	Internal consistency		Reproducibility			Content validity		Structural validity	Construct validity		Cross-cultural	Resp.	Floor-Ceiling		Interp
					Ag.	Reliability											
Comfort of paediatric patients																	
Ambuel	1992	CS	**	+	0	**	?	****	?	***	**	?			0	+	?
Blauer	1996	CS		0	0		0		0			0		**	?	0	+
de Carvalho	1999	CS		0	0		0		0		*	?			0	0	+
Van Djik	2000	CS	****	+	0	***	+		0	****	**	?			0	0	+
Courtman	2003	CS		0	0		0		0		**	?			0	0	+
Wielenga	2004	CS		0	0	**	+		0		**	?			0	0	0
Ista	2005	CS	***	+	0	***	+		0		**	?			0	0	+
Triltsch	2005	CS		0	0		0		0		**	?			0	0	+
Bear	2006	CS	***	+	0	**	?		0			0			0	0	?
Froom	2008	CS		0	0		0		0		*	?			0	0	+
Lamas	2008	CS		0	0		0		0		***	+			0	0	0
Franck	2011	CS		0	0	***	+		0	***		0		**	?	0	+
Cury	2013	CS		0	0		0		0		*	?		*	?	0	+
Nievas	2014	CS		0	0		0		0		*	?			0	0	?
Tschiedel	2015	CS		0	0		0		0		*	?			0	0	+
Lee	2005	Modified CS	*	+	0	*	?		0		**	0			0	0	?
Caljouw	2007	Adapted CS	***	+	0	***	+		0		***	+		***	+	0	?
Gjerstad	2008	Modified CS		0	0		0		0		*	?			0	0	+
Van Djik	2009	CNeo	***	+	0	***	+		0		**	?		**	?	0	+
Johansson	2009	CBS		0	0	**	+		0		**	?			0	0	+
de Jong	2010	CBS	***	+	0	**	+		0		**	?		**	+	0	?
Amigoni	2012	CBS		0	0	**	-		0		**	?		**	?	0	+
Bai	2012	CBS		0	0		0		0		**	?			0	0	?
Boerlage	2012	CBS		0	0	**	?		0			0			0	0	0
de Jong	2012	CBS		0	0		0		0		**	?			0	0	?
Valkenburg	2012	CBS	***	+	0		0		0	***	**	?			0	+	?
da Costa	2013	CBS		0	0	*	+		0		*	?			0	0	?
Tristão	2013	CBS		0	0		0		0		**	+		**	+	0	?
Boerlage	2014	CBS		0	0	**	+		0			0		****	+	0	+
Andersen	2015	CBS		0	0	**	+		0		**	+			0	-	+
Moriber	2009	PPCI	***	+	0		0	****	+		***	+			0	0	+
Comfort of adult patients																	
Jensen	1991	PDS	***	+	0	**	?	****	+	***	**	?			0	+	?
Kolcaba	1992	GCQ	***	+	0		0	****	?	***		0			0	0	0
Yen	1994	Psychosocial Comfort	***	+	0		0	****	+	***		0			0	0	+
Betemps	1999	PDS	***	+	0		0	***	?	***	**	?		**	?	0	0
Durnell	2003	CCQ	***	+	0		0	****	+	***	***	+			?	+	+
Alves-Apóstolo	2007	PICS	***	+	0		0	****	+	***	**	?			0	0	0
Kalfon	2010	IPREA	***	+	0	***	+	****	+	***		0			0	-	+
Ashkenazy	2011	CS	***	-	0	***	-		0		**	?		*	?	0	+

Table 2. Methodological quality. Continue.

Author	Year	Questionn.	Internal consistency		Reproducibility			Content validity		Structural validity	Construct validity		Cross-cultural	Resp.		Floor-Ceiling	Interp
					Ag.	Reliability											
Comfort of elderly patients with dementia																	
Hurley	1992	DS-DAT	**	?	0	**	?	****	+			0		**	?	0	?
Morrison	1998	Discomfort Scale		0	0	**	?	****	+		**	?			0	0	+
Novak	2001	HCQ	**	+	0	***	+	****	+		**	?			0	0	+
Volicer	2001	CAD-EOLD	***	+	0		0	****	+	***	**	?			0	0	?
Kiely	2006	CAD-EOLD	***	+	0		0		0		**	?			0	0	?
Stevenson	2006	DBS	***	+	0		0	****	+	***		0			0	0	+
Dello Russo	2008	DS-DAT	***	+	0	***	+	****	+			0	*		0	0	+
Tanatwanit	2011	HQC	***	+	0		0	****	+		**	?	*		0	+	+
Cohen-Mansfield	2013	SODS		0	0	**	?	****	+		***	+			0	0	?
Oliveira	2016	End of Life	***	+	0		0	****	+	**	**	?	*		0	0	0

Note: COSMIN Ratings: **** Excellent; ***Good; ** Fair; * Poor. Quality Criteria for Measurement Properties Ratings: **+** A positive rating indicates strong properties according to quality criteria using design and method;? An intermediate rating indicates some but not all aspects of psychometric are positive, or doubtful design or method: **-** A negative rating indicates psychometric properties do not meet criteria despite adequate design and method; **0** No information provided. Abbreviations: Questionn. = Questionnaire; Ag.= Agreement; Resp. =Responsiveness; Interp.= Interpretability.

Table 3. Utility matrix

Author	Year	Question.	Cost efficiency				Acceptability			Educational Impact		
			Cost Ef1	Cost Ef2	Cost Ef3	Cost Ef4	Accept1	Accept2	Accept3	Ed Imp1	Ed Imp2	Ed Imp3
Comfort of paediatric patients												
Ambuel	1992	CS	***	****	**	***	na	nr	****	****	****	****
Blauer	1996	CS	***	****	**	***	na	nr	****	****	****	****
Carvalho	1999	CS	****	****	**	****	na	nr	****	****	****	***
Van Djik	2000	CS	*	****	**	*	na	****	****	****	****	****
Courtman	2003	CS	***	****	**	***	na	nr	****	****	****	****
Wielenga	2004	CS	****	****	**	****	na	nr	****	****	****	****
Ista	2005	CS	**	****	**	**	na	nr	****	****	****	****
Triltsch	2005	CS	***	****	**	***	na	nr	****	****	****	****
Bear	2006	CS	**	****	**	**	na	nr	****	****	****	****
Froom	2008	CS	****	****	**	****	na	nr	****	****	****	****
Lamas	2008	CS	**	****	**	**	na	nr	****	****	****	****
Franck	2011	CS	**	****	**	**	na	nr	****	****	****	***
Cury	2013	CS	****	****	**	****	na	nr	****	****	****	****
Nievas	2014	CS	****	****	**	****	na	nr	****	****	****	***
Tschiedel	2015	CS	***	****	**	***	na	nr	****	****	****	****
Lee	2005	Modified CS	****	****	**	****	na	nr	****	****	****	****
Caljouw	2007	Adapted CS	**	****	**	**	na	nr	****	****	****	****
Gjerstad	2008	Modified CS	****	****	**	****	na	nr	****	****	****	****
Van Djik	2009	CNeo	*	****	**	*	na	nr	****	****	****	****
Johansson	2009	CBS	***	****	**	***	na	nr	****	****	****	***
de Jong	2010	CBS	*	****	**	*	na	nr	****	****	****	****
Amigoni	2012	CBS	***	****	**	***	na	nr	****	****	****	****
Bai	2012	CBS	*	****	**	*	na	nr	****	****	****	***
Boerlage	2012	CBS	**	****	**	**	na	nr	****	****	****	***
de Jong	2012	CBS	*	****	**	*	na	nr	****	****	****	****
Valkenburg	2012	CBS	**	****	**	**	na	nr	****	****	****	****
da Costa	2013	CBS	***	****	**	***	na	nr	****	****	****	****
Tristão	2013	CBS	***	****	**	***	na	nr	****	****	****	****
Boerlage	2014	CBS	*	****	**	*	na	***	****	****	****	****
Andersen	2015	CBS	***	****	**	***	na	nr	****	****	****	****
Moriber	2009	PPCI	*	nr	***	*	na	nr	****	****	****	****
Comfort of adult patients												
Jensen	1991	PDS	**	****	***	**	**	**	****	****	****	****
Kolcaba	1992	GCQ	*	nr	***	*	*	nr	****	****	****	****
Yen	1994	Psychosocial Comfort	*	nr	***	*	*	nr	****	****	****	****
Betemps	1999	PDS	*	nr	***	*	**	nr	****	****	****	***
Durnell	2003	CCQ	**	****	***	**	**	nr	****	****	****	****
Alves Apóstolo	2007	PICS	*	nr	***	*	****	nr	****	****	****	****
Ashkenazy	2011	CS	**	****	**	**	na	nr	****	****	****	***
Comfort of elderly patients with dementia												
Hurley	1992	DS-DAT	**	****	**	**	na	nr	****	****	****	****
Morrison	1998	Discomfort Scale	*	nr	***	*	**	nr	****	****	****	****
Novak	2001	HCQ	*	****	***	*	**	***	****	****	****	****
Volicer	2001	CAD-EOLD	**	nr	***	**	**	nr	****	****	****	****
Kiely	2006	CAD-EOLD	*	nr	***	*	**	nr	****	****	****	****
Stevenson	2006	DBS	*	nr	**	*	na	****	****	****	****	****
Dello Russo	2008	DS-DAT	**	****	**	**	na	nr	****	****	****	***
Tanatwanit	2011	HQC	*	**	***	*	**	nr	****	****	****	****
Cohen-Mansfield	2013	SODS	*	nr	**	*	na	nr	****	****	****	****
Oliveira	2016	End of Life	*	****	***	*	****	nr	****	****	****	****

Note: Ratings: * poor, **fair, ***good, ****excellent, **nr** not reported, **na** no applicable. Utility aspects: CostEf1=number of observations needed to reach the required level of reliability; CostEf2= time to complete the questionnaire; CostEf3=administrative costs; CostEf4=cost to complete a reliable sample;Accept1=subjects understanding of the instrument; Accept2=assessments not completed; Accept3=instrument tested in an appropriate context; EdImp1=evidence of instrument purpose is achieved;EdImp2=scoring system stated and/or available in an easy format; EdImp3=feedback from the results can be used for action. Abbreviations: Questionn. = Questionnaire; CostEf= Cost Efficiency; Accept=Acceptability; EdImp= Educational Impact.

3.3. TOOLS TO ASSESS THE MEASUREMENT PROPERTIES OF QUALITY OF LIFE INSTRUMENTS: A META-REVIEW PROTOCOL

TOOLS TO ASSESS THE MEASUREMENT PROPERTIES OF QUALITY OF LIFE INSTRUMENTS: A META-REVIEW PROTOCOL

Sonia LORENTE, Jaume VIVES, Carme VILADRICH, Josep-Maria LOSILLA,

Universitat Autònoma de Barcelona

Sonia LORENTE, Adjunct lecturer, PhD Student, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB; Paediatric nurse, PNP, Hospital de Terrassa, Consorci Sanitari de Terrassa, Sonia.Lorente@uab.cat; Jaume VIVES, PhD, Associate Professor, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB, Jaume.Vives@uab.cat; Carme VILADRICH, PhD, Associate Professor, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB, Carme.Viladrich@uab.cat; Josep-Maria LOSILLA, PhD, Associate Professor, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB, JosepMaria.Losilla@uab.cat.

The authors declare no conflict of interest. Protocol supported with the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness. Correspondence concerning this article should be addressed to Sonia Lorente, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB. E-mail: sonia.lorente@uab.cat

ABSTRACT

Introduction: Using specific tools to assess the measurement properties of health status instruments is recommended both to standardize the review process and to improve the methodological quality of systematic reviews. However, depending on the measurement standards upon which these tools are developed, the approach to appraise the measurement properties of instruments may vary. For this reason, the present meta-review aims to: 1) identify systematic reviews assessing the measurement properties of instruments evaluating Health-Related Quality of Life (HRQoL); 2) identify the tools applied to assess the measurement properties of HRQoL instruments; 3) describe the characteristics of the tools applied to assess the measurement properties of HRQoL instruments; 4) identify the measurement standards upon which these tools were developed or conform to, and 5) compare the similarities and differences among the identified measurement standards. **Methods and analysis**: A systematic review will be conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols Guidelines (PRISMA-P). Electronic search will be carried out on bibliographic databases, including PubMed, CINAHL, PsycINFO, SCOPUS, WOS, COSMIN database, and ProQuest Dissertations & Theses Global, being limited by time (2008-2018) and language (English). Descriptive analyses of different aspects of tools applied to evaluate the measurement properties of HRQoL instruments will be presented; the different measurement standards will be described, and some recommendations about the methodological and research applications will be made. **Ethics and dissemination**: Ethical approval is not necessary for systematic review protocols. The results will be disseminated by its publication in a peer-reviewed journal and presented at a relevant conference. Registration: PROSPERO, CRD42017065232. **Key words**: Meta-review, Quality of life, Health instruments, Measurement properties, Measurement standards.

STRENGTHS AND LIMITATIONS

- The search strategy has been designed to be comprehensive, following the Peer Review of Electronic Search Strategies (PRESS) guidelines and including filters for finding studies on measurement properties of measurement instruments.
- The systematic review protocol is developed using the Preferred Reporting Items for Systematic Reviews and Meta-analyses for Protocols guidelines (PRISMA-P).
- Inclusion of studies published in English only may lead to language bias.

INTRODUCTION

Systematic reviews of measurement properties critically appraise and compare the content and measurement properties of all instruments measuring a certain construct of interest in a specific study population (Mokkink et al., 2009). High quality systematic reviews can provide a comprehensive overview of the measurement properties of patient-reported outcome measures (PROM), and support evidence-based recommendations in the selection of the most suitable health status instrument for a given purpose (i.e. research or clinical practice) (Prinsen et al., 2018). To be confident that the design, conduct, analysis, and interpretation of the review results and conclusions are adequate, the methodological quality of systematic reviews should be appraised (Mokkink et al., 2009).

Because of this, different authors evaluate systematic reviews assessing the measurement properties of health status assessment instruments, as Mokkink et al. (Mokkink et al., 2009) or Terwee et al. (Terwee et al., 2016). In both cases, authors examine the search strategy, data extraction (two or more reviewers), data synthesis, and whether the measurement properties of health status instruments were assessed using specific tools that are recommended both to standardize the review process and to improve the methodological quality of systematic reviews (Terwee et al., 2016). However, depending on the measurement standards upon these tools were developed, the approach to analyse the measurement properties of instruments may vary. Given this, the present meta-review aims to discuss the methodological, research and practical applications of these tools in systematic reviews that assess the measurement properties of instruments evaluating the quality of life within the context of health and disease, i.e., Health-Related Quality of Life (HRQoL) instruments (Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives, 2010).

METHODS

Objectives

1. To identify systematic reviews assessing the measurement properties of Health-Related Quality of Life (HRQoL) instruments.
2. To identify the main tools applied to assess the measurement properties of HRQoL instruments.
3. To describe the most relevant characteristics of the tools applied to assess the measurement properties of HRQoL instruments (validity, reliability, feasibility, etc.).
4. To identify the measurement standards upon which these tools were developed or conform to.
5. To compare the similarities and differences among the identified measurement standards.

Study design

Where applicable, the present meta-review will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols Guidelines (PRISMA-P) (Moher et al., 2015). This meta-review was registered in PROSPERO, CRD42017065232, available from: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017065232.

Search strategy

A systematic review will be performed in PubMed, US National Library of Medicine, by National Center for Biotechnology Information (NCBI); CINAHL, Cumulative Index to Nursing and Allied Health Literature, by EBSCOhost; PsycINFO, Psychological Information, by APA PsycNET; SCOPUS by Elsevier; WOS (Web of Science CORE) by Thomson Reuters, and COSMIN database by COSMIN Initiative (http://www.cosmin.nl/). In addition, ProQuest Dissertations & Theses Global will be used for searching grey literature, and search alerts in all databases will be set. The search strategy will follow the Peer Review of Electronic Search Strategies (PRESS) guidelines recommendations (McGowan et al., 2016a, 2016b), and will consist of 3 filters composed of search terms for the following: (1) systematic review methodology; (2) HRQoL instruments; and (3) measurement properties. The latter filter was developed for the VU University Medical Center for finding studies on measurement properties of measurement instruments (Terwee et al., 2009). All filters will be adapted for all databases. The systematic search will be performed in July 2018, limited by time and language (English) (Table 1 shows the string of terms in PubMed).

Inclusion criteria

Time frame

We will limit our search to studies published between 2008-2018.

Study design

Systematic reviews aiming to report or to assess the measurement properties of instruments evaluating the quality of life within the context of health and disease, namely Health Related Quality of Life (HRQoL) instruments (Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives, 2010), including all studies examining at least two or more measurement properties of a HRQoL instrument. Systematic reviews were required to include the full results report and detailed information about the instruments used to assess the measurement properties.

Setting and Participants

We will include the whole range of ages (new-borns, toddlers, children, teenagers, young adults, middle-age adults and elderly people), in any healthcare setting.

Condition or domain being studied

The quality of health status and the quality of life instruments is essential to obtain accurate diagnoses and to assess the efficacy or effectiveness of a specific intervention in health care. Evaluating and improving the quality of life, as well, is considered a public health priority (Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives, 2010), and because of this the present meta-review is focused on systematic reviews that appraised the measurement properties of HRQoL instruments.

Context

To study the characteristics of tools assessing the measurement properties of HRQoL instruments in systematic reviews, and to compare the measurement standards upon which these tools were developed or conform to, with examples found in Viladrich et al. (Viladrich et al., 2015): Attributes and Criteria to assess Health Status and Quality of Life Instruments (Aaronson et al., 2002; Mokkink et al., 2018), the Standards for Educational and Psychological Measurement (American Educational Research Association et al., 2014; APA et al., 1999), or the Health Status Measures in Economic Evaluation (Brazier et al. 1999; 2017).

Primary Outcomes

Identification of the main specific tools applied to assess the measurement properties of HR-QoL instruments and comparison of their most relevant characteristics. Identification and comparison of the measurement standards upon which these tools were developed. Appraisal of how authors of the systematic reviews include the assessment of the quality of the HR-QoL instruments in their results and how they use this evaluation to come to an overall conclusion regarding the quality of each instrument.

Instruments

We will include tools aiming to assess the quality of measurement properties of HRQoL instruments.

Study screening

References identified by the search strategy will be entered into Mendeley bibliographic software, and duplicates will be removed. Titles and abstracts will be screened independently by two reviewers. When decisions are unable to be made from title and abstract alone, the full paper will be retrieved.

Full-text inclusion criteria will be checked independently by two reviewers. Discrepancies during the process will be resolved through discussion (with a third reviewer where necessary).

Data extraction

Extracted information of each selected systematic review and meta-analysis will include: general information (author, year, country of origin and papers, theoretical/conceptual framework); tools applied to assess the measurement properties of HRQoL instruments (title, purpose/use, number of items, response categories, criteria to assess the measurement properties upon specific measurement standard, ease and usefulness of interpretation, level of expertise required for scoring and interpreting, and time required to completion); reporting of the measurement properties assessed; use of the results from the evaluation of the measurement properties to come to an overall conclusion regarding the quality of each HRQoL instruments. Authors of eligible studies will be contacted to provide missing or additional data if necessary.

Strategy for data analysis

We will initially categorise the tools applied to assess the measurement properties of the HRQoL instruments according to the measurement standards upon they were developed or conform to. Next, we will detail the most relevant characteristics of these tools according to their measurement standards and their conceptual frameworks.

Strategy for data synthesis

Descriptive analyses of different aspects of the identified tools applied to evaluate the measurement properties of HRQoL instruments. The extracted information related to these tools will be reported in a table to facilitate their comparison. Some recommendations about the methodological, practical and the research applications of each tool will be made.

Patient and Public Involvement

No patient or public involvement.

Ethics and dissemination

Ethical approval is not necessary for systematic review protocols. The results will be disseminated by its publication in a peer-reviewed journal and presented at a relevant conference.

DISCUSSION

To date, there are not meta-reviews of tools assessing the measurement properties of HRQoL instruments and the different measurement standards upon which these tools were developed. The findings of this work will be useful, firstly, to compare the minimum criteria and attributes recommended to assess the measurement properties of HRQoL; secondly, to establish the most relevant differences and similarities among both the measurement standards and the assessment tools of measurement properties; and finally, to discuss the methodological, research and practical applications of these tools in systematic reviews. This information will facilitate and improve the work of researchers and clinicians that conduct systematic reviews of HRQoL instruments measurement properties.

Authors' contribution

All authors meet the criteria recommended by the International Committee of Medical Journal Editors, ICMJE. All authors made substantial contributions to conception and design, piloted the inclusion criteria and provided direction of the data extraction and analysis. SL draft the article, and JV, CV and JML critically revised the draft for important intellectual content. All authors agreed on the final version.

Funding statement

This work was supported by the Grant PSI2014-52962-P, Spanish Ministry of Economy and Competitiveness.

Competing interests' statement.

The authors declare no conflict of interest. The authors have no financial relationships relevant to this article to disclose, and the funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Correspondence

Correspondence concerning this article should be addressed to Sonia Lorente, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB. E-mail: sonia.lorente@uab.cat

REFERENCES

1. Mokkink L, Terwee C, Stratford P, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res*. 2009;18(3):313-333. doi:10.1007/s11136-009-9451-9.
2. Prinsen C, Mokkink L, Bouter L, et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018;0(0):1-11. doi:10.1007/s11136-018-1798-3.
3. Terwee C, Prinsen C, Ricci Garotti M, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25:767-779. doi:10.1007/s11136-015-1122-4.
4. Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives. *Health-Related Quality of Life and Well-Being*.; 2010. https://www.healthypeople.gov/sites/default/files/HRQoLWBFullReport.pdf.
5. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4(1):1-9. doi:10.1186/2046-4053-4-1.
6. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). *Cadth Methods Guidel*. 2016;(January):40-46. doi:10.1016/j.jclinepi.2016.01.021.
7. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40-46. doi:10.1016/j.jclinepi.2016.01.021.
8. Terwee CB, Jansma EP, Riphagen II, De Vet H. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123. doi:10.1007/s11136-009-9528-5.
9. Viladrich C, Doval E. *Measurement. Reliability and Validity.[Medición. Fiabilidad y Validez]*. Bellaterra: Laboratori d´ Estadística Aplicada i de Modelització (UAB); 2006.
10. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments and review criteria. *Qual Life Res*. 2002;11(3):193-215.
11. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN Methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). User manual. 2018:1-78. http://www.cosmin.nl/.

12. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing.* American Educational Research Association.; 1999.
13. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
14. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess (Rockv)*. 1999;3(9).
15. Brazier J, Ara R, Rowen D, Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. *Pharmacoeconomics*. 2017;35(s1):21-31. doi:10.1007/s40273-017-0545-x.

Table 1. Search string for Pubmed

1	("Quality of Life"[Mesh] OR HRQL [tiab] OR HRQoL[tiab] OR QoL[tiab] OR "quality of life"[tiab]
2	(instrument[tiab] OR instruments[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR scale[tiab] OR scales[tiab] OR tool[tiab] OR tools[tiab]
3	(Validation Studies[pt] OR "reproducibility of results"[MeSH Terms] OR reproducib*[tiab] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tiab] OR clinometr*[tiab] OR "observer variation"[MeSH] OR observer variation[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR "item correlation"[tiab] OR "item correlations"[tiab] OR "item selection"[tiab] OR "item selections"[tiab] OR "item reduction"[tiab] OR "item reductions"[tiab] OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest [tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR "kappa's"[tiab] OR kappas[tiab] OR "coefficient of variation"[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structure"[tiab] OR dimensionality[tiab] OR subscale*[tiab] OR "multitrait scaling analysis"[tiab] OR "multitrait scaling analyses"[tiab] OR "item discriminant"[tiab] OR "interscale correlation"[tiab] OR "interscale correlations"[tiab] OR ((error[tiab] OR errors[tiab]) AND (measure*[tiab] OR correlat*[tiab] OR evaluat*[tiab] OR accuracy[tiab] OR accurate[tiab] OR precision[tiab] OR mean[tiab])) OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR "variability analysis"[tiab] OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "minimal important change"[tiab] OR "minimal important difference"[tiab] OR "minimally important change"[tiab] OR "minimally important difference"[tiab] OR "minimal detectable change"[tiab] OR "minimal detectable difference"[tiab] OR "minimally detectable change"[tiab] OR "minimally detectable difference"[tiab] OR "minimal real change"[tiab] OR "minimal real difference"[tiab] OR "minimally real change"[tiab] OR "minimally real difference"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF [tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]
4	#1 AND #2 AND #3
5	("protocol"[ti] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference"[Publication Type] OR "practice guideline"[Publication Type])
6	#4 NOT #5
7	FILTER: Article Type (Review or Systematic Review)
8	FILTER: Subject (Systematic Review)
9	FILTER: Language (English)
10	FILTER: Period (2008-2018)

3.4. TOOLS TO ASSESS THE MEASUREMENT PROPERTIES OF QUALITY OF LIFE INSTRUMENTS: A META-REVIEW

TOOLS TO ASSESS THE MEASUREMENT PROPERTIES OF QUALITY OF LIFE INSTRUMENTS: A META-REVIEW

Sonia LORENTE, Carme VILADRICH, Jaume VIVES, Josep-Maria LOSILLA
Universitat Autònoma de Barcelona

Sonia LORENTE, PhD Student, Adjunct Lecturer, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB; Paediatric nurse, PNP, Hospital de Terrassa, Consorci Sanitari de Terrassa, Sonia.Lorente@uab.cat; Carme VILADRICH, PhD, Associate Professor, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB, Carme.Viladrich@uab.cat; Jaume VIVES, PhD, Associate Professor, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB, Jaume.Vives@uab.cat; Josep-Maria LOSILLA, PhD, Associate Professor, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB, JosepMaria.Losilla@uab.cat

The authors declare no conflict of interest. This work is supported by the Grant PGC2018-100675-B-I00, Spanish Ministry of Science, Innovation and Universities (Spain). Correspondence concerning this article should be addressed to Jaume Vives, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB. E-mail: Jaume.Vives@uab.cat

ABSTRACT

Objective: This meta-review aims to discuss the methodological, research and practical applications of tools that assess the measurement properties of instruments evaluating Health-Related Quality of Life (HRQoL) that have been reported in systematic reviews. **Design:** Meta-review. **Methods:** Electronic search from January 2008 to May 2020 was carried out on PubMed, CINAHL, PsycINFO, SCOPUS, WoS, Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) database, Google Scholar, and ProQuest Dissertations & Theses. **Results:** A total of 246 systematic reviews were assessed. Concerning the quality of the review process, some methodological shortcomings were found, such as poor compliance with reporting or methodological guidelines. Regarding the procedures to assess the quality of measurement properties, 164 (66.6 %) of reviewers applied one tool at least. Tool format and structure differed across standards or scientific traditions (i.e. psychology, medicine and economics), but most assess both measurement properties and the usability of instruments. As far as the results and conclusions of systematic reviews are concerned, only 68 (27.5 %) linked the intended use of the instrument to specific measurement properties (e.g. evaluative use to responsiveness). **Conclusions:** The reporting and methodological quality of reviews have increased over time, but there is still room for improvement regarding adherence to guidelines. The COSMIN would be the most widespread and comprehensive tool to assess both the risk of bias of primary studies, and the measurement properties of HRQoL instruments for evaluative purposes. Our analysis of other assessment tools and measurement standards can serve as a starting point for future lines of work on the COSMIN tool, such as considering a more comprehensive evaluation of feasibility, including burden and fairness; expanding its scope for measurement instruments with a different use than evaluative; and improving its assessment of the risk of bias of primary studies. **PROSPERO number:** CRD42017065232. **Key words**: Meta-review, Quality of life, Health instruments, Measurement properties, Measurement standards, HRQoL.

STRENGTHS AND LIMITATIONS

- The search strategy has been designed to be comprehensive, following the Peer Review of Electronic Search Strategies (PRESS) guidelines including specific filters for finding studies on psychometric properties of measurement instruments.
- A total of 246 systematic reviews were included and, to our knowledge, this meta-review provides the broadest overview of the most common tools used to assess measurement properties of HRQoL instruments and their relationship with measurement standards, scientific traditions and the intended use of the measures.
- Some of the included systematic reviews poorly reported the review process, outcomes, and conclusions, and this fact may have led to the loss of some data.
- Inclusion of studies published in English only may have led to language bias.

INTRODUCTION

The systematic reviews of measurement properties critically appraise the content and measurement properties of all instruments that assess a certain construct of interest in a specific study population[1]. These systematic reviews provide both a comprehensive overview of the measurement properties of health instruments and supportive evidence for the selection of instruments for a specific purpose (e.g. research, clinical practice, predictive)[2,3]. In this type of systematic review, different authors have evaluated not only the methodological quality of their key phases, -namely the search strategy, the bias risk assessment of the primary studies and the data synthesis- but also whether the measurement properties of the health status instruments have been appraised with standardized procedures or tools during the data extraction phase[1,2,4,5]. However, depending on the measurement standards upon which these tools were developed, the approach to analyse the measurement properties of instruments may vary.[6] This could lead to different conclusions and recommendations, in spite of the effort undertaken by the international Society for Quality of Life Research to set consensus-based minimum standards[7]. Besides, according to Rosenkoetter and Tate[6], the assessment tools commonly used by clinicians and researchers to select the appropriate outcome measures for specific purposes show a variety of forms and cover a mix of standards related to reporting, methodological quality and statistical outcome quality.

The aims of this present meta-review are to: 1) identify systematic reviews assessing the measurement properties of HRQoL instruments; 2) identify the main tools applied to assess their measurement properties; 3) describe the contents of the applied tools (validity, reliability, feasibility, etc.); 4) identify the measurement standards upon which these tools were developed or conform to, comparing their similarities and differences, and 5) appraise how authors of these systematic reviews include the assessment of the measurement quality in their results and conclusions, i.e., to what extent conclusions depend on the results of the evaluation of the measurement properties, as well as their relationship, if any, with the intended use of the HRQoL instrument (e.g. evaluative).

METHODS

The protocol of this review[8] was prospectively registered. We conducted this meta-review following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Guidelines (PRISMA)[9,10].

Search strategy

A systematic search was performed in PubMed, US National Library of Medicine, by National Center for Biotechnology Information (NCBI); CINAHL, Cumulative Index to Nursing and Allied Health Literature, by EBSCOhost; PsycINFO, Psychological Information, by APA PsycNET; SCOPUS by Elsevier; WoS ,Web of Science CORE, by Thomson Reuters; Consensus-based Standards for the selection of Health Measurement Instruments database, by COSMIN Initiative (http://www.cosmin.nl/); and Google Scholar (up to 400 links). ProQuest Dissertations & Theses Global was used for searching grey literature, and search alerts in all databases were set. The search strategy followed the Peer Review of Electronic Search Strategies (PRESS) guidelines recommendations[11,12], and consisted of 3 filters composed of search terms for the following: (1) systematic review methodology; (2) HRQoL instruments; and (3) measurement properties. The latter filter was developed by the Vrije University Medical Center for finding studies on measurement properties of measurement instruments[13]. All filters were adapted for all databases. The searches were completed in May 2020. Restrictions by language (English) and publication date (from January 2008) were applied. (See Supplementary File S1 for search strings for all databases).

Inclusion criteria

Systematic reviews specifically aiming to report or to assess the measurement properties of instruments evaluating the quality of life within the context of health and disease[14] were included. Systematic reviews were required to include the full results report, and detailed information about the procedures used to assess the measurement properties.

Exclusion criteria

Systematic reviews exclusively focused on evaluating clinical interventions were excluded. Systematic reviews specifically focused on assessing Patient-reported outcomes measures (PROMs) other than HRQoL for specific diseases, clinical conditions or populations, were excluded. Systematic reviews that did not report full information about the procedures to assess the measurement properties were also excluded (e.g. conference abstracts).

Study screening

References identified by the search strategy were entered to Mendeley reference management software, and duplicates were removed. Titles and abstracts were screened independently by two reviewers (SL and JV). When decisions were unable to be made from title and abstract alone, the full paper was retrieved. Full-text inclusion criteria were checked independently by two reviewers (SL and JV). Discrepancies during the process were resolved through discussion (with independent reviews of JML and CV when necessary).

Data extraction

Extracted information of each selected systematic review and meta-analysis included general information such as author, year, and quality of review process of systematic reviews (e.g. protocol registration, reporting guidelines, and use of flowchart). Information concerning the main identified tools applied to assess the measurement properties of HRQoL instruments included the title, intended use, number of items, response categories, instrument assessment criteria, and measurement properties assessed. Information on how authors included the assessment of the quality of HRQoL in their results and conclusions was also extracted. Authors of eligible studies were contacted to provide missing or additional data when necessary.

Study aim

To examine the methodological, research and practical applications of the reported tools in systematic reviews that assess the measurement properties of instruments evaluating quality of life within the context of health and disease, i.e. HRQoL.

RESULTS

Search results

Figure 1 shows the results of the search strategy, reported according to the PRISMA flow diagram. A total of 4320 references were identified through database searches. After removing duplicates, 3055 titles and abstracts were screened. After the assessment of 525 full-text documents for eligibility, a total of 246 systematic reviews were included in the qualitative analysis. These systematic reviews covered a wide range of HRQoL instruments, both generic and disease-specific. A total of 24 (9.8 %) of the systematic reviews assessed the quality of one measurement property only, such as the conceptual and measurement model or the content validity (See Supplementary File 2 for characteristics and references of studies).

Reporting and methodological quality of the studies

Table 1 shows the reporting and methodological quality of systematic reviews. Findings showed that 27 (10.9 %) of the reports registered the protocol prospectively, a figure that raised to 20.8 % when considering the reports from 2014 onwards; 78 (31.7 %) followed reporting guidelines such as PRISMA (50.8 % the last six years); 42 (17.0 % since 2008; 23.8 % for the last six years) assessed the reporting and/or the methodological quality of primary studies using recommended guides, such as Standards for the Reporting of Diagnostic Accuracy Studies (STARD) and Quality Assessment of Diagnostic Accuracy Studies (QUADAS), respectively; 238 (96.7 %) reported the search strategy; 116 (47.41%) reported the detailed syntax for one database at least; 134 (54.4 %) made the article selection by two or more independent reviewers;

166 (67.5 %) used a flowchart to report search outcomes, and 132 (53.7 %) stated the funding. These last percentages slightly increased when reducing the time frame to the last six years.

Assessment of measurement properties of HRQoL instruments

Assessment procedures of measurement properties varied considerably. A total of 164 (66.6 %) out of 246 systematic reviews applied one tool at least, that is, a published and well accepted list of criteria, to rate the evidence on measurement properties of instruments; 41 (16.6 %) applied their own author´s criteria only; 30 (12.2 %) followed literature recommendations included in very highly circulated books or papers only, and 14 (5.7 %) used an *ad hoc* checklist of criteria only. A total of 98 (39.8 %) systematic reviews did combine different procedures. Most usual combinations were the use of two tools or one tool and literature recommendations.

Tools to assess measurement properties of HRQoL instruments

The first twelve columns of Table 2 present the characteristics for the identified tools used to assess measurement properties using the last update we are aware of. Tools are reported in order of frequency of use, as pointed out in the last row of the table: 1) "COnsensus-based Standards for the selection of Health Measurement INstruments (COSMIN)", COSMIN initiative[15,16]; 2) "Quality Criteria for Measurement Properties", Terwee et al.[17]; 3) "Attributes and Criteria to assess Health Status and Quality of Life Instruments", Scientific Advisory Committee Medical Outcomes Trust (SACMOT)[18,19]; 4) "Health Status Measures in Economic Evaluation", Brazier et al.[20,21]; 5) "Guidance for Industry Patient-reported Outcomes Measures", Food and Drug Administration (FDA)[22,23]; 6) "Evaluating Patient-based Outcomes Measures for use in clinical trials", Fitzpatrick et al.[24] (also known as Fitzpatrick´s criteria); 7) "International Classification of Functioning" and "International Classification of Functioning for Children and Youth", World Health Organization[25]; 8) "Evaluating Measures of Patient Reported Outcomes (EMPRO)", Spanish Cooperative Investigation Network for Health and Health Service Outcomes Research (IRYSS)[26]; 9) "Spinal Cord Injury Criteria", Spinal Cord Injury Rehabilitation Evidence (SCIRE)[27,28]; 10) "Criteria for Assessing the Tools of Disability Outcomes Research", Andresen[29] (also known as Andresen´s tool); 11) "CanChild Outcomes Measures", CanChild Center for Childhood Disability Research[30]; and 12) "Outcomes Measures in Rheumatology Clinical Trials (OMERACT)", OMERACT initiative[31]. Table 2 also includes a final column showing the characteristics of Testing Standards by American Educational Research Association (AERA), American Psychological Association (APA) and National Council on Measurement in Education (NCME)[32,33] (hereinafter "Testing Standards") initially published in 1954 and regularly updated every decade using consensus based procedures. The Testing Standards are the source of most of the technical vocabulary for measurement properties in HRQoL instruments, therefore they will be used as a reference to compare the twelve

identified tools. In fact, these standards have already been recommended to establish a unified approach to validity and reliability of results derived from psychometric instruments in clinical medicine, research and education[34].

Different methodologies were used to develop the tools. The expert panel consensus and the literature review were the most usual methods, led by Steering Committees or Staff/Working Groups. The format and structure of these tools also vary. Whereas seven of them were itemized to allow the assignment of quality scores, the other six took the form of standards or guidelines. Tools with an itemized structure were the COSMIN, Quality Criteria for Measurement Properties, EMPRO, SCI Criteria, Criteria for Assessing the Tools of Disability Outcomes Research (Andresen´s tool), CanChild Outcomes Measures, and OMERACT.

Among all measurement properties considered in Testing Standards, eleven out of the twelve tools recommended to assess the conceptual and measurement model; content, structural, convergent, discriminant, concurrent and predictive validity; responsiveness or sensitivity to change; and internal consistency, test-retest and inter-rater reliability. However, the approach to analyse these measurement properties varied, with examples found in construct validity, criterion validity and reliability. Depending on the tool, the validity of the construct can be evaluated either by hypothesis confirmation in general (e.g. COSMIN or EMPRO), or by specific hypothesis based on correlations with other measures, i.e. convergent and discriminant validity (e.g. Andresen´s tool). Criterion validity can be assessed either exclusively by calculating the correlation coefficient with a gold standard (e.g. CanChild Outcomes Measures) or by obtaining variously correlation, specificity and sensitivity, or predictive values (e.g. FDA). Reliability can be analysed either by test retest reliability, inter-rater reliability and internal consistency (e.g. FDA), or only by test retest and inter-rater agreement (e.g. Economic evaluation). Despite the Testing Standards recommendations, just one tool includes additional criteria to assess consequential validity (SCI), and four assess fairness (e.g. accessible forms for subjects with vision impairment, or for specific populations) (SACMOT, FDA, SCI and Andresen's tool). None of them includes criteria to assess the validity of response processes. Other HRQoL instrument characteristics, such as feasibility (e.g. cost of obtaining a sample), acceptability (e.g. suitability from the patient perspective), or burden (e.g. the time or effort placed on the administration of the instrument) are assessed instead.

Finally, notice that some concepts have changed their place over time. The clearest case is evidence regarding cross-cultural equivalence, which was treated as an additional characteristic of the instruments in most tools released before 2014 (e.g., EMPRO or SCI), but was considered a proper measurement property in the COSMIN's 2018 update. It is also considered a

measurement property in Testing Standards where it is included as a particular case of differential item functioning when assessing the internal structure of the instruments (See Supplementary File S3 for more details).

Intended uses of instruments and their association to measurement properties

Some of the differences between tools can be attributed to the fact that they are devoted to the evaluation of instruments developed with different intended uses. For instance, COSMIN aims at assessing the quality of instruments for an evaluative purpose whereas the Economic Evaluation tool aims at the assessment of instruments for analytical purposes. Nevertheless, the relation between the intended use of the instruments and the measurement properties assessed is not usually included in the conclusions of the systematic reviews. Table 3 shows the intended use of instruments, based on the framework proposed by McDowell et al.[35], and the association to measurement properties that reviewers established in their conclusions. The instruments were most frequently used for evaluation (178, 72.3 %) and for assessment of impact of disease on HRQoL (138, 55.1 %), either alone or in conjunction. Other purposes were analytic (35, 14.2 %), diagnostic (16, 6.5 %), descriptive (4, 1.6 %), and predictive (2, 0.8 %). A total of 6 (2.4 %) systematic reviews did not report or did not clearly state the intended use of the instruments. As far as the assessment and conclusions is concerned, only 68 (27.6 %) systematic reviews linked the intended use of the instrument to measurement properties. The most common use was evaluative, generally associated to responsiveness, content validity or reliability, for example. When the purpose was the assessment of the impact of disease on HRQoL, the conceptual and measurement model and content validity were usually reported. The analytic purpose involved reporting preference-based valuation (e.g. utility scores) and evidence of agreement, and the diagnostic use was linked to known groups validity and test-retest reliability. To better understand these results, some examples are given. First, the evaluative purpose was associated to responsiveness, we found conclusions such as: “For use in longitudinal studies or clinical practice, where responsiveness is an issue, the Minnesota Living with Heart Failure Questionnaire (MLHFQ) and the Chronic Heart Failure Questionnaire (CFHQ) would be adequate”[36]. Second, the intended use was the assessment of the impact of disease on HRQoL, the usual association was to the measurement model and conclusions resembled this one: “None of the RLS specific QOL measures appears to have been informed by a conceptual model or a conceptual framework. Consequently, none can be considered comprehensive in terms of assessing the full impact of Rest Legs Syndrome on QOL”[37]. Third, an example illustrating general conclusions, i.e. conclusions that did not associate the intended use of the instrument to any specific measurement properties, was as follows: “None of the available instruments fulfils

the psychometric demands of reliability, validity and responsiveness to serve as a primary outcome measure in clinical trials"[38]

DISCUSSION

The present meta-review identified 246 systematic reviews assessing measurement properties of HRQoL instruments in order to analyse the quality of the review process, describe the most used tools to assess measurement properties and examine how reviewers included the assessment of the quality of HRQoL in their conclusions.

