



UNIVERSIDAD DE CHILE

FACULTAD DE ODONTOLOGÍA

DEPARTAMENTO DE ODONTOLOGÍA CONSERVADORA

ÁREA DE PERIODONCIA

**“REVISIÓN SISTEMÁTICA DE PROTOCOLOS Y PROPUESTA DE
VALIDACIÓN DEL EXAMEN DENTAL PREVENTIVO DEL ADULTO MAYOR”**

Pablo Esteban Donoso Schoebitz

TRABAJO DE INVESTIGACIÓN

REQUISITO PARA OPTAR AL TÍTULO DE

CIRUJANO-DENTISTA

TUTOR PRINCIPAL

Franco Cavalla Ruiz

TUTORES ASOCIADOS

Alicia Morales Chvets

Ximena Moreno Ruiz

Adscrito a Proyecto FONDEF ID18110034

Santiago – Chile

2021



UNIVERSIDAD DE CHILE

FACULTAD DE ODONTOLOGÍA

DEPARTAMENTO DE ODONTOLOGÍA CONSERVADORA

ÁREA DE PERIODONCIA

**“REVISIÓN SISTEMÁTICA DE PROTOCOLOS Y PROPUESTA DE
VALIDACIÓN DEL EXAMEN DENTAL PREVENTIVO DEL ADULTO MAYOR”**

Pablo Esteban Donoso Schoebitz

TRABAJO DE INVESTIGACIÓN

REQUISITO PARA OPTAR AL TÍTULO DE

CIRUJANO-DENTISTA

TUTOR PRINCIPAL

Franco Cavalla Ruiz

TUTORES ASOCIADOS

Alicia Morales Chvets

Ximena Moreno Ruiz

Adscrito a Proyecto FONDEF ID18110034

Santiago – Chile

2021

DEDICATORIA

A mi madre por el apoyo constante e incondicional, por los años que dedicó para que yo pudiera aprender, por la sonrisa constante a pesar de los problemas que la rodeaban.

A mi padre que me apoyaba monetariamente para yo que solamente me dedique a estudiar.

A mis amigos que hicieron de estos años los mejores que he vivido y me han hecho redefinir la palabra amigo y familia.

A Katherinna por todos los momentos que pasamos juntos, la ayuda y compañía que me dabas cuando más lo necesitaba.

A mis profesores que me orientaron durante estos años y entregaron lo mejor de ellos para que hoy pueda ser quien soy.

AGRADECIMIENTOS

Al Dr. Cavalla, por la paciencia y tiempo dedicado para que todo se pueda realizar de la forma más exitosa posible.

A la Dra. Moreno por guiarme y orientarme durante todo este proceso, por su tiempo y disposición a que yo aprenda.

A la Dra. Morales por la oportunidad de ingresar a este proyecto y por todos sus años que fue mi profesora de pregrado.

A los encargados del FONDEF por la oportunidad de participar en este proyecto y confiar en mis capacidades.

Índice

Introducción	1
Pregunta de investigación	6
Objetivo general.	6
Objetivos específicos.	6
Materiales y Métodos	6
Fuentes de información.....	7
Algoritmo de búsqueda	7
Criterios de inclusión y exclusión	8
Proceso de selección de artículos	8
Evaluación de la calidad de los artículos	9
Síntesis de datos.....	10
Resultados	10
Búsqueda y selección de artículos.....	10
Calidad de los artículos	11
Descripción de los estudios incluidos	11
Comité de ética	12
Validez aparente.....	12
Validez de contenido.....	14
Validez de criterio	16
Validez de constructo.....	17
Consistencia interna	18
Fiabilidad intra observador	18
Fiabilidad inter observador	20
Programas estadísticos	21

Discusión	22
Primera Fase: Elaboración	23
Segunda fase: Validación por expertos	24
Fase 3: Prueba piloto	25
Fase 4: Validez del instrumento	25
Fase 5: Fiabilidad del instrumento	26
Limitaciones de la revisión.....	29
Conclusión	30
Financiamiento	31
Bibliografía	31
Anexos.....	39
Anexo 1: Lista de verificación PRISMA.....	39
Anexo 2: Descripción de los términos MeSH utilizados	42
Anexo 3: Lista de verificación de calidad de los artículos	43

Resumen

Introducción

El examen dental preventivo del adulto mayor (EDePAM) es un examen oral integral con categorización de riesgo del adulto mayor. Este examen debe ser sometido a un proceso de validación para su implementación en la atención de salud.