Reporting and methodological quality of the studies

Findings showed how the reporting and methodological quality of systematic reviews has increased over time. Most reviewers reported the search strategy, stated the inclusion and exclusion criteria taking the judgement of two or more independent reviewers into account and used a flow chart to report search outcomes. However, some crucial methodological shortcomings were found. Practices such as registration of the protocol, reporting the detailed search syntax for one database at least, adherence to reporting guidelines, and assessing the reporting and the methodological quality of primary studies were quite sparse even in recent years. As Pussegoda et al.[4] suggested, this fact may be related to the percieved time-consuming task of using guidelines or to the lack of information about the most appropiate tool. According to our data, there is still large room for improvement in the assessment of the methodological quality of included studies in order to attend to Terwee et al.'s warning[2] of avoiding the risk of presenting biased results, leading to underestimation or overestimation of the quality of an instrument.

Assessment of measurement properties of HRQoL instruments

Assessment procedures of measurement properties of HRQoL instruments were diverse. Most of the reviewers used at least one tool. Nevertheless, there were reviewers that applied their own criteria, followed literature recommendations or applied different *ad hoc* devised checklists. The use of such diverse procedures is noticeable, even in recent years, when well-accepted tools to assess measurement properties are available.

Our meta-review identified up to twelve tools. Seven of them had an itemized structure, offering a comparable approach to rate the evidence on measurement properties. Length and scoring differed, but also the instrument assessment criteria. Actually, depending on the tool used, the approach to assess properties varied greatly, with potentially serious consequences. The fact that a single measurement property is or isn't required can change the status of quality of the evidence supporting the same measurement instrument. The variety of forms found were in concordance to results from related research, which also highlighted the complexity with regard

to definitions of measurement properties[6]. This complexity is also reflected in the search filter developed by the COSMIN initiative[13]. They recommend using 3 filters that sum up more than 100 search terms in order to get sensible and specific results. In addition, and also depending on the tool used, other characteristics, such as feasibility, acceptability, and burden were assessed. In spite of the diversity, a shared conclusion can be stated as follows: because these instruments are to be used in the daily practice, their usability should be always balanced with other characteristics considered as proper measurement properties[39,40]. For instance, an instrument needs to be long enough to ensure reliability and construct validity, but short enough to ensure the adequate response rate and sample size. Otherwise the instrument intended use and sustainability will be at hazard[39].

The differences between tools and their potentially serious consequences on the assessment of the quality of the primary studies may be better addressed in the light of three considerations: the date of publication, the main scientific tradition involved when developing the tools, and the intended uses of the instruments under assessment. Some differences can be simply explained by the date of publication of the tools. As an example, where older tools require specific forms of validity evidence related to external variables such as convergent and discriminant validity, recent tools incorporate the more general view of hypothesis testing. That is, when developing a new use for an instrument, hypotheses should be made regarding the expected relations with other relevant variables in their nomological network and these hypotheses and no other should be tested[32]. Regarding the scientific traditions, the assessment of outcomes is a constitutive part of the disciplines of Education and Psychology where the Testing Standards come from. In these contexts, participation is taken for granted as assessment practices result in high stakes decisions such as, for instance, certification or personnel selection. The main concern regarding integrity of the instrument purpose is its fakeability, which could distort the decision-making process, and this would explain the interest in response processes in this field[41,42]. By contrast, the main objective in the discipline of Medicine is to provide health care services. Evaluation of subjective views of patients was a late addition related to the inclusion of HRQoL in the accounting of health care outcomes, despite the instruments assessing the patient experience should be acceptable to both patients and clinicians, as Beattie et al. (2014) highlighted[39]. Specifically, in the context of disability research, the administrative and respondent burden requires additional consideration. The administrative burden may include the need for a Sign Language interpreter, and the respondent burden includes the length of the questionnaire, which is especially relevant when using HRQoL instruments with cognitively impaired subjects[29]. Balancing the traditional psychometric criteria, the practicalities of the instruments and patient preferences is a generic recommendation for health research, but becomes a special obligation

for research with people with specific needs[29]. Moreover, devising test accommodations or accessible forms when needed is expected to become a required psychometric criterion in the near future, given that it has already been included under the title "fairness in testing" as a new section next to validity and reliability in the chapter of measurement foundations in the most recent update of Testing Standards[32].

Another criterion is that of Economic evaluation, traditionally embedded in providing quantitative judgments able to be integrated into mathematical models such as those used in calculating quality-adjusted life years (QUALYs) and using preference-based methods to obtain their data. Due to that, some very popular measurement properties such as internal structure based on factor analysis are not relevant and thus not considered in their tools. In this tradition the main concerns regarding the integrity of the instrument purpose is whose values should be considered when determining preferences and how well the preferences of patients and decision makers are likely to conform to the main assumptions of the utility models[20,21].

Intended uses of instruments and their association to measurement properties

In our view, considering in the first place the intended use of the HRQoL instrument would help to reconcile the different requirements included in each tool. Tools for evaluating the measurement quality of instruments should be adapted or extended according to the different intended uses of these instruments, such as evaluative, impact of disease, analytic, diagnostic, descriptive or predictive. Notice that depending on the intended use of the measure, some domains of validity and reliability may be of greater or lesser relevance[6,16]. For instance, an instrument developed to assess longitudinal changes should demonstrate high responsiveness[6], but if used for diagnostic purposes, it should be able to distinguish among individuals or groups[6], i.e. known groups validity. Another example is the internal consistency reliability based on inter-item relationships, that may be not relevant for a preference-based instrument but is relevant for an instrument based on a unidimensional measurement model. However, our data showed that only a few authors established a clear link in their recommendations between the intended use of the measure and the reported evidence of measurement properties. The vast field of HRQoL offered a plethora of instruments but, as most reviewers did not take the intended use of the instrument into account, the overall rating of measurement properties was not consistent and thus the instrument may or may not have been adequate for its intended use. Because the evaluation and improvement of quality of life is considered a public health priority[14], we strongly encourage researchers to assess the quality of measurement properties of HRQoL instruments according to the intended use of the measure. Otherwise, there is a serious risk of biased results, which could lead to underrating the quality and suitability of the instrument.

CONCLUSIONS

The quality of the systematic review process has been increasing over time, but it should still improve with regard to the prospective registration of protocol, and with respect to the adoption of guidelines to improve both the methodological and reporting quality of the reviews. In the specific context of systematic reviews of measurement instruments, enhancing the quality of the process also involves the assessment of measurement properties by using a standardized tool. The selection of the most suitable tool may be addressed according to the coverage of the appraised measurement properties, but also according to other important criteria, such as the intended use of the HRQoL instruments, the format of the tool, and whether it assesses both usability (e.g. feasibility or burden) and accommodation (or accessible forms). First, the assessment methodology should be adapted when necessary, establishing the relation between the intended use of the HRQoL instruments and the measurement properties assessed. Second, to standardize the review process, the tool´s format should be itemized offering a comparable approach to rate the evidence on measurement properties. Those tools that take the form of guidelines, such as the SACMOT or the Economic Evaluation would be considerably upgraded if the structure is reconverted, since the current format only allows description rather than critical appraisal of the quality of an instrument, and furthermore it complicates comparison of results. Lastly, because systematic reviews on measurement properties aim to help professionals to select the best instrument for a clinical scenario, the feasibility, patient's preferences, administrator and respondent burden, and the accommodations (or accessible forms) should be addressed and evaluated. Otherwise the suitability and the intended use of instruments might be compromised, especially in the context of disability research. Tools identified in our meta-review that meet most of these criteria are the COSMIN, EMPRO, SCI criteria, Andresen´s tool, CanChild Outcomes, and OMERACT, since all of them cover a wide range of measurement properties, offer an item structure, and assess the usability of instruments.

Special mention is due to the COSMIN, the most widespread and comprehensive tool to assess measurement properties of health instruments designed for an evaluative purpose. The COSMIN standards were developed in a Delphy study[43] aiming to improve the selection of the most appropriate health instrument for a clinical scenario. The most recent version of the COSMIN consists of a manual for conducting systematic reviews of health instruments, providing different steps with respect to the literature search process, the assessment of measurement properties and feasibility of instruments, and the evaluation of the risk of bias (RoB) of studies according to the Cochrane methodology[16]. Additionally, the COSMIN initiative recently developed a guideline exclusively focused on assessing the content validity of health instruments, considered the most important property to ensure the adequate reflection of

the construct measured[44,45]. In the light of these considerations, we strongly recommend the application of the latest version of the COSMIN to conduct high quality systematic reviews on measurement properties of health instruments for an evaluative purpose, or for other purposes with appropriate adaptation.

Despite COSMIN's many strengths, our analysis of the other assessment tools and measurement standards allow us to suggest future lines of work on this tool. First, the current format of COSMIN is fairly complex, requiring high expertise in the field of psychometrics and specific training for its proper application. The reporting of the inter-rater agreement coefficients when reviewers use the last version of COSMIN may provide useful data about its reliability. Second, consideration should be given to the Testing Standards recommendation on the inclusion of the assessment of fairness (i.e., evaluation of accessible forms for specific populations). Third, the feasibility of the measurement instruments, merely described in COSMIN, and their burden, should be properly rated, with examples found in EMPRO or Andresen´s tool. Fourth, it must be considered that the RoB evaluation of studies is itself a productive field of research with a long tradition, with specific tools that have been developed for different research questions and study designs. Examples might be found in the Cochrane Collaboration´s Tool for Assessing the Risk of Bias of Clinical Trials[46], the Newcastle Ottawa Scale (NOS)[47] for nonrandomised studies, or the Quality Assessment Tool for Cohort Studies (Q-COH II)[48,49]. From our point of view, the COSMIN proposal could also be simplified and improved by guiding the reviewers towards the identification of the most appropriate RoB assessment tools instead of developing their own RoB appraisal guidelines, taking advantage of knowledge and innovations in that field of research. And last, but not least, improving the quality of systematic reviews encompasses researchers, sponsors and promoters, but also journals, which should require full compliance with reporting and methodological guidelines, and the use of assessment tools.

Patient and Public Involvement

No patient or public involvement.

Ethics and dissemination

Ethical approval is not necessary for meta-reviews.

Data availability statement

Data are available upon reasonable request to the authors.

Authors' contribution

All authors meet the criteria recommended by the International Committee of Medical Journal Editors, ICMJE. All authors made substantial contributions to conception and design, piloted the inclusion criteria and provided direction of the data extraction and analysis. SL drafted the article, and JV, CV and JML critically revised the draft for important intellectual content. All authors agreed on the final version.

Funding statement

This work was supported by the Grant PGC2018-100675-B-I00, Spanish Ministry of Science, Innovation and Universities (Spain).

Competing interests' statement.

The authors declare no conflict of interest. The authors have no financial relationships relevant to this article to disclose, and the funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Correspondence

Correspondence concerning this article should be addressed to Jaume Vives, Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB. E-mail: jaume.vives@uab.cat

REFERENCES

1. Mokkink L, Terwee C, Stratford P, et al. Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Qual Life Res*. 2009;18(3):313-333. doi:10.1007/s11136-009-9451-9
2. Terwee C, Prinsen C, Ricci Garotti M, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25:767-779. doi:10.1007/s11136-015-1122-4
3. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. *Qual Life Res*. 2018:1-11. doi:10.1007/s11136-018-1798-3
4. Pussegoda K, Turner L, Garritty C, et al. Identifying approaches for assessing methodological and reporting quality of systematic reviews: A descriptive study. *Syst Rev*. 2017;6(1):1-12. doi:10.1186/s13643-017-0507-6
5. Pussegoda K, Turner L, Garritty C, et al. Systematic review adherence to methodological or reporting quality. *Syst Rev*. 2017;6(1):1-14. doi:10.1186/s13643-017-0527-2
6. Rosenkoetter U, Tate RL. Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. *Brain Impair*. 2017:1-16. doi:10.1017/BrImp.2017.29
7. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res*. 2013;22(8):1889-1905. doi:10.1007/s11136-012-0344-y
8. Lorente S, Vives J, Viladrich C, Losilla JM. Tools to assess the measurement properties of quality of life instruments : a meta-review protocol. *BMJ Open*. 2018;8:1-4. doi:10.1136/bmjopen-2018-022829
9. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Ann Intern Med*. 2009;151(4):264-269. doi:10.1371/journal.pmed1000097
10. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions : Explanation and Elaboration. *Ann Intern Med*. 2009;151(4):W65-W94. doi:10.1371/journal.pmed.1000100
11. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). *Cadth Methods Guidel*. 2016;(January):40-46.

doi:10.1016/j.jclinepi.2016.01.021

12. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. *J Clin Epidemiol*. 2016;75:40-46. doi:10.1016/j.jclinepi.2016.01.021
13. Terwee CB, Jansma EP, Riphagen II, De Vet H. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123. doi:10.1007/s11136-009-9528-5
14. Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives. *Health-Related Quality of Life and Well-Being*.; 2010. https://www.healthypeople.gov/sites/default/files/HRQoLWBFullReport.pdf.
15. Mokkink L, Terwee C, Patrick D, et al. *The COSMIN Checklist Manual*.; 2012. http://www.cosmin.nl/images/upload/files/COSMIN checklist manual v9.pdf.
16. Mokkink LB, Prinsen C, Patrick D, et al. COSMIN Methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). User manual. 2018:1-78. http://www.cosmin.nl/.
17. Terwee C, Bot S, de Boer M, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012
18. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments and review criteria. *Qual Life Res*. 2002;11(3):193-215.
19. Lohr K, Aaronson N, Alonso J, Burnam M, Patrick D, Perrin E. Evaluating Quality of Life and Health status instruments: development of scientific review criteria. *Clin Ther*. 1996;18(5):979-992. doi:doi.org/10.1016/S0149-2918(96)80054-3
20. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess (Rockv)*. 1999;3(9).
21. Brazier J, Ratcliffe J. Measurement and Valuation of Health for Economic Evaluation. In: *International Encyclopedia of Public Health*. Vol 4. Elsevier; 2017:586-593. doi:10.1016/B978-0-12-803678-5.00457-4
22. Department of Health and Human Services. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims: Draft Guidance*. Vol 20.; 2006. doi:10.1186/1477-7525-4-79
23. Department of Health and Human Services. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*.; 2009. doi:10.1111/j.1524-4733.2009.00609.x
24. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. *Evaluating Patient-Based Outcome*

Measures for Use in Clinical Trials. Vol 2.; 1998. doi:9812244

25. World Health Organization. International Classification of Functioning (ICF). www.who.int/classifications/icf/en/. Published 2016. Accessed February 6, 2020.
26. Valderas JM, Ferrer M, Mendívil J, et al. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. *Value Heal*. 2008;11(4):700-708. doi:10.1111/j.1524-4733.2007.00309.x
27. SCIRE. Spinal Cord Injury Rehabilitation Evidence. https://scireproject.com.
28. Johnston M V, Graves DE. Towards guidelines for evaluation of measures: An introduction with application to spinal cord injury. *J Spinal Cord Med*. 2008;31(1):13-26. doi:10.1080/10790268.2008.11753976
29. Andresen EM. Criteria for assessing the tools of disability outcomes research. *Arch Phys Med Rehabil*. 2000;81(12 SUPPL. 2):15-20. doi:10.1053/apmr.2000.20619
30. Law M. *Outcome Measures Rating Form Guidelines*.; 2004. https://www.canchild.ca/system/tenon/assets/attachments/000/000/371/original/measguid.pdf.
31. OMERACT. Instrument selection for Core Outcome Measurement Sets. In: *OMERACT Handbook*. ; 2019. https://omeracthandbook.org/handbook.
32. APA, AERA, NCME. *Standards for Educational and Psychological Testing.* American Educational Research Association.; 2014.
33. APA, AERA, NCME. *Standards for Educational and Psychological Testing.* American Educational Research Association.; 1999.
34. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application. *Am J Med*. 2006;119(2):166.e7-166.e16. doi:10.1016/j.amjmed.2005.10.036
35. McDowell I, Spassoff RA, Kristjansson B. On the classification of population health measurements. *Am J Public Health*. 2004;94(3):1413-1448. doi:10.1142/s0217751x95000681
36. Garin O, Ferrer M, Pont À, et al. Disease-specific health-related quality of life questionnaires for heart failure: A systematic review with meta-analyses. *Qual Life Res*. 2009;18(1):71-85. doi:10.1007/s11136-008-9416-4
37. Speight J, Howarth A. Quality of life in restless legs syndrome: A systematic review of clinical trials and a critical review of instruments. *Patient*. 2010;3(3):185-203. doi:10.2165/11534390-000000000-00000
38. Chassany O, Holtmann G, Malagelada J, Gebauer U, Doerfler H, Devault K. Systematic review: health-related quality of life (HRQOL) questionnaires in gastro-oesophageal

reflux disease. *Aliment Pharmacol Ther*. 2008;27(11):1053-1070. doi:http://dx.doi.org/10.1111/j.1365-2036.2008.03683.x

39. Beattie M, Lauder W, Atherton I, Murphy DJ. Instruments to measure patient experience of health care quality in hospitals: a systematic review protocol. *Syst Rev*. 2014;3(1):4. doi:http://dx.doi.org/10.1186/2046-4053-3-4
40. Lorente S, Losilla JM, Vives J. Instruments to assess patient comfort during hospitalization: A psychometric review. *J Adv Nurs*. 2018;74(5):1001-1015. doi:10.1111/jan.13495
41. Ferrando PJ, Anguiano-Carrasco C. A Structural Model-Based Optimal Person-Fit Procedure for Identifying Faking. *Educ Psychol Meas*. 2012;73(2):173-190. doi:10.1177/0013164412460049
42. Ferrando PJ, Anguiano-Carrasco C. A structural equation model at the individual and group level for assessing faking-related change. *Struct Equ Model*. 2011;18(1):91-109. doi:10.1080/10705511.2011.532725
43. Mokkink LB, Terwee CB, Patrick 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(4):539-549. doi:10.1007/s11136-010-9606-8
44. Terwee CB, Prinsen CA, Chiarotto A, et al. *COSMIN Methodology for Assessing the Content Validity of PROMs: User Manual*. Vol 120.; 2018.
45. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient- reported outcome measures : a Delphi study. *Qual Life Res*. 2018;27(7). doi:10.1007/s11136-018-1829-0
46. Higgins J, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343(7829):1-9. doi:10.1136/bmj.d5928
47. Wells G, Shea B, O´Connell D, et al. The Newcastle-Otawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses 2000. www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
48. Jarde A, Losilla J-M, Vives J, Rodrigo MF. Q-Coh : A tool to screen the methodological quality of cohort studies in systematic reviews and meta-analyses. *Int J Clin Heal Psychol*. 2013;13(2):138-146. doi:10.1016/S1697-2600(13)70017-6
49. Jarde A, Losilla J-M, Oliveras I, Vives J. Quality assssment tool for cohort studies (Q-COH II) User´s manual. 2014:1-13.

Figura 1. PRISMA flowchart. Flow diagram for search results

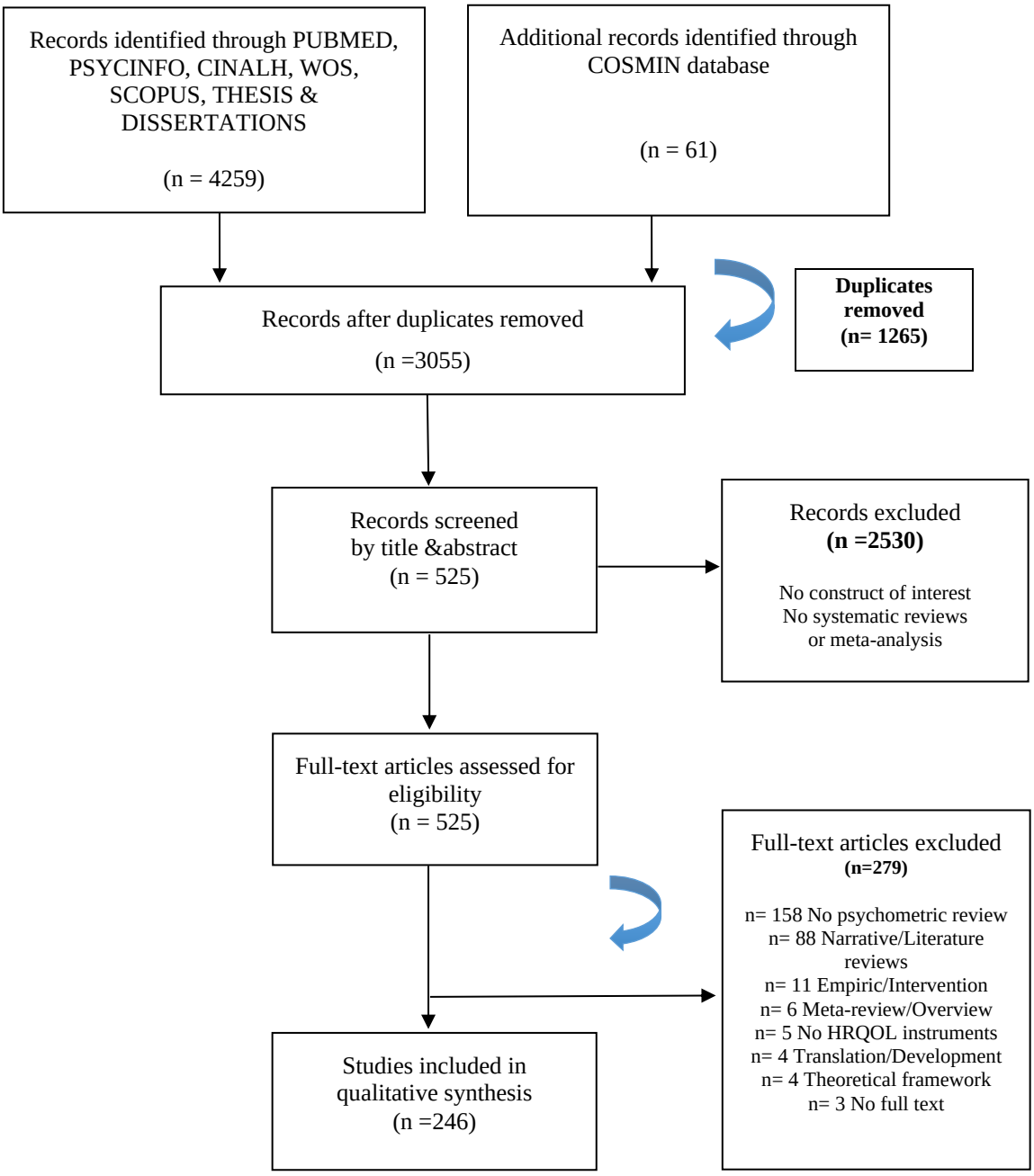


Table 1. Reporting and methodological quality of studies

	2008-2020		2014-2020	
	n	%	n	%
Protocol registered prospectively				
• Yes, PROSPERO	27	10.9	26	20.5
• No registered	219	89.1	100	79.3
Standards of systematic review reporting and/or quality assessment				
• Yes (AMSTAR, PRISMA, QUOROM…)	78	31.7	64	50.8
• No	168	68.3	62	49.2
Standards to assess reporting and/or quality assessment of primary studies				
• Yes (QUADAS, STARD…)	42	17.0	30	23.8
• No	204	83.0	96	76.2
Number of databases searched				
• 1-3	96	39.1	50	39.6
• 4-6	107	43.4	61	48.4
• 7-9	22	8.9	8	6.3
• >=10	18	7.3	6	4.7
• Not reported	3	1.2	1	0.8
Other sources				
• Official websites/Internet	25	10.1	7	5.5
• Virtual libraries	24	9.7	12	9.4
• Google/Google Scholar	25	10.1	14	11.0
• Scientific journals/Thesis	6	2.4	2	1.6
Search strategy				
Terms, databases, time period				
• Yes	238	96.7	123	97.6
• No	8	3.3	3	2.4
Search syntax				
• Detailed syntax reported (Truncations, Booleans…)	115	46.7	79	62.7
• Syntax not reported or not detailed enough to be replicable	125	50.8	46	36.5
• Supplementary file under request (not available)	5	2.1	1	0.8
Inclusion / Exclusion selection criteria				
• Reported and well-defined	229	93.1	122	96.8
• Not reported or not clearly stated	17	6.9	4	3.2
Article selection				
• By 2 or more independent reviewers	134	54.4	87	69.0
• Not reported or not clearly stated	112	45.6	39	31.0
Flow chart				
• Yes	166	67.5	108	85.7
• No	80	32.5	18	14.1
Funding				
• Reported	132	53.7	69	54.8
• Not reported or not clearly stated	114	46.3	57	45.2
TOTAL	246	100	126	100

PROSPERO= Prospective Register of Systematic Reviews; AMSTAR=Assessment of Multiple Systematic reviews; PRISMA= Preferred Reporting Items for Systematic Reviews and Meta-Analyses; QUOROM=Quality of Reporting of Meta-analysis; QUADAS= Quality Assessment of Diagnostic Accuracy Studies; STARD= Standards for the Reporting of Diagnostic Accuracy Studies; n= frequency; %= percentage

Table 2. Tools to assess measurement properties. Characteristics and comparison to Testing Standards

Tools	**1) COSMIN**	**2) TERWEE'S CRITERIA**	**3) ATTRIBUTES & CRITERIA**	**4) ECONOMIC EVALUATION**	**5) GUIDANCE FOR INDUSTRY**	**6) FITZPATRICK´S CRITERIA**	**7) ICF ICFCY**	**8) EMPRO**	**9) SCI CRITERIA**	**10) ANDRESEN´S TOOL**	**11) CANCHILD OUTCOMES**	**12) OMERACT**	**13) TESTING STANDARDS**
Development	Delphi	Author criteria	Expert panel	Literature review	Consensus	Literature review		Expert panel	Expert panel Literature review	Literature review	Expert panel	Expert Panel Delphi	Consensus
Sponsor/s	COSMIN Initiative	Author	SACMOT Working group	Standing Group Of Health Technology	FDA Staff	Standing Group of Health Technology	WHO Member States	IRYSS Committee	SCIRE Working group	Author	CanChild Center Staff	OMERACT Initiative	AERA, APA NCME
Approval Updates	2010 2018	2007	1996 2002, 2013	1999 2017	2006 2009	1998	2001 2019[a]	2008	2008, 2016	2000	1987[b] 2004	1992 1998,2007,2014 2019	1954 1966, 1974, 1985, 1999, 2014
Items (scoring)	5-18 items/box (+ / - / ?)	8-9 items total (+ / - / ?)	Not item structured (no scoring)	Not item structured (no scoring)	Not item structured (no scoring)	Not item structured (no scoring)	Not item structured (no scoring)	39 items total (strongly agree, agree, disagree, strongly disagree)	3-5 items/box (++++ / +++ / ++ / +)	11 items total (A, B, C)	2-6 items/box (excellent, adequate, poor)	2-5 items/box (Green, amber, red, white)	Not item structured (no scoring)
Measurement properties													
Validity	Content Construct (Int.Structure Cross-Cultural Hypotheses test) Criterion (Gold standard)	Content Construct (Hypotheses test) Criterion (Gold standard) Floor/Ceiling	Conceptual & measurement model Content Construct (Hypotheses test) Criterion (Gold standard)	Descriptive (Content Face Construct) Preference-based valuation Empirical (Criterion)	Conceptual model Content Construct (Hypothesis test, Discriminant, Convergent, Known groups) Criterion (Gold standard, sensitivity)	Purpose Content/Face Construct (Convergent, Discriminant, Int.Structure) Criterion (Predictive) Cut-score precision	Content	Conceptual & measurement model Content Construct (Hypotheses test) Criterion	Content Criterion (Concurrent Predictive "Discriminant") Clinical utility (Consequential validity)) Floor/Ceiling	Conceptual & measurement model Instrument bias Int.Structure Convergent Discriminant	Purpose Scale construction Content Construct (Hypotheses test) Criterion (Gold standard)	Content, Face Construct (Convergent, Divergent) Criterion (Accuracy) Discrimination (Sensitivity over time & over treatment)	Content Response process Int. Structure (Dimensions, DIF, Relations to other variables (Hypotheses test Convergent, Discriminant, Criterion, Responsiveness) Consequences
	Responsiveness	Responsiveness	Responsiveness		Responsiveness	Responsiveness		Responsiveness	Responsiveness	Responsiveness	Responsiveness		
Reliability	Int. Consistency Measurement error (Test retest, Agreement)	Int.Consistency Reproducibility (Agreement, Relative measurement error)	Int.Consistency Reproducibility (Test retest, inter-rater)	Test retest Inter-rater	Test retest Inter-rater Int.Consistency	Int.Consistency Reproducibility (Test retest)		Int.Consistency Reproducibility (Test retest, Inter-rater)	Int.Consistency Test retest	Int.Consistency Test retest	Int.Consistency Intra/Inter-rater Test retest	Reproducibility Test retest	Int.Consistency Test retest Alternate forms Scorers & Decision consistency/ accuracy
Fairness													Equivalence of accomodations
Other characteristics									Norms	Norms, Standard values	Norms Standardization		Scales, norms, and score comparability
	Interpretability	Interpretability	Interpretability		Interpretability	Interpretability		Interpretability					Test development and revision
			Burden		Burden	Acceptability (Burden)		Burden	Burden	Burden			
			Administration Accessible forms		Administration Accessible forms			Administration	Administration Accessible forms	Administration Accessible forms			
	Feasibility		Cultural Adaptations	Practicality		Feasibility Cultural Adaptations		Cultural Adaptations	Applicability Cultural Adaptations	Cultural Adaptations	Clinical utility (Feasibility)	Feasibility	
Frequency of use (%)	61 (30.4)	45 (22.4)	33 (16.4)	17 (8.4)	14 (6.9)	14 (6.9)	7 (3.4)	4 (2.0)	2 (1.0)	2 (1.0)	1(0.5)	1 (0.5)	0

Note: DIF= Differential Item Functioning; %=Percentage; Invariance=Measurement invariance; Int.Structure= Internal Structure; Int. Consistency= Internal Consistency; AERA= American Educational Research Association; APA= American Psychological Association; NCME= National Council on Measurement in Education; SACMOT= Scientific Advisory Committee Medical Outcomes Trust; FDA= Food and Drug Administration; WHO= World Health Organization; IRYSS= Spanish Cooperative Investigation Network for Health and Health Service Outcomes Research; SCIRE= Spinal Cord Injury Rehabilitation Evidence; COSMIN=Consensus Standards for Selection of Health Measurement Instruments; TERWEE´S CRITERIA= Quality Criteria for Measurement Properties; ATTRIBUTES&CRITERIA= Attributes and Criteria to assess Health Status and Quality of Life Instruments; ECONOMIC EVALUATION= Health Status Measures in Economic Evaluation; GUIDANCE FOR INDUSTRY=Guidance for Industry patient-reported outcomes measures; FITZPATRICK´S CRITERIA=Evaluating patient-based outcomes measures for use in clinical trials.; ICF= International Classification of Functioning; ICFCY= International Classification of Functioning for Children and Youth; EMPRO= Evaluating Measures of Patient Reported Outcomes; SCI CRITERIA= Spinal Cord Injury guidelines; ANDRESEN´S TOOL=Criteria for Assessing the Tools of Disability Outcomes Research; CANCHILD OUTCOMES= CanChild Outcomes Measures; OMERACT= Outcomes Measures in Rheumatology Clinical Trials; TESTING STANDARDS= Standards for Educational and Psychological Testing. See text for references. [a]Updated version at website, [b] Reference at 2004

Table 3. Intended use of instruments and their association to measurement properties

Intended use of instruments identified across the systematic reviews		**Frequency**	**% (Over 246)**
Evaluative (Change scores pre-post studies. Effectiveness of an intervention)		178	72.3
Impact of disease on HRQoL (disease symptoms, burden…)		138	55.1
Analytic (Health policies. Cost-effectiveness. Funding)		35	14.2
Diagnostic (Distinguish between groups, levels of severity…)		16	6.5
Descriptive (Health measures in surveys. Needs of groups of people)		4	1.6
Predictive (Anticipation of future health status. Risk factors. Risk profiles)		2	0.8
Intended use is no reported or no clearly stated		6	2.4
Conclusions according to the intended use of instruments		**n**	**% (Over 246)**
Yes, reviewers made specific conclusions		68	27.6
No, reviewers made general conclusions		178	72.4
Measurement properties associated to the intended use of the instrument		**n**	**% (Over 68)**
Evaluative	Responsiveness / Conceptual and Measurement Model / Content validity / Reliability (internal consistency, test retest) / Respondent Burden / Convergent validity / Cross cultural validity	41	60.3
Impact	Conceptual and Measurement Model / Content validity	29	42.6
Analytic	Preference-based valuation / Agreement	11	16.2
Diagnostic	Known groups validity / Test retest	7	10.3
Predictive	Sensivity and specificity	1	1.5

4. DISCUSIÓN

La presente tesis aporta un marco teórico y metodológico para la evaluación y la selección de las PROMs más apropiadas para la práctica clínica. Los resultados de la revisión psicométrica facilitan la elaboración de recomendaciones relacionadas con la evaluación y aplicabilidad de medidas de confort (Lorente et al., 2017; Lorente, Losilla, et al., 2018), y los resultados de la meta-revisión destacan el gran número de propuestas elaboradas desde diferentes aproximaciones para evaluar las propiedades de las medidas de salud, en concreto de la calidad de vida (Lorente, Vives, et al., 2018; Lorente et al. 2020).

4.1. CONTRASTE ENTRE APROXIMACIONES CON RELACIÓN A LA EVALUACIÓN DE LAS PROPIEDADES DE MEDIDA

Los resultados de nuestra meta-revisión indican que la mayoría de las herramientas que evalúan la calidad de la medida en el ámbito de la salud provienen de las tradiciones de la medicina y de la economía. La herramienta más aplicada serían los Estándares COSMIN (Mokkink et al., 2012; Mokkink et al., 2018), derivados de la tradición médica. Las diferencias más relevantes entre las herramientas revisadas estarían relacionadas con la selección y clasificación de los criterios para evaluar la calidad de la medida, el procedimiento de análisis y el formato de la herramienta, como detallamos a continuación (Lorente et al. 2020).

En primer lugar, y con respecto a los criterios para evaluar la calidad de la medida, observamos que la mayoría de las herramientas evalúan las evidencias de validez y de fiabilidad recomendadas por los Estándares psicométricos (APA et al., 2014). Observamos, sin embargo, ciertas diferencias entre aproximaciones. Por un lado, tanto las herramientas que se desarrollan desde la tradición de la medicina, provenientes de SACMOT (Aaronson et al., 2002; Lohr et al., 1996), como de la economía (Brazier et al., 1999, 2017), difieren de los Estándares psicométricos en la valoración de determinadas propiedades de la medida, como la equidad, definida como la acomodación de las medidas para su administración en diferentes poblaciones (e.g. poblaciones con necesidades especiales) (APA et al., 2014). Además, la herramienta que se desarrolla desde la tradición de la Evaluación económica también difiere de los Estándares psicométricos en la importancia de evaluar determinadas propiedades, como la consistencia interna. En este caso, la consistencia interna se consideraría una propiedad prescindible cuya evaluación podría ser, incluso, motivo de conflicto, ya que su estricto cumplimiento podría derivar en la exclusión ítems de contenido relevante para la adecuada valoración de las preferencias de los pacientes (Brazier et al., 1999,

2017). También hallamos numerosas diferencias entre los tres estándares respecto a la nomenclatura para referirse a algunos criterios de evaluación. Por ejemplo, la validez discriminante en algunos casos se puede identificar como validez divergente (Lorente et al. 2020).

En segundo lugar, y con respecto al procedimiento de análisis, destacamos las diferencias entre métodos de evaluación de las propiedades. Estas diferencias las podemos hallar entre herramientas de la misma tradición, es decir, la medicina, y también entre diferentes tradiciones. Los ejemplos más representativos serían la validez de constructo, la validez discriminante y la fiabilidad. Con relación a la validez de constructo observamos dos posibles enfoques para su evaluación. En algunos casos la validez de constructo se evalúa con relación a variables externas de forma más o menos estándar, es decir, analizando correlaciones de puntuaciones entre medidas y aportando pruebas de validez convergente y/o discriminante, mientras que en otros casos se comprueban todas las hipótesis posibles relacionadas con variables externas, ya sea mediante correlaciones de puntuaciones, comparación de medias, análisis de sensibilidad y/o especificidad o mediante aquellos procedimientos de análisis que se consideren necesarios. Con respecto a la validez discriminante distinguimos, principalmente, dos métodos de análisis. Cuando la validez discriminante hace referencia a la relación entre medidas se analiza mediante correlaciones, pero si hace referencia a diferencias entre grupos de individuos se analiza mediante la comparación de medias o análisis de la varianza. Y, con respecto a la fiabilidad, podemos ver que las herramientas suelen utilizar los mismos métodos o índices de evaluación, como el coeficiente Kappa para establecer acuerdo entre jueces, o alfa de Cronbach para consistencia interna, pero hallamos variabilidad en cuanto a la tipología de evidencias valoradas. Por ejemplo, las herramientas enmarcadas en la tradición de la medicina evalúan, en general, la fiabilidad test-retest, la fiabilidad entre jueces y la consistencia interna, mientras que la herramienta desarrollada desde la tradición de la economía sólo evalúa la fiabilidad test-retest y entre jueces (Lorente et al. 2020).

Finalmente, y con respecto al formato de las herramientas, podemos distinguir dos modelos. Por un lado, tenemos las herramientas que adoptan el formato de guía y, por otro lado, las que adoptan un formato de escala. Las herramientas con formato de guía se limitarían a establecer recomendaciones sobre las propiedades que deben valorarse, y no tanto sobre cómo deben valorarse. En cambio, las herramientas que presentan un formato de escala se basan en una estructura de ítems, facilitando la evaluación de la calidad de la medida mediante escalas Likert o categorías. En este caso, las tradiciones no establecen criterios de formato, de forma que desde cualquier perspectiva podemos hallar cualquiera de estas presentaciones (Lorente et al. 2020).

A pesar de las diferencias señaladas, los resultados de nuestra meta-revisión sugieren que los criterios de evaluación que siguen las herramientas revisadas se aproximan, cada vez más, a las recomendaciones de los Estándares psicométricos. Sin embargo, propiedades como la equidad, las consecuencias derivadas de la aplicación de las medidas y los procesos de respuesta todavía no se evalúan en las tradiciones de la medicina y la economía. En este sentido, y de acuerdo con Cook & Beckman (2006), consideramos que se deberían unificar criterios de evaluación de la calidad de las medidas según recomendaciones de los Estándares psicométricos.

La evaluación de las propiedades de la medida depende también de otros factores, como el modelo teórico que enmarca el diseño de las PROMs (Coltman et al., 2008; Lorente et al., 2017; Lorente, Losilla, et al., 2018). Por ejemplo, cuando la medida se diseña desde un modelo reflectivo, la consistencia interna se consideraría una propiedad necesaria, ya que el constructo subyace en la inter-correlación de los ítems. En cambio, cuando la medida se diseña desde un modelo formativo los indicadores conforman el constructo, y no a la inversa, de forma que la consistencia interna no sería una propiedad tan relevante como en el caso de las medidas reflectivas (Coltman et al., 2008). De acuerdo con este criterio, pues, la herramienta diseñada desde la tradición de la economía parece ser que evaluaría, en la mayoría de los casos, medidas desarrolladas en el marco de un modelo formativo, mientras que las herramientas diseñadas desde la tradición de la medicina evaluarían, en general, medidas desarrolladas en el marco de modelos reflectivos o mixtos. En conclusión, evaluar la calidad de la medida precisaría de la unificación de criterios en el diseño de las herramientas, pero también de la consideración de otros factores, como el modelo teórico en el que se basa el diseño de las PROMs.

Tabla 2. Contraste entre aproximaciones con relación a la evaluación de las propiedades de la medida

TESTING STANDARDS	COSMIN	TERWEE´S CRITERIA	ATTRIBUTES & CRITERIA	ECONOMIC EVALUATION	GUIDENCE FOR INDUSTRY	FITZPATRICK CRITERIA	ICF ICFCY	EMPRO	SCI CRITERIA	ANDRESEN TOOL	CANCHILD OUTCOMES	OMERACT
EVIDENCIAS DE VALIDEZ												
Contenido											Uso de la medida	
Proceso de respuesta												
Estructura Interna DIF/Trans-cultural			Adaptación cultural		Adaptación cultural	Adaptación cultural		Adaptación cultural	Adaptación cultural	Adaptación cultural		
Convergente Discriminante				Constructo								
Criterio		Gold standard	Gold standard		Gold standard			Gold standard	Gold standard		Gold standard	
Sensibilidad al cambio												
Consecuencias												
EVIDENCIAS DE FIABILIDAD												
Consistencia interna												
Test re-test												
Entre jueces												
Formas alternativas												
EQUIDAD												
Acomodaciones Equivalencia												
OTRAS CARACTERÍSTICAS												
Normas/Escalas Diseño/ Revisión	Interpret. Factibilidad	Interpret. Techo/Suelo	Interpret. Carga	Preferencias Factibilidad	Factibilidad Carga	Interpret. Factibilidad		Interpret. Carga	Utilidad clínica	Normas Factibilidad	Normas/Uso Factibilidad Utilidad clínica	Factibilidad
Frecuencia de uso (n, %)	61 (30.4)	45 (22.4)	33 (16.4)	17 (8.4)	14 (6.9)	14 (6.9)	7 (3.4)	4 (2.0)	2 (1.0)	2 (1.0)	1 (0.5)	1 (0.5)
Uso y población	PROMs evaluativos Población general/esp	PROMs evaluativos Población general/esp	PROMs de varios usos Población general/esp	PROMs analíticos Población general/esp	PROMs evaluativos Población general/esp	PROMs evaluativos Población general/esp	PROMs de varios usos Población general/esp	PROMs evaluativos Población general/esp	PROMs de varios usos Población específica	PROMs de varios usos Población específica	PROMs de varios usos Población específica	PROMs evaluativos Población específica

Nota. La herramienta ICF/ICFCY se diseña específicamente para evaluar validez de contenido. DIF= Funcionamiento diferencial del ítem; Interpret. = Interpretabilidad; esp= específica; Prof= Profesionales; Pac= Pacientes. Las anotaciones "Población esp" o "Población específica" se refieren a poblaciones con características específicas (idioma, edad, necesidades especiales, ...). Código de colores: Turquesa= criterios evaluados siguiendo Estándares APA; Gris claro= criterios no evaluados; Verde= otros criterios o características evaluadas.

4.2. IMPORTANCIA DE LA EVALUACIÓN DE LAS MEDIDAS SEGÚN SU USO

El uso de las medidas sería un criterio de relevancia para evaluar la calidad de sus propiedades. De acuerdo con este criterio, sólo las propiedades necesarias para el uso de una medida en un contexto determinado deberían demostrar calidad elevada. Es más, el resto de las propiedades evaluadas podrían presentar calidad moderada y no ser, por ello, motivo de descalificación o infravaloración de la medida.

Evaluar las medidas según su uso no es un planteamiento reciente, ni tampoco relacionado con una aproximación o una herramienta determinadas. Una de las primeras propuestas en este sentido sería la de Guyatt y colaboradores (1992), quienes señalan la necesidad de evaluar las medidas según si su uso es evaluativo o discriminativo y resaltando la importancia de determinadas propiedades, como la sensibilidad al cambio o la fiabilidad, respectivamente. Esta propuesta se recoge, actualmente, en el manual de la Cochrane para revisiones sistemáticas de intervenciones (Johnston et al., 2019). En concordancia, los Estándares psicométricos también señalan la relevancia de diseñar y evaluar las medidas según su uso para favorecer la toma de mejores decisiones, tanto a nivel individual (i.e. diagnóstico) como comunitario (i.e. programas para la salud y la educación). En la misma línea, los autores de los Estándares COSMIN, diseñados específicamente para analizar la calidad de medidas de uso evaluativo, advierten del potencial incremento en el riesgo de sesgo de resultados y conclusiones si no se toma en consideración en el contexto de aplicación de las medidas. Por último, la aproximación económica también destaca la importancia de diseñar PROMs que permitan evaluar costes y beneficios de las intervenciones sanitarias mediante el índice AVAC.

A pesar de estas recomendaciones, los resultados de nuestra meta-revisión indican que es poco frecuente que los revisores evalúen las medidas según su uso, lo cual dificulta la selección de las PROMs más apropiadas. Cuando los revisores sí toman en consideración el uso de los instrumentos, basan sus recomendaciones acerca de las PROMs más apropiadas en la evaluación de la calidad de aquellas propiedades que consideran más relevantes para su aplicación en un contexto determinado (Lorente et al. 2020).

4.3. EVALUACIÓN DE LA FACTIBILIDAD

Parte de las barreras identificadas en la implementación de las PROMs en la práctica clínica están relacionadas con la carga del administrador y del usuario, es decir, con aspectos relacionados con la factibilidad de la medida (Foster et al., 2018). Estas barreras se asocian con características del entorno y, quizá por este motivo, las herramientas revisadas no consideran la factibilidad como una propiedad psicométrica.

La factibilidad, empero, es un factor que puede incidir en el riesgo de sesgo de resultados y conclusiones, ya que el contexto, el formato y el modo de administración pueden modificar la capacidad de la medida para evaluar el constructo deseado (Frost et al., 2007). Por ejemplo, la carga que representa para el paciente la administración de una medida excesivamente larga o no adaptada a sus necesidades podría conllevar un número alto de no respuestas en los ítems y derivar en cuestionarios incompletos, afectando a la validez de las puntuaciones y a los resultados. El modo de administración de la medida, e.g. papel y lápiz, internet, o entrevista personal, también puede influir en los resultados, ya que los formatos podrían no ser totalmente equivalentes, o podrían no estar adaptados a las especificidades de la población diana (Beattie et al., 2014; Frost et al., 2007).

Dada su relevancia en el contexto de la evaluación de la calidad de la medida, tanto los Estándares psicométricos (APA et al., 2014), como la última versión de los Estándares COSMIN (Mokkink et al., 2018) incorporan la valoración de la factibilidad. En este sentido, destacaríamos que los Estándares psicométricos consideran algunos aspectos de la factibilidad como una propiedad psicométrica, bajo la denominación “imparcialidad de la medida” o “equidad” (en inglés, “fairness”), resaltando la importancia de realizar acomodaciones de la medida para garantizar que no existen factores que puedan afectar a la validez de los resultados, como la forma de administración, el formato de respuesta o cualquier otra característica del instrumento de medida que pueda dificultar las respuestas (APA et al., 2014). En cambio, la última revisión de los Estándares COSMIN considera la factibilidad como una característica de la medida en un contexto determinado. Además, al igual que sucede en casi todas las herramientas que incorporan la evaluación de la factibilidad, su valoración es meramente descriptiva y no permite obtener una puntuación (Mokkink et al., 2018).

En este sentido, y como propuesta de mejora para la evaluación de la factibilidad, destacaríamos el sistema de puntuación de van der Vleuten (1996). Este sistema consiste en una escala de cuatro puntos que valora los costes (e.g. coste de obtener una muestra representativa, costes administrativos), la aceptabilidad (e.g. idoneidad de la medida desde perspectiva del paciente), y el impacto (e.g. utilidad de los datos obtenidos), tal y como se detalla en Beattie et al. (2015) y Lorente, Losilla, et al. (2018). Su aplicación es sencilla y permite la comparabilidad de resultados.

5. CONCLUSIONES

La rigurosidad metodológica de las revisiones sistemáticas (RS) de las propiedades de la medida, así como la adecuada evaluación de la calidad de la medida mediante herramientas estandarizadas, son factores que minimizan el riesgo de sesgo en los resultados de dichas revisiones al tiempo que facilitan la selección de las PROMs más adecuadas para su aplicación en un contexto determinado. Asimismo, tanto el diseño como la evaluación de las PROMs precisan de la unificación de criterios que valoren validez y fiabilidad, pero también la factibilidad. Basándonos en los resultados de la presente tesis, a continuación, estructuramos nuestras conclusiones en relación con actuales y futuras líneas de investigación.

5.1. CALIDAD DEL REPORTE Y CALIDAD METODOLÓGICA DE LAS RS DE PROPIEDADES DE LA MEDIDA

Las guías para evaluar la calidad de reporte y la calidad metodológica de las revisiones sistemáticas (RS) se diseñan con el objetivo de facilitar la implementación de la práctica basada en la evidencia. Diferenciamos entre la calidad del reporte, que consiste en la adecuada descripción de la metodología utilizada en el proceso de revisión; y la calidad metodológica, que hace referencia a cómo se realiza la RS (e.g. estrategia de búsqueda, criterios de inclusión, evaluación del riesgo de sesgo de los estudios primarios, evaluación del sesgo de publicación, análisis de la heterogeneidad, etc.) (Pussegoda et al., 2017a, 2017b). Con respecto a la calidad del reporte de las RS de intervenciones en el ámbito de la salud, hallamos las guías PRISMA (Liberati et al., 2009; Moher et al., 2009) y PRISMA-P, para protocolos (Moher et al 2015). Con respecto a la calidad metodológica hallamos, entre otras propuestas, las guías publicadas por el Centre for Reviews and Dissemination (CRD), cuyas recomendaciones de buenas prácticas incluyen el registro del protocolo de la RS y la evaluación de la calidad de los estudios primarios (CRD, 2009), y la herramienta AMSTAR 2 (Shea et al., 2017).

Cuando nos referimos a RS de propiedades de la medida hallamos la propuesta publicada por la institución Joanna Briggs Institute (JBI) (Aromataris & Munn, 2017), la cual establece criterios relacionados tanto con el reporte como con la calidad metodológica. Entre otros aspectos, la JBI propone la adaptación del acrónimo PICOS (Patients, Intervention, Comparison, Outcomes, Study design), habitual en las RS de intervenciones, por el acrónimo PICO (Population, Instruments, Construct, Outcomes of measurement properties), la inclusión del filtro propuesto por Terwee (2009) en la estrategia de búsqueda, la evaluación del riesgo de sesgo de los estudios según la

propuesta COSMIN (Mokkink et al., 2018), y también indica cómo realizar el proceso de síntesis de resultados y gradación de la calidad de la evidencia aportada por la RS mediante la aplicación del sistema GRADE (Guyatt et al., 2011).

Siguiendo los criterios establecidos por la propuesta JBI, una línea de investigación futura sería la elaboración de guías PRISMA específicas para RS de propiedades de la medida, pues actualmente el reporte de este tipo de RS se realiza según las recomendaciones PRISMA para RS de intervenciones, con las dificultades de adaptación, estandarización y comparabilidad de resultados que ello implica. Asímismo, para estandarizar la evaluación de la calidad metodológica de las RS y meta-análisis de propiedades de la medida, también sería útil desarrollar una herramienta específica en la línea de AMSTAR 2 (Shea et al., 2017).

5.2. INCORPORACIÓN DEL RIESGO DE SESGO DE LOS ESTUDIOS EN REVISIONES PSICOMÉTRICAS

La incorporación de la evaluación del riesgo de sesgo (RdS) de los estudios primarios se considera una necesidad para la mejora de procesos de revisión y meta-análisis (Oliveras, 2018; Viswanathan et al., 2012). Mientras que la evaluación del RdS de estudios experimentales (Sterne et al., 2019), cuasi-experimentales (Sterne et al., 2016), no experimentales (Jarde et al., 2013; Losilla et al., 2018; Oliveras et al., 2017) y diagnósticos (Whiting et al., 2011) parecen ser líneas bien establecidas, no es el caso para los estudios de evaluación de propiedades de la medida. En este sentido, hallamos la propuesta COSMIN, que incide en la evaluación del riesgo de sesgo de los estudios primarios en las RS de propiedades de la medida así como en la gradación de la evidencia (Mokkink et al., 2018). La perspectiva de los Estándares COSMIN, sin embargo, es genérica, pues se plantea independientemente del diseño del estudio y del uso de la medida. Futuras líneas de investigación deberían evaluar si las herramientas utilizadas actualmente en los estudios de intervención se pueden adaptar para evaluar el RdS de los estudios de evaluación de calidad de la medida, valorando tanto su diseño como el uso de la medida de salud.

5.3. MÉTODOS CUALITATIVOS EN LA APLICACIÓN DE LAS PROMS

El manual Cochrane para revisiones sistemáticas de intervenciones recomienda la aplicación de técnicas de investigación cualitativa para el desarrollo de las PROMs (Johnston et al., 2019). A pesar del extenso marco teórico y cualitativo que respalda el diseño de estas medidas, sería posible que las PROMs no captaran suficientemente la experiencia del paciente con relación a la sintomatología relacionada con la calidad de vida.

En primer lugar, debido al método tradicional de desarrollo de las PROMs. Este método se basa en la elaboración de listas de signos y síntomas por parte de los profesionales clínicos, que posteriormente se interpretan y validan por los pacientes. Este proceso, sin embargo, no permite valorar realmente la perspectiva de los pacientes. Actualmente existen propuestas que impulsan nuevos modelos para desarrollar estas medidas de forma que tanto profesionales como pacientes pueden realizar sus aportaciones independientemente, ya sea mediante entrevistas u otros métodos, como focus group. Tractenberg et al. (2017) denominan a estas medidas PROMs centrados en la perspectiva del paciente (PC-PROMs). Y, en segundo lugar, porque los resultados de las PROMs se basan en puntuaciones cuantitativas. La incorporación de preguntas abiertas para la obtención de resultados permitiría complementar e interpretar de forma más comprehensiva la información que aportan las puntuaciones numéricas (Neale & Strang, 2015). Por ejemplo, Corner et al. (2013), en un estudio de calidad de vida que realiza en pacientes de cáncer mediante la aplicación de PROMs junto con preguntas abiertas, pone de manifiesto las numerosas dificultades que experimentan los pacientes en el manejo y cuidado de bolsas para estomas, pero también revela las causas que ocasionan esta problemática. Como ventajas añadidas, este formato de medida permite a los pacientes no sólo responder los diferentes ítems, sino también expresar sus necesidades y justificar su respuesta (Neale & Strang, 2015). Futuras líneas de investigación deberían orientarse, pues, al diseño de PROMs que incorporen técnicas cualitativas que permitan obtener información relevante e interpretar los datos cuantitativos. Este tipo de medidas podrían ser óptimas para captar la perspectiva del paciente.

5.4. IMPLEMENTACIÓN DE LAS PROMS EN LA PRÁCTICA CLÍNICA

La implementación de las PROMs precisa de la eliminación de ciertas barreras, como ya se ha mencionado en el apartado 4.3 al tratar la evaluación de la factibilidad (Foster et al., 2018). La adaptación de las PROMs tradicionales a los dispositivos electrónicos actuales, ePROMs, parece ser una opción viable para salvar parte de estas barreras, y facilitar así su implementación en la práctica clínica. Este formato de aplicación presenta numerosas ventajas en la recogida de datos (Coons et al., 2015). Por ejemplo, dado que las ePROMs facilitan la respuesta del paciente en cualquier lugar y momento, su uso permite una recogida de datos más exhaustiva, superando así dificultades de movilidad y/o de accesibilidad a los servicios sanitarios, y aumentando el tamaño de la muestra. La aplicación de las ePROMS, además, permite obtener información en tiempo real, formando parte de un sistema de métodos de recogida de datos denominados de evaluación ecológica momentánea (en inglés ecological momentary assessment, EMA) (McKay et al., 2016).

Estos métodos ofrecen diferentes ventajas, como la posibilidad de obtener medidas repetidas mediante un mensaje de aviso en un momento determinado, superando así barreras de comunicación y problemas de pérdida de datos. El uso de las ePROMs también incrementa la precisión de la información obtenida, pues reduce la probabilidad de ocurrencia de errores secundarios derivados de la entrada de datos, así como también disminuye la carga de los profesionales, de los pacientes, y de los costes económicos.

Sin embargo, también hallamos inconvenientes, como la necesidad del paciente de acceder a dispositivos adecuados y/o a la conectividad necesaria. En algunos casos, y dependiendo de estas condiciones, existiría riesgo de sesgo de selección de la muestra. Asimismo, la adaptación del formato de la medida también puede ocasionar riesgo de sesgo de resultados. Por un lado, existen aspectos que se modifican cuando las PROMs tradicionales se adaptan a los dispositivos electrónicos o plataformas web, como son el vocabulario y la disposición de la pregunta, implicando diferentes procesos cognitivos que pueden afectar la interpretabilidad y, por tanto, la elección de la respuesta. Por otro lado, las habilidades de los pacientes, la edad o el nivel educativo serían aspectos implicados en el adecuado manejo de los dispositivos electrónicos que podrían condicionar el número de respuestas y la validez y fiabilidad de las puntuaciones obtenidas (Coons et al., 2015; Eremenco et al., 2014; Meirte et al., 2020).

En definitiva, las ePROMs presentan numerosas ventajas en la práctica clínica, pero su uso también puede incrementar el error de medida si la calidad de sus propiedades no se evalúa adecuadamente. Por este motivo, la International Society for Pharmaeconomics and Outcomes Research (ISPOR) (Coons et al., 2009, 2015; Eremenco et al., 2014) recomienda la evaluación de las posibles fuentes de sesgo relacionadas con la factibilidad de su implementación, la usabilidad de los dispositivos, y su validez y fiabilidad mediante estudios de equivalencia entre modos de administración. Dos medidas se consideran equivalentes si el rango y distribución de sus puntuaciones, medias y desviaciones estándar son similares (APA, 1986; Gwaltney et al., 2008). Según esta definición, estudios previos concluyen que existe equivalencia entre el formato tradicional y el formato aplicado en soporte electrónico (ordenadores) (Gwaltney et al., 2008), y respecto a otros formatos como dispositivos móviles o plataformas web, siempre y cuando los cambios realizados sean moderados según los criterios ISPOR (Campbell et al., 2015; Coons et al., 2009; Muehlhausen et al., 2015). Destacan, sin embargo, la diversidad de técnicas estadísticas empleadas en los estudios primarios para la evaluación de la equivalencia que dificultan la comparabilidad de resultados, como el coeficiente intraclase (ICC) o el método de Bland-Altman (Campbell et al., 2015).

Futuras líneas de investigación deberían promover la unificación de criterios para evaluar la equivalencia entre medidas, analizar la invariancia de la medida entre grupos en el marco de la Teoría de Respuesta al Ítem (TRI) y/o en el marco del modelado de ecuaciones estructurales (Putnick & Bornstein, 2016), y valorar otras propiedades psicométricas relacionadas con la interpretabilidad (procesos cognitivos de respuesta) y la adaptación de la medida (equidad), siguiendo recomendaciones de los Estándares psicométricos (APA et al., 2014).

6. FINANCIACIÓN

Esta tesis se ha realizado gracias a la financiación del Ministerio de Economía y Competitividad, Proyecto PSI2014-52962-P, y del Ministerio de Ciencia, Innovación y Universidades, Proyecto PGC2018-100675-B-I00.

7. AGRADECIMIENTOS

En primer lugar, quiero agradecer a mis directores, los doctores Josep-Maria Losilla y Jaume Vives, su inestimable dedicación, supervisión y apoyo durante todo el proceso de elaboración de esta tesis. Sus aportaciones, tanto profesionales como personales, hacen posible esta tesis por compendio. Excelentes profesores, compañeros y amigos.

En segundo lugar, quiero agradecer a mi familia, a mi madre y a mis hermanos, la paciencia que han mostrado durante todos estos años, ya que he invertido mucho tiempo para finalizar este proyecto. A pesar de las dificultades y complicaciones, siempre me han apoyado y me han animado a continuar.

También quiero agradecer la colaboración de todos los coautores de los artículos que conforman este compendio. A mis compañeras de la profesión de enfermería, Rebeca y Sandra, por su entusiasmo por la investigación. Y a la doctora Carmen Viladrich, por sus aportaciones y sugerencias. Su visión de la psicometría es, sin duda, excepcional.

Finalmente, quiero agradecer a mis amigos su apoyo incondicional durante este largo camino.

8. REFERENCIAS

Aaronson, N., Alonso, J., Burnam, A., Lohr, K. N., Patrick, D., Perrin, E., & Stein, R. E. K. (2002). Assessing health status and quality-of-life instruments and review criteria. Quality of Life Research, 11(3), 193–215.

Aaronson, N., Choucair, A., Elliott, T., Greenhalgh, J., Halyard, M., Hess, R., Miller, D., Reeve, B., Santana, M., & Snyder, C. (2011). User's Guide to Implementing Patient-Reported Outcomes Assessment in Clinical Practice (p. 57). https://pdfs.semanticscholar.org/9f86/43e764e9ef42dd9715312e0f91d631a530f7.pdf

Alonso, J., Bartlett, S. J., Rose, M., Aaronson, N. K., Chaplin, J. E., Efficace, F., Leplège, A., LU, A., Tulsky, D. S., Raat, H., Ravens-Sieberer, U., Revicki, D., Terwee, C. B., Valderas, J. M., Cella, D., Forrest, C. B., Aaronson, N. K., BjØrner, J., Brazier, J., … Cella, D. (2013). The case for an international patient-reported outcomes measurement information system (PROMIS®) initiative. Health and Quality of Life Outcomes, 11(1), 1–5. https://doi.org/10.1186/1477-7525-11-210

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for Educational and Psychological Testing. American Educational Research Association.

APA. (1954). Technical recommendations for psychological tests and diagnostic techniques. Psychological Bulletin, 51, 201–238.

APA. (1986). Guidelines for Computer-Based Tests and Interpretations. American Psychological Association.

APA, AERA, & NCME. (1999). Standards for educational and psychological testing. American Educational Research Association.

APA, AERA, & NCME. (2014). Standards for educational and psychological testing. American Educational Research Association.

Aromataris, E., & Munn, Z. (2017). JBI Reviewer ' s Manual. The Joanna Briggs Institute, March, 32. https://reviewersmanual.joannabriggs.org/

Baker, A. (2014). Integrating the use of patient-reported outcomes for both clinical practive and performance measurement: views of experts from 3 countries. The Milbank Quarterly, 92(4), 754–775.

Beattie, M., Lauder, W., Atherton, I., & Murphy, D. J. (2014). Instruments to measure patient experience of health care quality in hospitals: a systematic review protocol. Systematic Reviews, 3(1), 4. https://doi.org/10.1186/2046-4053-3-4

Beattie, M., Lauder, W., Atherton, I., & Murphy, D. J. (2015a). Instruments to measure patient experience of health care quality in hospitals: a systematic review. Systematic Reviews, 4(97). https://doi.org/10.1186/s13643-015-0089-0

Beattie, M., Lauder, W., Atherton, I., & Murphy, D. J. (2015b). Instruments to measure patient experience of health care quality in hospitals: a systematic review. Systematic Reviews, 3(1), 4. https://doi.org/10.1186/2046-4053-3-4

Boaz, A., Ashby, D., & Young, K. (2002). Systematic Reviews : What have they got to offer evidence based policy and practice? In ESRC UK Centre for Evidence Based Policy and Practice (Issue January, pp. 1–26).

Bombardier, C., & Tugwell, P. A. (1982). A methodological framework to develop and select indices for clinical trials: statistical and judgmental approaches. The Journal of Reumatologý, 9, 753–757.

Brazier, J., Ara, R., Rowen, D., & Chevrou-Severac, H. (2017). A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. PharmacoEconomics, 35, 21–31. https://doi.org/10.1007/s40273-017-0545-x

Brazier, J., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. In Health Technology Assessment (Vol. 3, Issue 9).

Campbell, N., Ali, F., Finlay, A. Y., & Salek, S. S. (2015). Equivalence of electronic and paper-based patient-reported outcome measures. Quality of Life Research, 24(8), 1949–1961. https://doi.org/10.1007/s11136-015-0937-3

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research, 61(12), 1250–1262. https://doi.org/10.1016/j.jbusres.2008.01.013

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. American Journal of Medicine, 119(2), 166.e7-166.e16. https://doi.org/10.1016/j.amjmed.2005.10.036

Coons, S. J., Eremenco, S., Lundy, J. J., O'Donohoe, P., O'Gorman, H., & Malizia, W. (2015). Capturing Patient-Reported Outcome (PRO) Data Electronically: The Past, Present, and Promise of ePRO Measurement in Clinical Trials. Patient, 8(4), 301–309. https://doi.org/10.1007/s40271-014-0090-z

Coons, S. J., Gwaltney, C. J., Hays, R. D., Lundy, J. J., Sloan, J. A., Revicki, D. A., Lenderking, W. R., Cella, D., & Basch, E. (2009). Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value in Health, 12(4), 419–429. https://doi.org/10.1111/j.1524-4733.2008.00470.x

Corner, J., Wagland, R., Glaser, A., & Richards, M. (2013). Qualitative analysis of patients' feedback from a PROMs survey of cancer patients in England. BMJ Open, 3(4), 1–9. https://doi.org/10.1136/bmjopen-2012-002316

CRD. (2009). CRD's guidance for undertaking reviews in health care. University of York. https://www.york.ac.uk/crd/guidance/

Department of Health and Human Services. (2006). Guidance for Industry Patient-reported Outcome measures: Use in Medical Product Development to Support Labeling Claims: draft guidance. In Health and Quality of Life Outcomes (Vol. 20). https://doi.org/10.1186/1477-7525-4-79

Department of Health and Human Services. (2009). Guidance for Industry Patient-reported Outcome measures: Use in Medical Product Development to Support Labeling Claims. In Guidance for Industry (Issue December). https://doi.org/10.1111/j.1524-4733.2009.00609.x

Eremenco, S., Coons, S. J., Paty, J., Coyne, K., Bennett, A. V, Mcentegart, D., Pro, I., Modes, M., & Force, T. (2014). ISPOR TASK FORCE REPORT PRO Data Collection in Clinical Trials Using Mixed Modes : Report of the ISPOR PRO Mixed Modes Good Research Practices Task Force. Value in Health, 17(5), 501–516. https://doi.org/10.1016/j.jval.2014.06.005

Foster, A., Croot, L., Brazier, J., Harris, J., & O'Cathain, A. (2018). The facilitators and barriers to implementing patient reported outcome measures in organisations delivering health related services: a systematic review of reviews. Journal of Patient-Reported Outcomes, 2(1), 1–16. https://doi.org/10.1186/s41687-018-0072-3

Frost, M. H., Reeve, B. B., Liepa, A. M., Stauffer, J. W., Hays, R. D., & Sloan, J. A. (2007). What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value in Health, 10(SUPPL. 2), 94–105. https://doi.org/10.1111/j.1524-4733.2007.00272.x

Guyatt, G. H., Kirshner, B., & Jaeschke, R. (1992). Measuring health status: What are the necessary measurement properties? Journal of Clinical Epidemiology, 45(12), 1341–1345. https://doi.org/10.1016/0895-4356(92)90194-R

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., Debeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011). GRADE guidelines: 1. Introduction - GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology, 64(4), 383–394. https://doi.org/10.1016/j.jclinepi.2010.04.026

Gwaltney, C. J., Shields, A. L., & Shiffman, S. (2008). Equivalence of Electronic and Paper-and-Pencil Administration of Patient-Reported Outcome Measures : A Meta-Analytic Review. Value in Health, 11(2), 322–333. https://doi.org/10.1111/j.1524-4733.2007.00231.x

Jarde, A., Losilla, J.-M., Vives, J., & Rodrigo, M. F. (2013). Q-Coh : A tool to screen the methodological quality of cohort studies in systematic reviews and meta-analyses. International Journal of Clinical and Health Psychology, 13(2), 138–146. https://doi.org/10.1016/S1697-2600(13)70017-6

Johnston, B. C., Patrick, D. L., Devji, T., Maxwell, L. J., Bingham, C. O., Beaton, D. E., Boers, M., Briel, M., Busse, J. W., Carrasco-Labra, A., Christensen, R., da Costa, B. R., El Dib, R., Lyddiatt, A., Ostlo, R. W., Shea, B., Singh, J., Terwee, C. B., Williamson, P. R., … Guyatt, G. H. (2019). Patient-reported outcomes. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), Cochrane Handbook for Systematic Reviews of Interventions version 6.0. https://training.cochrane.org/handbook/current/chapter-18

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions : Explanation and Elaboration. Annals of Internal Medicine, 151(4), W65–W94. https://doi.org/10.1371/journal.pmed.1000100

Lohr, K. N., Aaronson, N. K., Alonso, J., Burnam, M. A., Patrick, D. L., Perrin, E. B., & Roberts, J. S. (1996). Evaluating quality-of-life and health status instruments: Development of scientific review criteria. Clinical Therapeutics, 18(5), 979–984. https://doi.org/10.1016/S0149-2918(96)80054-3

Lorente, S., Gimeno, R., Losilla, J. M., Garzón, S., & Vives, J. (2017). Benefits of the humidified low-flow oxygen therapy in infants with mild-moderate bronchiolitis. Journal of Clinical Nursing, 27(5–6), 1125–1133. https://doi.org/10.1111/jocn.14140

Lorente, S., Losilla, J. M., & Vives, J. (2018). Instruments to assess patient comfort during hospitalization: A psychometric review. Journal of Advanced Nursing, 74(5), 1001–1015. https://doi.org/10.1111/jan.13495

Lorente, S., Vives, J., & Losilla, J. M. (2017). Instruments to assess the patient comfort during hospitalization: a psychometric review protocol. Journal of Advanced Nursing, 73(3), 735–741. https://doi.org/10.1111/jan.13180

Lorente, S., Vives, J., Viladrich, C., & Losilla, J. M. (2018). Tools to assess the measurement properties of quality of life instruments : a meta-review protocol. BMJ Open, 8, e022829. https://doi.org/10.1136/bmjopen-2018-022829

Lorente, S., Viladrich, C., Vives, J., & Losilla, J.M. (2020). Tools to assess the measurement properties of quality of life instruments:a meta-review. BMJ Open. 10, e036038. https://doi.org/http://dx.doi.org/10.1136/bmjopen-2019-036038

Losilla, J. M., Oliveras, I., Marin-Garcia, J. A., & Vives, J. (2018). Three risk of bias tools lead to opposite conclusions in observational research synthesis. Journal of Clinical Epidemiology, 101, 61–72. https://doi.org/10.1016/j.jclinepi.2018.05.021

McDowell, I., Spassoff, R. A., & Kristjansson, B. (2004). On the classification of population health measurements. American Journal of Public Health, 94(3), 1413–1448. https://doi.org/10.1142/s0217751x95000681

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016a). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Explanation and Elaboration (PRESS E&E). In Cadth Methods and Guidelines (Issue January, pp. 40–46). https://doi.org/10.1016/j.jclinepi.2016.01.021

McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016b). PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. Journal of Clinical Epidemiology, 75, 40–46. https://doi.org/10.1016/j.jclinepi.2016.01.021

McKay, D., Przeworski, A., & Neill, S. O. (2016). Emerging Technologies for Clinical Practice. In Computer-Assisted and Web-Based Innovations in Psychology, Special Education, and Health (pp. 365–378). Elsevier Inc. https://doi.org/10.1016/B978-0-12-802075-3.00014-0

Meirte, J., Hellemans, N., Anthonissen, M., Denteneer, L., & Maertens, K. (2020). Benefits and Disadvantages of Electronic Patient-reported Outcome Measures : Systematic Review Corresponding Author : JMIR Perioperative Medicine, 3(1), e15588. https://doi.org/10.2196/15588

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. Annals of Internal Medicine, 151(4), 264–269. https://doi.org/10.1371/journal.pmed1000097

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., Altman, D. G., Booth, A., Chan, A. W., Chang, S., Clifford, T., Dickersin, K., Egger, M., Gøtzsche, P. C., Grimshaw, J. M., Groves, T., Helfand, M., … Whitlock, E. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Systematic Reviews, 4(1), 1–9. https://doi.org/10.1186/2046-4053-4-1

Mokkink, L. B., Prinsen, C., Patrick, D., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018). COSMIN Methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). User manual. (pp. 1–78). http://www.cosmin.nl/

Mokkink, L., Terwee, C., Knol, D., Stratford, P. W., Alonso, J., Patrick, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. Medical Research Methodology, 10(1), 22. https://doi.org/10.1186/1471-2288-10-22

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2012a). The COSMIN checklist manual. https://doi.org/10.1007/s13398-014-0173-7.2

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2012b). The COSMIN checklist manual. http://www.cosmin.nl/images/upload/files/COSMIN checklist manual v9.pdf

Mokkink, L., Terwee, C., Stratford, P., Alonso, J., Patrick, D., Riphagen, I., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. Quality of Life Research, 18(3), 313–333. https://doi.org/10.1007/s11136-009-9451-9

Muehlhausen, W., Doll, H., Quadri, N., Fordham, B., O'Donohoe, P., Dogar, N., & Wild, D. J. (2015). Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. Health and Quality of Life Outcomes, 13, 167. https://doi.org/10.1186/s12955-015-0362-x

NHS. (2017). National Patient Reported Outcome Measures (PROMs) Programme Consultation Report. https://www.england.nhs.uk/wp-content/uploads/2017/10/proms-consultation-report.pdf

OECD. (2017). Recommendations to OECD Ministers of Health from the High Level Reflection on the future of health statistics (Issue January). http://www.oecd.org/health/health-systems/Recommendations-from-high-level-reflection-group-on-the-future-of-health-statistics.pdf

Oliveras, I. (2018). Evaluación e incorporación del riesgo de sesgo de estudios no experimentales en revisiones sistemáticas y metaanálisis. Universidad Autònoma de Barcelona (UAB).

Oliveras, I., Losilla, J. M., & Vives, J. (2017). Methodological quality is underrated in systematic reviews and meta-analyses in health psychology. Journal of Clinical Epidemiology, 86, 59–70. https://doi.org/10.1016/j.jclinepi.2017.05.002

Øvretveit, J., Zubkoff, L., Nelson, E. C., Frampton, S., Knudsen, J. L., & Zimlichman, E. (2017). Using patient-reported outcome measurement to improve patient care. International Journal for Quality in Health Care, 29(6), 874–879. https://doi.org/10.1093/intqhc/mzx108

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. Quality of Life Research, 1–11. https://doi.org/10.1007/s11136-018-1798-3

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. Quality of Life Research, 1–11. https://doi.org/10.1007/s11136-018-1798-3

Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., Boutron, I., Sarkis-Onofre, R., Bjerre, L. M., Hróbjartsson, A., Altman, D. G., & Moher, D. (2017a). Identifying approaches for assessing methodological and reporting quality of systematic reviews: A descriptive study. Systematic Reviews, 6(1), 1–12. https://doi.org/10.1186/s13643-017-0507-6

Pussegoda, K., Turner, L., Garritty, C., Mayhew, A., Skidmore, B., Stevens, A., Boutron, I., Sarkis-Onofre, R., Bjerre, L. M., Hróbjartsson, A., Altman, D. G., & Moher, D. (2017b). Systematic review adherence to methodological or reporting quality. Systematic Reviews, 6(1), 1–14. https://doi.org/10.1186/s13643-017-0527-2

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of art and future directions for psychological research. Development Review, 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.Measurement

Rosenkoetter, U., & Tate, R. L. (2017). Assessing Features of Psychometric Assessment Instruments: A Comparison of the COSMIN Checklist with Other Critical Appraisal Tools. Brain Impairment, 1–16. https://doi.org/10.1017/BrImp.2017.29

Secretary's Advisory Committee on National Health Promotion and Disease Prevention Objectives. (2010). Health-Related Quality of Life and Well-Being. In Healthy People 2020. https://www.healthypeople.gov/sites/default/files/HRQoLWBFullReport.pdf

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ, 358, 1–9. https://doi.org/10.1136/bmj.j4008

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A. W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ (Online), 355, 4–10. https://doi.org/10.1136/bmj.i4919

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. The BMJ, 366, 1–8. https://doi.org/10.1136/bmj.l4898

Streiner, D. L. (2003). Clinimetrics vs. psychometrics: An unnecessary distinction. Journal of Clinical Epidemiology, 56(12), 1142–1145. https://doi.org/10.1016/j.jclinepi.2003.08.011

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & De Vet, H. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Quality of Life Research, 18(8), 1115–1123. https://doi.org/10.1007/s11136-009-9528-5

Terwee, C., Bot, S., de Boer, M., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007a). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C., Bot, S., de Boer, M., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. W. (2007b). Quality criteria were proposed for measurement properties of health status questionnaires. Journal of Clinical Epidemiology, 60(1), 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012

Terwee, C. C. B., Mokkink, L. L. B., Knol, D. L. D., Ostelo, R. W. J. G., Bouter, L. M., & de Vet, H. C. W. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: A scoring system for the COSMIN checklist. Quality of Life Research, 21(4), 651–657. https://doi.org/10.1007/s11136-011-9960-1

Terwee, C., Mokkink, L., & Patrick, D. (2011). COSMIN checklist with 4-point scale.

Terwee, C., Prinsen, C., Ricci Garotti, M., Suman, A., de Vet, H. C. W., & Mokkink, L. B. (2016). The quality of systematic reviews of health-related outcome measurement instruments. Quality of Life Research, 25, 767–779. https://doi.org/10.1007/s11136-015-1122-4

Tractenberg, R. E., Garver, A., Ljungberg, I. H., Schladen, M., & Groah, S. L. (2017). Maintaining primacy of the patient perspective in the development of patient- centered patient reported outcomes. Plos ONE, 12(3), 1–20. https://doi.org/10.1371/journal.pone.0171114

Tugwell, P., & Bombardier, C. (1982). A methodologic framework for developing and selecting endpoints in clinical trials. Journal of Rheumatology, 9, 758–762.

Valderas, J. M., & Alonso, J. (2008). Patient reported outcome measures: A model-based classification system for research and clinical practice. Quality of Life Research, 17(9), 1125–1135. https://doi.org/10.1007/s11136-008-9396-4

Van der Vleuten, C. (1996). The assessment of professional competence: developments, research and practical implications. Advances in Health Sciences Education, 1, 41–67.

van der Wees, P. J., Verkerk, E. W., Verbiest, M. E. A., Zuidgeest, M., Bakker, C., Braspenning, J., de Boer, D., Terwee, C. B., Vajda, I., Beurskens, A., & van Dulmen, S. A. (2019). Development of a framework with tools to support the selection and implementation of patient-reported outcome measures. Journal of Patient-Reported Outcomes, 3(1). https://doi.org/10.1186/s41687-019-0171-9

Viladrich, C., & Doval, E. (2007). Medición: Fiabilidad y Validez (Laboratori d´Estadística Aplicada i de Modelització (ed.)). UAB.

Viladrich, C., Doval, E., & Penelo, E. (2015). Medición: Fiabilidad y Validez (8th ed.).

Viladrich, C., Doval, E., & Penelo, E. (2019). Medición. Fiabilidad y Validez (Laboratori d´Estadística i de Modelització (ed.); 12th ed.). UAB.

Viswanathan, M., MT, A., Berkman, N., Chang, S., Hartling, L., & McPheeters, M. et al. (2012). Assessing the risk of bias of individual studies in systematic reviews of health care interventions. Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews. www.effectivehealthcare.ahrq.gov/

Weldring, T., & Smith, S. M. S. (2013). Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Services Insights, 6, 61–68. https://doi.org/10.4137/hsi.s11093

Whiting, P. F., Rutjes, A. W., Wetwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A. C., & Bossuyt, P. M. M. (2011). QUADAS-2. A revised tool for the quality assessment of diagnostic studies. Annals of Internal Medicine, 155(8), 529–536.

Williams, K., Sansoni, J., Morris, D., Grootemaat, P., & Thompson, C. (2016). Patient-reported outcome measures:literature review. In Sidney (Ed.), ACSQHC.

Wilson, I. B., & Cleary, P. D. (1995). Linking Clinical Variables With Health-Related Quality of Life: A Conceptual Model of Patient Outcomes. JAMA, 273(1), 59–65. https://doi.org/10.1001/jama.1995.03520250075037

World Health Organization. (2016). International Classification of Functioning (ICF). www.who.int/classifications/icf/en/

9. ANEXOS

9.1. ANEXOS RELACIONADOS CON LA REVISIÓN PSICOMÉTRICA

9.1.1. REGISTRO PROSPECTIVO DEL PROTOCOLO DE LA REVISIÓN PSICOMÉTRICA

Instruments to assess the patient comfort during hospitalization: a psychometric review protocol

Sonia Lorente, Jaume Vives, Josep Maria Losilla

Citation

Sonia Lorente, Jaume Vives, Josep Maria Losilla. Instruments to assess the patient comfort during hospitalization: a psychometric review protocol. PROSPERO 2016 CRD42016036290 Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42016036290

Review question

To identify the health instruments used for the measurement of patient comfort as a holistic experience during hospitalization.
To examine the psychometric properties of each instrument, assessing their reliability and validity.
To examine the outcomes quality, including the reproducibility, responsiveness, floor-ceiling effects and interpretability.
To examine the utility of each instrument, assessing their cost-efficiency, acceptability and educational impact in different healthcare settings.
To classify the different instruments according to their psychometric properties, outcomes and utility.

Searches

We aim to identify published instruments measuring the patient comfort during the hospitalization. We will define different combinations of keywords (using MeSH and Thesaurus, where available), in relation to the concept (e.g., comfort, theory), the setting (hospitalization or admission) and the instruments (e.g. questionnaires, scales).
The following databases will be included: Medical Literature Analysis and Retrieval System (MEDLINE), by ProQuest, Cumulative Index to Nursing and Allied Health Literature (CINAHL), by EBSCOhost, Psychological Information (PsycINFO), by APA PsycNET, Thesis & Dissertations, by ProQuest, and ISI Web of Knowledge (WoS, Web of Science CORE), by Thomson Reuters. To include grey literature we will also search records in Google, and will review up to 400 links. The search will be limited by Population (humans), by time (1990 to 2015) and by language (English). In addition, search alerts in CINALH & PsycINFO will be set.

Types of study to be included

Studies developing or validating questionnaires and/or scales measuring the holistic patient comfort during the hospitalization. Protocols and guidelines, as well as those studies exclusively qualitative designed, will be excluded.

Condition or domain being studied

Patient comfort is considered a direct indicator of quality in health care. The General Comfort Questionnaire (GCQ) (Kolcaba, 1992) is likely the first instrument specifically developed to measure the patient comfort as a holistic experience in four contexts (physical, psychospiritual, sociocultural and environmental). It was framed within the Theory of Comfort, by Kolcaba (Kolcaba, 1992), validated with different patients and pathologies, and registered as a multidisciplinary outcome indicator by the National Quality Measures Clearinghouse (NQMC, 2002). Hence, numerous and varied instruments have been developed, adapted or validated, to assess patient comfort in different health care settings and situations. However, as far as we are concerned, and despite this increasing trend, there are no systematics reviews or meta-analyses to evaluate the methodological quality and the psychometric properties of these health questionnaires/scales.
Whether these instruments are to be used in practical or research projects it is essential to evaluate how they perform, that means to assess the reliability, the validity and the quality of their results, as Keszei, Novak, & Streiner (2010) and Terwee et al. (2007) suggested. Therefore, this systematic revision aims to search, identify and analyse the methodological quality of each instrument measuring the patient comfort and, eventually, to discuss the practical and research applications of these instruments in the current health context.

Participants/population
We will include the whole range of ages (newborns, toddlers, children, teenagers, young adults, middle-age adults and elderly people).

Intervention(s), exposure(s)
No intervention

Comparator(s)/control
None

Main outcome(s)
The systematic review of the methodological quality and practical and research applications of each instrument developed to measure the patient comfort in the health care setting

Additional outcome(s)
None

Data extraction (selection and coding)
A reviewer will apply the inclusion criteria to all titles and abstracts. If no decision can be taken based solely on title and abstract alone, the full paper will be retrieved. Authors of eligible studies will be contacted to provide missing or additional data if necessary. Full-text inclusion criteria will be screened independently by two review authors. Discrepancies will be resolved through discussion (with a third author where necessary). A pre-piloted form will be used to extract data from the included studies to assess the study quality and to synthesize the evidence. Extracted information of each selected instrument will include: general information (author, year, country of origin and papers); instrument detail (outcome measures, purpose/use, number of items, response categories, scale design, type of patients); utility characteristics (theoretical/conceptual framework, validity tests conducted and results, reliability tests conducted and results, response rate, sample size, setting, respondents population and demographics, ease and usefulness of interpretation, cost-efficiency, level of expertise required for scoring, acceptability, time required to completion, mode of administration, acceptability by managers and users, educational impact). Two review authors will extract data independently, and discrepancies will be identified and resolved through discussion (with a third author where necessary).

Risk of bias (quality) assessment
To minimize the risk of bias during the methodological analysis of each instrument we will apply the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) (Mokkink, Terwee, Knol, et al. 2010; Mokkink et al. 2012; Terwee et al. 2012). With the same purpose, and to improve the quality of reporting of results of evaluation of health care instruments, we will apply the Quality Criteria for Measurement Properties (Terwee et al. 2007).