Objetivos

Determinar por medio de una revisión sistemática de la literatura los métodos más eficientes para la validación del EDePAM como instrumento diagnóstico y de determinación de riesgo de enfermedades orales en personas mayores.

Materiales y Métodos

La búsqueda se realizó en las siguientes bases de datos: Medline a través de Pubmed, Embase, Colaboración Cochrane y Scielo, incluyendo artículos desde el año 2000 hasta junio 2020. Los términos de búsqueda (Mesh) fueron "Health Care Evaluation Mechanisms", "Validation Study", "Evaluation Study" "Reproducibility of Results", "Surveys and Questionnaires". Se incluyeron estudios en humanos tipo ensayos clínicos aleatorizados, estudios transversales y longitudinales.

Resultados

De un total de 647 artículos evaluados por dos revisores independientes, se incluyeron 48 artículos en la revisión. El tipo de estudio se distribuyó en 2 grandes grupos, 38 artículos fueron descriptivos transversales y 10 longitudinales prospectivos.

Discusión

El proceso de validación de los instrumentos diagnósticos presenta dentro de su estructura: la validación de contenido del instrumento, validación de criterio y de constructo, además de la consistencia interna y fiabilidad intra e inter observador. En base a esto se propone una estructura de 2 dimensiones para la elaboración y validación del EDePAM

Conclusión

La validación del EDePAM se debe realizar en el ámbito de criterio y constructo junto con consistencia interna y fiabilidad intra e inter observador.

Introducción

En Chile la esperanza de vida al nacer se ha triplicado, siendo la población de personas mayor a 65 años la que ha aumentado en mayor proporción. Se ha visto un incremento de casi siete veces más personas mayores en los últimos sesenta años. El aumento de las expectativas de vida trae consigo la aparición de mayor prevalencia de enfermedades en el último tramo de vida, esto por el proceso normal de envejecimiento de los tejidos. Dentro de los problemas asociados al envejecimiento de los tejidos de la cavidad oral encontramos la enfermedad de caries, las enfermedades periodontales, xerostomía, hiposialia, trastornos del gusto y problemas asociados al uso de prótesis dentales sin posteriores controles clínicos (Chile. Ministerio de Salud, 2010). Como solución a estas complicaciones, los servicios de salud pública ofrecen como tratamiento aquello que está determinado por las Garantías Explícitas de Salud (GES). A las personas mayores corresponde el GES de Salud Oral Integral del Adulto de 60 años cuyo fin es ofrecer tratamiento preventivo y rehabilitar mediante prótesis dentales a personas de 60 años. El GES de los 60 años ha demostrado una disminución del promedio del componente caries en el índice COPD y ha aumentado sustancialmente los tratamientos protésicos de tipo removible en la población beneficiaria, sin embargo no demuestra capacidad para mejorar los índices periodontales ni para manejar los factores de riesgo (Quezada Peña et al., 2013) Este programa de salud oral integral 60 años considera actividades de prevención y hasta de rehabilitación, pero debe profundizar en estrategias efectivas para lograr cambios de conductas, ofrecer prestaciones con enfoque de riesgo y diagnósticos en patología oral, pues lo que se hace actualmente resulta insuficiente (Jara et al., 2019).

Con el objetivo de mejorar la atención dental en este tramo de la población es que nace el examen dental preventivo del adulto mayor (EDePAM) que ofrece una atención basada en categorías de riesgo. El EDePAM pretende ser un examen complementario al examen médico preventivo del adulto mayor (EMPAM). El EMPAM es un monitoreo periódico que evalúa el estado de salud y la funcionalidad de las personas mayores que en el 2014 alcanzó una cobertura del 41,5% de la

población inscrita. Como comparación, la cobertura que ofrece el GES odontológico de los 60 años que fue de un 27,7% (MINSAL, 2014).

El EDePAM está dirigido a todas las personas mayores que tengan su EMPAM al día, con el objetivo de detectar problemas de salud oral, funcionalidad masticatoria y deglución, así como factores de riesgo. Su objetivo es desarrollar un plan de atención y seguimiento en función de una clasificación de riesgo odontológico del adulto mayor. La terapia contempla una fase preventiva, una fase de rehabilitación y una fase de seguimiento, procesos que cambiarán el nivel de riesgo del paciente y, por lo tanto, contribuirán a mejorar su calidad de vida.

El EDePAM se aplica anualmente en conjunto con el EMPAM en adultos de 65-69 años como población objetivo, de tal manera que el beneficiario tenga la oportunidad, dentro de ese período de vida, de ser atendido a través del EDePAM y con la cartera de prestaciones adecuada según su carga de enfermedad, para posteriormente continuar en control de acuerdo con el riesgo individual atribuido.

Los objetivos del EDePAM son evaluar la salud dental, la funcionalidad masticatoria y la deglución en adultos de 65-69 años, también identificar y controlar los factores de riesgo de la patología dental junto a otras enfermedades crónicas no transmisibles y en base a esto, elaborar un plan de atención dental y seguimiento, para ser ejecutado por el cirujano-dentista que trabaja en la atención primaria de salud y por el equipo de salud.

El EDePAM fue construido mediante una estrategia iterativa participativa basada en el criterio de expertos en cada una de las dimensiones evaluadas (funcionalidad masticatoria, lesiones de la mucosa oral, caries y enfermedad periodontal). Las opiniones de expertos fueron recogidas, evaluadas, resumidas y sistematizadas utilizando la metodología e-DELPHI en cuatro etapas. La construcción fue dirigida por un comité coordinador que contactó a los expertos en cuatro fases consecutivas, culminando con un taller de discusión presencial para alcanzar consensos:

1. El panel de expertos recibió un cuestionario con 2 preguntas amplias

2. El panel de expertos recibió las respuestas anónimas de la primera etapa y un segundo cuestionario
3. El panel de expertos recibió un informe integrando las respuestas del segundo cuestionario y un tercer cuestionario
4. El panel de expertos recibió un informe con los resultados

En cada una de las etapas se fueron acotando incrementalmente las posibilidades de respuesta, dando la oportunidad a los expertos de calificar y corregir sus respuestas iniciales. El resultado de esta estrategia iterativa fue alcanzar amplios consensos respecto a las mejores estrategias para evaluar objetiva y eficazmente cada una de las dimensiones del examen.

Una vez alcanzados los consensos de los expertos, se pasó a una segunda fase de confección de los instrumentos: ficha clínica y manual de aplicación. Estos instrumentos, a su vez, fueron sometidos a un proceso de validación inicial por expertos externos al grupo responsable de EDePAM, sin embargo, para poder implementar el EDePAM es necesario realizar la validación de este instrumento, siendo el principal objetivo de esta investigación.

El proceso de validación de un instrumento diagnóstico consiste en adaptarlo culturalmente si es necesario y comprobar sus características de fiabilidad y validez para valorar la precisión del instrumento (Gómez-Benito & Hidalgo, 2005).

La fiabilidad hace referencia al hecho de medir una variable de manera constante y, a su vez, la validez, a que el instrumento mida lo que se quiere medir y cumpla con el objetivo para lo cual fue construido. No todo instrumento que sea fiable es válido. Un instrumento puede ser fiable porque mide una variable de manera constante, pero inválido si no mide el fenómeno que quiere medir (Luján-Tangarife & Cardona-Arias, 2015). En conjunto, validez y fiabilidad determinan la calidad de un instrumento diagnóstico.

La fiabilidad se puede analizar respecto a la estabilidad y la consistencia interna que presente el instrumento. La estabilidad, nos habla de la consistencia de las respuestas obtenidas en repetidas ocasiones con los mismos sujetos. Una posibilidad para comprobar la estabilidad de un instrumento es la prueba repetida que nos permite establecer la estabilidad de las medidas obtenidas a partir de los mismos sujetos con el mismo instrumento en dos momentos diferentes. La razón de su selección radica en la simplicidad y bajo coste que supone la prueba repetida (Peña-Casanova, 2006).

Por otra parte, la consistencia interna mide el grado de correlación entre los distintos ítems del instrumento. Puede ser evaluada por medio de cuatro métodos. El primer método es mediante mitad dividida, el segundo, por correlaciones totales, el tercer método es el coeficiente de Kuder-Richardson y por último, el coeficiente alfa de Cronbach. Usualmente se utiliza Alfa de Cronbach como cálculo estadístico para determinar esta consistencia (Oviedo & Campo-Arias, 2005), pese a que posee críticas que deben ser consideradas, es uno de los métodos más versátiles para evaluar la consistencia interna y, además, sus resultados son de fácil obtención por ser basados en la aplicación de un test (Cascaes da Silva, Gonçalves, Valdivia, Grazielle. Da Silva y cols., 2015).

Como se había mencionado anteriormente, la validez indica el grado de exactitud con el que un instrumento mide el constructo teórico que pretende medir y si se puede utilizar con el fin previsto. Se puede decir que la validez tiene tres grandes componentes: validez de contenido, de criterio y de constructo (Juárez-Hernández y Tobón, 2018).