Strategy for data synthesis
We will initially categorise the instruments according to the theoretical model used to design and validate the instrument: reflective vs. formative. To decide whether the model is reflective or formative we will take into account the theoretical and practical considerations detailed in Coltman, Devinney, Midgley, & Venaik (2008). Second, we will apply the COSMIN checklist in four steps, and we will use the four-point scoring system (excellent, good, fair or poor) (Mokkink et al. 2010, 2012, Terwee et al. 2011, 2012). We will also apply the Quality Criteria for Measurement Properties for assessing the design, the methods and the outcomes of each instrument. These criteria consist of assessing the content validity, the internal consistency, the criterion validity, the construct validity, the reproducibility, the responsiveness, the floor and ceiling effects and the interpretability (Terwee et al. 2007).
Finally, we will assess the cost-efficiency, the acceptability and the educational impact of each instrument to discuss their utility in the real practice world, according to Van der Vleuten´s Utility Index Matrix (Van der Vleuten, 1996, cited in Beattie et al. 2015). The extracted information using the pre-pilot form and the results of the whole process will be showed in tables and they will be synthesized in a narrative way.

Analysis of subgroups or subsets
None

Contact details for further information
Sonia Lorente Sánchez
sonia.lorente@uab.cat

Organisational affiliation of the review
None

Review team members and their organisational affiliations
Assistant/Associate Professor Sonia Lorente. Universitat Autònoma de Barcelona
Dr Jaume Vives. Universitat Autònoma de Barcelona
Dr Josep Maria Losilla. Universitat Autònoma de Barcelona

Type and method of review
Systematic review

Anticipated or actual start date
26 October 2015

Anticipated completion date
19 December 2016

Funding sources/sponsors
This research was supported by Grant PSI2014-52962-P from the Spanish Ministry of Economy and Competitiveness.

Conflicts of interest
None known

Language
English

Country
Spain

Stage of review
Review Completed published

Details of final report/publication(s) or preprints if available
LORENTE S., LOSILLA J. M. & VIVES J. (2018) Instruments to assess patient comfort during hospitalization: a psychometric review. Journal of Advanced Nursing,74(5), 1001-1015.doi 10.1111/jan.13495
http://europepmc.org/abstract/med/29098701

DOI: 10.1111/jan.13495

Subject index terms status
Subject indexing assigned by CRD

Subject index terms
Emotions; Hospitalization; Humans

Date of registration in PROSPERO
11 March 2016

Date of first submission
13 June 2018

Stage of review at time of this submission

Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	Yes
Risk of bias (quality) assessment	Yes	Yes
Data analysis	Yes	Yes

The record owner confirms that the information they have supplied for this submission is accurate and complete and they understand that deliberate provision of inaccurate information or omission of data may be construed as scientific misconduct.

The record owner confirms that they will update the status of the review when it is completed and will add publication details in due course.

Versions

11 March 2016
21 September 2016
20 December 2016
19 June 2018

9.1.2. ESTRATEGIA DE BÚSQUEDA EN LAS BASES DE DATOS

CINAHL

521 TI (comfort* OR discomfort* OR uncomfort*) AND AB (theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission*) NOT (chair* OR ergonomic* OR thermal* OR cloth*) 1990 to 2015 Filters; Humans, English

GOOGLE

400 (n=20) All of these words: comfort* OR discomfort* OR uncomfort* Any of these words: theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission* None of these words: chair* OR ergonomic* OR thermal* OR cloth* 1990 to 2015. Filters; English

PROQUEST MEDLINE & THESES AND DISSERTATIONS

1302 ti (comfort* OR discomfort* OR uncomfort*) AND ab (theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission*) NOT (chair* OR ergonomic* OR thermal* OR cloth*) 1990 to 2015 Filters; Humans, English

PSYCINFO

610 results for Title: comfort* OR discomfort* OR uncomfort* AND Abstract: theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission* NOT Any Field: chair* OR ergonomic* OR thermal* OR cloth* AND Language: english AND Population Group:Human AND Year: 1990 To 2015

WEB OF SCIENCE (WoS) (CORE)

562 Tittle: (comfort* OR discomfort* OR uncomfort*) AND Tittle: (theor* OR questionnaire* OR instrument* OR measure* OR scale* OR assessment* OR hospital* OR admission*) NOT Tittle: (chair* OR ergonomic* OR thermal* OR cloth*) Language: (ENGLISH) Year: 1990-2015 Index: SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH.

9.1.3. CONSENSUS-BASED STANDARDS FOR THE SELECTION OF HEALTH MEASUREMENT INSTRUMENTS

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

9.1.4. QUALITY CRITERIA FOR MEASUREMENT PROPERTIES OF HEALTH INSTRUMENTS

Property	Definition	Quality criteria
1.Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	+ A clear description is provided of the measurement aim, the target population, the concepts that are being measured, and the item selection AND target population and (investigators OR experts) were involved in item selection. ? A clear description of above-mentioned aspects is lacking OR only target population involved OR doubtful design or method. - No target population involvement. 0 No information found on target population involvement.
2.Internal consistency	The extent to which items in a (sub)scale are inter-correlated, thus measuring the same construct	+ Factor analyses performed on adequate sample size (7 * # items and ≥100) AND Cronbach's alpha(s) calculated per dimension AND Cronbach's alpha(s) between 0.70 and 0.95. ? No factor analysis OR doubtful design or method. - Cronbach's alpha(s) <0.70 or >0.95, despite adequate design and method; 0 No information found on internal consistency.
3.Criterion validity	The extent to which scores on a particular questionnaire relate to a gold standard	+ Convincing arguments that gold standard is ''gold'' AND correlation with gold standard ≥0.70. ? No convincing arguments that gold standard is ''gold'' OR doubtful design or method. - Correlation with gold standard <0.70, despite adequate design/method. 0 No information found on criterion validity.
4.Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	+ Specific hypotheses were formulated AND at least 75% of the results are in accordance with these hypotheses. ? Doubtful design or method (e.g., no hypotheses). - Less than 75% of hypotheses were confirmed, despite adequate design/ methods. 0 No information found on construct validity
5. Reproducibility		
5.1 Agreement	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	+ MIC<SDC or MIC outside the LOA or convincing arguments that agreement is acceptable. ? Doubtful design or method or (MIC not defined AND no convincing arguments that agreement is acceptable). - MIC≥SDC or MIC equals or inside LOA, despite adequate design and method. 0 No information found on agreement.
5.2 Reliability	The extent to which patients can be distinguished from each other, despite measurement errors (relative measurement error)	+ ICC or weighted Kappa≥0.70. ? Doubtful design or method (e.g., time interval not mentioned). - ICC or weighted Kappa<0.70, despite adequate design and method. 0 No information found on reliability
6. Responsiveness	The ability of a questionnaire to detect clinically important changes over time	+ SDC or SDC< MIC or MIC outside the LOA or RR>1.96 OR AUC≥0.70. ? Doubtful design or method. - SDC or SDC≥MIC or MIC equals or inside LOA or RR≤1.96 OR AUC<0.70, despite adequate design/methods. 0 No information found on responsiveness.
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score	+ <15% of the respondents achieved the highest or lowest possible scores. ? Doubtful design or method; - >15% of the respondents achieved the highest or lowest possible scores, despite adequate design /methods. 0 No information found on interpretation.
8. Interpretability	The degree to which one can assign qualitative meaning to quantitative scores	+ Mean and SD scores presented of at least four relevant subgroups of patients and MIC defined. ? Doubtful design or method or less than four subgroups or no MIC defined. 0 No information found on interpretation.

Note: AUC= area under the curve; MIC= minimal important change; SDC=smallest detectable change; LOA=limits of agreement; ICC= Intraclass correlation; SD=standard deviation. Ratings: + positive rating; ? indeterminate rating; - = negative rating; 0 = no information available. Table from "Quality criteria were proposed for measurement properties of health status questionnaires", *Journal of Clinical Epidemiology*, 60(1), pp. 34–42, by Terwee et al. (2007)

9.1.5. MODELO TEÓRICO PARA EVALUAR MODELOS REFLECTIVOS Y FORMATIVOS

Considerations	Reflective model	Formative model
Theoretical considerations		
1.Nature of construct	Latent construct exists ✓ Latent construct exists independent of the measures used	Latent construct is formed ✓ Latent constructs is a combination of its indicators
2.Direction of causality between items and latent construct	Causality from construct to items ✓ Variation in the construct causes variation in the item measures ✓ Variation in item measures does not cause variation in the construct	Causality from items to construct ✓ Variation in the construct does not cause variation in the item measures ✓ Variation in item measures causes variation in the construct
3.Characteristics of items used to measure the construct	Items are manifested by the construct ✓ Items share a common theme ✓ Items are interchangeable ✓ Adding or dropping an item does not change the conceptual domain of the construct	Items define the construct ✓ Items need not share a common theme ✓ Items are not interchangeable ✓ Adding or dropping an item may change the conceptual domain of the construct
Empirical considerations		
4.Item intercorrelation	Items should have high positive intercorrelations ✓ Empirical tests: assessing internal consistency and reliability by Cronbach alpha, average variance extracted and factors loadings (e.g. from common or confirmatory factor analysis)	Items can have any pattern of intercorrelation but should possess the same directional relationship ✓ Empirical tests: no empirical assessment of indicator reliability is possible; various preliminary analyses are useful to check directionality between items and construct
5.Item relationships with construct antecedents and consequences	Items have a similar sign and significance of relationship with the antecedents/consequences as the construct ✓ Empirical tests: establishing content validity by theoretical considerations, assessing convergent and discriminant validity empirically.	Items may not have similar significance of relationships with the antecedents/consequences as the construct ✓ Empirical tests: assessing nomological validity by using a MIMIC model, and/or structural linkage with another criterion variable
6.Measurement error collinearity	Identifying the error in items is possible ✓ Empirical test: identifying and extracting measurement error by common factor analysis	Identifying the error is not possible if the formative measurement model is estimated in isolation ✓ Empirical test: using the vanishing tetrad test to determine if the formative items behave as predicted. Collinearity should be ruled out by standard diagnostics such as the condition index

9.1.6. CRITERIOS PARA EVALUAR LA FACTIBILIDAD DE LA MEDIDA

	Excellent (****)	Good (***)	Fair (**)	Poor (*)
Cost-Efficiency				
1. What are the number of observations (patients, raters, times) needed to reach the required level of reliability for the purpose of the instrument?	Only a small sample needed (<30)	A moderate sample size (30–49)	Not explicit but can be assumed or (50–99 assessments needed)	No details given or (≥100 assessments needed)
2. How long does an assessment take to complete.	≤15 min	≤30 min	30–60 min	>60 min
3. What are the administrative costs of completing the assessment?	Easily embedded within existing resource. Little additional support required	Some administrative resource but no specialist resource required	Large amount of resource to assess and administer	Significant specialist expertise and administrative time required to assess and administer
Acceptability				
1. Is there evidence of subjects understanding of the instrument/assessment?	Investigations of subjects understanding (i.e. cognitive testing of instruments)	Estimated evidence of subjects understanding (i.e. high number of questions missed)	Subject understanding not explicitly stated but some can be assumed (i.e. student guide to OSCE)	No evidence of subject understanding
2. How many assessments are not completed?	There are low numbers of missing items (<10 %) & adequate response rates (>40 %)	There are a high number of missing items (>10 %) & adequate response rates (>40 %)	There are low numbers of missing items or poor (<10 %) & inadequate response rate (<40 %)	There are high numbers of missing items (>10 %) & poor response rates (<40 %)
3. Has the instrument/assessment been tested in an appropriate context?	Evidence of successful administration/use within an appropriate setting	Tested in vivo and changes recommended would be achievable	Testing in vivo and changes recommended would be difficult or only partially tested in vivo	Testing has only been conducted in vitro/simulation
Educational Impact				
1. There is evidence of the instruments intended purpose being achieved (i.e. if aim is to enable hospital ranking for patient selection, is there evidence that the results are actually influencing patient choice?)	Clear evidence of intended purpose being fulfilled	Explanatory or theoretical link between intended and actual use but no clear evidence	Evidence of theoretical work but relationship between intended and actual purpose poorly or not described	No evidence of intended purpose becoming actual
2. The scoring system is easily translated or available in an easy to use format?	Explicitly stated and easy to calculate Feedback	Explicitly stated but not easy to calculate	Scoring only calculated by resource with statistical knowledge	Scoring not explained well enough to calculate
3. The feedback from the results can be readily used for action where necessary?	Feedback is readily available in a format that enables necessary action	Feedback is readily available but not drilled down enough to enable targeted action	Minimal feedback available or delay results in limited impact	No explanation to determine adequacy of feedback. No direct feedback could be readily used without additional expertise

Note: Table from "Instruments to measure patient experience of healthcare quality in hospitals: a systematic review" *Systematic Reviews*, 3(1), p. 4 by Beattie et al. (2015)

9.1.7. CARACTERÍSTICAS DE LOS INSTRUMENTOS

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Ambuel et al. (1992) U.S.A	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	37 children, 50 observations in PICU, with cardiac respiratory diseases. Female 20, male 17; mean age 37.1 months (sd 52.1)	Two factors: Behavioural and Physiological distress (84%) Comparison by correlations CS with VAS r=.75 (p<.01)	a=.90	Inter-rater r= .84	
Blauer (1996) U.S.A	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	33 newborns, 1428 observations in NICU. Age 24-40 weeks, weight 600- 4000 grams				Scores at 2 minutes before (T0), during (T1), and 3 min after procedure (T2), Anova. Intubation: T0-T1/T2-T3(p<.001) Intravenous insertion: T0-T1/T2-T3(p<.001)(p<.01) Endotracheal suctioning: T0-T1/T2-T3(p<.01)(p<.05) Diaper change: T0-T1 (p<.01)
Bruno de Carvalho et al. (1999) Brazil	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	18 children, 30 observations in PICU, with respiratory and infectious diseases. Mean age 16,4 months (sd 17,2), mechanically ventilated,with opiates/benzodiazepines, and arterial line for MAP	Comparison by Kappa CS with Hartwig k= .35			
Van Djik et al.(2000) NED	Comfort Scale (CS)	Reflective IRT/CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	158 neonates, with congenital anomalies, under surgery, mechanically ventilated in NICU. Male (59%), age 0-3 years, weight >1500grams	Three latent variables by Rasch: Behavioural Comfort,MAP,HR. Comparison by Kappa, CS with VAS, postoperative T1(3h) k=.96 T2(6h) k=.89 T3(9h) k=.90	T1 a=.90 T2 a=.92 T3 a=.92	Inter-rater Kw= .70	
Courtman et al. (2003) U.K.	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	40 children, 373 observations in PICU. Neurologically normal 25; with neurological focus 15. Mean age 3.9 years (sd 4.5). Male: female=25:15. Ventilated, with midazolam	Comparison by correlations CS with BIS total group r=.50 CS with BIS normal r=.51 CS with BIS neurologic r=.26			
Wielenga et al.(2004) NED	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	19 preterm, 30 observations, NICU. Female 9, male 10. Mean gestional age 30 weeks, mean birth weight 1385 grams.	Comparison by Pearson CS with expert opinion r=.84 Cutoff point 20. AUC= .95 Sensivity 1.00 Specificity .77		Inter-rater Kw= .84 ICC= .94	

Note: CTT: Classical Test Theory; IRT: Item Response Theory; PICU: Paediatric Intensive Care Unit; a: Cronbach's alpha; r: correlation coefficient; Kw: Weighted Kappa; VAS: Visual Analogic Scale; NICU: Neonatal Intensive Care Unit; BIS: Bispectral Index; MAP: Mean Arterial Pressure; HR: Heart Rate; AUC: Area Under Curve; U.S.A: United States of America; U.K: United Kingdom; NED: Netherlands

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Ista et al. (2005) NED	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	78 children, 843 observations, in PICU, cardiac-respiratory diseases. Female 29, male 49, median age 17 months. Most ventilated, with midazolam.	Comparison by Kruskal-Wallis CS with NISS categories (insuficient sedation, adequate sedation, oversedation) p <.01	CS a=.78 CS deleted MAP&HR a=.84	Inter-rater Kw=.77-1.00 ICC=.99	
Triltsch et al.(2005) Germany	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	40 children, PICU, most cardiac. Female 19, male 21, mean age 5.6 months(sd 21 days-16 years)Ventilated, with opioids and benzodiazepines	Comparison by Spearman CS with BIS total r=.65 CS with BIS <6months r=.78 CS with BIS >6months r=.47			
Bear & Ward-Smith (2006) U.S.A	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	55 children, PICU, most postoperative. Female 30, male 25. Mean age 27 months (ranged 1 month to 18 years)		a=.85	Inter-rater r=.79	
Froom et al.(2008) U.K	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	19 children, 28 observations, PICU, most respiratory and cardiac diseases. Female 9, male 10, mean age 1.5 years (sd 0.5-9.7), ventilated, with midazolam and tiopental	Comparison between light, moderate and deep sedation Comfort scores p < .001 Comparison by Spearman CS with BIS Unestimulated period r=.60 Stimulated period r=.30			
Lamas et al.(2008) Spain	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	77 children, 234 observations in PICU, most cardiac surgery Female 32, male 45, median age 8 months (15 days to 228 months). Ventilated, sedation with midazolam and fentanyl	Comparison by Spearman CS with BIS r=.48 CS with AEP r=.53 CS with Ramsay r=.73			

Note: CTT: Classical Test Theory; IRT: Item Response Theory; PICU: Paediatric Intensive Care Unit; a: Cronbach's alpha; r: correlation coefficient; Kw: Weighted Kappa; ICC: Intraclass Correlation; NISS: Nurse Interpretation of Sedation Score; MAP: Mean Arterial Pressure; HR: Heart Rate; NICU: Neonatal Intensive Care Unit; BIS: Bispectral Index; AEP: Auditory-Evoked Potential index; NED: Netherlands; U.K: United Kingdom

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Franck et al. (2011) U.S.A	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	81 patients, most with cardiac malformations, NICU. Mean birth weight 3.3 kg(sd .5), mean gestational age 39.6 weeks (sd 1.4). Sedated, ventilated.	Two factors: Behavioural and Physiological distress (60%)		Inter-rater ICC=.84	Scores prior and 3minutes after different painful and care procedures, p< .001
Cury et al.(2013) Brazil	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	16 neonates with congenital cardiac disease, ventilated. Median age 60days (4-410) Female 56%. Sedation with midazolam and fentanyl.	Comparison by Pearson CS with CAAS baseline r=.80 CS with CAAS painful r=.70 Comparison by Kappa CS with CAAS cutoff 3 k=.37 CS with CAAS cutoff 4 k=.73			Scores before (T0) and after tracheal suctioning (T1): 1st day after cardiac surgery T1 vs T0 p=.02 2nd and 3rd day after surgery T1 vs T0 p=.9 and p=.36
Nievas et al.(2014) U.S.A	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	11 children, 111 observations, PICU, respiratory diseases, ventilated. Female 4, male 7, age 1month to 13 years.	Comparison by regression CS with SNAP II r= .02 Comparison by Bland-Altman Limits of Agreement = 95%			
Tschiedel et al.(2015) Germany	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	30 children, 204 observations, under muscle biopsy. Female 7, male 23, age 9months-18 years.	Comparison by Pearson CS with BIS r=. 58			
Lee&Young (2005) China	Modified Comfort Scale (CS)	Reflective CTT	7/Likert 5p Rating 7-35 No “Muscle Tone” item	By trained professionals 2 minutes	15 neonates, 27 observations, NICU. Age 23-54 weeks, weight 560-3740 grams. Ventilated, sedation with fentanyl or midazolam		CS a=.80 CS deleted MAP a=.83	Inter-rater r =.76 r =.62 r =.77	
Caljouw et al.(2007) NED	Adapted Comfort Scale (CS)	Reflective CTT	7/Likert 5p Rating 7-35 No “MAP” item	By trained professionals 2 minutes	57 neonates, NICU. Female 26, male 31,mean age 3.3days (sd (1.8)	Comparison by Pearson CS with VAS pre-post test: Pre heel puncture r=.09 to .49 Post heel puncture r=.44 to .74 Cut-off point 17. AUC .97 Sensivity .93 Specificity .80	Pre-test a=.76 Post-test a=.86	Inter-rater Pre ICC=.85 Post ICC=.93	Scores 10 minutes before and 1 minute after the heel puncture: Mean scores increased from 13.2 (sd 2.7) to 23.3 (sd 5.5)
Gjerstad et al.(2008) Norway	Modified Comfort Scale (CS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR” items	By trained professionals 2 minutes	20 children, PICU. Female 11, male 9, aged 0-11 years. Ventilated, sedation with midazolam and thiopental	Comparison by regression CS with Skin Conductance T1 2minutes before tracheal suctioning r=.61 T2 from during to after the tracheal suctioning r=.46			

Note: CTT: Classical Test Theory; IRT: Item Response Theory; a: Cronbach’s alpha; r: correlation coefficient; ICC: Intraclass Correlation; CAAS: Cardiac Analgesic Assessment Scale; PICU: Paediatric Intensive Care Unit; NICU: Neonatal Intensive Care Unit; SNAP II: Index derived from a single-lead electroencephalogram device; BIS: Bispectral Index; MAP: Mean Arterial Pressure; VAS: Visual Analogic Scale; AUC: Area Under Curve; U.S.A: United States of America; NED: Netherlands

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Van Djik et al.(2009) NED	Comfort Neo Scale (CNeo)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR” items	By trained professionals 2 minutes	286 neonates and prematures, 3600 observations,NICU, most respiratory diseases. Female 112, male 174, mean gestional ages 24.6-42.6 weeks. Most ventilated, with opioids, benzodiazepines, barbiturates	Comparison by Pearson CNeo with NRS pain r=.54 CNeo with NRS distress r=.83 Cut-off points 14. Sensivity .81 Specificity .90	Ventilated a=.84 No ventilated a=.88	Inter-rater Kw=.79	(T1)Scores before and after intervention p<.001 (T2) Scores in situations suspected for pain p<.001 (T3) Scores in suspected over sedation p<.001
Johansson & Kokinsky (2009) Sweden	Comfort Behavioural Scale (CB)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	40 children, 119 observations, PICU, most with cardiac malformations. Female18, male 22, mean age 4 months. Ventilated, with midazolam	Comparison by Kruskal-Wallis CB with NISS r=.57 CB with VAS r=.49 CB with FLACC r=.76		Inter-rater Kw=.71	
de Jong et al.(2010) NED	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	154 children with burns, most scalds. Female 53, male 101, mean age 20 months (sd 11)	Comparison by Spearman CBS with POCIS T1 Background pain r=.45 T2 Procedural pain r=.88	T1 a=.77 T2 a=.86	Inter-rater T1 ICC=.83 T2 ICC=.82	T1 1 hour before wound care T2 Just after the wound care T1 vs T2 p<.001
Amigoni et al.(2012) Italy	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	46 children, most respiratory, neurological,sepsis diseases. PICU. Female 17, male 29. Mechanically ventilated with invasive devices.	Comparison by Kappa CBS with BIS CBSdoctor with BIS k=.13 CBSnurses with BIS k=.08		Inter-rater k= .43	CBSdoctor and CBSnurses before tracheal suctioning 9.22±3.2 and 9.28±2.5 CBSdoctor and CBSnurses after tracheal suctioning 15.24±5.4 and 14.4±5.6
Bai et al.(2012) China	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	170 children, most surgical, in PICU. Female 72, male 98. Median age 8 months (0.5-72), with opioids or morphine	Comparison by Pearson CBS with FLACC r= .51 CBS with VAS r= .31 Cut-off CBS 13. AUC .93 Sensivity .86 Specificity .83			
Boerlage et al.(2012) NED	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes(min) 30 seconds(sec)	80 children, 236 observations, PICU, 75% were surgical	Sensivity, + predictive values CBS 30sec =.44 CBS 2min= .80 Specificity, - predictive values CBS 30sec=.97 CBS 2min= .88		Inter-rater CBS2min vs CBS30sec by t test p<.001	
de Jong et al.(2012) NED	Comfort Behavioural Scale (CBS)	Reflective IRT/CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	154 children with burns, 3884 observations, most scalds. Female 53, male 101, mean age 20 months (sd 11)	Unidimensional, by Rasch			

Note: CTT: Classical Test Theory; IRT: Item Response Theory; a: Cronbach’s alpha; Kw: Weighted Kappa; k: Cohen´s kappa; ICC: Intraclass Correlation; MAP: Mean Arterial Pressure; HR: Heart Rate; NRS: Numerical Rating Scale; NICU: Neonatal Intensive Care Unit; PICU: Paediatric Intensive Care Unit; NISS: Nurse Interpretation of Sedation Score; FLACC; Face, Legs, Activity, Cry and Consolability; VAS: Visual Analogic Scale; POCIS: Pain observation scale for young children; BIS: Bispectral Index; NED: Netherlands

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability	Responsiveness
Valkenburg (2012) NED	Comfort Behavioural Scale (CBS)	Reflective IRT/CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	76 children with Down, most surgical, male 45, mean age 81 days (42-273). 466 children control group, male 273, mean age 119 days (22-355).Total of 7439 observations in PICU.	Unidimensional, by Rasch Comparison by Pearson CBS with NRS Down r= .45 CBS with NRS Control r= .57 Cut-off point CBS Down 17 Sensivity .82 Specificity .92	Down a= .84 a= .86	
da Costa Silva et al. (2013) Brazil	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	11 children, 35 observations, PICU, most with bronchiolitis. Female 4, male 7, mean age 7 months (4-36), ventilated, with midazolam and fentanyl	Comparison by Pearson CBS with BIS, four raters r= .49 ; r= .42 ; r= .51; r= .48	Interrater from k= .56 to k= .75	
Tristão et al.(2013) Brazil	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	36 neonates, in maternity ward. Apgar 9.2 (sd 0.7), female 17, male 19, gestional ages 37-41 weeks, mean 38.9 weeks	Comparison by Spearman CBS with SCA, at 15, 30, 180 sec after heel puncture: r= .42; r= .38; r=.50		Scores 3 min before (T0), 3 min after (T1), 6 min after heel puncture (T2): T0 vs. T1 p<.001 T1 vs. T2 p<.01
Boerlage et al.(2014) NED	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	180 children, 747 observations in PICU, ventilated, most surgical. Female 73, male 107, median age 0.4years(0.1-2.0)		Interrater kw = .79 ICC= .96	Scores before-after analgesia or sedation. In 63% observations the scores decreased at least 5 points; p<.001
Andersen et al.(2015) Norway	Comfort Behavioural Scale (CBS)	Reflective CTT	6/Likert 5p Rating 6-30 No “MAP” “HR”items	By trained professionals 2 minutes	45 children, 307 observations, under minor surgery. Male 31.7%, mean age 24.8months (sd 7.0), ranged 12-36 months	Difference between baseline-pain/distress p=.039 baseline-light sedation p<.001 light-deep sedation p= .002	Interrater from kw= .75 to kw =.96 ICC =.96	
Moriber (2009) U.S.A	Pediatric Perioperative Comfort Instrument (PPCI)	Reflective IRT/CTT	7/Likert 5p No reported	By trained professionals No reported	150 children, outpatient surgery. Female 42.7%, male 57.3%, mean age 4.3 years (sd 2.2). Most without preoperative sedation.	Unidimensional, by Rasch Comparison by Spearman PPCI pre with YPAS r=-.96 PPCI post with FPS r=-.71	Pre a=.98 Post a=.93	

Note: CTT: Classical Test Theory; IRT: Item Response Theory; a: Cronbach’s alpha; Kw: Weighted Kappa; k: Cohen´s kappa; MAP: Median Arterial Pressure; HR: Heart Rate; PICU: Paediatric Intensive Care Unit;; SCA: Skin Conductance Activity YPAS; Yale Preoperative Anxiety Scale; FPS: Faces Pain Scale; NRS: Numerical Rating Scale; BIS: Bispectral Index; U.S.A: United States of America; NED: Netherlands

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Jensen et. al.(1991) U.S.A	Pain Discomfort Scale (PDS)	Reflective CTT	10/Likert 5p Rating 0-40	Self reported or by interview Brief	59 patients with chronic pain. Female 40, male 19; mean age 49.3 years (sd 14.5)	Two factors: Pain intensity and Pain affect (72%) Comparison by correlations PDS with Beck r=.58 PDS with MPQ r=.38	a=.77	Test-retest T1 r= .64 T2 r= .76	
Kolcaba (1992) U.S.A	General Comfort Questionnaire (GCQ)	Formative Reflective CTT	48/Likert 5p Rating 48-240	Self reported No reported	256 patients >21years, from medical,surgical, psychiatric, acute care oncology.	Twelve factors in three subscales: Ease, Relief, Transcendence (63.4%)	a=.88		
Yen (1994) U.S.A	Psychosocial Comfort	Reflective IRT/CTT	21/Dichotomy No reported	Self reported No reported	1667 patients, 61% surgical. Female 950, male 717. Age 19-96 years	Unidimensional, by Rasch	a=.79		
Betemps (1999) U.S.A	Psychiatric Discomfort Scale (PDS)	? CTT	23/Likert 5p No reported	Self-reported No reported	292 patients from a VA, with schizophrenia or schizoaffective disorder	Three factors: Disordered thinking, Irritability,Akathisia Comparison Inpatients and Outpatients scores, Anova (F= 4.07,df=54.47, p<.001)	a= .93		Scores at admission, at 3rd day, and at discharge, by Anova (F=3.23,df =2.52,p =.047)
Durnell (2003) U.S.A	Childbirth Comfort Questionnaire (CCQ)	Formative Reflective CTT	14/Likert 5p Rating 14-70	Self reported 5 minutes	64 primigravida women, 18-40 years, admitted when 2-3cm dilated, the 71% with pitocin and electronic fetal monitoring.	T1 Labor phase (64.8%) T2 Active phase (67%) Comparison by correlations CCQ with NRS (T1) r=.28 CCQ with NRS (T2) r=.55	T1 a=.67 T2 a=.75		Between T1-T2, with and without comfort measures: Women with *Freedom of movement* and *Massage* showed higher levels of comfort (p<.0001/p<.05)
Alves-Apóstolo et al.(2007) Portugal	Psychiatric In-Patients Comfort Scale (PICS)	Formative Reflective CTT	38/Likert 5p No reported	Self reported or by interview. No reported	322 psychiatric patients (schizophrenia, schizotypal, delusional disorders), female 148, male125. Mean age 41 years (sd 14,4)	Three factors: Relief, ease, Trascendence (38.6%) Comparison by Pearson PICS with ISSEI r=-.55 PICS with SPWB r=.47	a=.89		
Kalfon et al.(2010) France	Inconforts des Patients de REAnimation (IPREA)	Formative IRT/CTT	16/VAS Rating 0-100	Self reported or by interview 10±7minutes	868 patients, most surgical, ICU. Female 330, male 538, mean age 55.7 years (sd 19.1) Mechanically ventilated 439, non-invasive ventilation 153.	Unidimensional, by Rasch	a=.78	Inter-rater ICC=.91	

Note: CTT: Classical Test Theory; IRT: Item Response Theory; a: Cronbach's alpha; ICC: Intraclass correlation; r: correlation coefficient; Beck: Beck Depression Inventory; MPQ: McGill Plain Questionnaire; VA: Department of Veterans Affairs; NRS: Numerical Rating Scale; ISSEI: Inventory of Subjective Suffering Experiences in Illness; SPWB: Scale of Psychological Well-Being; ICU: Intensive Care Unit; U.S.A: United States of America.

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Ashkenazy& DeKeyser-Ganz (2011) Israel	Comfort Scale (CS)	Reflective CTT	8/Likert 5p Rating 8-40	By trained professionals 2 minutes	88 patients, ICU, with sepsis, respiratory failure or surgery. Female 31, male 57, mean age 58 years (sd 22)most ventilated.	Comparison by Pearson CS with VAS r=.63 Comparison by Kappa on first middle, last day intubation CS with RASS r=.65,r=.61,r=.74 CS with SAS r=.68,r=.56,r=.73 CS with RS r=.58,r=.34,r=.63	a=.66	Test-retest ICC=.47	Scores of 12 patients before and after suction, position and dress change, bath p= .052
Hurley et al. (1992) U.S.A	Discomfort in Alzheimer Type (DS-DAT)	Formative Reflective CTT	9/Likert 4p Rating 0-27	By trained professionals 5 minutes	82 patients with DAT in VA. Female 5, male 77, dependent of nursing staff, scoring 5.6 out of 6 in ADL		a=.86 a=.89	Inter-rater T1 r= .98 T2 r= .89 T3 r= .86 T4 r= .91	Scores before, during, and after peak of fever, by Anova F(1,19)= 167.02 (p<.01)
Morrison et al. (1998) U.S.A	Discomfort Scale	Formative CTT	16/Likert 5p No reported	By interview No reported	165 patients with Alzheimer, 3 groups (100, 30, 35). Mean age 65, 73, 55 years. Most female.	Comparison between pain-discomfort during procedures, by Anova F=19.7 (p<.001)		Inter-rater r=.75	
Novak et al. (2001) U.S.A	Hospice Comfort Questionnaire (HCQ)	Formative Reflective CTT	49/ Likert 6p Rating 49-294 49/ Likert 4p Rating 49-196	By interview 12 minutes	206 dayds (103 patients/103 caregivers). Patients were female/male 1:1, most with respiratory/digestive terminal diseases or other cancer	Comparison by correlations HQC 6p with TCv r=.45 to .48 HQC 4p with TCh r=.31 to .45	6p a=.98 4p a=.83	Inter-rater TCv ICC=.64 TCh ICC=.42	
Volicer et al. (2001) U.S.A	Comfort Assessment Dying with Dementia (CAD_EOLD)	Formative Reflective CTT	14/Likert 3p Rating 14-42	By interview to caregivers No reported	159 caregivers, white 97%, female (75%). 159 patients female (45%), male (55%), mean age of die 81.2 years (sd 7.5), with severe dementia, and 42% completely dependent.	Four factors: Physical and Emotional distress, Dying Symptoms,WellBeing (64.7%) Comparison by correlations. CAD-EOLD with SWC-EOLD CAD-EOLD with SM-EOLD Total scores (p<.005)	a=.86		
Kiely et al.(2006) U.S.A	Comfort Assessment Dying with Dementia (CAD-EOLD)	Formative Reflective CTT	14/Likert 3p Rating 14-42	By interview to caregivers No reported	189 dyads in Nursing Homes. Caregivers, most female, mean age 59.7 years. Patients, most female, white and mean age 84.8 years (sd 7.9)	Comparison by correlations CAD-EOLD with QUALID Pearson r = -.50 Spearman r = -.50	a=.82		
Stevenson et al. (2006) U.S.A	Discomfort Behaviour Scale (DBS)	Formative Reflective IRT/CTT	17/Likert 6p Rating 0-102	Professionals, by dataset. No reported	29120 patients with dementia, three groups (9960,9665,9672), with moderate-severe cognitive impairment. Female 69%, male 31%, mean age 84.4 years (sd 10.7), ranged 18 to 111 years.	Unidimensional (42.4%)	a=.77		

Note: CTT: Classical Test Theory; IRT: Item Response Theory; a: Cronbach's alpha; ICC: Intraclass correlation; k: Cohen´s kappa; ICU: Intensive Care Unit; VAS: Visual Analogic Scale; RASS: Richmond Agitation Sedation Scale; SAS: Sedation Agitation Scale; RS: Ramsay Scale; TCv: Total Comfort Line vertical line; TCh: Total Comfort horizontal line; SWC-EOLD: Satisfaction with Care at End-of-Life in Dementia; SM-EOLD: Symptom Management at the End-of-Life in Dementia; QUALID: Quality of Life in Late-Stage Dementia; U.S.A; United States of America

Author	Questionnaire	Framework	Items/ Rating	Administration	Patients	Construct (% variance)	Reliability		Responsiveness
Dello Russo et al.(2008) Italy	Discomfort Scale-Dementia of Alzheimer Type in Italian (DS-DAT)	Formative Reflective CTT	9/Likert 4p Rating 0-27	By trained professionals 5 minutes	71 patients with severe dementia, female 53, male 18, mean age 81.5 years (sd 7.7)	Comparison between patients with low level of discomfort (mean score 6) and high level of discomfort (mean score 19) p < .0001	a=.81	Inter-rater k=.84	
Tanatwanit et al.(2011) Thailand	Hospice Comfort Questionnaire (HCQ)	Formative Reflective CTT	49/Likert 6p Rating 49-294	Self reported or by interview 42±15minutes	111patients with cancer, female 47.7%, male 52.3%. Mean age 68 years (sd 6.1). Most married with adequate lucidity and capacity to participate in study.	Comparison by correlations HCQ with VRS Pearson r=.55 Kendall´s r=.35 Spearman r=.47	a=.89		
Cohen-Mansfield et al.(2013) U.S.A	Source of Discomfort Scale (SODS)	Formative CTT	20/Dichotomy No reported	By interview No reported	179 patients with dementia, MMSE, Mini Mental State Examination 8.8 (sd 6.4). Female 72%, mean age 86.0years (sd 8.2). Most caucasian and widowed.	Comparison by Pearson SODS total with PAINE r=.28 SODS position with PAINE r=.42		Inter-rater r=.51	
Oliveira et al.(2016) Portugal	End of Life Comfort Planning Questionnaire	Reflective CTT	28/Likert 6p Rating 28-168	Self reported or by interview 8 minutes	141 patients with incurable disease in palliative care unit. Male 60.6 %, mean age 59.9 years, ranged 21 to 91 years most married, retired, catholic.	Five factors (57.3%) Comparison by Pearson End of Life with HCQ r=.74	a=. 84		

Note: CTT: Classical Test Theory; IRT: Item Response Theory; a: Cronbach's alpha; r: correlation coefficient; VRS: Verbal Rating Scale; PAINE; Pain Assessment in Non-Communicative Elderly Persons; U.S.A; United States of America

9.2. ANEXOS RELACIONADOS CON LA META-REVISIÓN

9.2.1. REGISTRO PROSPECTIVO DEL PROTOCOLO DE LA META-REVISIÓN

Systematic review

1. * Review title.

Give the title of the review in English

Tools to assess the measurement properties of quality of life instruments: a meta-review protocol

2. Original language title.

For reviews in languages other than English, give the title in the original language. This will be displayed with the English language title.

3. * Anticipated or actual start date.

Give the date the systematic review started or is expected to start.

10/05/2017

4. * Anticipated completion date.

Give the date by which the review is expected to be completed.

31/12/2018

5. * Stage of review at time of this submission.

Tick the boxes to show which review tasks have been started and which have been completed. Update this field each time any amendments are made to a published record.
Reviews that have started data extraction (at the time of initial submission) are not eligible for inclusion in PROSPERO. If there is later evidence that incorrect status and/or completion date has been supplied, the published PROSPERO record will be marked as retracted.
This field uses answers to initial screening questions. It cannot be edited until after registration.

The review has not yet started: No

Review stage	Started	Completed
Preliminary searches	Yes	Yes
Piloting of the study selection process	Yes	Yes
Formal screening of search results against eligibility criteria	Yes	Yes
Data extraction	Yes	Yes
Risk of bias (quality) assessment	Yes	Yes
Data analysis	Yes	Yes

Provide any other relevant information about the stage of the review here.

6. * Named contact.

The named contact is the guarantor for the accuracy of the information in the register record. This may be any member of the review team.

Mrs Sonia Lorente

Email salutation (e.g. "Dr Smith" or "Joanne") for correspondence:

Mrs Lorente

7. * Named contact email.

Give the electronic email address of the named contact.

sonia.lorente@uab.cat

8. Named contact address

Give the full institutional/organisational postal address for the named contact.

Sonia Lorente

Universitat Autònoma de Barcelona.

Campus de la UAB · 08193 Bellaterra

(Cerdanyola del Vallès) · Barcelona · Spain

9. Named contact phone number.

Give the telephone number for the named contact, including international dialling code.

10. * Organisational affiliation of the review.

Full title of the organisational affiliations for this review and website address if available. This field may be completed as 'None' if the review is not affiliated to any organisation.

Universitat Autònoma de Barcelona

Organisation web address:

https://www.uab.cat/psicologia/

11. * Review team members and their organisational affiliations.

Give the personal details and the organisational affiliations of each member of the review team. Affiliation refers to groups or organisations to which review team members belong. **NOTE: email and country now MUST be entered for each person, unless you are amending a published record.**

Mrs Sonia Lorente. Universitat Autònoma de Barcelona
Dr Jaume Vives. Universitat Autònoma de Barcelona
Dr Carme Viladrich. Universitat Autònoma de Barcelona
Dr Josep-Maria Losilla. Universitat Autònoma de Barcelona

12. * Funding sources/sponsors.

Details of the individuals, organizations, groups, companies or other legal entities who have funded or sponsored the review.

This research was supported by Grant PSI2014-52962-P from the Spanish Ministry of Economy and Competitiveness.

Grant number(s)

State the funder, grant or award number and the date of award

13. * Conflicts of interest.

List actual or perceived conflicts of interest (financial or academic).

None

14. Collaborators.

Give the name and affiliation of any individuals or organisations who are working on the review but who are not listed as review team members. **NOTE: email and country must be completed for each person, unless you are amending a published record.**

15. * Review question.

State the review question(s) clearly and precisely. It may be appropriate to break very broad questions down into a series of related more specific questions. Questions may be framed or refined using PI(E)COS or similar where relevant.

To identify systematic reviews assessing the measurement properties of Health-Related Quality of Life (HRQoL) instruments.

To identify the main tools applied to assess the measurement properties of HRQoL instruments.

To describe the most relevant characteristics of the tools applied to assess the measurement properties of HRQoL instruments (validity, reliability, feasibility, etc.).

To identify the measurement standards upon which these tools were developed or conform to, comparing their similarities and differences.

To appraise how authors of these systematic reviews include the assessment of the quality of HRQOL in their results and conclusions.

16. * Searches.

State the sources that will be searched (e.g. Medline). Give the search dates, and any restrictions (e.g. language or publication date). Do NOT enter the full search strategy (it may be provided as a link or attachment below.)

A systematic review will be performed in PubMed, US National Library of Medicine, by National Center for Biotechnology Information (NCBI); CINAHL, Cumulative Index to Nursing and Allied Health Literature, by EBSCOhost; PsycINFO, Psychological Information, by APA PsycNET; Scopus by Elsevier; WOS (Web of Science CORE) by Thomson Reuters, and COSMIN database by COSMIN Initiative (http://www.cosmin.nl/). In addition, ProQuest Dissertations &Theses Global will be used for searching grey literature, and search alerts in all databases will be set. The search strategy will follow the Peer Review of Electronic Search Strategies (PRESS) guidelines, and will consist of 3 filters composed of search terms for the following: (1) systematic review methodology; (2) HRQoL instruments; and (3) measurement properties. The latter filter

was developed for the VU University Medical Center for finding studies on measurement properties of measurement instruments. All filters will be adapted for all databases. The systematic search will be performed in July, 2018, limited by time and language (English).

17. URL to search strategy.

Upload a file with your search strategy, or an example of a search strategy for a specific database, (including the keywords) in pdf or word format. In doing so you are consenting to the file being made publicly accessible. Or provide a URL or link to the strategy. Do NOT provide links to your search **results**.

Alternatively, upload your search strategy to CRD in pdf format. Please note that by doing so you are consenting to the file being made publicly accessible.

Do not make this file publicly available until the review is complete

18. * Condition or domain being studied.

Give a short description of the disease, condition or healthcare domain being studied in your systematic review.

The quality of health status and the quality of life instruments is essential to obtain accurate diagnoses and to assess the efficacy or effectiveness of a specific intervention in health care. Evaluating and improving the quality of life is considered, as well, a public health priority, and because of this the present meta-review is focused on systematic reviews that appraised the measurement properties of HRQoL instruments.

19. * Participants/population.

Specify the participants or populations being studied in the review. The preferred format includes details of both inclusion and exclusion criteria.

We will include the whole range of ages (newborns, toddlers, children, teenagers, young adults, middle-age adults and elderly people) in any healthcare setting.

20. * Intervention(s), exposure(s).

Give full and clear descriptions or definitions of the interventions or the exposures to be reviewed. The preferred format includes details of both inclusion and exclusion criteria.

None

21. * Comparator(s)/control.

Where relevant, give details of the alternatives against which the intervention/exposure will be compared (e.g. another intervention or a non-exposed control group). The preferred format includes details of both inclusion and exclusion criteria.

None

22. * Types of study to be included.

Give details of the study designs (e.g. RCT) that are eligible for inclusion in the review. The preferred format includes both inclusion and exclusion criteria. If there are no restrictions on the types of study, this should be stated.

Systematic reviews aiming to report or to assess the measurement properties of instruments evaluating the

quality of life within the context of health and disease, namely Health Related Quality of Life (HRQoL) instruments, including all studies examining at least one or more measurement properties of a HRQoL instrument. Systematic reviews that include the full results report and detailed information about the instruments used to assess the measurement properties.

23. Context.

Give summary details of the setting or other relevant characteristics, which help define the inclusion or exclusion criteria.

To study the characteristics of tools assessing the measurement properties of HRQoL instruments in systematic reviews, and to compare the measurement standards upon which these tools were developed or conform to, with examples found in Viladrich & Doval (2006): Attributes and Criteria to assess Health Status and Quality of Life Instruments, the Standards for Educational and Psychological Measurement, or the Health Status Measures in Economic Evaluation.

24. * Main outcome(s).

Give the pre-specified main (most important) outcomes of the review, including details of how the outcome is defined and measured and when these measurement are made, if these are part of the review inclusion criteria.

Identification of the main specific tools applied to assess the measurement properties of HRQoL instruments and comparison of their most relevant characteristics.

Identification and comparison of the measurement standards upon which these tools were developed.

Appraisal of how authors of the systematic reviews include the assessment of the quality of the HRQoL instruments in their results and how they use this evaluation to come to an overall conclusion regarding the quality of each instrument.

Appraisal of how this conclusion concern to the pychometric properties evaluated as well as the relation stablished, if any, with the purpose of the HRQOL instrument (e.g. evaluative)

* Measures of effect

Please specify the effect measure(s) for you main outcome(s) e.g. relative risks, odds ratios, risk difference, and/or 'number needed to treat.

Not applicable.

25. * Additional outcome(s).

List the pre-specified additional outcomes of the review, with a similar level of detail to that required for main outcomes. Where there are no additional outcomes please state ‘None’ or ‘Not applicable’ as appropriate to the review

None.

* Measures of effect

Please specify the effect measure(s) for you additional outcome(s) e.g. relative risks, odds ratios, risk

difference, and/or 'number needed to treat.

Not applicable.

26. * Data extraction (selection and coding).

Describe how studies will be selected for inclusion. State what data will be extracted or obtained. State how this will be done and recorded.

References identified by the search strategy will be entered into Mendeley bibliographic software, and duplicates will be removed. Titles and abstracts will be screened independently by two reviewers. When decisions are unable to be made from title and abstract alone, the full paper will be retrieved. Full-text inclusion criteria will be screened independently by two reviewers. Discrepancies during the process will be resolved through discussion (with a third reviewer where necessary).

Extracted information of each selected systematic review will include: general information (author, year, country of origin and papers, theoretical/conceptual framework); tools applied to assess the measurement properties of HRQoL instruments (title, purpose/use, number of items, response categories, criteria to assess the measurement properties upon specific measurement standard, ease and usefulness of interpretation, level of expertise required for scoring and interpreting, and time required to completion); reporting the measurement properties assessed; use of the results from the evaluation of the measurement properties to come an overall conclusion regarding the quality of each HRQoL instruments. Authors of eligible studies will be contacted to provide missing or additional data if necessary.

27. * Risk of bias (quality) assessment.

State which characteristics of the studies will be assessed and/or any formal risk of bias/quality assessment tools that will be used.

To maximize the quality of reporting of the present meta-review we will follow the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols Guidelines (PRISMA-P)(Moher et al. 2015).

28. * Strategy for data synthesis.

Describe the methods you plan to use to synthesise data. This **must not be generic text** but should be **specific to your review** and describe how the proposed approach will be applied to your data. If meta-analysis is planned, describe the models to be used, methods to explore statistical heterogeneity, and software package to be used.

Descriptive analyses of the use of the identified tools applied to evaluate the measurement properties of HRQoL instruments. The extracted information related to these tools will be reported in a table to facilitate their comparison. Some recommendations about the methodological, practical and the research applications of each tool will be made.

29. * Analysis of subgroups or subsets.

State any planned investigation of ‘subgroups’. Be clear and specific about which type of study or participant will be included in each group or covariate investigated. State the planned analytic approach.

None.

30. * Type and method of review.

Select the type of review, review method and health area from the lists below.

Type of review

Cost effectiveness
No

Diagnostic
No

Epidemiologic
No

Individual patient data (IPD) meta-analysis
No

Intervention
No

Meta-analysis
No

Methodology
No

Narrative synthesis
No

Network meta-analysis
No

Pre-clinical
No

Prevention
No

Prognostic
No

Prospective meta-analysis (PMA)
No

Review of reviews
Yes

Service delivery
No

Synthesis of qualitative studies
No

Systematic review
Yes

Other
No

Health area of the review

Alcohol/substance misuse/abuse
No

Blood and immune system
No

Cancer
No

Cardiovascular

No

Care of the elderly
No

Child health
No

Complementary therapies
No

COVID-19
No

Crime and justice
No

Dental
No

Digestive system
No

Ear, nose and throat
No

Education
No

Endocrine and metabolic disorders
No

Eye disorders
No

General interest
No

Genetics
No

Health inequalities/health equity
No

Infections and infestations
No

International development
No

Mental health and behavioural conditions
No

Musculoskeletal
No

Neurological
No

Nursing
No

Obstetrics and gynaecology
No

Oral health
No

Palliative care
No

Perioperative care
No

Physiotherapy
No

Pregnancy and childbirth
No

Public health (including social determinants of health)
No

Rehabilitation
No

Respiratory disorders
No

Service delivery
No

Skin disorders
No

Social care
No

Surgery
No

Tropical Medicine
No

Urological
No

Wounds, injuries and accidents
No

Violence and abuse
No

31. Language.

Select each language individually to add it to the list below, use the bin icon to remove any added in error.
English

There is an English language summary.

32. * Country.

Select the country in which the review is being carried out. For multi-national collaborations select all the countries involved.

Spain

33. Other registration details.

Name any other organisation where the systematic review title or protocol is registered (e.g. Campbell, or The Joanna Briggs Institute) together with any unique identification number assigned by them. If extracted data will be stored and made available through a repository such as the Systematic Review Data Repository (SRDR), details and a link should be included here. If none, leave blank.

34. Reference and/or URL for published protocol.

If the protocol for this review is published provide details (authors, title and journal details, preferably in Vancouver format)

LORENTE S., VIVES J., VILADRICH C. & LOSILLA J.M. (2018) Tools to assess the measurement properties of quality of life instruments: A meta-review protocol. BMJ Open, 8, e022829. doi 10.1136/bmjopen-2018-02289

Add web link to the published protocol.

Or, upload your published protocol here in pdf format. Note that the upload will be publicly accessible.

No I do not make this file publicly available until the review is complete

Please note that the information required in the PROSPERO registration form must be completed in full even if access to a protocol is given.

35. Dissemination plans.

Do you intend to publish the review on completion?

Yes

Give brief details of plans for communicating review findings.?

The results will be disseminated by its publication in a peer-reviewed journal and presented at a relevant conference.

36. Keywords.

Give words or phrases that best describe the review. Separate keywords with a semicolon or new line. Keywords help PROSPERO users find your review (keywords do not appear in the public record but are included in searches). Be as specific and precise as possible. Avoid acronyms and abbreviations unless these are in wide use.

Meta-review, Quality of life, Health instruments, Measurement properties, Measurement standards.

37. Details of any existing review of the same topic by the same authors.

If you are registering an update of an existing review give details of the earlier versions and include a full bibliographic reference, if available.

None

38. * Current review status.

Update review status when the review is completed and when it is published.New registrations must be ongoing.
Please provide anticipated publication date

Review_Completed_published

39. Any additional information.

Provide any other information relevant to the registration of this review.

40. Details of final report/publication(s) or preprints if available.

Leave empty until publication details are available OR you have a link to a preprint. List authors, title and journal details preferably in Vancouver format.

LORENTE, S., VILADRICH, C., VIVES, J. & LOSILLA, J.M. (2020). Tools to assess the measurement properties of quality of life instruments: a meta-review. BMJ Open, 10, e036038.

http://doi.org/10.1136/bmjopen-2019-036038

Give the link to the published review or preprint.

9.2.2. ESTRATEGIA DE BÚSQUEDA EN LAS BASES DE DATOS

Search strings for Pubmed

1	("Quality of Life"[Mesh] OR HRQL [tiab] OR HRQoL[tiab] OR QoL[tiab] OR "quality of life"[tiab]
2	(instrument[tiab] OR instruments[tiab] OR questionnaire[tiab] OR questionnaires[tiab] OR scale[tiab] OR scales[tiab] OR tool[tiab] OR tools[tiab]
3	(Validation Studies[pt] OR "reproducibility of results"[MeSH Terms] OR reproducib*[tiab] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tiab] OR clinometr*[tiab] OR "observer variation"[MeSH] OR observer variation[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR valid*[tiab] OR coefficient[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR "item correlation"[tiab] OR "item correlations"[tiab] OR "item selection"[tiab] OR "item selections"[tiab] OR "item reduction"[tiab] OR "item reductions"[tiab] OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest [tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR "kappa's"[tiab] OR kappas[tiab] OR "coefficient of variation"[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structure"[tiab] OR dimensionality[tiab] OR subscale*[tiab] OR "multitrait scaling analysis"[tiab] OR "multitrait scaling analyses"[tiab] OR "item discriminant"[tiab] OR "interscale correlation"[tiab] OR "interscale correlations"[tiab] OR ((error[tiab] OR errors[tiab]) AND (measure*[tiab] OR correlat*[tiab] OR evaluat*[tiab] OR accuracy[tiab] OR accurate[tiab] OR precision[tiab] OR mean[tiab])) OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR "variability analysis"[tiab] OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "minimal important change"[tiab] OR "minimal important difference"[tiab] OR "minimally important change"[tiab] OR "minimally important difference"[tiab] OR "minimal detectable change"[tiab] OR "minimal detectable difference"[tiab] OR "minimally detectable change"[tiab] OR "minimally detectable difference"[tiab] OR "minimal real change"[tiab] OR "minimal real difference"[tiab] OR "minimally real change"[tiab] OR "minimally real difference"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF [tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]
4	#1 AND #2 AND #3
5	("protocol"[ti] OR "addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference"[Publication Type] OR "practice guideline"[Publication Type])
6	#4 NOT #5
7	FILTER: Article Type (Review or Systematic Review)
8	FILTER: Subject (Systematic Review)
9	FILTER: Language (English)
10	FILTER: Period (2008-2018)

Search strings for CINAHL

1	TI "quality of life" OR "HRQOL" OR AB "quality of life" OR "HRQOL"
2	TI (instrument OR instruments OR questionnaire OR questionnaires OR scale OR scale OR tool OR tools) OR AB (instrument OR instruments OR questionnaire OR questionnaires OR scale OR scale OR tool OR tools)
3	TI ("Validation Studies" OR "reproducibility of results" OR reproducib* OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "observer variation" OR observer variation OR "discriminant analysis" OR reliab* OR valid* OR coefficient OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR "item correlation" OR "item correlations" OR "item selection" OR "item selections" OR "item reduction" OR "item reductions" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR "kappa's" OR kappas OR "coefficient of variation" OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structure" OR dimensionality OR subscale* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation" OR "interscale correlations" OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR (limit AND detection) OR "minimal detectable concentration" OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "minimal important change" OR "minimal important difference" OR "minimally important change" OR "minimally important difference" OR "minimal detectable change" OR "minimal detectable difference" OR "minimally detectable change" OR "minimally detectable difference" OR "minimal real change" OR "minimal real difference" OR "minimally real change" OR "minimally real difference" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence") OR AB ("Validation Studies" OR "reproducibility of results" OR reproducib* OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "observer variation" OR observer variation OR "discriminant analysis" OR reliab* OR valid* OR coefficient OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR "item correlation" OR "item correlations" OR "item selection" OR "item selections" OR "item reduction" OR "item reductions" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR "kappa's" OR kappas OR "coefficient of variation" OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structure" OR dimensionality OR subscale* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation" OR "interscale correlations" OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR (limit AND detection) OR "minimal detectable concentration" OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "minimal important change" OR "minimal important difference" OR "minimally important change" OR "minimally important difference" OR "minimal detectable change" OR "minimal detectable difference" OR "minimally detectable change" OR "minimally detectable difference" OR "minimal real change" OR "minimal real difference" OR "minimally real change" OR "minimally real difference" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
4	TI review OR AB review
5	#1 AND #2 AND #3 AND #4
6	TI ("protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "directory" OR "editorial" OR "festschrift" OR "interview" OR "lectures" OR "legal cases" OR "legislation" OR "letter" OR "news" OR "newspaper article" OR "patient education handout" OR "popular works" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
7	#5 NOT #6
7	FILTER: Language (English)
8	FILTER: Period (2008-2018)

Search strings for PsycInfo

1	it=Quality of life
2	it=Questionnaires OR it="Rating Scales" OR it=Screening OR it= "Screening Tests" OR it="Psychological Assessment" OR it=Inventories OR it="Individual Testing" OR it="Human Factors Measures" OR it="Checklist Testing" OR it=Psychometrics
3	#1 AND #2
4	FILTER: Methodology (Literature Review)
5	FILTER: Language (English)
6	FILTER: Period (2008-2018)

Search strings for Scopus

1	TITLE-ABS-KEY("Quality of life" OR "HRQOL")
2	TITLE-ABS-KEY(instrument OR instruments OR questionnaire OR questionnaires OR scale OR scale OR tool OR tools)
3	TITLE-ABS-KEY("Validation Studies" OR "reproducibility of results" OR reproducib* OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "observer variation" OR observer variation OR "discriminant analysis" OR reliab* OR valid* OR coefficient OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR "item correlation" OR "item correlations" OR "item selection" OR "item selections" OR "item reduction" OR "item reductions" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR "kappa's" OR kappas OR "coefficient of variation" OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structure" OR dimensionality OR subscale* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation" OR "interscale correlations" OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR (limit AND detection) OR "minimal detectable concentration" OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "minimal important change" OR "minimal important difference" OR "minimally important change" OR "minimally important difference" OR "minimal detectable change" OR "minimal detectable difference" OR "minimally detectable change" OR "minimally detectable difference" OR "minimal real change" OR "minimal real difference" OR "minimally real change" OR "minimally real difference" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
4	#1 AND #2 AND #3
5	TITLE("protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "directory" OR "editorial" OR "festschrift" OR "interview" OR "lectures" OR "legal cases" OR "legislation" OR "letter" OR "news" OR "newspaper article" OR "patient education handout" OR "popular works" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
6	#4 NOT #5
7	FILTER: Document Type (Review)
8	FILTER: Language (English)
9	FILTER: Period (2008-2018)

Search strings for Web of Science (WoS)

1	TI=("Quality of Life" OR “HRQOL”)
2	TS=(instrument OR instruments OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools)
3	TS=(“Validation Studies” OR “reproducibility of results” OR reproducib* OR “psychometrics” OR psychometr* OR clinimetr* OR clinometr* OR “observer variation” OR observer variation OR “discriminant analysis” OR reliab* OR valid* OR coefficient OR “internal consistency” OR (cronbach* AND (alpha OR alphas)) OR “item correlation” OR “item correlations” OR “item selection” OR “item selections” OR “item reduction” OR “item reductions” OR agreement OR precision OR imprecision OR “precise values” OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR “kappa’s” OR kappas OR “coefficient of variation” OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR “known group” OR “factor analysis” OR “factor analyses” OR “factor structure” OR “factor structure” OR dimensionality OR subscale* OR “multitrait scaling analysis” OR “multitrait scaling analyses” OR “item discriminant” OR “interscale correlation” OR “interscale correlations” OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR “individual variability” OR “interval variability” OR “rate variability” OR “variability analysis” OR (uncertainty AND (measurement OR measuring)) OR “standard error of measurement” OR sensitiv* OR responsive* OR (limit AND detection) OR “minimal detectable concentration” OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR “meaningful change” OR “minimal important change” OR “minimal important difference” OR “minimally important change” OR “minimally important difference” OR “minimal detectable change” OR “minimal detectable difference” OR “minimally detectable change” OR “minimally detectable difference” OR “minimal real change” OR “minimal real difference” OR “minimally real change” OR “minimally real difference” OR “ceiling effect” OR “floor effect” OR “Item response model” OR IRT OR Rasch OR “Differential item functioning” OR DIF OR “computer adaptive testing” OR “item bank” OR “cross-cultural equivalence”)
4	#1 AND #2 AND #3
5	TI=(“protocol” OR “addresses” OR “biography” OR “case reports” OR “comment” OR “directory” OR “editorial” OR “festschrift” OR “interview” OR “lectures” OR “legal cases” OR “legislation” OR “letter” OR “news” OR “newspaper article” OR “patient education handout” OR “popular works” OR “congresses” OR “consensus development conference” OR “consensus development conference” OR “practice guideline”)
6	#4 NOT #5
7	FILTER: Document Type (Literature Review)
8	FILTER: Language (English)
9	FILTER: Period (2008-2018)

Search strings for ProQuest Dissertations & Theses Global

1	ti("Quality of life" OR HRQOL) OR ab("Quality of life" OR HRQOL)
2	ti(instrument OR instruments OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools) OR ab(instrument OR instruments OR questionnaire OR questionnaires OR scale OR scales OR tool OR tools)
3	ti("Validation Studies" OR "reproducibility of results" OR reproducib* OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "observer variation" OR observer variation OR "discriminant analysis" OR reliab* OR valid* OR coefficient OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR "item correlation" OR "item correlations" OR "item selection" OR "item selections" OR "item reduction" OR "item reductions" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR "kappa's" OR kappas OR "coefficient of variation" OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structure" OR dimensionality OR subscale* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation" OR "interscale correlations" OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR (limit AND detection) OR "minimal detectable concentration" OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "minimal important change" OR "minimal important difference" OR "minimally important change" OR "minimally important difference" OR "minimal detectable change" OR "minimal detectable difference" OR "minimally detectable change" OR "minimally detectable difference" OR "minimal real change" OR "minimal real difference" OR "minimally real change" OR "minimally real difference" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence") OR ab("Validation Studies" OR "reproducibility of results" OR reproducib* OR "psychometrics" OR psychometr* OR clinimetr* OR clinometr* OR "observer variation" OR observer variation OR "discriminant analysis" OR reliab* OR valid* OR coefficient OR "internal consistency" OR (cronbach* AND (alpha OR alphas)) OR "item correlation" OR "item correlations" OR "item selection" OR "item selections" OR "item reduction" OR "item reductions" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR (test AND retest) OR (reliab* AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR "kappa's" OR kappas OR "coefficient of variation" OR repeatab* OR ((replicab* OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza* OR generalisa* OR concordance OR (intraclass AND correlation*) OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structure" OR dimensionality OR subscale* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation" OR "interscale correlations" OR ((error OR errors) AND (measure* OR correlat* OR evaluat* OR accuracy OR accurate OR precision OR mean)) OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR sensitiv* OR responsive* OR (limit AND detection) OR "minimal detectable concentration" OR interpretab* OR (small* AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR "minimal important change" OR "minimal important difference" OR "minimally important change" OR "minimally important difference" OR "minimal detectable change" OR "minimal detectable difference" OR "minimally detectable change" OR "minimally detectable difference" OR "minimal real change" OR "minimal real difference" OR "minimally real change" OR "minimally real difference" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")
4	#1 AND #2 AND #3
5	ti(Systematic Review) OR ab(Systematic Review)
5	#4 AND #5
6	ti("protocol" OR "addresses" OR "biography" OR "case reports" OR "comment" OR "directory" OR "editorial" OR "festschrift" OR "interview" OR "lectures" OR "legal cases" OR "legislation" OR "letter" OR "news" OR "newspaper article" OR "patient education handout" OR "popular works" OR "congresses" OR "consensus development conference" OR "consensus development conference" OR "practice guideline")
7	#5 NOT #6
8	FILTER: Language (English)
9	FILTER: Period (2008-2018)

9.2.3. CARACTERÍSTICAS Y REFERENCIAS DE LOS ESTUDIOS

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
1	2008	Barbosa & Gaviao	Oral health-related quality of life in children: Part III. Is there agreement between parents in rating their children´s oral health-related quality of life? A systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
2	2008	Basra et al.	The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results	Studies on the quality of one instrument to measures HRQoL in general population	Disease-specific	One instrument	Multiple properties
3	2008	Carabin et al.	Quality of life measurement tools for people living with HIV/AIDS.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
4	2008	Chassany et al.	Systematic review: health-related quality of life (HRQOL) questionnaires in gastro-oesophageal reflux disease.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
5	2008	El Achhab et al.	Disease-specific health-related quality of life instruments among adults diabetic: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
6	2008	Finger et al.	Quality of life in age-related macular degeneration: a review of available vision-specific psychometric tools.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
7	2008	Franic & Bothe	Psychometric evaluation of condition-specific instruments used to assess health-related, quality of life, attitudes, and related constructs in stuttering	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
8	2008	Janssens et al.	Health-related quality-of-life measures for long-term follow-up in children after major trauma.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
9	2008	Klassen et al.	Clinical research in Pediatric plastic surgery and Systematic review of quality of life questionnaires	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
10	2008	Kluivers et al.	Systematic review on recovery specific quality of life instruments	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
11	2008	Langham et al.	Health-related quality of life instruments in studies of adult men with testosterone deficiency syndrome: a critical assessment.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
12	2008	Pearce et al.	Measuring quality of life in cancer survivors: a methodological review of existing scales	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
13	2008	Price et al.	Measures of functional status and quality-of-life in schizophrenia	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
14	2008	Quittner et al.	Systematic review of health-related quality of life measures for children with respiratory conditions.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
15	2008	Reaney et al.	Understanding and assessing the impact of alcoholism on quality of life. A systematic review of the content validity of instruments used to assess health related quality of life in alcoholism	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
16	2008	Schalarman et al.	The use of health-related quality of life (HRQOL) in children and adolescents as an outcome criterion to evaluate family oriented support for young carers in Germany: an integrative review of the literature	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
17	2008	Solans et al.	Health-related quality of life measurement in children and adolescents: A systematic review of generic and disease-specific instruments.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
18	2008	Tschiesner et al.	Content comparison of quality of life questionnaires used in head and neck cancer based on the international classification of functioning, disability and health: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	One specific property
19	2008	Upton et al.	Parent-child agreement across child health-related quality of life instruments: a review of the literature.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
20	2009	Davies N.	Measuring health-related quality of life in cancer patients.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
21	2009	Derret et al.	Outcome after injury-a systematic literature search of studies using the EQ-5D	Studies on the quality of one instrument to measures HRQoL in general population	Generic	One instrument	Multiple properties
22	2009	Epton et al.	Quality of life in amyotrophic lateral sclerosis/motor neuron disease: a structured review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
23	2009	Fitzsimmons et al.	A systematic review of the use and validation of health-related quality of life instruments in older cancer patients.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
24	2009	Garin et al.	Disease-specific health-related quality of life questionnaires for heart failure: A systematic review with meta-analyses.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
25	2009	Garvie et al.	Quality of life measurement in paediatric and adolescent populations with HIV: A review of the literature	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
26	2009	Guo et al.	Measuring health-related quality of life in tuberculosis: a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
27	2009	Jay et al.	A review of quality of life instruments used in liver transplantation.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
28	2009	Speight et al	Not all roads lead to Rome-a review of quality of life measurement in adults with diabetes.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
29	2009	Street et al	Health related quality of life assessment in metastatic disease of the spine: a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
30	2009	Waters et al.	Quality of life instruments for children and adolescents with neurodisabilities: How to choose the appropriate instrument.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
31	2009	Wettergren et al.	The use, feasibility and psychometric properties of an individualised quality-of-life instrument: A systematic review of the SEIQoL-DW.	Studies on the quality of one instrument to measures HRQoL in general population	Generic	One instrument	Multiple properties
32	2010	Albers et al.	Evaluation of quality-of-life measures for use in palliative care: a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
33	2010	Branski et al.	Measuring quality of life in dysphonic patients: a systematic review of content development in patient-reported outcomes measures.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
34	2010	Bronsard et al.	What are the best outcome measures for assessing quality of life in plaque type psoriasis? A systematic review of the literature.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
35	2010	Carlon et al.	A systematic review of the psychometric properties of Quality of life measures for school children with cerebral palsy	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
36	2010	Chen et al.	Measuring Quality of Life in Oncologic Breast Surgery: A Systematic Review of Patient-Reported Outcome Measures.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
37	2010	Forhan et al.	A systematic review of the quality of psychometric evidence supporting the use of an obesity-specific quality of life measure for use with persons who have class III obesity: Diagnostic in Obesity and Complications	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
38	2010	Danquah et al.	Quality of life measures for patients on hemodialysis: a review of psychometric properties.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
39	2010	Hill et al.	Quality of life instruments and definitions in individuals with spinal cord injury: a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
40	2010	Kamalski et al.	Measuring disease-specific health-related quality of life to evaluate treatment outcomes in tinnitus patients: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
41	2010	Klassen et al.	Quality of life questionnaires for children with cancer and childhood cancer survivors: a review of the development of available measures	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
42	2010	Kwon et al.	Quality of life of women with urinary incontinence: a systematic literature review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
43	2010	Luckett et al.	Assessing health-related quality of life in gynecologic oncology: a systematic review of questionnaires and their ability to detect clinically important differences and change.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
44	2010	Palfreyman et al.	Assessing current health-related quality of life questionnaires administered to patients with venous ulcers: Can they be used in economic evaluations?	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
45	2010	Palfreyman et al.	A systematic review of health-related quality of life instruments used for people with venous ulcers: an assessment of their suitability and psychometric properties.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
46	2010	Passarelli et al.	Validity Studies of Quality of Life Instruments for Eating Disorders	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
47	2010	Riordain & McCreary	The use of quality of life measures in oral medicine: A review of the literature	Studies on the quality of all available validated instruments to measure HRQoL in general population	Both	More than two instruments	Multiple properties
48	2010	Speight & Howarth	Quality of life in restless legs syndrome: A systematic review of clinical trials and a critical review of instruments.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
49	2010	Street et al.	Introducing a New Health Related Quality of Life Outcome tool for mestastatic disease of the spine	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
50	2010	Zeng et al.	Quality of life measurement in women with cervical cancer: implications for Chinese cervical cancer survivors	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
51	2011	Carlton & Kaltenthaler	Health-related quality of life measures (HRQoL) in patients with amblyopia and strabismus: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
52	2011	Carlton & Kaltenthaler	Amblyopia and quality of life: a systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
53	2011	da Silva et al.	Quality of life assessment after Acute Coronary Syndrome: Systematic Review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
54	2011	Eckstein et al.	Measuring Quality of Life in Cleft Lip and Palate patients: currently available patient reported outcomes	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
55	2011	Fayed et al.	Health status and QOL instruments used in childhood cancer research: deciphering conceptual content using World Health Organization definitions	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
56	2011	Glover et al.	Understanding and assessing the impact of End-Stage renal disease on QOL. A systematic review of the content validity…	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
57	2011	Hounsome et al.	EQ-5D as a Quality of Life measure in people with dementia and their carers: evidence and key issues	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiple properties
58	2011	Janssen et al.	The use of the EQ-5D preference based health status measure in adults with type 2 diabetes mellitus	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiples properties
59	2011	Kowal-Bielecka	Analysis of the validation status of WOL and Functional measures in Pulmonary Arterial Hypertension.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	Comparison of two instruments	Multiples properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
60	2011	Lien et al.	Comparison of the EORTC QLQ-C15-PAL and the FACIT-Pal for assessment of quality of life in patients with advanced cancer	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	Comparison of two instruments	One specific property
61	2011	Luckett et al.	Choosing between the EORTC QLQ-C30 and FACT for measuring health related quality of life in cancer clinical research: issues, evidence and recommendations	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	Comparison of two instruments	Multiple properties
62	2011	Mordiffi et al.	Quality of life tools for adult patients with cancer undergoing chemotherapy: a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
63	2011	Noyes et al.	EQ-5D for the Assessment of Health-Related Quality of Life and Resource Allocation in Children: A Systematic Methodological Review	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiple properties
64	2011	Papaioannou et al.	How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
65	2011	Reavey et al.	Measuring quality of life and patient satisfaction after body contouring: a systematic review of patient-reported outcome measures.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
66	2011	Schiarti et al.	Content comparison of health-related quality of life measures for cerebral palsy based on the International Classification of Functioning Quality of life in people with venous leg ulcers: an integrative review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
67	2011	Tayyem et al	Analysis of Health-Related quality of life instruments measuring the impact of bariatric surgery: systematic review of instruments and their content validity	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
68	2011	Virginia et al	Quality of life in people with venous leg ulcers: an integrative review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
69	2011	Whitehurst et al.	Systematic review and empirical comparison of contemporanous EQ-5D and SF-6D group mean scores	Studies on the quality of a selection of instruments to measure HRQoL in general population	Generic	Comparison of two instruments	One specific property
70	2011	Wilson et al.	Spinal cord injury and quality of life: a systematic review of outcome measures	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
71	2012	Bhatt et al.	Health outcome measures for diabetes mellitus: a review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
72	2012	Chopra & Kamal	A systematic review of quality of life instruments in long-term breast cancer survivors.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
73	2012	Cormier et al.	Health related quality of life in patients with melanoma. Overeview of instruments and outcomes	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
74	2012	Correia & De Carlo	Evaluation of quality of life in a palliative care context: an integrative literature review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
75	2012	Gräske et al.	Dementia-Specific Quality of Life Instruments and Their Appropriateness in Shared-Housing Arrangements--A Literature Study.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
76	2012	Ho et al.	Measuring Quality of life and patient satisfaction in facial paralysis patients: a systematic review of patient reported outcome measures	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
77	2012	Hogg et al.	Measures of health related quality of life in diabetes-related foot disease: a systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
78	2012	Luquiens et al.	Quality of life among alcohol-dependent patients: how satisfactory are the available instruments? A systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
79	2012	Madureira et al.	Quality of life measurements is patients with osteoporosis and fractures	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
80	2012	Milne et al.	Measuring Health-Related Quality of Life for Patients with Diabetic Retinopathy	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
81	2012	Ojo et al.	A Systematic Review of Head and Neck Cancer Quality of Life Assessment Instruments	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
82	2012	Popovic et al.	Comparison of the EORTC QLQ-BM22 and the FACT-BP for assessment of quality of life in cancer patients with bone metastases	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	Comparison of two instruments	Multiple properties
83	2012	Quintanilla et al.	Comparison of disease-specific quality of life instruments in the assessment of chronic rhinosinusitis	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
84	2012	Rajmil et al.	Health-related quality of life measurement in children and adolescents in Ibero-American countries, 2000 to 2010.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
85	2012	Shin & Shin	Measurement of quality of life in menopausal women: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
86	2012	Smith et al	Measuring health-related quality of life in diabetic peripheral neuropathy: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
87	2012	Tosh et al	A review of generic preference-based measures of health-related quality of life in visual disorders.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
88	2012	Townsend-White et al.	Review: a systematic review of quality of life measures for people with intellectual disabilities and challenging behaviours	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
89	2012	Walker et al.	Are they worth it? A systematic review of QOL instruments for use with mentally disordered offended	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
90	2012	Whitehurst et al.	A review of preference-based health-related quality of life questionnaires in spinal cord injury research.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
91	2012	Yip et al.	Reliability, validity and feasibility of quality of life instruments for adult patients with cancer undergoing chemotherapy: Result from a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
92	2013	Al Sayah et al.	Health related quality of life measures in Arabic speaking populations: A systematic review on cross-cultural adaptation and measurement properties	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
93	2013	Basra et al.	Infants' Dermatitis Quality of Life Index: a decade of experience of validation and clinical application.	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
94	2013	Castelino et al.	Comparison of the psychometric properties of health-related quality of life measures used in adults with systemic lupus erythematosus: a review of the literature.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
95	2013	Chandratre et al.	Health-related quality of life in gout: a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
96	2013	Chow et al.	Condition-specific quality of life questionnaires for caregivers of children with pediatric conditions: A systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
97	2013	Davis et al.	A review of the psychometric performance of the EQ-5D in people with urinary incontinence.	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiples properties
98	2013	de Almeida et al	Quality of life instruments for skull base pathology: Systematic review and methodologic appraisal	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
99	2013	Djan et al.	A systematic review of questionnaires to measure the impact of appearance on quality of life for head and neck cancer patients.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
100	2013	Gakhar et al.	Health-related quality of life assessment after antiretroviral therapy: A review of the literature	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
101	2013	Hitzig et al.	Identifying and classifying quality-of-life tools for assessing pressure ulcers after spinal cord injury.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
102	2013	Jabir et al.	Assessing Improvement in Quality of Life and Patient Satisfaction following Body Contouring Surgery in Patients with Massive Weight Loss: A Critical Review of Outcome Measures Employed.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
103	2013	Lee et al.	A systematic review of patient-reported outcome instruments of nonmelanonma skin cancer in the dermatologic population	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
104	2013	Levterova et al.	Instruments for disease-specific quality-of-life measurement in patients with type 2 diabetes mellitus--a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
105	2013	Li et al.	Psychometric properties of self-reported quality of life measures for people with intellectual disabilities: A systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
106	2013	Lin et al.	Evaluation of content on EQ-5D as compared to disease-specific utility measures.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
107	2013	Mitera et al.	Quality of life measures used in radiation therapy trials for patients with metastatic spinal cord compresssion (MSCC)	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
108	2013	Mogos et al.	A Systematic Review of Quality of Life Measures in Pregnant and Postpartum Mothers.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
109	2013	Mousavi et al.	Assessment of Questionnaires Measuring Quality of Life in Infertile Couples: A Systematic Review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
110	2013	Moyle et al.	Health-related quality of life in older people with severe dementia: challenges for measurement and management	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
111	2013	Muzzatti et al.	Assessing quality of life in long-term cancer survivors: a review of available tools.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
112	2013	Paltzer et al.	Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: a review of existing measures.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
113	2013	Perales. et al.	Health-related quality-of-life instruments for Alzheimer's disease and mixed dementia.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
114	2013	Pusic et al.	Quality of life among breast cancer patients with lymphedema: A systematic review of patient-reported outcome instruments and outcomes.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
115	2013	Roncada et al.	Specific instruments to assess quality of life in children and adolescents with asthma.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
116	2013	Salek et al.	Clinical experience and psychometric properties of the Children's Dermatology Life Quality Index (CDLQI), 1995-2012	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
117	2013	Testart et al.	Quality of life and other outcome measures in caregivers of patients with schizophrenia	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
118	2013	Weldam et al.	Evaluation of Quality of Life instruments for use in COPD care and research: A systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
119	2013	Wheelright et al.	A systematic review of health-related quality of life instruments in patients with cancer cachexia	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
120	2013	Yang et al.	An assessment of validity and responsiveness of generic measures of health-related quality of life in hearing impairment.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
121	2014	Anthony et al.	Considering quality of life for children with cancer: a systematic review of patient-reported outcome measures and the development of a conceptual model	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
122	2014	Aspden et al.	Quality-of-life measures for use within care homes: a systematic review of their measurement properties.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
123	2014	Balioussis et al.	Identifying and classifying quality of life tools for assessing spasticity after spinal cord injury.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
124	2014	Brazier et al.	A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
125	2014	Chiu et al.	Comparison of three shortened questionnaires for assessment of quality of life in advanced cancer.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
126	2014	Chow et al.	Comparison of the EORTC QLQ-BN20 and the FACT-Br quality of life questionnaires for patients with primary brain cancers: a literature review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	Comparison of two instruments	Multiple properties
127	2014	Garin et al.	Assessing health-related quality of life in patients with heart failure: a systematic, standardized comparison of available measures.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
128	2014	Gilchrist et al.	Assessment of the quality of measures of child oral health-related quality of life.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
129	2014	Grubbs et al.	A review of quality of life measures in dry eye questionnaires.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
130	2014	Gupta et al.	The COPD assessment test: a systematic review.	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
131	2014	Hawkins et al.	A Systematic Review of Functional and Quality of Life Assessment after Major Lower Extremity Amputation	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
132	2014	Hewison et al.	An evaluative review of questionnaires recommended for the assessment of quality of life and symptom severity in women with urinary incontinence.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
133	2014	Ikeda et al.	Assessment of quality of life in children and youth with autism spectrum disorder: a critical review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
134	2014	Jardine et al.	Self-reported quality of life of young children with conditions from early infancy: a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
135	2014	Kuspinar et al.	A review of the psychometric properties of generic utility measures in multiple sclerosis.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
136	2014	Lee et al.	Measurement properties of rheumatoid arthritis-specific quality-of-life questionnaires: Systematic review of the literature.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
137	2014	Lieu et al.	Pediatric quality of life in children with otolaryngologic disease: what inventories are available and what is still needed?	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
138	2014	Longworth et al.	Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: a systematic review, statistical modelling and survey.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
139	2014	Makai et al.	Quality of life instruments for economic evaluations in health and social care for older people: A systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
140	2014	Niu et al.	Health-related quality of life in women with breast cancer: a literature-based review of psychometric properties of breast cancer-specific measures.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
141	2014	Salvilla et al.	Disease-specific health-related quality of life instruments for IgE-mediated food allergy	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
142	2014	Schmidt et al.	Assessing quality of life in patients with prostate cancer: a systematic and standardized comparison of available instruments.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
143	2014	Smith et al.	Most domains of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 are reliable.	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	One specific property
144	2014	Souza et al.	Tools used for evaluation of Brazilian children's quality of life	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
145	2014	Swigris et al.	The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review.	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
146	2014	Timmerman et al.	Psychometric characteristics of health-related quality-of-life questionnaires in oropharyngeal dysphagia.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
147	2014	Treanor & Donelly	A methodological review of the Short Form Health Survey 36 (SF-36) and its derivatives among breast cancer survivors	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
148	2014	Watt et al.	Assessing health-related quality of life in patients with benign non-toxic goitre	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
149	2014	Wolpe et al.	Assessing the impact of urinary incontinence on quality of life: systematic review of instruments in Portuguese.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
150	2015	Alrubaiy et al.	Systematic review of health-related quality of life measures for inflammatory bowel disease	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
151	2015	Aspesberro et al.	Health-related quality of life following pediatric critical illness.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
152	2015	Bédard et al.	Systematic review of vision-related quality of life questionnaires for older institutionalised seniors with dementia	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
153	2015	Bowling et al.	Quality of life in dementia: a systematically conducted narrative review of dementia-specific measurement scales.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
154	2015	Conijn et al.	Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
155	2015	de Climens et al.	Review of patient-reported outcome instruments measuring health-related quality of life and satisfaction in patients with type 2 diabetes treated with oral therapy.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
156	2015	Dronavalli & Thompson	A systematic review of measurement tools of health and well-being for evaluating community-based interventions.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
157	2015	Hamoen et al.	Measuring health-related quality of life in men with prostate cancer: A systematic review of the most used questionnaires and their validity.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
158	2015	Hu et al.	How Quality of Life as Patient-Reported Outcome Has Been Studied for Rheumatoid Arthritis in Chinese-Speaking Population	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
159	2015	Janssens et al.	Measurement properties of multidimensional patient-reported outcome measures in neurodisability: a systematic review of evaluation studies.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
160	2015	Launois et al.	Health-related quality-of-life scales specific for chronic venous disorders of the lower limbs.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
161	2015	Monticone et al.	Measurement properties of translated versions of the Scoliosis Research Society-22 Patient Questionnaire, SRS-22: A systematic review.	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
162	2015	Nguyen et al.	EORTC QLQ-BR23 and FACT-B for the assessment of quality of life in patients with breast cancer: a literature review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
163	2015	Oliveira et al.	Evaluation of cross-cultural adaptation and measurement properties of breast cancer-specific quality-of-life questionnaires: A systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
164	2015	Polinder et al.	Health-related quality of life after TBI: a systematic review of study design, instruments, measurement properties, and outcome.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
165	2015	Taghavi et al.	Health-related quality of life in polycystic ovary syndrome patients: A systematic review.fit	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiples properties
166	2015	Wong et al.	Systematic review recommends the European Organization for Research and Treatment of Cancer colorectal cancer-specific module for measuring quality of life in colorectal cancer patients	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
167	2016	Algar et al.	Measuring the quality of life and well-being of people with dementia: A review of observational measures.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
168	2016	Bryant et al.	A Systematic Review of Psychometric Properties of Health-Related Quality-of-Life and Symptom Instruments in Adult Acute Leukemia Survivors.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
169	2016	Coombes et al.	Health-related quality-of-life outcome measures in paediatric palliative care: A systematic review of psychometric properties and feasibility of use.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
170	2016	Dichter et al.	Linguistic validation and reliability properties are weak investigated of most dementia-specific quality of life measurements-a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
171	2016	Ganesh et al.	Comparison of the FACT-C, EORTC QLQ-CR38, and QLQ-CR29 quality of life questionnaires for patients with colorectal cancer: a literature review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
172	2016	Gutiérrez-Vargas et al.	Instruments to measure the quality of life in patients with oral mucositis undergoing oncological treatment: a systematic review of the literature.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
173	2016	Hand et al.	Measuring health-related quality of life in adults with chronic conditions in primary care settings: Critical review of concepts and 3 tools	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
174	2016	Heinl et al.	Measurement properties of adult quality of life measurement instruments for eczema: a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
175	2016	Kotecha et al.	Patient-Reported Outcomes for Quality of Life Assessment in Atrial Fibrillation: A Systematic Review of Measurement Properties.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
176	2016	Lee et al.	A systematic review of measurement properties of the instruments measuring health-related quality of life in patients with irritable bowel syndrome.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
177	2016	Maratia et al.	Assessing health-related quality of life in patients with breast cancer: a systematic and standardized comparison of available instruments using the EMPRO tool.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
178	2016	Mestre et al.	Rating scales for behavioral symptoms in Huntington's disease: Critique and recommendations.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
179	2016	Spinou et al.	The validity of health-related quality of life questionnaires in bronchiectasis: a systematic review and meta-analysis	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
180	2016	Tapia et al.	Health-Related Quality-of-Life Instruments for Pediatric Patients with Diverse Facial Deformities: A Systematic Literature Review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
181	2016	Wong et al.	A systematic review of quality of thyroid-specific health related quality of life instruments recommends ThyPRO for patients with benign thyroid diseases	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
182	2016	Woo et al.	Comparison of the EORTC STO-22 and the FACT-Ga quality of life questionnaires for patients with gastric cancer.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	Comparison of two instruments	Multiple properties
183	2017	Ahmadi et al.	Acceptability, reliability, and validity of the Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39) across languages: A systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
184	2017	Baghdadli et al.	Measurement properties of screening and diagnostic tools for autism spectrum adults of mean normal intelligence: A systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
185	2017	Best et al.	Identifying and classifying quality of life tools for neurogenic bladder function after spinal cord injury: A systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
186	2017	Chen et al.	Inflammatory bowel disease-specific health-related quality of life instruments: a systematic review of measurement properties.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
187	2017	Frew et al.	Disease-specific health related quality of life patient reported outcome measures in Genodermatoses: a systematic review and critical evaluation.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
188	2017	Heaney et al.	A review of the psychometric properties and use of the Rheumatoid Arthritis Quality of Life Questionnaire (RaQoL) in clinical research	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiple properties
189	2017	Heinl et al.	Measurement properties of quality-of-life measurement instruments for infants, children and adolescents with eczema: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
190	2017	Kandel et al.	Patient-reported Outcomes for Assessment of Quality of Life in Refractive Error: A Systematic Review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
191	2017	Kao et al.	Scoping Review of Pediatric Tonsillectomy Quality of Life Assessment Instruments	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
192	2017	Khan et al.	Health Status and Quality of Life in Tuberculosis: Systematic Review of Study Design, Instruments, Measuring Properties and Outcomes.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
193	2017	Kwan et al.	A systematic review of quality-of-life domains and items relevant to patients with spondyloarthritis	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	One specific property
194	2017	Limpberg et al.	Health-related quality of life questionnaires in individuals with haemophilia: a systematic review of their measurement properties	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
195	2017	Lucendo et al.	Systematic review: health-related quality of life in children and adults with eosinophilic oesophagitis-instruments for measurement and determinant factors.Systematic review: health-related quality of life in children and adults with eosinophilic oesophagitis-instruments for measurement and determinant factors.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
196	2017	Page et al.	Instruments measuring the disease-specific quality of life of family carers of people with neurodegenerative diseases: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
197	2017	Poku et al.	Systematic review assessing the measurement properties of patient-reported outcomes for venous leg ulcers.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
198	2017	Roydhouse et al.	Systematic review of caregiver responses for patient health-related quality of life in adult cancer care.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
199	2017	Strada et al.	Measuring quality of life in opioid-dependent people: a systematic review of assessment instruments.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
200	2017	Sullivan et al.	Assessing quality of life of patients with hypospadias: A systematic review of validated patient-reported outcome instruments.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
201	2017	Tang et al.	Assessing quality of life in diabetes: II - Deconstructing measures into a simple framework.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
202	2017	Tax et al.	Measuring health-related quality of life in cervical cancer patients: a systematic review of the most used questionnaires and their validity.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	Comparison of two instruments	Multiple properties
203	2017	Xin & McIntosh	Assessment of the construct validity and responsiveness of preference-based quality of life measures in people with Parkinson's: a systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
204	2018	Aber et al.	Themes that Determine Quality of Life in Patients with Peripheral Arterial Disease: A Systematic Review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
205	2018	Chiarotto et al.	Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiple properties
206	2018	Cornelissen et al.	Quality of Life Questionnaires in Breast Cancer-Related Lymphedema Patients: Review of the Literature	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
207	2018	de Vries et al.	Recommendations on the most suitable quality-of-life measurement instruments for bariatric and body contouring surgery: a systematic review.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
208	2018	Dow et al.	How best to assess quality of life in informal carers of people with dementia; A systematic review of existing outcome measures	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
209	2018	Grobet et al.	Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review.	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiple properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
210	2018	Haywood et al.	Assessing health-related quality of life (HRQoL) in survivors of out-of-hospital cardiac arrest: A systematic review of patient-reported outcome measures	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
211	2018	Luan et al.	A Review of Studies of Quality of Life for Chinese-Speaking Patients with Ischemic Heart Disease	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
212	2018	Mason et al.	Evaluating patient-reported outcome measures (PROMs) for bladder cancer: a systematic review using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN)	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
213	2018	Mohammed et al.	Pharmaceutical care and health related quality of life outcomes over the past 25 years: Have we measured dimensions that really matter?	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	One specific property
214	2018	Mpundu-Kaambwa et al.	A review of preference-based measures for the assessment of quality of life in children and adolescents with cerebral palsy.	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
215	2018	Pollo et al.	Evaluation Instruments for Quality of Life Related to Melasma: An Integrative Review.	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiples properties
216	2018	Tian & Cao	Systematic review of the psychometric properties of disease-specific, quality-of-life questionnaires for patients with hepatobiliary or pancreatic cancers	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
217	2018	van Ierssel et al.	Identifying the concepts contained within health-related quality of life outcome measures in concussion research using the International Classification of Functioning, Disability, and Health as a reference: a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	One specific property
218	2018	van Roij et al.	Measuring health-related quality of life in patients with advanced cancer: a systematic review of self-administered measurement instruments.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
219	2018	Yarlas et al.	Psychometric validation of the SF-36®Health Survey in ulcerative colitis: results from a systematic literature review	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiple properties
220	2018	Yazdani et al.	Psychometric Properties of Quality of Life Assessment Tools in Morbid Obesity: A Review of Literature.	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
221	2018	Zaror et al.	Assessing oral health-related quality of life in children and adolescents: a systematic review and standardized comparison of available instruments	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
222	2018	Hettiarachchi et al.	Pediatric Quality of Life Instruments in Oral Health Research: A Systematic Review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
223	2019	Balk et al.	Psychometric properties of functional, ambulatory, and quality of life instruments in lower limb amputees: a systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
224	2019	Belayneh et al.	A systematic review of the psychometric properties of the cross-cultural adaptations and translations of the prolapse quality of life questionarie	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiples properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
225	2019	Bull et al.	Systematic review: measurement properties of patient-reported otucome measures evaluated wwith childhood brain tumor survivores or othe acquired brain injury	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
226	2019	Choukou et al.	Identifiying and calssifiying quality of life tools for assessing neruogenic bowel dysfunction after spinal cord injury	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
227	2019	Daliya et al.	A systematic review of patient reported outcome measures (PROMs) and quality of life reporting in patients undergoing laparoscopic cholecystectomy	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
228	2019	Gabes et al.	Measurement properties of quality-of-life outcome measures for children and adults with eczema: An updated systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
229	2019	Gondivkar et al.	Assessment of oral health-related quality of life instruments for oral submucous fibrosis: A systematic review using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
230	2019	Hasanvand et al.	A Critical Review of Instruments Measuring the Quality of Life of Cancer Patients in Iranian Studies and Their Psychometrics Properties	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
231	2019	Hughes et al.	Psychometric properties and feasibility of use of dementia specific quality of life instruments for use in care settings: a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
232	2019	Hunt et al.	Quality of life in older adults after traumatic brain injury: a systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
233	2019	Kamilu et al.	Quality of life assessment scales in polio survivors: a scoping review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
234	2019	Lamsal et al.	Generic preference-based health-related quality of life in children with neurodevelopmental disorders: a scoping review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Generic	More than two instruments	Multiples properties
235	2019	Moshki et al.	The content comparison of health-related quality of life measures in heart failure based on the international classification of functioning, disability, and health: a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	One specific property
236	2019	Speyer et al.	"Measurement properties of self-report questionnaires on health-related quality of life and functional health status in dysphonia: a systematic review using the COSMIN taxonomy"	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiple properties
237	2019	van der Hout et al.	Measuring health-related quality of life in colorectal cancer patients: systematic review of measurement properties of the EORTC QLQ-CR29	Studies on the quality of one instrument to measures HRQoL in a particular population	Generic	One instrument	Multiple properties
238	2019	Vasconcelos et al.	Quality of Life in Women with Defecatory Dysfunctions: Systematic Review of Questionnaires Validated in the Portuguese Language	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiple properties
239	2020	Crudgington et al.	Epilepsy-specific patient-reported outcome measures of children´s health related quality of life: a systematic review of measurement properties	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiples properties

ID	Year	Author	Title	Type of review	Instruments assessed	Number Instruments	Measurement Properties
240	2020	Furtado et al.	Cross-cultural adaptaions and measurement properties of the WORC (Western Ontario rotator cuff index): a systematic review	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiples properties
241	2020	Jones et al.	A feasibility assessment of functioning and quality of life patient reported outcome measures in adult epilepsy clinics. A systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties
242	2020	Killian et al.	Measurement of health-related quality of life in pediatric organ transplantation recipients: a systematic review of the PedsQL transplant module	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	One instrument	Multiples properties
243	2020	Losada-Puente	A systematic review of the assessment of quality of life in adolescents	Studies on the quality of a selection of instruments to measure HRQoL in a particular population	Disease-specific	More than two instruments	Multiples properties
244	2020	Moller et al.	Specific measures of quality of life in patients with multimorbidity in primary healthcare: a systematic review on patient-reported outcomes measures´adequacy of measurement	Studies on the quality of one instrument to measures HRQoL in a particular population	Disease-specific	More than two instruments	Multiples properties
245	2020	Qian et al.	Measurement properties of commonly used generic preference-based measures in East and South-East Asia: a systematic review	Studies on the quality of a selection of instruments to measure HRQoL in general population	Generic	More than two instruments	Multiples properties
246	2020	Santana-Berlanga et al.	Instruments to measure quality of life in institutionalised older adults: systematic review	Studies on the quality of all available validated instruments to measure HRQoL in a particular population	Both	More than two instruments	Multiples properties

Note: The concept “Both” is referred to generic and disease-specific instruments.

REFERENCES

1. Barbosa T, Gavião M. Oral health-related quality of life in children: Part III. Is there agreement between parents in rating their children's oral health-related quality of life? A systematic review. Int J Dent Hyg. 2008;6(2):108–13.
2. Basra MKA, Fenech R, Gatt RM, Salek MS, Finlay AY. The Dermatology Life Quality Index 1994-2007: A comprehensive review of validation data and clinical results. Br J Dermatol. 2008;159(5):997–1035.
3. Carabin H, Sonleitner N, Keesee M, Shinault K. Quality of life measurement tools for people living with HIV/AIDS. J HIV/AIDS Soc Serv. 2008;7(1):71–82.
4. Chassany O, Holtmann G, Malagelada J, Gebauer U, Doerfler H, Devault K. Systematic review: health-related quality of life (HRQOL) questionnaires in gastro-oesophageal reflux disease. Aliment Pharmacol Ther. 2008;27(11):1053–70.
5. El Achhab Y, Nejjari C, Chikri M, Lyoussi B. Disease-specific health-related quality of life instruments among adults diabetic: A systematic review. Diabetes Res Clin Pract. 2008;80(2):171–84.
6. Finger RP, Fleckenstein M, Holz FG, Scholl HPN. Quality of life in age-related macular degeneration: a review of available vision-specific psychometric tools. Qual Life Res. 2008;17(4):559–74.
7. Franic DM, Bothe AK. Psychometric Evaluation of Condition-Specific Instruments Used to Assess Health-Related Quality of Life, Attitudes, and Related Constructs in Stuttering. Am J Speech. 2008;17(1):60–80.
8. Janssens L, Gorter JW, Ketelaar M, Kramer WLM, Holtslag HR. Health-related quality-of-life measures for long-term follow-up in children after major trauma. Qual Life Res. 2008;17(5):701–13.
9. Klassen AF, Stotland MA, Skarsgard ED, Pusic AL. Clinical research in pediatric plastic surgery and systematic review of quality-of-life questionnaires. Clin Plast Surg. 2008;35:251–7.
10. Kluivers KB, Riphagen I, Vierhout ME, Brölmann HAM, de Vet HCW. Systematic review on recovery specific quality-of-life instruments. Surgery. 2008;143(2):206–15.
11. Langham S, Maggi M, Schulman C, Quinton R, Uhl-Hochgraeber K. Health-related quality of life instruments in studies of adult men with testosterone deficiency syndrome: A critical assessment. J Sex Med. 2008;5(12):2842–52.
12. Pearce NJM, Sanson-Fisher R, Campbell HS. Measuring quality of life in cancer survivors: a methodological review of existing scales. Psychooncology. 2008;17(7):629–40.
13. Price MA, Hill CD, Williams VS, Morlock RJ, Leewenkamp O, Patterson T. Measures of functional status and quality-of-life in schizophrenia. Curr Psychiatry Rev. 2008;4(1):28–38.
14. Quittner AL, Modi A, Cruz I. Systematic review of health-related quality of life measures for children with respiratory conditions. Paediatr Respir Rev. 2008;9(3):220–32.
15. Reaney MD, Martin C, Speight J. Understanding and Assessing the Impact of Alcoholism on Quality of Life. Patient. 2008;1(3):151–63.
16. Schlarmann J, Metzing-Blau S, Schnepp W, Schlarmann JG, Metzing-Blau S, Schnepp W, et al. The use of health-related quality of life (HRQOL) in children and adolescents as an outcome criterion to evaluate family oriented support for young carers in Germany: an integrative review of the literature. BMC Public Health. 2008;8(1):414.
17. Solans M, Pane S, Estrada M-D, Serra-Sutton V, Berra S, Herdman M, et al. Health-related quality of life measurement in children and adolescents: A systematic review of generic and disease-specific instruments. Value Heal. 2008;11(4):742–64.
18. Tschiesner U, Rogers SN, Harréus U, Berghaus A, Cieza A. Content comparison of quality of life questionnaires used in head and neck cancer based on the international classification of functioning, disability and health: A systematic review. Eur Arch Oto-Rhino-Laryngology. 2008;265(6):627–37.
19. Upton P, Lawford J, Eiser C. Parent-child agreement across child health-related quality of life instruments: a review of the literature. Qual Life Res. 2008;17(6):895–913.
20. Davies N. Measuring health-related. Nurs Stand. 2009;23(30):42–50.
21. Derrett S, Black J, Herbison GP. Outcome after injury-a systematic literature search of studies using the EQ-5D. J Trauma. 2009;67(4):883–90.
22. Epton J, Harris R, Jenkinson C. Quality of life in amyotrophic lateral sclerosis/motor neuron disease: a structured review. Amyotroph Lateral Scler. 2009;10(1):15–26.
23. Fitzsimmons D, Gilbert J, Howse F, Young T, Arrarras JI, Brédart A, et al. A systematic review of the use and validation of health-related quality of life instruments in older cancer patients. Eur J Cancer. 2009;45(1):19–32.
24. Garin O, Ferrer M, Pont À, Rué M, Kotzeva A, Wiklund I, et al. Disease-specific health-related quality of life questionnaires for heart failure: A systematic review with meta-analyses. Qual Life Res. 2009;18(1):71–85.
25. Garvie PA, Lawford J, Banet MS, West RL. Quality of life measurement in paediatric and adolescent populations with HIV: A review of the literature. Childcare, Heal Dev. 2009;35(4):440–53.
26. Guo N, Marra F, Marra CA. Measuring health-related quality of life in tuberculosis: a systematic review. Health Qual Life Outcomes. 2009;7:14.

27. Jay CL, Butt Z, Ladner DP, Skaro AI, Abecassis MM. A review of quality of life instruments used in liver transplantation. J Hepatol. 2009;51(5):949–59.

28. Speight J, Reaney MD, Barnard KD. Not all roads lead to Rome-a review of quality of life measurement in adults with diabetes. Diabet Med. 2009;26(4):315–27.

29. Street J, Berven S, Fisher C, Ryken T. Health Related Quality of Life Assessment in Metastatic Disease of the Spine. Spine (Phila Pa 1976). 2009;34(Supplement):S128–34.

30. Waters E, Davis E, Ronen GM, Rosenbaum P, Livingston M, Saigal S. Quality of life instruments for children and adolescents with neurodisabilities: How to choose the appropriate instrument. Vol. 51, Developmental Medicine & Child Neurology. 2009. p. 660–9.

31. Wettergren L, Kettis-Lindblad Å, Sprangers M, Ring L. The use, feasibility and psychometric properties of an individualised quality-of-life instrument: A systematic review of the SEIQoL-DW. Qual Life Res. 2009;18(6):737–46.

32. Albers G, Echteld MA, de Vet HCW, Onwuteaka-Philipsen BD, van der Linden MHM, Deliens L. Evaluation of quality-of-life measures for use in palliative care: a systematic review. Palliat Med. 2010;24(1):17–37.

33. Branski RC, Cukier-Blaj S, Pusic A, Cano SJ, Klassen A, Mener D, et al. Measuring quality of life in dysphonic patients: a systematic review of content development in patient-reported outcomes measures. J Voice. 2010;23(2):193–8.

34. Bronsard V, Paul C, Prey S, Puzenat E, Gourraud P-A, Aractingi S, et al. What are the best outcome measures for assessing quality of life in plaque type psoriasis? A systematic review of the literature. J Eur Acad Dermatology Venereol. 2010;24 Suppl 2:17–22.

35. Carlon S, Shields N, Yong K, Gilmore R, Sakzewski L, Boyd R. A systematic review of the psychometric properties of Quality of Life measures for school aged children with cerebral palsy. BMC Pediatr. 2010;10:81.

36. Chen CM, Cano SJ, Klassen AF, King T, McCarthy C, Cordeiro PG, et al. Measuring Quality of Life in Oncologic Breast Surgery: A Systematic Review of Patient-Reported Outcome Measures. Breast J. 2010;16(6):587–97.

37. Danquah FVN, Wasserman J, Meininger J, Bergstrom N. Quality of life measures for patients on hemodialysis: a review of psychometric properties. Nephrol Nurs J. 2010;37(3):255–69; quiz 270.

38. Forhan M, Vrkljan B, MacDermid J. A systematic review of the quality of psychometric evidence supporting the use of an obesity-specific quality of life measure for use with persons who have class III obesity: Diagnostic in Obesity and Complications. Obes Rev. 2010;11(3):222–8.

39. Hill MR, Noonan VK, Sakakibara BM, Miller WC, SCIRE. Quality of life instruments and definitions in individuals with spinal cord injury: A systematic review. Spinal Cord. 2010;48(6):438–50.

40. Kamalski DM, Hoekstra CE, van Zanten BG, Grolman W, Rovers MM. Measuring disease-specific health-related quality of life to evaluate treatment outcomes in tinnitus patients: a systematic review. Otolaryngol Neck Surg. 2010;143(2):181–5.

41. Klassen AF, Strohm SJ, Maurice-Stam H, Grootenhuis MA. Quality of life questionnaires for children with cancer and childhood cancer survivors: a review of the development of available measures. Support Care Cancer. 2010 Sep;18(9):1207–17.

42. Kwon BE, Kim GY, Son YJ, Roh YS, You MA. Quality of life of women with urinary incontinence: a systematic literature review. Int Neurourol J. 2010;14(3):133–8.

43. Luckett T, King M, Butow P, Friedlander M, Paris T. Assessing health-related quality of life in gynecologic oncology: a systematic review of questionnaires and their ability to detect clinically important differences and change. Int J Gynecol Cancer. 2010;20(4):664–84.

44. Palfreyman SJ, Shackley P, Brazier JE. Assessing current health-related quality of life questionnaires administered to patients with venous ulcers: Can they be used in economic evaluations? Vol. 19, Journal of Clinical Nursing. 2010. p. 892–7.

45. Palfreyman SJ, Tod AM, Brazier JE, Michaels JA, SJ P, AM T, et al. A systematic review of health-related quality of life instruments used for people with venous ulcers: an assessment of their suitability and psychometric properties. J Clin Nurs. 2010;19(19–20):2673–703.

46. Passarelli P, Stefano SC, Blay SL. Validity Studies of Quality of Life Instruments for Eating Disorders. J Nerv Ment Dis. 2010;198(12):854–9.

47. Riordain RN, McCreary C. The use of quality of life measures in oral medicine: A review of the literature. Oral Dis. 2010;16(5):419–30.

48. Speight J, Howarth A. Quality of life in restless legs syndrome: A systematic review of clinical trials and a critical review of instruments. Vol. 3, The Patient. 2010. p. 185–203.

49. Street J, Lenehan B, Berven S, Fisher C. Introducing a New Health-Related Quality of Life Outcome Tool for Metastatic Disease of the Spine. Spine (Phila Pa 1976). 2010;35(14):1377–86.

50. Zeng YC, Ching SSY, Loke AY. Quality of life measurement in women with cervical cancer: implications for Chinese cervical cancer survivors. Vol. 8, Health and Quality of Life Outcomes. 2010. p. 30.

51. Carlton J, Kaltenthaler E. Amblyopia and quality of life: a systematic review. Eye. 2011;25(4):403–13.

52. Carlton J, Kaltenthaler E. Health-related quality of life measures (HRQoL) in patients with amblyopia and strabismus: A systematic review. Br J Ophthalmol. 2011;95(3):325–30.

53. da Silva SA, Passos SRL, Carballo MT, Figueiró M. Quality of Life Assessment after Acute Coronary Syndrome : Systematic Review. Arq Bras Cardiol. 2011;97(6):526–40.

54. Eckstein DA, Wu RL, Akinbiyi T, Silver L, Taub PJ. Measuring Quality of Life in Cleft Lip and Palate Patients. Plast Reconstr Surg. 2011;128(5):518e-526e.

55. Fayed N, Schiariti V, Bostan C, Cieza A, Klassen A. Health status and QOL instruments used in childhood cancer research: deciphering conceptual content using World Health Organization definitions. Qual Life Res. 2011;20(8):1247–58.

56. Glover C, Banks P, Carson A, Martin CR, Duffy T. Understanding and Assessing the Impact of Alcoholism on Quality of Life. Patient. 2011;4(1):10–30.

57. González-Consuegra RV, Verdú J. Quality of life in people with venous leg ulcers: An integrative review. J Adv Nurs. 2011;67(5):926–44.

58. Hounsome N, Orrell M, Edwards RT. EQ-5D as a quality of life measure in people with dementia and their carers: evidence and key issues. Value Heal. 2011;14(2):390–9.

59. Janssen MF, Lubetkin EI, Sekhobo JP, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. Diabet Med. 2011;28(4):395–413.

60. Kowal-Bielecka O, Avouac J, Pittrow D, Huscher D, Behrens F, Denton CP, et al. Analysis of the validation status of quality of life and functional disability measures in pulmonary arterial hypertension related to systemic sclerosis: Results of a systematic literature analysis by the expert panel on outcomes measures in pulmonary art. J Rheumatol. 2011;38(11):2419–27.

61. Lien K, Zeng L, Nguyen J, Cramarossa G, Culleton S, Caissie A, et al. Comparison of the EORTC QLQ-C15-PAL and the FACIT-Pal for assessment of quality of life in patients with advanced cancer. Expert Rev Pharmacoecon Outcomes Res. 2011;11(5):541–6.

62. Luckett T, King MT, Butow PN, Oguchi M, Rankin N, Price MA, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: Issues, evidence and recommendations. Ann Oncol. 2011;22(10):2179–90.

63. Mordiffi SZ, Kin YW, NK EA. Quality of life tools for adult patients with cancer undergoing chemotherapy: a systematic review. JBI. 2011;9(57):2482–532.

64. Noyes J, Edwards RT. EQ-5D for the assessment of health-related quality of life and resource allocation in children: A systematic methodological review. Value Heal. 2011;14(8):1117–29.

65. Papaioannou D, Brazier J, Parry G. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in Schizophrenia? A systematic review. Value Heal. 2011;14(6):907–20.

66. Reavey PL, Klassen AF, Cano SJ, McCarthy C, Scott A, Rubin JP, et al. Measuring quality of life and patient satisfaction after body contouring: a systematic review of patient-reported outcome measures. Aesthetic Surg J. 2011;31(7):807–13.

67. Schiariti V, Fayed N, Cieza A, Klassen A, O'Donnell M. Content comparison of health-related quality of life measures for cerebral palsy based on the International Classification of Functioning. Vol. 33, Disability and Rehabilitation. Informa Healthcare; 2011. p. 1330–9.

68. Tayyem R, Ali A, Atkinson J, Martin CR. Analysis of Health-Related Quality-of-Life Instruments Measuring the Impact of Bariatric Surgery. Patient. 2011;4(2):73–87.

69. Whitehurst DGT, Bryan S, Lewis M. Systematic review and empirical comparison of contemporaneous EQ-5D and SF-6D group mean scores. Med Decis Making. 2011;31(6):E34-44.

70. Wilson JR, Hashimoto RE, Dettori JR, Fehlings MG. Spinal cord injury and quality of life: a systematic review of outcome measures. Evid Based Spine Care J. 2011;2(1):37–44.

71. Bhatt JK, Thomas S, Nanjan MJ. Health outcome measures for diabetes mellitus: A review. Vol. 7, Applied Research in Quality of Life. Springer; 2012. p. 413–43.

72. Chopra I, Kamal KM. A systematic review of quality of life instruments in long-term breast cancer survivors. Health Qual Life Outcomes. 2012;10(1):1–15.

73. Cormier JN, Cromwell DD, Ross MI. Health-related quality of life in patients with melanoma: overview of instruments and outcomes. Dermatol Clin. 2012;30(2):245–54.

74. Correia FR, De Carlo MMR do P. Evaluation of quality of life in a palliative care context: an integrative literature review. Rev Lat Am Enfermagem. 2012;20(2):401–10.

75. Gräske J, Fischer T, Kuhlmey A, Wolf-Ostermann K. Dementia-Specific Quality of Life Instruments and Their Appropriateness in Shared-Housing Arrangements-A Literature Study. Geriatr Nurs (Minneap). 2012;33(3):204–16.

76. Ho AL, Scott AM, Klassen AF, Cano SJ, Pusic AL, Van Laeken N. Measuring Quality of Life and Patient Satisfaction in Facial Paralysis Patients. Plast Reconstr Surg. 2012 Jul;130(1):91–9.

77. Hogg FRA, Peach G, Price P, Thompson MM, Hinchliffe RJ. Measures of health-related quality of life in diabetes-related foot disease: A systematic review. Diabetologia. 2012;55(3):552–65.

78. Luquiens A, Reynaud M, Falissard B, Aubin HJ. Quality of life among alcohol-dependent patients: How satisfactory are the available instruments? A systematic review. Vol. 125, Drug and Alcohol Dependence. 2012. p. 192–202.

79. Madureira MM, Ciconelli RM, Pereira RMR. Quality of life measurements in patients with osteoporosis and fractures. Clinics. 2012;67(11):1315–20.

80. Milne A, Johnson JA, Tennant M, Rudniski C, Dryden DM. Measuring Health-Related Quality of Life for Patients With Diabetic Retinopathy. Technol Assess. 2012;267–309.

81. Ojo B, Genden EM, Teng MS, Milbury K, Misisukiewicz KJ, Badr H. A systematic review of head and neck cancer quality of life assessment instruments. Oral Oncol. 2012;48(10):92–937.

82. Popovic M, Nguyen J, Chen E, Di Giovanni J, Zeng L, Chow E. Comparison of the EORTC QLQ-BM22 and the FACT-BP for assessment of quality of life in cancer patients with bone metastases. Expert Rev Pharmacoeconomics Outcomes Res. 2012;12(2):213–9.

83. Quintanilla-Dieck L, Litvack JR, Made JC, Smith TL. Comparison of disease-specific quality-of-life instruments in the assessment of chronic rhinosinusitis. Int Forum Allergy Rhinol. 2012;2(6):437–43.

84. Rajmil L, Roizen M, Psy AU, Hidalgo-Rasmussen C, Fernandez G, Dapueto JJ. Health-related quality of life measurement in children and adolescents in Ibero-American countries, 2000 to 2010. Value Heal. 2012;15(2):312–22.

85. Shin H, Shin HS. Measurement of quality of life in menopausal women: a systematic review. West J Nurs Res. 2012;34(4):475–503.

86. Smith SC, Lamping DL, Maclaine GDH. Measuring health-related quality of life in diabetic peripheral neuropathy: a systematic review. Diabetes Res Clin Pract. 2012;96(3):261–70.

87. Tosh J, Brazier J, Evans P, Longworth L. A review of generic preference-based measures of health-related quality of life in visual disorders. Value Heal. 2012;15(1):118–27.

88. Townsend-White C, Pham ANT, Vassos M V. Review: a systematic review of quality of life measures for people with intellectual disabilities and challenging behaviours. J Intellect Disabil Res. 2012;56(3):270–84.

89. Walker H, Tulloch L, Martin C. Are they worth it? A systematic review of QOL instruments for use with mentally disordered offenders who have a diagnosis of psychosis. Vol. 14, The British Journal of Forensic Practice. 2012. p. 252–68.

90. Whitehurst DGT, Noonan VK, Dvorak MFS, Bryan S. A review of preference-based health-related quality of life questionnaires in spinal cord injury research. Spinal Cord. 2012;50(9):646–54.

91. Yip WK, Mordiffi SZ, Ang E. Reliability, validity and feasibility of quality of life instruments for adult patients with cancer undergoing chemotherapy: Result from a systematic review. Int J Evid Based Healthc. 2012;10(1):27–52.

92. Al Sayah F, Ishaque S, Lau D, Johnson JA. Health related quality of life measures in Arabic speaking populations: A systematic review on cross-cultural adaptation and measurement properties. Qual Life Res. 2013;22(1):213–29.

93. Basra MKA, Gada V, Ungaro S, Finlay AY, Salek SM. Infants' Dermatitis Quality of Life Index: a decade of experience of validation and clinical application. Br J Dermatol. 2013;169(4):760–8.

94. Castelino M, Abbott J, McElhone K, Teh LS. Comparison of the psychometric properties of health-related quality of life measures used in adults with systemic lupus erythematosus: A review of the literature. Rheumatology. 2013;52(4):684–96.

95. Chandratre P, Roddy E, Clarson L, Richardson J, Hider SL, Mallen CD. Health-related quality of life in gout: A systematic review. Rheumatology. 2013;52(11):2031–40.

96. Chow MYK, Morrow AM, Cooper Robbins SC, Leask J. Condition-specific quality of life questionnaires for caregivers of children with pediatric conditions: A systematic review. Qual Life Res. 2013;22(8):2183–200.

97. Davis S, Wailoo A. A review of the psychometric performance of the EQ-5D in people with urinary incontinence. Health Qual Life Outcomes. 2013;11:20.

98. de Almeida JR, Witterick IJ, Gullane PJ, Gentili F, Lohfeld L, Ringash J, et al. Quality of life instruments for skull base pathology: Systematic review and methodologic appraisal. Head Neck. 2013;36(10):1391.

99. Djan R, Penington A. A systematic review of questionnaires to measure the impact of appearance on quality of life for head and neck cancer patients. J Plast Reconstr Aesthetic Surg. 2013;66(5):647–59.

100. Gakhar H, Kamalli A, Holodniy M. Health-related Quality of Life Assessment after Antiretroviral Therapy: A Review of the Literatu. Drugs. 2013;73(7):651–72.

101. Hitzig SL, Balioussis C, Nussbaum E, McGillivray CF, Catharine Craven B, Noreau L. Identifying and classifying quality-of-life tools for assessing pressure ulcers after spinal cord injury. J Spinal Cord Med. 2013;36(6):600–15.

102. Jabir S. Assessing Improvement in Quality of Life and Patient Satisfaction following Body Contouring Surgery in Patients with Massive Weight Loss: A Critical Review of Outcome Measures Employed. Plast Surg Int. 2013;

103. Lee EH, Klassen AF, Nehal KS, Cano SJ, Waters J, Pusic AL. A systematic review of patient-reported outcome instruments of nonmelanoma skin cancer in the dermatologic population. J Am Acad Dermatol. 2013;69(2):e59–67.

104. Levterova BA, Dimitrova DD, Levterov GE, Dragova EA. Instruments for Disease-Specific Quality-of-Life Measurement in Patients with Type 2 Diabetes Mellitus - A Systematic Review / Инструменты Для Оценки Специфического Качества Жизни Пациентов, Больных Сахарным Диабетом Типа 2. Folia Med (Plovdiv). 2013;55(1):83–92.

105. Li C, Tsoi EWS, Zhang AL, Chen S, Wang CKJ. Psychometric properties of self-reported quality of life measures for people with intellectual disabilities: A systematic review. Vol. 25, Journal of Developmental and Physical Disabilities. Springer; 2013. p. 253–70.

106. Lin FJ, Longworth L, Pickard AS. Evaluation of content on EQ-5D as compared to disease-specific utility measures. Qual Life Res. 2013;22(4):853–74.

107. Mitera G, Zeiadin N, Sahgal A, Finkelstein J, Chow E, Loblaw A. Quality of life measures used in radiation therapy trials for patients with Metastatic Spinal Cord Compression (MSCC). In: Advanced Cancer. 2013. p. 97–106.

108. Mogos MF, August EM, Salinas-Miranda AA, Sultan DH, Salihu HM. A Systematic Review of Quality of Life Measures in Pregnant and Postpartum Mothers. Appl Res Qual Life. 2013;8(2):219–50.

109. Mousavi SA, Masoumi SZ, Keramat A, Pooralajal J, Shobeiri F, Abbas Mousavi S, et al. Assessment of Questionnaires Measuring Quality of Life in Infertile Couples: A Systematic Review. J Reprod Infertil. 2013 Jul;14(3):110–9.

110. Moyle W, Murfield JE. Health-related quality of life in older people with severe dementia: challenges for measurement and management. Expert Rev Pharmaeconomics Outcomes Res. 2013;13(1):109–22.

111. Muzzatti B, Annunziata MA. Assessing quality of life in long-term cancer survivors: a review of available tools. Support Care Cancer. 2013;21(11):3143–52.

112. Paltzer J, Barker E, Witt WP. Measuring the health-related quality of life (HRQoL) of young children in resource-limited settings: a review of existing measures. Qual Life Res. 2013;22(6):1177–87.

113. Perales J, Cosco TD, Stephan BCM, Haro JM, Brayne C. Health-related quality-of-life instruments for Alzheimer's disease and mixed dementia. Int Psychogeriatrics. 2013;25(5):691–706.

114. Pusic AL, Cemal Y, Albornoz C, Klassen A, Cano S, Sulimanoff I, et al. Quality of life among breast cancer patients with lymphedema: A systematic review of patient-reported outcome instruments and outcomes. J Cancer Surviv. 2013;7(1):83–92.

115. Roncada C, Mattiello R, Pitrez PM, Sarria EE. Specific instruments to assess quality of life in children and adolescents with asthma. J Pediatr (Rio J). 2013;89(3):217–25.

116. Salek MS, Jung S, Brincat-Ruffini LA, MacFarlane L, Lewis-Jones MS, Basra MKA, et al. Clinical experience and psychometric properties of the Children's Dermatology Life Quality Index (CDLQI), 1995-2012. Br J Dermatol. 2013;169(4):734–59.

117. Testart J, Richieri R, Caqueo-Urízar A, Lancon C, Auquier P, Boyer L. Quality of life and other outcome measures in caregivers of patients with schizophrenia. Expert Rev Pharmacoeconomics Outcomes Res. 2013;13(5):641–9.

118. Weldam SWM, Schuurmans MJ, Liu R, Lammers JWJ. Evaluation of Quality of Life instruments for use in COPD care and research: A systematic review. Int J Nurs Stud. 2013;50(5):688–707.

119. Wheelwright S, Darlington A-S, Hopkinson JB, Fitzsimmons D, White A, Johnson CD. A systematic review of health-related quality of life instruments in patients with cancer cachexia. Support Care Cancer. 2013;21(9):2625–36.

120. Yang Y, Longworth L, Brazier J. An assessment of validity and responsiveness of generic measures of health-related quality of life in hearing impairment. Vol. 22, Quality of Life Research. 2013. p. 2813–28.

121. Anthony SJ, Selkirk E, Sung L, Klaassen RJ, Dix D, Scheinemann K, et al. Considering quality of life for children with cancer: a systematic review of patient-reported outcome measures and the development of a conceptual model. Qual Life Res. 2014;23(3):771–89.

122. Aspden T, Bradshaw SA, Playford ED, Riazi A. Quality-of-life measures for use within care homes: A systematic review of their measurement properties. Age Ageing. 2014;43(5):596–603.

123. Balioussis C, Hitzig S, Flett H, Noreau L, Craven B. Identifying and Classifying Quality of Life Tools for Assessing Spasticity After Spinal Cord Injury. Top Spinal Cord Inj Rehabil. 2014;20(3):208–24.

124. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess (Rockv). 2014;18(34):1–188.

125. Chiu L, Chiu N, Chow E, Cella D, Beaumont JL, Lam H, et al. Comparison of Three Shortened Questionnaires for Assessment of Quality of Life in Advanced Cancer. J Palliat Med. 2014;17(8):918–23.

126. Chow R, Lao N, Popovic M, Chow E, Cella D, Beaumont J, et al. Comparison of the EORTC QLQ-BN20 and the FACT-Br quality of life questionnaires for patients with primary brain cancers: a literature review. Support Care Cancer. 2014;22(9):2593–8.

127. 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(3):359–67.

128. Gilchrist F, Rodd H, Deery C, Marshman Z. Assessment of the quality of measures of child oral health-related quality of life. BMC Oral Health. 2014;14:40.

129. Grubbs JR, Tolleson-Rinehart S, Huynh K, Davis RM. A review of Quality of Life Measures in Dry Eye Questionnaires. 2014;33(2):215–8.

130. Gupta N, Pinto LM, Morogan A, Bourbeau J. The COPD assessment test: a systematic review. Eur Respir J. 2014;44(4):873–84.

131. Hawkins AT, Henry AJ, Crandell DM, Nguyen LL. A systematic review of functional and quality of life assessment after major lower extremity amputation. Ann Vasc Surg. 2014;28(3):763–80.

132. Hewison A, Mccaughan D, Watt I. An evaluative review of questionnaires recommended for the assessment of quality of life and symptom severity in women with urinary incontinence. J Clin Nurs. 2014;23(21–22):2998–3011.

133. Ikeda E, Hinckson E, Krageloh C. Assessment of quality of life in children and youth with autism spectrum disorder: a critical review. Qual Life Res. 2014;23(4):1069–85.

134. Jardine J, Glinianaia S V, McConachie H, Embleton ND, Rankin J. Self-reported quality of life of young children with conditions from early infancy: a systematic review. Pediatrics. 2014;134(4):e1129-48.

135. Kuspinar A, Mayo NE. A review of the psychometric properties of generic utility measures in multiple sclerosis. Vol. 32, PharmacoEconomics. Springer; 2014. p. 759–73.

136. Lee J, Kim SH, Moon SH, Lee EH. Measurement properties of rheumatoid arthritis-specific quality-of-life questionnaires: systematic review of the literature. Qual Life Res. 2014;23(10):2779–91.

137. Lieu JEC, Chalivendra V, Ead B. Pediatric quality of life in children with otolaryngologic disease: what inventories are available and what is still needed? Curr Opin Otolaryngol Head Neck Surg. 2014;22(6):506–20.

138. Longworth L, Yang Y, Young T, Mulhern B, Hernández Alava M, Mukuria C, et al. Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: A systematic review, statistical modelling and survey. Health Technol Assess (Rockv). 2014;18(9):1–224.

139. Makai P, Brouwer WBF, Koopmanschap MA, Stolk EA, Nieboer AP. Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. Soc Sci Med. 2014;102:83–93.

140. Niu H-Y, Niu C-Y, Wang J-H, Zhang Y, He P. Health-related quality of life in women with breast cancer: a literature-based review of psychometric properties of breast cancer-specific measures. Asian Pacific J Cancer Prev. 2014;15(8):3533–6.

141. Salvilla SA, Dubois AEJ, Flokstra-De Blok BMJ, Panesar SS, Worth A, Patel S, et al. Disease-specific health-related quality of life instruments for IgE-mediated food allergy. Allergy. 2014;69(7):834–44.

142. 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 Oct;23(8):2169–81.

143. Smith AB, Cocks K, Taylor M, Parry D. Most domains of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 are reliable. J Clin Epidemiol. 2014;67(8):952–7.

144. Souza JGS, Pamponet MA, Souza TCS, Pereira AR, Souza AGS, Martins AME d. BL. Tools used for evaluation of Brazilian children's quality of life. Rev Paul Pediatr. 2014;32(2):272–8.

145. Swigris JJ, Esser D, Conoscenti CS, Brown KK. The psychometric properties of the St George's Respiratory Questionnaire (SGRQ) in patients with idiopathic pulmonary fibrosis: a literature review. Health Qual Life Outcomes. 2014;12:124.

146. Timmerman AA, Speyer R, Heijnen BJ, Klijn-Zwijnenberg IR. Psychometric characteristics of health-related quality-of-life questionnaires in oropharyngeal dysphagia. Dysphagia. 2014;29(2):183–98.

147. Watt T, Cramon P, Frendl DM, Ware JE. Assessing health-related quality of life in patients with benign non-toxic goitre. Best Pract Res Clin Endocrinol Metab. 2014;28(4):559–75.

148. Wolpe RE, Toriy AM, da Silveira GF, Cardoso FL, Sperandio FF. Assessing the impact of urinary incontinence on quality of life: systematic review of instruments in Portuguese. Man Ther Posturology Rehabil J. 2014;12(1):273–80.

149. Alrubaiy L, Rikaby I, Dodds P, Hutchings HA, Williams JG. Systematic review of health-related quality of life measures for inflammatory bowel disease. J Crohn's Colitis. 2015;9(3):284–92.

150. Aspesberro F, Mangione-Smith R, Zimmerman JJ. Health-related quality of life following pediatric critical illness. Intensive Care Med. 2015;41(7):1235–46.

151. Bédard E, Kergoat HH, Kergoat M-J, Leclerc B-S, Bedard E, Kergoat HH, et al. Systematic review of vision-related quality of life questionnaires for older institutionalised seniors with dementia. Ophthalmic Physiol Opt. 2015;35(4):377–87.

152. Bowling A, Rowe G, Adams S, Sands P, Samsi K, Crane M, et al. Quality of life in dementia: a systematically conducted narrative review of dementia-specific measurement scales. Aging Ment Health. 2015;19(1):13–31.

153. Conijn AP, Jens S, Terwee CB, Breek JC, Koelemay MJW. Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist. Eur J Vasc Endovasc Surg. 2015;49(3):316–34.

154. de Climens AR, Tunceli K, Arnould B, Germain N, Iglay K, Norquist J, et al. Review of patient-reported outcome instruments measuring health-related quality of life and satisfaction in patients with type 2 diabetes treated with oral therapy. Curr Med Res Opin. 2015 Apr;31(4):643–65.

155. Dronavalli M, Thompson SC. A systematic review of measurement tools of health and well-being for evaluating community-based interventions. Vol. 69, Journal of Epidemiology and Community Health. 2015. p. 805–15.

156. Hamoen EHJ, De Rooij M, Witjes JA, Barentsz JO, Rovers MM. Measuring health-related quality of life in men with prostate cancer: A systematic review of the most used questionnaires and their validity. Urol Oncol. 2015;33(2):69.e19-69.e28.

157. Hu H, Luan L, Li S-C. How Quality of Life as Patient-Reported Outcome Has Been Studied for Rheumatoid Arthritis in Chinese-Speaking Population. Value Heal Reg Issues. 2015;6:98–102.

158. Janssens A, Rogers M, Gumm R, Jenkinson C, Tennant A, Logan S, et al. Measurement properties of multidimensional patient-reported outcome measures in neurodisability: a systematic review of evaluation studies. Dev Med Child Neurol. 2015;58(5):437–51.

159. Launois R. Health-related quality-of-life scales specific for chronic venous disorders of the lower limbs. J Vasc Surg Venous Lymphat Disord. 2015;3(2):219-227.e3.

160. Monticone M, Nava C, Leggero V, Rocca B, Salvaderi S, Ferrante S, et al. Measurement properties of translated versions of the Scoliosis Research Society-22 Patient Questionnaire, SRS-22: a systematic review. Qual Life Res. 2015;24(8):1981–98.

161. Nguyen J, Popovic M, Chow E, Cella D, Beaumont JL, Chu D, et al. EORTC QLQ-BR23 and FACT-B for the assessment of quality of life in patients with breast cancer: A literature review. J Comp Eff Res. 2015;4(2):157–66.

162. Oliveira IS, da Cunha Menezes Costa L, Fagundes FRC, Cabral CMN. Evaluation of cross-cultural adaptation and measurement properties of breast cancer-specific quality-of-life questionnaires: a systematic review. Qual Life Res. 2015;24(5):1179–95.

163. Polinder S, Haagsma JA, van Klaveren D, Steyerberg EW, van Beeck EF. Health-related quality of life after TBI: a systematic review of study design, instruments, measurement properties, and outcome. Popul Health Metr. 2015;13:4.

164. Taghavi SA, Bazarganipour F, Montazeri A, Kazemnejad A, Chaman R, Khosravi A. Health-related quality of life in polycystic ovary syndrome patients: A systematic review. Iran J Reprod Med. 2015;13(8):473–82.

165. Treanor C, Donnelly M. A methodological review of the Short Form Health Survey 36 (SF-36) and its derivatives among breast cancer survivors. Qual Life Res. 2015;24(2):339–62.

166. Wong CKH, Chen J, Yu CLY, Sham M, Lam CLK. Systematic review recommends the European Organization for Research and Treatment of Cancer colorectal cancer-specific module for measuring quality of life in colorectal cancer patients. J Clin Epidemiol. 2015;68(3):266–78.

167. Algar K, Woods RT, Windle G. Measuring the quality of life and well-being of people with dementia: A review of observational measures. Dementia. 2016;15(4):832–57.

168. Bryant AL, Walton A, Shaw-Kokot J, Mayer DK, Reeve BB. A systematic review of psychometric properties of health-related quality-of-life and symptom instruments in adult acute leukemia survivors. Cancer Nurs. 2016;39(5):375–82.

169. Coombes LH, Wiseman T, Lucas G, Sangha A, Murtagh FE. Health-related quality-of-life outcome measures in paediatric palliative care: A systematic review of psychometric properties and feasibility of use. Palliat Med. 2016;30(10):935–49.

170. Dichter MN, Schwab CGG, Meyer G, Bartholomeyczik S, Halek M. Linguistic validation and reliability properties are weak investigated of most dementia-specific quality of life measurements - A systematic review. J Clin Epidemiol. 2016;70:233–45.

171. Ganesh V, Agarwal A, Popovic M, Cella D, McDonald R, Vuong S, et al. Comparison of the FACT-C, EORTC QLQ-CR38, and QLQ-CR29 quality of life questionnaires for patients with colorectal cancer: a literature review. Support Care Cancer. 2016;24(8):3661–8.

172. Gutiérrez-Vargas R, Díaz-García ML, Villasís-Keever MÁ, Portilla-Robertson J, Zapata-Tárres M. Instruments to measure the quality of life in patients with oral mucositis undergoing oncological treatment: a systmatic review of the literature. Bol Med Hosp Infant Mex. 2016;73(6):457–66.

173. Hand C. Measuring health-related quality of life in adults with cronic conditions in primary care settings. Canadian Family Physician. Can Fam Physician. 2016;62(7):375–83.

174. Heinl D, Prinsen CAC, Deckert S, Chalmers JR, Drucker AM, Ofenloch R, et al. Measurement properties of adult quality-of-life measurement instruments for eczema: a systematic review. Allergy. 2016;71(3):358–70.

175. Kotecha D, Ahmed A, Calvert M, Lencioni M, Terwee CB, Lane DA. Patient-Reported Outcomes for Quality of Life Assessment in Atrial Fibrillation: A Systematic Review of Measurement Properties. PLoS One. 2016;11(11):e0165790.

176. Lee J, Lee E-H, Moon SH. A systematic review of measurement properties of the instruments measuring health-related quality of life in patients with irritable bowel syndrome. Qual Life Res. 2016;25(12):2985–95.

177. Maratia S, Cedillo S, Rejas J. Assessing health-related quality of life in patients with breast cancer: a systematic and standardized comparison of available instruments using the EMPRO tool. Qual Life Res. 2016;25(10):2467–80.

178. Mestre TA, van Duijn E, Davis AM, Bachoud-Lévi AC, Busse M, Anderson KE, et al. Rating scales for behavioral symptoms in Huntington's disease: Critique and recommendations. Mov Disord. 2016;31(10):1466–78.

179. Spinou A, Fragkos KC, Lee KK, Elston C, Siegert RJ, Loebinger MR, et al. The validity of health-related quality of life questionnaires in bronchiectasis: a systematic review and meta-analysis. Thorax. 2016;71(8):683–94.

180. Tapia VJ, Epstein S, Tolmach OS, Hassan AS, Chung NN, Gosman AA. Health-related quality-of-life instruments for pediatric patients with diverse facial deformities: A systematic literature review. Plast Reconstr Surg. 2016;138(1):175–87.

181. Wong CKH, Lang BHH, Lam CLK. A systematic review of quality of thyroid-specific health-related quality-of-life instruments recommends ThyPRO for patients with benign thyroid diseases. J Clin Epidemiol. 2016;78:63–72.

182. Woo A, Fu T, Popovic M, Chow E, Cella D, Wong CS, et al. Comparison of the EORTC STO-22 and the FACT-Ga quality of life questionnaires for patients with gastric cancer. Ann Palliat Med. 2016;5(1):13–21.

183. Ahmadi A, Tohidast SA, Mansuri B, Kamali M, Krishnan G. Acceptability, reliability, and validity of the Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39) across languages: a systematic review. Clin Rehabil. 2017;31(9):1201–14.

184. Baghdadli A, Russet F, Mottron L. Measurement properties of screening and diagnostic tools for autism spectrum adults of mean normal intelligence: A systematic review. Eur Psychiatry. 2017;44:104–24.

185. Best KL, Ethans K, Craven BC, Noreau L, Hitzig SL. Identifying and classifying quality of life tools for neurogenic bladder function after spinal cord injury: A systematic review. J Spinal Cord Med. 2017;40(5):505–29.

186. Chen X-L, Zhong L-H, Wen Y, Liu T-W, Li X-Y, Hou Z-K, et al. Inflammatory bowel disease-specific health-related quality of life instruments: a systematic review of measurement properties. Health Qual Life Outcomes. 2017;15(1):177.

187. Frew JW, Davidson M, Murrell DF. Disease-specific health related quality of life patient reported outcome measures in Genodermatoses: a systematic review and critical evaluation. Orphanet J Rare Dis. 2017;12(1):189.

188. Heaney A, Stepanous J, Rouse M, McKenna SP. A review of the psychometric properties and use of the rheumatoid arthritis quality of life questionnaire (Raqol) in clinical research. Curr Rheumatol Rev. 2017;13(3):197–205.

189. Heinl D, Prinsen CAC, Sach T, Drucker AM, Ofenloch R, Flohr C, et al. Measurement properties of quality-of-life measurement instruments for infants, children and adolescents with eczema: a systematic review. Br J Dermatol. 2017;176(4):878–89.

190. Kandel H, Khadka J, Goggin M, Pesudovs K. Patient-reported Outcomes for Assessment of Quality of Life in Refractive Error: A Systematic Review. Optom Vis Sci. 2017;94(12):1102–19.

191. Kao SS, Peters MDJ, Ooi EH. Pediatric tonsillectomy quality of life assessment instruments: A scoping review protocol. Laryngoscope. 2017;127:2399–406.

192. Khan S, Tangiisuran B, Imtiaz A, Zainal H. Health Status and Quality of Life in Tuberculosis: Systematic Review of Study Design, Instruments, Measuring Properties and Outcomes. Heal Sci J. 2017;11(1):1–10.

193. Kwan YH, Fong W, Tan VIC, Lui NL, Malhotra R, Østbye T, et al. A systematic review of quality-of-life domains and items relevant to patients with spondyloarthritis. Semin Arthritis Rheum. 2017;47(2):175–82.

194. Limperg PF, Terwee CB, Young NL, Price VE, Gouw SC, Peters M, et al. Health-related quality of life questionnaires in individuals with haemophilia: a systematic review of their measurement properties. Haemophilia. 2017;23(4):497–510.

195. Lucendo AJ, Arias-González L, Molina-Infante J, Arias Á. Systematic review: health-related quality of life in children and adults with eosinophilic oesophagitis-instruments for measurement and determinant factors. Aliment Pharmacol Ther. 2017;

196. Page TE, Farina N, Brown A, Daley S, Bowling A, Basset T, et al. Instruments measuring the disease-specific quality of life of family carers of people with neurodegenerative diseases: A systematic review. BMJ Open. 2017;7(3).

197. Poku E, Aber A, Phillips P, Essat M, Buckley Woods H, Palfreyman S, et al. Systematic review assessing the measurement properties of patient-reported outcomes for venous leg ulcers. BJS open. 2017;1(5):138–47.

198. Roydhouse J, Wilson I b. Systematic review of caregiver responses for patient health- related quality of life in adult cancer care. Qual Life Res. 2017;68(8):1925–54.

199. Strada L, Vanderplasschen W, Buchholz A, Schulte B, Muller AE, Verthein U, et al. Measuring quality of life in opioid-dependent people: a systematic review of assessment instruments. Qual Life Res. 2017;26:3187–200.

200. Sullivan KJ, Hunter Z, Andrioli V, Guerra L, Leonard M, Klassen A, et al. Assessing quality of life of patients with hypospadias: A systematic review of validated patient-reported outcome instruments. J Pediatr Urol. 2017;13(1):19–27.

201. Tang TS, Yusuf FLA, Polonsky WH, Fisher L. Assessing quality of life in diabetes: II – Deconstructing measures into a simple framework. Diabetes Res Clin Pract. 2017;126:286–302.

202. Tax C, Steenbergen ME, Zusterzeel PLM, Bekkers RLM, Rovers MM. Measuring health-related quality of life in cervical cancer patients: a systematic review of the most used questionnaires and their validity. BMC Med Res Methodol. 2017;17(1):15.

203. Xin Y, Mcintosh E. Assessment of the construct validity and responsiveness of preference-based quality of life measures in people with Parkinson's: a systematic review. Qual Life Res. 2017;26(1):1–23.

204. Aber A, Lumley E, Phillips P, Woods HB, Jones G, Michaels J. Themes that Determine Quality of Life in Patients with Peripheral Arterial Disease: A Systematic Review. Patient. 2018;11(5):489–502.

205. Chiarotto A, Terwee CB, Kamper SJ, Boers M, Ostelo RW. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain, a systematic review. J Clin Epidemiol. 2018;102:23–7.

206. Cornelissen AJ, Kool M, Keuter XH, Hetus EM, Grzymala AAP de, van der Hulst RR, et al. Concerning Quality of Life Questionnaires in Breast Cancer-Related Lymphedema Patients: Review of the Literature by Cornelissen et al. Lymphat Res Biol. 2018;16(2):134–9.

207. de Vries CEE, Kalff MC, Prinsen CAC, Coulman KD, den Haan C, Welbourn R, et al. Recommendations on the most suitable quality-of-life measurement instruments for bariatric and body contouring surgery: a systematic review. Obes Rev. 2018;19(10):1395–411.

208. Dow J, Robinson J, Robalino S, Finch T, McColl E, Robinson L. How best to assess quality of life in informal carers of people with dementia; A systematic review of existing outcome measures. PLoS One. 2018;13(3).

209. Grobet C, Marks M, Tecklenburg L, Audige L. Application and measurement properties of EQ-5D to measure quality of life in patients with upper extremity orthopaedic disorders: a systematic literature review. Arch Orthop Trauma Surg. 2018;138:953–61.

210. Haywood KL, Pearson N, Morrison LJ, Castrén M, Lilja G, Perkins GD. Assessing health-related quality of life (HRQoL) in survivors of out-of-hospital cardiac arrest: A systematic review of patient-reported outcome measures. Resuscitation. 2018;123:22–37.

211. Luan L, Hu H, Li S-C. A Review of Studies of Quality of Life for Chinese-Speaking Patients with Ischemic Heart Disease. Value Heal Reg Issues. 2018;15:82–90.

212. Mason SJ, Catto JWF, Downing A, Bottomley SE, Glaser AW, Wright P. Evaluating patient-reported outcome measures (PROMs) for bladder cancer: a systematic review using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist. BJU Int. 2018;

213. Mohammed MA, Moles RJ, Chen TF. Pharmaceutical care and health related quality of life outcomes over the past 25 years: Have we measured dimensions that really matter? Int J Clin Pharm. 2018;40(1):3–14.

214. Mpundu-Kaambwa C, Chen G, Huynh E, Russo R, Ratcliffe J. A review of preference-based measures for the assessment of quality of life in children and adolescents with cerebral palsy. Qual Life Res. 2018;27(7):1781–99.

215. Pollo CF, Meneguin S, Miot HA. Evaluation Instruments for Quality of Life Related to Melasma: An Integrative Review. Clinics. 2018;73:e65.

216. Tian L, Cao XY. Systematic review of the psychometric properties of disease-specific, quality-of-life questionnaires for patients with hepatobiliary or pancreatic cancers. Japan J Nurs Sci. 2018;15(2):99–112.

217. van Ierssel J, Sveistrup H, Marshall S. Identifying the concepts contained within health-related quality of life outcome measures in concussion research using the International Classification of Functioning, Disability, and Health as a reference: a systematic review. Qual Life Res. 2018;27(12):3071–86.

218. van Roij J, Fransen H, van de Poll-Franse L, Zijlstra M, Raijmakers N. Measuring health-related quality of life in patients with advanced cancer: a systematic review of self-administered measurement instruments. Qual Life Res. 2018;27(8):1937–55.

219. Yarlas A, Bayliss M, Cappelleri JC, Maher S, Bushmakin AG, Chen LA, et al. Psychometric validation of the SF-36®Health Survey in ulcerative colitis: results from a systematic literature review. Qual Life Res. 2018;27(2):273–90.

220. Yazdani N, Sharif F, Elahi N, Ebadi A, Hosseini SV. Psychometric Properties of Quality of Life Assessment Tools in Morbid Obesity: A Review of Literature. J Evidence-based Care. 2018;7(4):7–21.

221. Zaror C, Pardo Y, Espinoza-Espinoza G, Pont A, Munoz-Millan P, Martinez-Zapata MJ, et al. Assessing oral health-related quality of life in children and adolescents: a systematic review and standardized comparison of available instruments. Clin Oral Investig. 2018;

222. Hettiarachchi RM, Kularatna S, Byrnes J, Scuffham PA. Pediatric Quality of Life Instruments in Oral Health Research: A Systematic Review. Value Heal. 2019;22(1):129–35.

223 Balk, E. M., Gazula, A., Markozannes, G., Kimmel, H. J., Saldanha, I. J., Trikalinos, T. A., & Resnik, L. J. (2019). Psychometric Properties of Functional , Ambulatory , and Quality of Life Instruments in Lower Limb Amputees : A Systematic Review. Archives of Physical Medicine and Rehabilitation, 100(12), 2354–2370. https://doi.org/10.1016/j.apmr.2019.02.015

224 Belayneh, T. (2019). A systematic review of the psychometric properties of the cross-cultural adaptations and translations of the Prolapse Quality of Life (P-QoL) questionnaire. Belayneh Abebaw Gebeyu Mulat Adefris Guri Rortveit, 30, 1989–2000. https://doi.org/10.1007/s00192-019-03920-1

225 Bull, K. S., Hornsey, S., Kennedy, C. R., Darlington, A.-S. E., Grootenhuis, M. A., Hargrave, D., Liossi, C., Jonathan, P., Walker, D. A., & Morris, C. (2019). Systematic review: measurement properties of patient reported outcome measures evaluated with chidlhood brain tumor survivors or other acquired brain injury. Neuro-Oncology Practice, December. https://doi.org/10.1093/nop/npz064

226 Choukou, M.-A., Best, K. L., Catharine, B., & Hitzig, S. L. (2019). Identifying and Classifying Quality of Life Tools for Assessing Neurogenic Bowel Dysfunction After Spinal Cord Injury. Spnial Cord Injury Rehabiliation, 25(1), 1–22. https://doi.org/10.1310/sci18-00019

227 Daliya, P., Gemmill, E. H., Lobo, D. N., & Parsons, S. L. (2019). A systematic review of patient reported outcome measures (PROMs) and quality of life reporting in patients undergoing laparoscopic cholecystectomy. Hepatobiliary Surgery and Nutrition, 8(3), 228–245. https://doi.org/10.21037/hbsn.2019.03.16

228 Gabes, M., Tischer, C., & Apfelbacher, C. (2020). Measurement properties of quality - of - life outcome measures for children and adults with eczema : An updated systematic review. Pediatric Allergy and Immunology, 31, 66–77. https://doi.org/10.1111/pai.13120

229 Gondivkar, S. M., Bhowate, R. R., Gadbail, A. R., Sarode, S. C., & Gondivkar, R. S. (2019). Assessment of oral health-related quality of life instruments for oral submucous fi brosis : A systematic review using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. Oral Oncology, 93(January), 39–45. https://doi.org/10.1016/j.oraloncology.2019.04.009

230 Hasanvand, S., Rassouli, M., Mandegari, Z., & Salmani, N. (2019). A Critical Review of Instruments Measuring the Quality of Life of Cancer Patients in Iranian Studies and Their Psychometrics Properties. Asian Pacific Journal of Cancer Prevention, 20, 333–343. https://doi.org/10.31557/APJCP.2019.20.2.333

231 Hughes LJ, Farina N, Page TE, Tabet N, Banerjee S. Psychometric properties and feasibility of use of dementia specific quality of life instruments for use in care settings: a systematic review. Int Psychogeriatrics. 2019;1–15.

232 Hunt, C., Zahid, S., Ennis, N., Michalak, A., Masanic, C., Vaidyanath, C., Bhalerao, S., Cusimano, M. D., Baker, A., & Hunt, C. (2019). Quality of life measures in older adults after traumatic brain injury : a systematic review. Quality of Life Research, 28(12), 3137–3151. https://doi.org/10.1007/s11136-019-02297-4

233 Kamilu, S., Heather, S., Aldersey, M., Fayed, N., Kaka, B., & Okyere, C. (2019). Quality of life assessment scales in polio survivors : a scoping review. Quality of Life Research, 28(9), 2341–2357. https://doi.org/10.1007/s11136-019-02185-x

234 Lamsal, R., Finlay, B., Whitehurst, D. G. T., & Zwicker, J. D. (2019). Generic preference-based health-related quality of life in children with neurodevelopmental disorders : a scoping review. Development Medicine and Child Neurology, 169–177. https://doi.org/10.1111/dmcn.14301

235 Moshki, M., Khajavi, A., Vakilian, F., Minaee, S., & Hashemizadeh, H. (2019). The content comparison of health-related quality of life measures in heart failure based on the international classification of functioning , disability , and health : a systematic review. Tabriz University of Medical Sciences, 11(3), 167–175. https://doi.org/10.15171/jcvtr.2019.29

236. Speyer R, Kim JH, Doma K, Chen YW, Denman D, Phyland D, et al. Measurement properties of self-report questionnaires on health-related quality of life and functional health status in dysphonia: a systematic review using the COSMIN taxonomy. Qual Life Res. 2019;28(2):283–96.

237 van der Hout A, Neijenhuijs KI, Jansen F, van Uden-Kraan C, Aaronson NK, Groenvold M, et al. Supportive Care in Cancer Measuring Health-Related Quality of Life in Colorectal Cancer Patients : Systematic Review of Measurement Properties of the EORTC QLQ-CR29. Support Care Cancer. 2019;

238 Vasconcelos Neto JA, Vasconcelos CTM, Karbage SAL, Farias HDCDAR, Machado SGDM, Saboia DM. Quality of Life in Women with Defecatory Dysfunctions: Systematic Review of Questionnaires Validated in the Portuguese Language. Rev Bras Ginecol e Obstet. 2019;41(3):191–8.

239 Crudgington, H., Pal, D. K., Rogers, M., Morris, C., & Morris, H. (2020). Epilepsy-specific patient-reported outcome measures of children ' s health-related quality of life : A systematic review of measurement properties. Epilepsia, 61, 230–248. https://doi.org/10.1111/epi.16430

240 Furtado, R., Macdermid, J. C., Nazari, G., Bryant, D. M., Faber, K. J., & Athwal, G. S. (2020). Cross-cultural adaptions and measurement properties of the WORC (Western Ontario rotator cuff index): a systematic review. Health and Quality of Life Outcomes, 9, 1–9. https://doi.org/10.1186/s12955-020-1276-9

241 Jones, F. J. S., Ezzeddine, F. L., Herman, S. T., Buchhalter, J., Fureman, B., & Moura, L. M. V. R. (2020). A feasibility assessment of functioning and quality-of-life patient-reported outcome measures in adult epilepsy clinics : A systematic review. Epilepsy & Behavior, 102, 106704. https://doi.org/10.1016/j.yebeh.2019.106704

242 Killian, M. O., Triplett, K. N., Masood, S. S., Boehler, J., & Mayersohn, G. S. (2020). Measurement of health - related quality of life in pediatric organ transplantation recipients : a systematic review of the PedsQL transplant module. Quality of Life Research, 29(5), 1137–1146. https://doi.org/10.1007/s11136-019-02398-0

243 Losada-Puente, L., Araujo, A. M., & Muñoz-Cantero, J. M. (2020). A Systematic Review of the Assessment of Quality of Life in Adolescents. Social Indicators Research, 147(3), 1039–1057. https://doi.org/10.1007/s11205-019-02171-3

244 Møller, A., Bissenbakker, K. H., Arreskov, A. B., & Brodersen, J. (2020). Specific Measures of Quality of Life in Patients with Multimorbidity in Primary Healthcare : A Systematic Review on Patient-Reported Outcome Measures Adequacy of Measurement.Patient Related Outcome Measures, 11, 1–10. https://doi.org/10.2147/PROM.S226576.

245 Qian, X., Lee, R., Tan, Y., Hsiang, L., & Nan, C. (2020). Measurement Properties of Commonly Used Generic Preference - Based Measures in East and South - East Asia : A Systematic Review. PharmacoEconomics, 38(2), 159–170. https://doi.org/10.1007/s40273-019-00854-w

246 Santana-Berlanga, N. R., Porcel-Gálvez, A. M., Botello-Hermosa, A., & Barrientos-Trigo, S. (2020). Instruments to measure quality of life in institutionalised older adults : Systematic review. Geriatric Nursing, 000. https://doi.org/10.1016/j.gerinurse.2020.01.018

9.2.4. HERRAMIENTAS PARA EVALUAR LAS PROPIEDADES DE LA MEDIDA

[1]COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) (2010, 2018)	
1. Reliability	The degree to which an instrument is free from random error.
1.1. Internal consistency	The degree of the interrelatedness among the items. *In COSMIN (2018) internal consistency is derived from internal structure evaluation.*
1.2. Reliability	Scores for patients who have not changed are the same for repeated measurement under several conditions
1.3. Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
2. Validity	The degree to which a Health Related-Patient Reported Outcome (HR-PRO) instrument measures the construct(s) it purports to measure. *Concept with major changes in COSMIN (2018) the definition and classification changed to content, structural, cross-cultural validity/measurement invariance, criterion, and hypothesis testing for construct validity (convergent, discriminative or known groups)*
2.1. Content (including Face validity)	The degree to which the content of an HR-PRO instrument is an adequate reflection of the construct to be measured (or looks as though the items are an adequate reflection)
2.2. Construct (Structural, Hypothesis, Cross-cultural)	The degree to which the scores of an HR-PRO instrument are an adequate reflection of the dimensionality of the construct to be measured. Scores of an HR-PRO instrument are consistent with hypotheses. Performance of the items on a translated or culturally adapted HR-PRO instrument is an adequate reflection of the performance of the items of the original version of the HR-PRO instrument
2.3. Criterion	The degree to which the scores of an HR-PRO instrument are an adequate reflection of a "gold standard"
3. Responsiveness	The instrument's ability to detect change over time in the construct to be measured
4. Interpretability	The degree to which one can assign easily understood meaning to an instrument's quantitative scores. *A complementary attribute, not a measurement property in COSMIN (2018), plus feasibility*
[2] Quality Criteria for Measurement Properties (Terwee et al. 2007)	
1. Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire
2. Internal consistency	The extent to which items in a (sub)scale are inter correlated, thus measuring the same construct
3. Criterion validity	The extent to which scores on a particular questionnaire relate to a gold standard
4. Construct validity	The extent to which scores on a particular questionnaire relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured
5. Reproducibility	
5.1. Agreement	The extent which the scores on repeated measures are close to each other (absolute measurement error)
5.2. Reliability	The extent to which patients can be distinguish from each other (relative measurement error)
6. Responsiveness	The ability of a questionnaire to detect clinically important changes over time
7. Floor and ceiling effects	The number of respondents who achieved the lowest or highest possible score
8. Interpretability	The degree to which one can assign qualitative meaning to quantitative scores

[1]Prinsen C, Mokkink L, Bouter L, et al. COSMIN guideline for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res. 2018;0(0):1-11. doi:10.1007/s11136-018-1798-3. Mokkink L, Terwee C, Patrick D, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-745. doi: 10.1016/j.jclinepi.2010.02.006. [2]Terwee C, Bot S, de Boer M, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34-42. doi: 10.1016/j.jclinepi.2006.03.012.

[3]Attributes and Criteria to assess Health Status and Quality of Life Instruments (1996, 2002)	
1. Conceptual and measurement model	The rationale for a description of the concepts and the populations that a measure is intended to assess and the expected relationship between these concepts
2. Reliability	The degree to which an instrument is free from random error
2.1. Internal consistency	The precision of a scale, homogeneity (inter correlations) of items at one point in time
2.2. Reproducibility	Stability of an instrument over time (test-retest) and inter-rater agreement
3. Validity	The degree to which the instrument measures what it purports to measure.
3.1. Content validity	The domain of an instrument is appropriate relative to its intended use
3.2. Construct-related validity	Interpretation of scores based on theoretical implications associated with the construct to be measured
3.3. Criterion-related validity	The extent to which scores of the instrument are related to a criterion measure (gold standard).
4. Responsiveness	The instrument's ability to detect change overtime
5. Interpretability	The degree to which one can assign easily understood meaning to an instrument's quantitative scores
6. Respondent and administrative burden	The time, effort, and other demands placed on those to whom the instrument is administered (respondent burden) or on those who administer the instrument (administrative burden)
7. Administration/Accessible forms	Data collection method, including self-report, interviewer-administered, trained observer rating, computer-assisted interviewer-administered, performance-based measures. Accommodations (e.g. Braille)
8. Cultural and language adaptations	Assessment of conceptual and linguistic equivalence.
[4]Health Status Measures in Economic Evaluation (1999, 2017)	
1. Practicality	Time to complete the instrument. Response rate. Rate of completion
2. Reliability	The degree to which an instrument is free from random error
2.1. Test-retest	Ability to reproduce results over repeated measurements with the minimum amount of random error
2.2. Inter-rater	Reliability between places of administration
3. Validity 3.1. Descriptive validity (Content, Face, Construct)	Dimensions covered. Items relevant for population. Ability of an instrument to reflect known or expected differences and changes in health to reflect preferences.
3.2. Valuation	Values used. Main assumptions of the model and how well the preferences of the patients and decision makers are likely to conform to these assumptions.
3.3. Empirical	Evidence regarding whether or not a measure could generate values which reflect people's preferences using revealed preferences; stated preferences or hypothetical preferences as criteria

[3] Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB, et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. Clin Ther. 1996;18(5):979–92. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments and review criteria. Qual Life Res. 2002;11(3):193-215. [4] Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. Health Technol Assess (Rockv). 1999;3(9). Brazier J, Ara R, Rowen D, Chevrou-Severac H. A Review of Generic Preference-Based Measures for Use in Cost-Effectiveness Models. Pharmacoeconomics. 2017;35(s1):21-31. doi:10.1007/s40273-017-0545-x..

[5] Guidance for Industry patient-reported outcomes measures (2006, 2009)	
1. Conceptual model	Conceptual framework.
2. Administration/Accessible forms	Data collection method, including self-report or interviewer, format and scoring. Adaptations for children and adolescents, patients cognitively impaired, or unable to communicate, culture and language subgroups
3. Respondent/Administrator Burden	Length, formatting, font size, instructions for items, privacy, time, need for physical support in responding.
4. Reliability	
4.1. Test retest	Stability of scores over time when no change has occurred in the concept of interest
4.2. Internal consistency	Whether the items in a domain are inter correlated, as evidenced by an internal consistency statistic
4.3. Inter interviewer reproducibility	Agreement between responses when the PRO is administered by two or more different interviewers
5. Validity	
5.1. Content validity	Whether items and response options are relevant and are comprehensive measures of the domain or concept
5.2. Construct validity (Hypotheses testing, including discriminant, convergent, known groups validity)	Ability to measure the concept. Whether relationships among items, domains, and concepts conform to what is predicted by the conceptual framework for the PRO instrument itself and its validation hypotheses
6. Criterion	Scores of a PRO instrument are related to a known gold standard. When the gold standard is not possible to be evaluated, criterion measure assesses sensitivity specificity, and predictive values
7. Responsiveness. Ability to detect change	Evidence that the instrument is equally sensitive to gains and losses in the measurement concept and to change at all points within the entire range expected for the clinical trial population
[6] Evaluating patient-based outcomes measures for use in clinical trials (1998) (Fitzpatrick's criteria)	
1. Reliability	The extent to which the instrument is free from random error and may be considered as the amount of a score that is a signal rather than noise
1.1. Internal consistency	The extent to which individual items in a questionnaire scale measure the same construct (homogeneity of items in the scale)
1.2. Reproducibility (test retest)	Whether and instrument yields the same results on repeated applications, when respondents have not changed on the domain being measured. Stability of the questionnaire over time
2. Validity	The extent to which it measures what it purports to measure
2.1. Criterion and Predictive validity	When a new measure correlates with other measures generally accepted as a more accurate variable. When the new measure correlates with future values of the criterion variable
2.2. Face and content validity	Face validity refers to what an item appears to measure based on tis manifest content. Content validity refers to how well a measurement battery covers important parts of the health components to be measured
2.3. Construct validity	A health status measure is intended to assess a postulated underlying construct.
2.3.1.Convergent validity	Correlations are expected to be strongest with the most related constructs
2.3.2.Discriminant validity	Correlations are expected to be weakest with most distally related constructs
2.3.3.Internal structure	A set of assumed relationships between underlying constructs
2.3.4 Validity for specific purposes	Measures need to be assessed for health status, personal preferences and utilities, and social values.
3. Responsiveness (sensitivity to change)	Ability to detect changes over time. Effect size, sensitivity and specificity of scores.
4. Precision	How precise are the distinctions between levels of health and illness (sensitivity). Format categories.
5. Interpretability	How meaningful are the scores from an instrument
6. Acceptability	Evidence of acceptability is associated with high response rates. Respondent burden.
7. Cultural applicability	Rigorous translation can by itself establish the appropriateness of an instrument
8. Feasibility	Impact of different patient-based outcome measures upon staff and researchers. Administrator burden.

[7]International Classification of Functioning (ICF) & International Classification of Functioning for Children and Youth (ICFCY)) (2019)		
1.	Content validity	Health and Health-related domains.

[8]Evaluating Measures of Patient Reported Outcomes (EMPRO) (2008)		
1.	Conceptual and Measurement model	The rationale for description of the concept and the populations that a measure is intended to assess and the relationship between these concepts
2.	Reliability	The degree to which an instrument is free from random error
2.1.	Internal consistency	The precision of a scale, based on the homogeneity of the scale´s items at one point in time
2.2.	Reproducibility	The stability of an instrument over time (test retest) and inter-rater agreement at one point in time
3.	Validity (including content, criterion, hypotheses testing and construct)	The degree to which the instrument measures what it purports to measure
4.	Responsiveness	The ability to detect change over time
5.	Interpretability	The degree to which one can assign meaning to an instrument´s quantitative scores
6.	Burden (Respondent/Administrator burden)	Time, effort and other demands placed on the administration of the instrument
7.	Administration mode	Data collection method. For each mode of administration, the information about validity, reliability, responsiveness, interpretability and burden should be assessed.
8.	Cultural and language adaptations	Methods to achieve linguistic equivalence are adequately described and appropriate. Differences from the original are adequately described and appropriate.

[9]Spinal Cord Injury Criteria (2008, 2016)		
1.	Content	Description. Items. Scale development. Internal structure or subscales
2.	Administration/Accessible forms	Data collection method. Items, time, training, burden of administering. Disability adaptation (e.g. Braille)
3.	Reliability (test retest, internal consistency)	Degree to which an instrument is consistent or free from random error
4.	Criterion oriented validity (concurrent, predictive, discriminant, and clinical validity)	Scale predicts other measures of the same construct. Gold standard and/or sensitivity and specificity. Scale distinguish between scores and/ or groups. Clinical utility, also called prescriptive and consequential validity
5.	Responsiveness, sensivity to change	Evidence of change in expected direction using methods such as standardized effect sizes
6.	Floor and ceiling effects	Floor and ceiling issues can determine whether change is detected or obscured by the measure
7.	Population application (Applicability in SCI groups, languages, norms)	Description of use in people with spinal cord injury (vs other people). Information of norms are available. Available in other languages

[5]Department of Health and Human Services. Guidance for Industry Patient-reported Outcome measures: Use in Medical Product Development to Support Labeling Claims: draft guidance. Health Qual Life Outcomes. 2006;20:1-20. doi:10.1186/1477-7525-4-79. Department of Health and Human Services. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.; 2009. doi:10.1111/j.1524-4733.2009.00609.x.[6]Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating Patient-Based Outcome Measures for Use in Clinical Trials. Vol 2.; 1998. doi:9812244.[7]World Health Organization. International Classification of Functioning (ICF). www.who.int/classifications/icf/en/. [8]Valderas JM, Ferrer M, Mendívil J, et al. Development of EMPRO: A tool for the standardized assessment of patient-reported outcome measures. Value Heal. 2008;11(4):700-708. doi:10.1111/j.1524-4733.2007.00309.x. [9]Johnston M V., Graves DE. Towards Guidelines for Evaluation of Measures: An Introduction With Application to Spinal Cord Injury. J Spinal Cord Med. 2016;31(1):13-26. doi:10.1080/10790268.2008.11753976. Spinal Cord. Spinal Cord Injury Rehabilitation Evidence. https://scireproject.com

[10] Criteria for Assessing the Tools of Disability Outcomes Research (2000) (Andresen´s Tool)		
1.	Conceptual model	Relevant domains are completely covered
2.	Norms, standard values	Published data (or public-domain data) are available for both general population and with disabilities
3.	Measurement model	Tool captures the detail and breadth of real differences among persons, includes floor/ceiling effects
4.	Instrument bias	In practical or statistical terms, individual questions (or scores) are biased for the population
5.	Respondent burden	Length and content are acceptable to the intended subjects
6.	Administrative burden	Ease to administer, score and interpret
7.	Reliability (test retest and internal consistency)	Instrument gives a consistent answer
8.	Validity (discriminant, convergent, structure)	The tool measures what it purports to measure. It distinguish among different levels of mobility
9.	Responsiveness	Instrument is sensitive to changes in interventions
10.	Administration/Accessible forms	Data collection method, as interviews, self-administration, computer surveys. Adaptations (e.g. Braille)
11.	Culture/language adaptations	Tested versions of the tool for subgroups (including ethnicity, gender, disability)
[11]CanChild Outcomes Measures (2004)		
1.	Focus. Purpose	Focus of measurement (using the International Classification of Functioning Framework, ICF). Rating attributes measured. List the primary purpose for which the scales have been designed (discriminative, predictive, evaluative, etc.). Describe population. Evaluation of the context
2.	Clinical utility	Clarity of instructions, format, time to complete the assessment, administration, scoring and interpretation. Specify whether formal training is required. Cost of the manual and score sheets.
3.	Scale construction	Item selection, weighting, level of measurement
4.	Standardization	Manual (published, specific procedures for administration, scoring) Norms.
5.	Reliability	
5.1.	Internal consistency	The degree of homogeneity of test items to the attribute being measured. Measured at one point in time
5.2.	Intra/Inter observer	Measures variation within an observer; measures variation between two or more observers
5.3.	Test retest	Measures variation in the test over a period of time
6.	Validity	
6.1.	Content	The instrument is comprehensive and fully represents the domain of the characteristics it claims to measure
6.2.	Construct	Measurements of the attribute conform to prior theoretical relationships among characteristics or individuals
6.3.	Criterion	Measurements obtained by the instrument agree with another more accurate instrument (gold standard)
6.4.	Responsiveness	Ability to detect minimal clinically important change over the time

[10]Andresen EM. Criteria for assessing the tools of disability outcomes research. Arch Phys Med Rehabil. 2000;81(12 SUPPL. 2):15-20. doi:10.1053/apmr.2000.20619. [11]Law M. Outcome Measures Rating Form Guidelines.; 2004. Available from: https://www.canchild.ca/system/tenon/assets/attachments/000/000/371/original/measguid.pdf

[12]Outcomes Measures in Rheumatology Clinical Trials (OMERACT) (2019)	
1. Truth	
1.1. Face validity (credibility)	Overall appropriateness of the method to be used for evaluation of the outcome, as assessed by the investigators and clinicians
1.2. Content validity (comprehensiveness)	Ability of the outcome measure to include or predict all those components of health status that are relevant to the intervention being assessed
1.3. Criterion validity (accuracy)	Ability of the outcome measure to reflect the best available estimate of the true clinical status of the patient. Comparison with the "gold standard"
1.4. Construct validity (convergent/divergent)	Ability of the outcome measure to match with the hypothesized expectations of the investigator when compared with other indirect assessments
2. Discrimination	
2.1. Sensitivity to change over time	Based on calculation of the standardized response mean (SRM) using repeated measures performed in a given population at 2 different time-points without therapeutic intervention
2.2. Discrimination capacity over treatment	Based on calculation of effect size (ES) in randomized controlled trials or SRM in open-label trials
2.3. Reliability (reproducibility)	Based on evaluation of intra- and interclass correlations
3. Feasibility	The measure's ease of use, cost-effectiveness, availability in different centres, and overall usefulness. Practicalities of using the instrument, as cost, burden, length, translations, equipment needs.
[13]Testing Standards (1999, 2014)	
1. Evidences of Validity	
1.1. Test Content	Themes, tasks, format of the items, wording, and processes of administration and scoring
1.2. Response Processes	Cognitive processes engaged in by test takers with consequences in the scores.
1.3. Internal Structure (Dimensionality, Differential item functioning)	The degree to which the relationships among test items and components conform to the construct on which the proposed test score interpretations are based including equivalence of scores among different populations.
1.4. Relations to other variables (Convergent, Discriminant, Criterion, nomological network including responsiveness)	The degree to which relationships with other variables are consistent with expectations derived from theory underlying the construct
1.5. Consequences of testing	Value judgement about unintended positive and negative consequences of test use
2. Reliability	*Revised Standard (2014) also includes Decision consistency/accuracy*
2.1. Internal consistency, Test- retest, Alternate forms, *Scorers Consistency, Decision consistency, Accuracy*	The degree to which an instrument is free from random error. The precision of a scale, homogeneity (inter correlations) of items. Replicability of the testing procedure.
3. Fairness	Characteristics of all individuals must be considered throughout all stages of development, administration, scoring, interpretation and use of test. *Revised Standards (2014) emphasize the role of the Fairness as a measurement property*
4. Scales, Norms and Score Comparability	Reference points should be documented based on population norms and/or expert criteria. Linking procedures devised to guarantee comparability of different measures of similar constructs should be described
5. Test development and revision	Tests and their supporting documents should be periodically reviewed. New forms such as those derived from translation to other languages should be thoroughly tested for equivalence

[12] OMERACT. Instrument selection for Core Outcome Measurement Sets. In: OMERACT Handbook [Internet]. 2019. Available from: https://omeracthandbook.org/handbook . [13]American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. American Educational Research Association.; 1999. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. American Educational Research Association; 2014.

10. ADDENDA

10.1. BENEFITS OF THE HUMIDIFIED LOW-FLOW OXYGEN THERAPY IN INFANTS WITH MILD-MODERATE BRONCHIOLITIS

BENEFITS OF THE HUMIDIFIED LOW-FLOW OXYGEN THERAPY IN INFANTS WITH MILD-MODERATE BRONCHIOLITIS

Sonia Lorente, Rebeca Gimeno, Josep-Maria Losilla,
Sandra Garzón, Jaume Vives

Sonia Lorente, Psychologist, Pediatric Nurse, MSc (sonia.lorentes@e-campus.uab.cat) Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB; Pediatric Department, Hospital de Terrassa, Consorci Sanitari de Terrassa; Rebeca Gimeno, Pediatric Nurse, MN (rebeca4tonda@hotmail.com), and Sandra Milena Garzón, Pediatric Nurse (sandradb61@hotmail.com), Pediatric Department, Hospital de Terrassa, Consorci Sanitari de Terrassa; Josep Maria Losilla, PhD, Statician (JosepMaria.Losilla@uab.cat) and Jaume Vives, PhD, Statician (jaume.vives@uab.cat), Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB.

This work was partially supported by the Spanish Ministry of Science and Innovation (grant number: PSI2014-52962-P). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no conflict of interest. Authors have no financial relationships relevant to this article to disclose.

We would like to sincerely thank Antonia Villalba, the Nursing Supervisor of Pediatric Department, for the financial support, Abel Martínez, the Medical Chief of Pediatric Department, for his guidance, as well as the nurses, the auxiliary staff and the doctors, for their collaboration in the collecting data process.

Correspondence concerning this article should be addressed to Sonia Lorente, Paediatric Department, Hospital de Terrassa, Consorci Sanitari de Terrassa; Department of Psychobiology and Methodology of Health Science, Universitat Autònoma de Barcelona, UAB (sonia.lorentes@e-campus.uab.cat) +34619287300

ABSTRACT

Background: There is no evidence to support the use of humidification in low-flow oxygen therapy as a usual clinical practice in the management of bronchiolitis**.** *Aims and objectives*: To investigate the clinical benefits of using humidification in low-flow oxygen therapy. Specific objectives were to investigate via an assessment of the number of nasal lavages whether humidification can help to decrease the nasal mucus viscosity, to determine whether it can relieve feeding difficulties by comparing the weight gain in infants, to ascertain whether it can relieve respiratory distress by assessing the heart and respiratory rates and contribute to improved clinical outcomes, measured by the length of stay and oxygen requirements. *Design*: A controlled quasi-experimental study. *Methods*: A total of 97 infants were included, aged ≤ 6 months, with bronchiolitis, low-flow oxygen therapy and bronchodilators nebulized with hypertonic saline 3%. Data from the control group (non-humidified) were gathered from 2010 - 2012 (49 infants), and data from the group with humidification from 2012 - 2014 (48 infants). Linear and Poisson regressions were performed adjusting for relevant characteristics of patients. *Results*: Humidification was shown to be associated with significant reductions in the number of nasal lavages in infants with Sant Joan de Déu Bronchiolitis Scores of BROSJOD≤7, in the heart rate of infants with mixed bronchodilators treatment, and in the length of stay and oxygen requirements of infants with Score BROSJOD≤5. No differences in weight and respiratory rate were found. *Conclusions*: Humidification in low-flow oxygen therapy is an effective nursing intervention to improve the clinical outcomes of infants with mild-moderate bronchiolitis. *Clinical implications:* Humidifying the nasal mucosa can help to reduce the need for professional procedures, oxygen requirements and hospitalization length. Further research into the economic savings involved is recommended. *Key words*: Bronchiolitis, Humidification, Low-Flow Oxygen Therapy, Infant, Nursing.

SUMMARY STATEMENT

'What does this paper contribute to the wider global clinical community?'

- There are no prior recommendations, nor any conclusive results related to the clinical benefits of humidification in low-flow oxygen therapy in infants.
- Our findings support the use of humidification in the management of mild to moderate bronchiolitis as a usual clinical practice to improve the clinical outcomes and comfort of infants.

INTRODUCTION

Bronchiolitis is the most common lower respiratory infection in infants younger than two years of age, and it is mainly caused by the respiratory syncytial virus, RSV. The normal epidemic season for bronchiolitis lasts from December to March, and the pathology includes acute inflammation, bronchiole edema, and increased mucus production. The signs and symptoms begin with rhinitis and cough, which may progress to tachypnea, wheezing, rales, use of accessory muscles and/or nasal flaring (American Academy of Pediatrics, 2014). The maintenance of hydration and oxygenation, as a supportive treatment, is generally recommended to manage acute bronchiolitis, since different trials of other treatments such as corticosteroids or antibiotics have not demonstrated conclusive benefits (American Academy of Pediatrics, 2014; Krilov, 2011).

Background

With regard to oxygenation management, noninvasive ventilation is commonly implemented during hospitalization, in the form of either high-flow (HF) oxygen therapy or low-flow (LF) oxygen therapy. When the HF system is implemented, the humidification of oxygen is a usual clinical practice, undertaken in order to prevent some complications of dry oxygen delivery, including nasal dryness and discomfort (Cuquemelle et al., 2012). However, despite the complications of the dry oxygen conditioning, there is neither a widespread consensus nor any standard recommendations about the use of humidified oxygen in LF rates.

On the one hand, some hospital guidelines always recommend the humidification of oxygen in infants, regardless of the flow rates, as the Saskatoon Health Region Nursing Practice Committee policies (2006). On the other hand, Nath, Ponnusany, Willis, Bisset, & Clarke (2010), reported a wide variety of clinical practices in the neonatal care with regard to humidification measures in LF therapy, thus complicating any effort to define standard policies, as stated above. Lastly, Umoren, Odey, & Meremikwu (2011) concluded that there have not been enough well-designed trials to examine the benefits of humidified oxygen in the treatment of acute bronchiolitis in children of up to three years of age. Because of this lack of evidence and consensus, both Nath et al. (2010) and Umoren, Odey, & Meremikwu (2011), recommend further research to explore the effectiveness of humidified vs. non-humidified oxygen therapy. In addition, the humidification may be helpful to prevent the thickening of secretions (Walsh & Smallwood, 2017) and relieving the main symptoms of respiratory distress. Finally, no adverse effects have been reported, either with steam humidification (Umoren, Odey, & Meremikwu, 2011), or by using prefilled bubble humidifiers (Yamashita, Nishiyama, Yokoyama, Abe, & Manabe, 2005), and only benefits may be expected.

Thus, given this call for an examination of the benefits of humidified oxygen in low-flow therapy, the humidification of oxygen in LF was implemented as a general practice in the management of all infants with bronchiolitis in our paediatric ward, and we compared its effects to non-humidified oxygen therapy. The overall objective was to investigate the clinical benefits of using humidification as a usual clinical practice in low-flow oxygen therapy in the management of infants with acute bronchiolitis. Our specific hypotheses were that the use of the humidification in low-flow oxygen therapy in infants with mild-moderate bronchiolitis would reduce their nasal mucus viscosity, thus alleviating the infants' feeding difficulties and easing their respiratory distress, and that it would improve their clinical outcomes.

METHODOLOGY

Design, setting and participants

A quasi-experimental study with a non-concurrent control group was conducted at our general pediatric ward, in a Spanish non-tertiary hospital. A total of 97 infants with mild-moderate bronchiolitis with low-flow oxygen therapy, defined as ≤ 2 L/min, administered via nasal cannula, were enrolled. The control (non-humidified) group data were gathered during the annual RSV incidence peak from November to March of the years 2010 to 2012 and consisted of observations of 49 infants. The study group (humidified) data were collected in the same seasonal period, 2012-2014, with this group consisting of 48 infants. The prevalence of RSV in the groups was between 70%-83%, similar to the reference population with bronchiolitis in Spain (Flores-González et al., 2017).

We included all infants younger than six months old, with diagnoses of bronchiolitis based on an initial episode of respiratory distress, with LF oxygen therapy, and bronchodilators (salbutamol and/or epinephrine), nebulized with hypertonic saline 3%. We excluded infants with previous respiratory tract infections and co-infections at admission (pneumonia, bordetella pertussis, tract urinary infection), and infants with risk factors for severe disease, such as chronic diseases (congenital heart disease, chronic lung disease, immunodeficiency, neuromuscular diseases) and premature birth (American Academy of Pediatrics, 2014; Krilov, 2011). Those infants who needed different management due to co-infections detected during the bronchiolitis treatment, as well as those who ultimately needed to be admitted in the pediatric intensive care unit (PICU), were also excluded (Figure 1).

Study variables

To evaluate the degree to which humidification can contribute to a decrease in nasal mucus viscosity, the number of nasal lavages needed to alleviate the nasal obstruction was recorded as an indirect indicator. To assess feeding difficulties, data were gathered on the infants' weight gain, with each infant´s weight in grams recorded at admission and at discharge, in each case with the baby naked and weighed on the same digital scale both times. In order to evaluate respiratory distress, measurements were taken of the infants' respiratory rate (RR) (counted by professionals during one minute) and of their heart rate (HR) and oxygen saturation (SpO2) (measured with pulse-oximetry) at admission and then after 24, 48, and 72 hours, and again at discharge. The distress score was also recorded at admission by using the Sant Joan de Déu Bronchiolitis Score (BROSJOD) (Balaguer et al., 2017). The bronchodilator therapy was recorded, because the use of epinephrine rather than salbutamol may be slightly more effective (Modaressi, Asadian, & Faghihinia, 2012). To assess the length of stay (LoS), we defined the period of days between the admission in our ward and the validation of the discharge criteria, generally when the oral fluid intake was adequate, the SpO2 was >95% without oxygen supply for at least twenty-four hours, and the infant exhibited minimal wheezes, rales, and chest retractions. To measure the length of oxygen requirement (LOR) we recorded the period of days necessary to decrease the work of breathing to the minimum with the supply of oxygen.

Potential confounders

Some potential confounders were controlled for this study; the infants' age, because the most severe bronchiolitis episodes are associated with infants aged less than 12 weeks (American Academy of Pediatrics, 2014); the infants' feeding (categories: breastfeeding, formula or mixed), since exclusive breastfeeding may have protective effects on infectious diseases, as bronchiolitis (American Academy of Pediatrics, 2014; Duijts, Jaddoe, Hofman, & Moll, 2010); the use of antibiotic agents (categories: presence or absence), given that despite the lack of evidence of their benefits in treating this viral infection (American Academy of Pediatrics, 2014), the administering of antibiotics might relieve some symptomatology; the findings in radiography and RSV diagnoses (García et al., 2010) (categories: presence or absence), since both variables may lead to worsened outcomes. Finally, pre-natal maternal smoking and the household smoking exposure were also recorded (categories: yes or no), because household tobacco exposure may increase clinical severity in bronchiolitis (Semple, Taylor-Robinson, Lane, & Smyth, 2011).

Procedure

The choices of management practices for infants with bronchiolitis at our hospital always depend on each infant's clinical status. These interventions range from the maintenance of oxygenation through the use of LF when the bronchiolitis episode is mild or moderate and the SpO2 <94%, to the maintenance of hydration by encouraging the oral fluid intake in infants diagnosed with a mild-moderate episode, and to the prescription of intravenous fluids or tube feeding, only in very exceptional and critical situations. They also include the administration of bronchodilators by nebulizing salbutamol and/or epinephrine with 3% hypertonic saline and antibiotic therapy. In addition, the nasal lavages are also recommended to alleviate the nasal obstruction due to mucus viscosity. This practice consists of the instillation of 2cc to 5cc of normal saline into the infant's nostrils, a process that is much less aggressive than nasal suctioning, although it still causes discomfort. Thus, to reduce the discomfort of infants associated to nasal lavages, we implemented the humidification in infants with bronchiolitis and LF oxygen therapy using the bubble sterile device Respiflo©, Tyco Healthcare (Tyco Healthcare Group, 2004a), which has been available in our pediatric ward since November 2012.

The general procedure during this study was to administer humidification to all infants with bronchiolitis who needed LF oxygen therapy, and at the time of admission to register the distress score as measured by the Sant Joan de Déu Bronchiolitis Score (BROSJOD) (Balaguer et al., 2017), as well as each infant's vital signs, weight, and potential confounding variables. The patients were monitored during their stay, with records taken of their vital signs, and weight at discharge. Infants with co-infections and complications were excluded from the study, as mentioned above (Figure 1). Because this practice was implemented as a usual prescription, a concurrent group without humidification was not possible.

Data collection

Data from the from the control, non-humidified group (2010-2012) were collected via a review of medical records, and data from the humidified group (2012-2014) were gathered using a specific form, designed for the purposes of this study. Demographic data from both groups were also recorded.

Instruments

Respiflo© cold bubble device

This is a closed system with a special adapter, suitable for the administration of humidified respiratory gases or/and nebulized therapy. The closed system and the adapters reduce the risk of contamination associated with other conventional open systems. This device

is suitable for the conditioning of dry respiratory gases in short-term low-flow oxygen therapy administered with nasal cannulas (Tyco Healthcare Group, 2004; Yamashita et al., 2005).

Sant Joan de Déu Bronchiolitis Score (BROSJOD)

The Sant Joan de Déu Bronchiolitis Score (BROSJOD score) was originally developed in 1999 (Caritg et al. 1999, cited in Balaguer et al. 2017), and it was recently validated as a diagnostic severity tool in infants aged less than two years, mean age 52.5 days, admitted to Hospital Sant Joan de Déu with acute bronchiolitis (Emergency Department, ward, or PICU) (Balaguer et al. 2017). At admission, the BROSJOD showed good internal consistency (Cronbach´s alpha 0.75 [95% CI 0.71- 0.82]), high inter-rater reliability (Concordance Correlation Coefficient, CCC 0.96 [95% CI 0.94-0.97]), and good discriminative validity (Volume Under Surface, VUS 0.80 [95% CI 0.70-0.90]). This score was specifically developed to address the severity of bronchiolitis in children, while other measures as the Wood-Downes (WD) scale were adapted from asthma scores and modified for younger children. In addition, the BROSJOD score considers RR and HR classified according to the age of infants, and it measures data not only on cyanosis but also on oxygen saturation, using pulse-oximetry. Severity score values ranged from 1 to 16. Higher scores indicate a more serious condition, with cases classified as mild crises (values 0 to 5), moderate crises (values 6 to 10) or severe crises (values 11 to 16) (Balaguer et al. 2017) (Table 1).

Ethical aspects

This study was conducted according to the principles expressed in the Declaration of Helsinki of 1975, revised in 1983, and was approved by the Hospital de Terrassa Ethics Committee (CEIC, Comité de Ética e Investigación Clínica). The parents' oral consent was obtained, and the data were gathered anonymously. Non-adverse effects were reported by professionals and parents.

Analysis

Linear and Poisson regressions were performed to investigate the impact of the use of humidification in low-flow oxygen therapy in infants with mild-moderate bronchiolitis, adjusting for the patients' relevant characteristics. Furthermore, we tested the interaction of humidification with the BROSJOD score to assess whether the impact of humidification on the number of nasal lavages, the length of stay (LoS) and the length of oxygen requirement (LOR) was the same at different severity levels of bronchiolitis. We also tested the interaction of humidification with the type of bronchodilator treatment used in order to assess the impact of humidification on heart rate (HR) and respiratory rate (RR). Regression backward model selection was conducted, fitted using IBM SPSS Statistics v19.0 package (IBM Corp, 2011).

The results of the association between treatment (humidification/no humidification) and each response investigated are presented as non-linear Poisson regression coefficients (Exp(B)) for count responses, and as linear regression coefficients (B) for quantitative responses, reporting 95% confidence intervals (95% CI), and P-values (P) in both cases.

RESULTS

Demographic characteristics and descriptive statistics of study variables

Table 2 shows the infants' clinical and demographic characteristics. Our groups were comparable in age, admission weight, BROSJOD scores, RSV diagnoses, antibiotic therapy, radiography findings and family smoking habits. However, the groups showed differences in terms of feeding and bronchodilator treatment, so these variables were adjusted in the statistical analysis, in regression models. Concerning the group characteristics, 71% of infants in the control group (non-humidified) and 83% in the humidified group were diagnosed as RSV positive, while most subjects had mild to moderate bronchiolitis scores, were of less than ninety days of age and had been exclusively breastfed. Table 3 shows the descriptive statistics and the results of the bivariate comparisons for each study variable.

Effects of humidification

In order to analyze the effect of humidification on the number of nasal lavages, a Poisson regression model was tested. This model included the interaction of humidification with the BROSJOD score adjusting for breastfeeding and age. Table 4 shows the results of the final model, where only the interaction of humidification with the BROSJOD score was kept. The humidified group showed a significant reduction in the number of nasal lavages. As the score decreases, the number of nasal lavages is also reduced, with some cases showing a reduction by as much as 78.5% (Score BROSJOD= 3).

Concerning feeding difficulties, the infants' weight gain during the stay was analyzed using a linear regression model adjusted for age, but no significant differences were found (B=-2.456 grams [95% CI: -64.5 – 59.6], p= 0.938).

To analyze the effect of humidification on respiratory distress, linear regression models were tested, with the HR and RR as response variables. This model included the interaction of humidification with the BROSJOD score and adjusted for bronchodilators and age. Table 5 shows the results of the final model, where only the interaction of humidification with bronchodilators remains. Infants with humidification and mixed bronchodilators treatment (salbutamol and epinephrine) showed a reduction in HR (B=-11.49 beats [95% CI: -19.68 – -3.30], p= 0.06). No significant differences in respiratory rates were observed.

To analyze the effect of humidification on clinical outcomes, Poisson regression models were applied, with LoS and LOR as response variables. These models also included the interaction of humidification with the BROSJOD score and were adjusted for admission weight, breastfeeding and age. Table 6 shows the final model, where only the interaction of humidification with the BROSJOD score was kept. As this score decreases, there is a corresponding reduction in the LoS and LOR, with the values for the latter two variables decreasing by as much as 58.2% and 69.5%, respectively (BROSJOD score= 3).

DISCUSSION

Bronchiolitis is an active area of research. Ninety percent of infants are infected with RSV in the first two years of life, and the treatment of this virus comes with a huge health care cost (American Academy of Pediatrics, 2014). For this reason, numerous reviews are constantly being conducted in order to improve the management of these infants (Øymar, Skjerven, & Mikalsen, 2014; Schroeder & Mansbach, 2014; Zentz, 2011). In this sense, the use of humidification in HF rates is almost an unquestionable clinical practice that helps enhance infants' comfort levels (Cuquemelle et al., 2012) and reduce the risk of complications (Dysart, Miller, Wolfson, & Shaffer, 2009). However, as mentioned previously, there are discrepancies in the research on the use of humidification in LF, because to date there is no conclusive evidence of its benefits (Nath et al., 2010; Umoren et al., 2011). To our knowledge, this is the first study reporting clinical benefits of the use of humidification in LF oxygen therapy in mild-moderate bronchiolitis.

Number of nasal lavages

Humidification seems to be helpful in reducing the number of nasal lavages needed to alleviate the nasal obstruction. The data confirm our hypothesis, since the humidification in LF may lighten secretions, as Umoren et al. (2011) suggested. In addition, reducing the number of nasal lavages may minimize the impact of nursing procedures. This "minimal handling" approach may be especially beneficial in infants aged less than three months, as Øymar et al. (2014) suggested. As expected, the effect of the humidification in LF decreased the number of nasal lavages in infants with lower BROSJOD scores, since infants scoring higher may need other supportive measures, such as heated humidification.

Feeding difficulties

The difference in weight gain was not statistically significant. These results, however, could have been significantly different if we had measured and analyzed their weight daily to examine the fluctuations. In addition, we could have examined how well the infant was feeding (i.e., how long it takes to complete a feed, and if the infant is breastfeeding, enquire as to whether the mother feels that her breast is being emptied), as Fitzgerald & Kilham (2004) recommended as ways to assess the clinical severity of bronchiolitis episodes.

Respiratory distress

Humidification decreased the mean HR significantly, and no statistical differences in the RR were found. These findings partially confirm our hypothesis, since decreasing the thickness of secretions may relieve respiratory distress, as was pointed out by Umoren et al. (2011). HR is considered a distress indicator in some instruments assessing bronchiolitis severity, including the BROSJOD score and other generally accepted measures like the Wood-Downes Clinical Scoring System, modified by Ferrés (Flores-González et al., 2016). When these instruments are used, lower HR generally contributes to lower scores and thus to decreased severity and improvements in infants' clinical evolution. The RR slightly decreased, but results were statistically not significant. This fact may be due to the measuring system used, because whereas the HR was measured with a pulse oxymeter, which brings valid and reliable data, RR was measured by professionals, who gathered this data by counting. In this sense, the recording of RR may vary from one professional to another, and so the analysis may be altered. So, no conclusive results were found concerning RR, but our results regarding the redution of HR were statistically significant and may be helpful for future research.

Length of stay and length of oxygen requirement

Humidification significantly reduced the LoS and the LOR. These results confirm our hypothesis, since lightening secretions and relieving respiratory distress seem to improve infants' clinical outcomes. This effect was associated with lower BROSJOD scores, as well as with reduction in the number of nasal lavages required, both these results likely owing to the same explanation. Apart from the relationship of humidification with the BROSJOD score, no other associations were found, although we initially expected an association between length of stay and breastfeeding.

The lack of association between these two variables could be explained by the fact that most of the infants in our groups were of less than ninety days of age, whereas the contribution of exclusive breastfeeding to improve outcomes in respiratory tract infections, including reductions in the length of stay, has been described for infants who have been exclusively breastfed for at least four months (American Academy of Pediatrics, 2014; Duijts et al., 2010).

Limitations

Some limitations should be noted and taken into account for further research. Our design did not include a concurrent control group without humidification, because this system was implemented as a general practice at our ward. While this non-experimental design does not ensure the comparability between groups, differences detected between groups regarding potential confounders were controlled for using regression models, as mentioned in the results section. We did not have a control group without bronchodilator treatment because our clinicians administer bronchodilators as a general rule, as detailed in the procedure section. We recorded the clinical distress score at admission to determine the severity of the bronchiolitis episode, but we did not register the score evolution during the stay. We did not quantitatively register the oral fluid intake of infants during the study, so we could not control for their hydration status as would have been desirable. Finally, we could not analyze the interaction of humidification with BROSJOD scores over 8, due to the insufficient number of infants in the humidified group.

Implications for clinical practice and further research

Humidification in LF may be an effective nursing intervention to improve the management of infants with mild-moderate bronchiolitis, reducing the discomfort associated with the dry oxygen delivery and nursing handling, relieving respiratory distress and improving clinical outcomes. Thus, humidification in low-flow oxygen therapy should be recommended as a regular clinical practice, since this viral infection is widespread and highly prevalent.

In addition, our results may be generalizable to other clinical scenarios in Spain where the management of infants with bronchiolitis is very similar, such as in bronchodilators nebulized with 3% hypertonic saline and when antibiotics are also prescribed (Flores-González et al., 2016, 2017). In other countries, clinicians also prescribe antibiotics, bronchodilators and corticosteroids, with examples found in Italy (De Brasi et al., 2010) or Finland (Elenius et al., 2017), despite the fact that administering these treatments has not been shown to have conclusive benefits (American Academy of Pediatrics, 2014). Along the same lines, and although nasal suctioning is not routinely recommended by the American Academy of Pediatrics guidelines (2014), other protocols and guidelines have highlighted the relevance of

suctioning nasal secretions before any inhaled therapy application or prior to feeding (Black & Brennan, 2011; Cincinnati Children's Hospital Medical Center, 2006). In these cases, we usually do nasal lavages, which are much less aggressive than nasal suctioning.

Moreover, some other issues concerning the economic costs should also be highlighted. Regarding the humidification system implementation, the bubble device in our public hospital costs less than one euro per infant/week, whereas a one-day stay costs nearly 300 euros. Although the costs may vary depending on the device brand used and the clinical scenario, the humidification system still remains economical (Villanueva & Bayón, 2013). Additionally, other cost savings should be taken into account if the humidification in LF is implemented, including the cost-time of the nursing procedures and the medical costs resulting from the possible complications of dry oxygen delivery. In this sense, further research is needed, because in this study we did undertake an analysis of the cost-effectiveness of humidification in LF. We also recommend conducting prospective studies with a control group to clarify the role of bronchodilators in association with the humidification, registering the exact dose-frequency of epinephrine and salbutamol. Finally, to follow infants' clinical evolution more accurately, we strongly recommend registering distress scores daily, because this measure seems to be a good predictor of clinical outcomes, as highlighted by Balaguer et al. (2017).

CONCLUSIONS

The use of humidification in LF oxygen therapy seems to be clinically beneficial. Humidifying the nasal and oral mucosa can help to lighten secretions, reduce the number of nasal lavages and minimize infants' discomfort associated with mucosa dryness, nasal obstruction and professional procedures. Humidification also relieves respiratory distress and reduces the LoS and the LOR. Our findings support the use of humidification in LF as a regular clinical practice in the management of mild-moderate bronchiolitis.

REFERENCES

American Academy of Pediatrics. (2014). Clinical Practice Guideline: The Diagnosis, Management, and Prevention of Bronchiolitis 2014. *Pediatrics*, *134*, e1474–e1502. http://doi.org/10.1542/peds.2014-2742

Balaguer, M., Alejandre, C., Vila, D., Esteban, E., Carrasco, J. L., Cambra, F. J., & Jordan, I. (2017). Bronchiolitis Score of Sant Joan de Déu: BROSJOD Score, validation and usefulness. *Pediatric Pulmonology*, *52*, 533–539. http://doi.org/10.1002/ppul.23546

Black, A., & Brennan, R. A. (2011). Breathing Easy : Implementing A Bronchiolitis Protocol. *Pediatric Nursing*, *37*(3), 129–135.

Cincinnati Children's Hospital Medical Center. (2006). Evidence based clinical practice guideline for medical management of bronchiolitis in infants less than 1 year of age presenting with a first time episode. *Cincinnati Children's Hospital Medical Center*, 1–13.

Cuquemelle, E., Pham, T., Papon, J.-F., Louis, B., Danin, P.-E., & Brochard, L. (2012). Heated and Humidified High-Flow Oxygen Therapy Reduces Discomfort During Hypoxemic Respiratory Failure. *Respiratory Care*, *57*(10), 1571–1577. http://doi.org/10.4187/respcare.01681

De Brasi, D., Pannuti, F., Antonelli, F., de Seta, F., Siani, P., & de Seta, L. (2010). Therapeutic approach to bronchiolitis: why pediatricians continue to overprescribe drugs? *Italian Journal of Pediatrics*, *36*, 1–8. http://doi.org/10.1186/1824-7288-36-67

Duijts, L., Jaddoe, V. W. V, Hofman, A., & Moll, H. A. (2010). Prolonged and exclusive breastfeeding reduces the risk of infectious diseases in infancy. *Pediatrics*, *126*(1), e18-25. http://doi.org/10.1542/peds.2008-3256

Dysart, K., Miller, T. L., Wolfson, M. R., & Shaffer, T. H. (2009). Research in high flow therapy: Mechanisms of action. *Respiratory Medicine*, *103*(10), 1400–1405. http://doi.org/10.1016/j.rmed.2009.04.007

Elenius, V., Bergroth, E., Koponen, P., Remes, S., Piedra, P. A., Espinola, J. A., … Jartti, T. (2017). Marked variability observed in inpatient management of bronchiolitis in three Finnish hospitals. *Acta Paediatrica*, 1–7. http://doi.org/10.1111/apa.13931

Fitzgerald, D. A., & Kilham, H. A. (2004). Bronchiolitis: Assessment and evidence-based management. *Medical Journal of Australia*, *180*(8), 399–404.

Flores-González, J. C., Mayordomo-Colunga, J., Jordan, I., Miras-Veiga, A., Montero-Valladares, C., Olmedilla-Jodar, M., … Goñi-Orayen, C. (2017). Prospective Multicentre

Study on the Epidemiology and Current Therapeutic Management of Severe Bronchiolitis in Spain. *Biomed Research International*, *2017*. http://doi.org/10.1155/2017/2565397

Flores-González, J. C., Rodríguez-Campoy, P., Pérez-Guerrero, J., Serrano-Moyano, B., Palma-Zambrana, E., Comino-Vázquez, P., … Flores González, J. C. (2016). Effect of Nebulized 3% Hypertonic Saline on Intensive Care Unit Admission Rates of Infants with Moderate Acute Bronchiolitis. *Journal of Clinical Research & Bioethics*, *7*(4), 3–7. http://doi.org/10.4172/2155-9627.1000281

García, C. G., Bhore, R., Soriano-Fallas, A., Trost, M., Chason, R., Ramilo, O., & Mejias, A. (2010). Risk factors in children hospitalized with RSV bronchiolitis versus non-RSV bronchiolitis. *Pediatrics*, *126*(6), e1453-60. http://doi.org/10.1542/peds.2010-0507

Krilov, L. R. (2011). Respiratory syncytial virus disease: update on treatment and prevention. *Expert Review of Anti-Infective Therapy*, *9*(1), 27–32. http://doi.org/10.1586/eri.10.140

Modaressi, M., Asadian, A., & Faghihinia, J. (2012). Comparison of Epinephrine to Salbutamol in Acute Bronchiolitis. *Iranian Journal of Pediatrics*, *22*(2), 241–244.

Nath, P., Ponnusany, V., Willis, K., Bisset, L., & Clarke, P. (2010). Current practices of high and low flow oxygen therapy and humidification in UK neonatal units. *Pediatrics International*, *52*(6), 893–4. http://doi.org/10.1111/j.1442-200X.2010.03274.x

Øymar, K., Skjerven, H. O., & Mikalsen, I. B. (2014). Acute bronchiolitis in infants , a review. *Scandinavian Journal of Trauma, Ressucitation and Emergency Medicine*, *22*(1), 23. http://doi.org/10.1186/1757-7241-22-23

Saskatoon Health Region Nursing Practice Committee. (2006). *Oxygen Administration Policies & Procedures*. Retrieved from https://www.saskatoonhealthregion.ca/about/NursingManual/1115.pdf

Schroeder, A. R., & Mansbach, J. M. (2014). Recent evidence on the management of bronchiolitis. *Current Opinion in Pediatrics*, *26*(3), 328–333. http://doi.org/10.1097/MOP.0000000000000090

Semple, M. G., Taylor-Robinson, D. C., Lane, S., & Smyth, R. L. (2011). Household tobacco smoke and admission weight predict severe bronchiolitis in infants independent of deprivation: prospective cohort study. *PloS One*, *6*(7), e22425. http://doi.org/10.1371/journal.pone.0022425

Tyco Healthcare Group. (2004). *Humidification Nebulisation Oxygen Therapy. What is Respiflo?* Retrieved from http://www.gulcanlarmedikal.com.tr/pdf/coviden/Oxygen Therapy.pdf

Umoren, R., Odey, F., & Meremikwu, M. M. (2011). Steam inhalation or humidified oxygen for acute bronchiolitis in children up to three years of age (Review). *Cochrane Database of Systematic Reviews*, (1), 1–33. http://doi.org/10.1002/14651858.CD006435.pub2.

Villanueva, G., & Bayón, J. (2013). Evaluation of bubble humidification systems in low-flow oxygen therapy. Systematic review and cost analysis. *Ministerio de Sanidad, Servicios Sociales E Igualdad. Servicio de Evaluación de Tecnologías Sanitarias Del País Vasco. Informes de Evaluación de Tecnologías Sanitarias. Osteba* (Abstract in English)

Walsh, B. K., & Smallwood, C. D. (2017). Pediatric Oxygen Therapy: A Review and Update. *Respiratory Care*, *62*(6), 645–661. http://doi.org/10.4187/respcare.05245

Yamashita, K., Nishiyama, T., Yokoyama, T., Abe, H., & Manabe, M. (2005). A comparison of the rate of bacterial contamination for prefilled disposable and reusable oxygen humidifiers. *Journal of Critical Care*, *20*(2), 172–175. http://doi.org/10.1016/j.jcrc.2005.01.002

Zentz, S. E. (2011). Care of infants and children with bronchiolitis: a systematic review. *Journal of Pediatric Nursing*, *26*(6), 519–29. http://doi.org/10.1016/j.pedn.2010.07.008

Figure 1. Flow diagram of selection of study patients

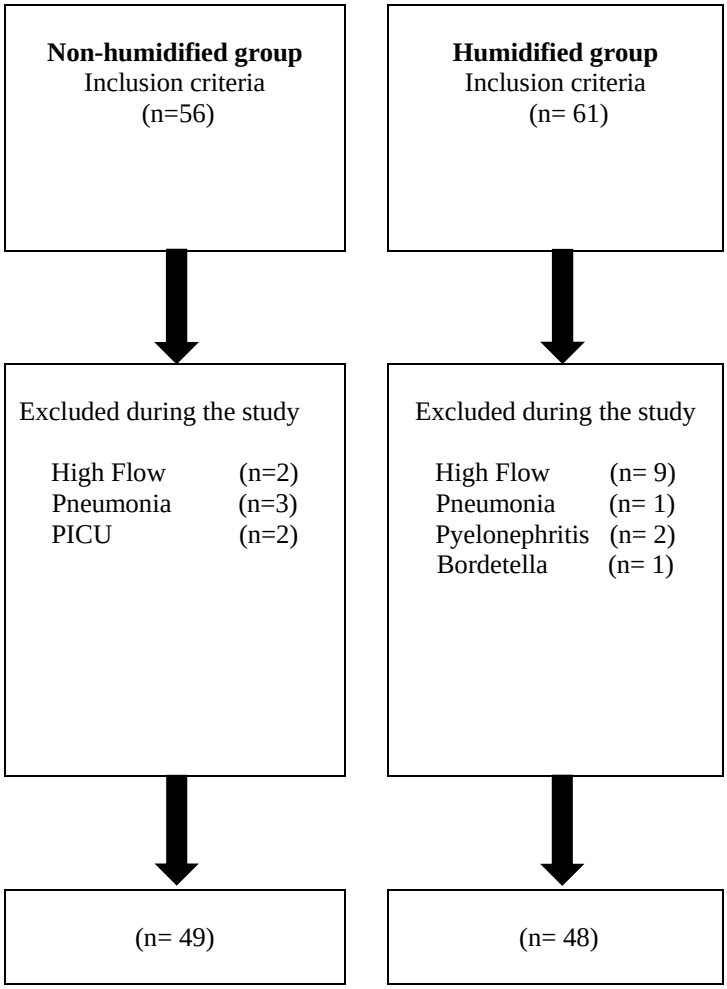


Table 1. Bronchiolitis Score of San Joan de Déu, BROSJOD (Balaguer et al., 2017)

Wheezing/rales	0	No
	1	Expiratory wheezes /Inspiratory rales
	2	Expiratory and Inspiratory wheezes/rales
Indrawing	0	No
	1	Subcostal + lower intercostal
	2	Previous + supraclavicular + nasal flaring
	3	Previous + upper intercostal + tracheal tug
Air entry	0	Normal
	1	Regular and symmetric
	2	Asymmetric
	3	Very reduced

Oxygen saturation		
	With oxygen	Without oxygen
0	≥ 95%	≥ 95%
1	91-94%	> 94% with FIO_2 ≤40%
2	≤90%	≤ 94% with FIO_2 ≤40%

Respiratory Rate (rpm)				
	0	1	2	3
< 3 months	< 40x’	40-59x’	60-70x’	>70x’
3-12 months	< 30x’	30-49x’	50-60x’	>60x’
12-24 months	< 30x’	30-39x’	40-50x’	>50x’

Heart Rate (bpm)				
	0	1	2	3
< 1 year	< 130/x’	130-149/x’	150-170/x’	>170/x’
1-2 years	< 110/x’	110-120/x’	120-140/x’	>140/x’

Note: Fraction of inspired oxygen (FIO_2); respirations per minute (rpm); beats per minute (bpm); Overall Score ranked into Minor crisis (0-5), Moderate crisis (6-10) and Severe crisis (11-16)

Table 2. Demographic and clinical characteristics of infants

	Non-humidified	Humidified	
	% (n)	% (n)	P
Age			0.768
<90 days	78 (38)	75 (36)	
>90 days	22 (11)	25 (12)	
Feeding			0.046
Breastfeeding	59 (29)	73 (35)	
Formula	16 (8)	21 (10)	
Mixed	26 (12)	6 (3)	
Admission weight (grams)			0.641
2500-3000	2 (1)	0 (0)	
3001-3500	6 (3)	10 (5)	
3500-4000	14 (7)	17 (8)	
>4001	78 (38)	73 (35)	
Smoking pregnant			0.090
yes	23 (10)	9 (4)	
no	77 (34)	91 (39)	
Smoking household			0.458
yes	31 (11)	40 (17)	
no	69 (24)	60 (26)	

Note: Pearson Chi-squared test to compare groups. Sample (n); p-values (P). Data are missing on smoking pregnant (n=10, 10.3%) and smoking household (n=19, 19.6%)

Table 3. Descriptive statistics of study variables

		No Humidified		Humidified	
	n	M ± Sd	n	M ± Sd	P
Weight gain (grams)					
	49	0.20 ± 154.9	48	13.5 ± 139.6	.657
HR (bpm)					
HR 24h	49	147.8 ± 16.7	48	145.1 ± 14.7	.406
HR 48h	49	145.5 ± 17.3	48	139.2 ± 15.1	.063
HR 72h	47	140.9 ± 15.6	47	138.7 ± 13.3	.465
RR (rpm)					
RR 24h	49	44.0 ± 6.6	48	44.1 ± 8.5	.852
RR 48h	49	44.2 ± 7.5	48	43.0 ± 8.7	.405
RR 72h	47	41.1 ± 6.5	47	42.5 ± 7.2	.336
	n	Md (IQR)	n	Md (IQR)	
Nasal lavages (count)					
	49	3 (2 – 9)	48	3 (1 – 5)	.467
Length of Stay and length of oxygen requirement					
LoS	49	5 (4 – 7)	48	5 (4 – 6)	.467
LOR	49	3 (2 – 5)	48	3 (2 – 4)	.174

Note: Sample (n), mean (M), Standard deviation (Sd), median (Md), Interquartile Range (IQR), P-value (P), Heart rate (HR), respiratory rate (RR), beats per minute (bpm), respirations per minute (rpm).

Table 4. Effects of humidification on number of nasal lavages due to lightening secretions across different bronchiolitis severity Score BROSJOD levels

	ExpB (95% CI)	(ExpB - 1) * 100	P
Humidification (Yes)			
Score 3 (n=4)	.341 (.215 – .539)	-65.9% (-78.5% − -46.1%)	< .001
Score 4 (n=11)	.406 (.286 – .577)	-59.4% (-71.4% − -42.3%)	< .001
Score 5 (n=18)	.485 (.375 – .627)	-55.1% (-62.5% − -37.3%)	< .001
Score 6 (n=24)	.579 (.473 – .708)	-42.1% (-52.7% − -29.2%)	< .001
Score 7 (n=19)	.691 (.557 – .855)	-30.9% (-44.3% − -14.5%)	< .001
Score 8 (n=13)	.824 (.619 – 1.09)	-17.6% (-38.1% − 9.8%)	.186

Note: Poisson regression coefficients (Exp. (B)), 95% Confidence Intervals (95% CI), p-values (P), sample (n). (ExpB - 1) * 100 shows the effect of humidification on the number of nasal lavages expressed in percentage of change. The initial model included the interaction of humidification with Score BROSJOD, and age and breastfeeding as potentials confounders, but only the interaction with Score BROSJOD was kept in the final model.

Table 5. Effects of humidification on heart rate and respiratory rate in the different bronchodilator treatments.

	Heart Rate (HR)		Respiratory Rate (RR)	
	B (95% CI)	P	B (95% CI)	P
Humidification (Yes)				
Salbutamol (n=48)	-.649 (-6.30 – 5.00)	.822	-7.25 (-14.76 – .26)	.058
Epinephrine (n=28)	.509 (-7.67 – 8.69)	.903	-6.50 (-13.67 – .67)	.076
Mixed (n=21)	-11.49 (-19.68 – -3.30)	.006	-.028 (-1.98 – 1.92)	.978

Note: Linear regression coefficients (B), 95% mean confidence intervals (95% CI), and p values (P), HR in beats per minute (bpm), RR in respirations per minute(rpm). The initial model included the interaction of humidification with Score BROSJOD, and bronchodilators treatment and age as potential confounders, but only the interaction with bronchodilators was kept in the final model.

Table 6. Effects of humidification on length of stay and length of oxygen requirement across different bronchiolitis severity Score BROSJOD levels

	Length of Stay (LoS)			Length of oxygen requirement (LOR)		
	ExpB (95% CI)	(ExpB - 1) * 100	P	ExpB (95% CI)	(ExpB - 1) * 100	P
Humidification						
Score 3 (n=4)	.613 (.418 – .899)	-38.7% (-58.2% − -10.1%)	.012	.502 (.305 – .827)	-49.8% (-69.5% − -10.1%)	.007
Score 4 (n=11)	.687 (.512 – .920)	-31.3% (-48.8% − -8.0%)	.012	.610 (.415 – .895)	-39.0% (-58.5% − -10.5%)	.011
Score 5 (n=18)	.770 (.620 – .955)	-23.0% (-38.0%− -4.5%)	.017	.740 (.558 – .983)	-26.0% (-44.2% − -1.7%)	.038
Score 6 (n=24)	.863 (.726 – 1.02)	-13.7% (-27.4% − 2.0%)	.092	.899 (.721 – 1.121)	-10.1% (-27.9% − 12.1%)	.345
Score 7 (n=19)	.967 (.803 – 1.16)	-3.3% (-19.7% − 16.0%)	.723	1.092 (.868 – 1.374)	9.2% (13.2% − 37.4%)	.453
Score 8 (n=13)	1.08 (.845 – 1.39)	8.0% (15.5% − 39.0%)	.527	1.326 (.979 – 1.796)	32.6% (-2.4% − 79.6%)	.069

Note: Poisson regression coefficients (Exp(B)), 95% Confidence Intervals (95% CI), and p values (P). LoS and LOR in days. (ExpB - 1) * 100 shows the effect of humidification on length of stay and length of oxygen requirement expressed in percentage of change. The initial model included the interaction of humidification with Score BROSJOD, and admission weight, breastfeeding and age as potential confounders, but only the interaction with Score was kept in the final model.