La validez de contenido hace referencia al grado en que se presentan los contenidos que la prueba trata de evaluar, sin omisiones. Descansa generalmente en el juicio de expertos. Por tanto, la validez de contenido se basa en la definición precisa del dominio y en el juicio sobre el grado de suficiencia con que ese dominio se evalúa. La validez de contenido se establece en diferentes situaciones, siendo dos de las más frecuentes: el diseño de una prueba, y la validación de un instrumento que fue construido para una población diferente, pero que se adaptó mediante un

procedimiento de traducción, conocido como equivalencia semántica (Escobar-Pérez, 2008). Dentro de los test que se usan para medir la validez de contenido está el método Delphi, metodología Q o el modelo de Lawshe modificado por Tristán en 2008 (Tristán-López, 2008). Este último método consiste en organizar un panel de expertos en el área los cuales van a analizar y deberán emitir su opinión en categorías obteniendo un valor que determinará la validez del contenido del instrumento. Cuando la validez de contenido se basa solo en el juicio de expertos se define como validez aparente y los datos obtenidos están basados en la objetividad del análisis del experto o el grupo formado para este fin (Zulay & Rayo, 2018).

En el caso de validez de criterio se busca relacionar las puntuaciones del instrumento con otras variables externas al instrumento. Dentro de la validez de criterio se conocen la validez concurrente y la validez predictiva. La diferencia entre ambas radica en la temporalidad. Si las puntuaciones de la prueba se utilizan para predecir alguna medida del criterio que se va a realizar a futuro, sería validez predictiva. Si, por el contrario, relacionamos las puntuaciones del test con alguna medida del criterio tomada en el mismo momento, sería validez concurrente (Argibay, 2006a).

Por último, la validez de constructo está relacionada a si una prueba refleja la teoría subyacente del constructo que se quiere medir (Luján-Tangarife & Cardona-Arias, 2015). Por lo tanto, la meta final de la validación es la explicación y comprensión de los resultados, para ello existen distintos procedimientos, entre los que se cuentan el análisis factorial, test convergente-divergente y validez discriminante (Antonio Pérez-Gil et al., 2000).

La validación de los instrumentos es un paso necesario para otorgar diagnósticos y tratamientos con la mayor evidencia disponible, midiendo las características necesarias y que cumpla con el objetivo con el que fue construido. El EDePAM tiene como objetivo detectar de forma temprana patologías en la cavidad oral, alteraciones en la funcionalidad masticatoria y de deglución, así como factores de riesgo que presente el adulto mayor para posteriormente desarrollar un plan de

atención y seguimiento acorde a la clasificación de riesgo odontológico. La validación de este instrumento es el objetivo de esta investigación determinando cual es la metodología más eficiente para validarla y así contar con la mejor evidencia científica disponible, para aumentar la efectividad diagnóstica y reducir los efectos adversos que puedan tener estas intervenciones en la salud de las personas mayores (Carvajal, Centeno, Watson y cols, 2011).

Pregunta de investigación.

¿Cuál es la metodología más eficiente para validar el instrumento EDePAM en las dimensiones de validez y fiabilidad?

Objetivo general.

Determinar el método más eficiente de validación del EDePAM para la su implementación como instrumento diagnóstico.

Objetivos específicos.

- Evaluar los protocolos existentes para validación de los cuestionarios e instrumentos diagnósticos.
- Determinar criterios de elaboración y confección de cuestionarios e instrumentos diagnósticos
- Determinar el método de validez más efectivo para EDePAM
- Determinar el método de fiabilidad más efectivo para EDePAM

Materiales y Métodos

Con el objetivo de obtener el protocolo más eficiente de validación se realizará una búsqueda en la literatura basada en las directrices PRISMA para obtener una revisión sistemática de tipo cualitativa (Urrutia & Bonfill, 2010). Las revisiones sistemáticas recopilan toda la información que se ajuste a criterios de selectividad previamente especificados y mediante el diseño PRISMA se busca explicitar y sistematizar los métodos utilizados con el fin de reducir al máximo el sesgo,

proporcionando resultados confiables y reproducibles. PRISMA comprende una lista de 27 pasos (Anexo 1) y un diagrama de flujo de cuatro fases.

Fuentes de información

La búsqueda se realizó por dos revisores de forma independiente (PD, XM) en las siguientes bases de datos: Medline a través de Pubmed, Embase, Colaboración Cochrane y Scielo de aquellos disponibles en inglés desde el año 2000 hasta junio 2020. La literatura gris quedó excluida. El software de gestión bibliográfica utilizado fue Mendeley © (2020 Mendeley Ltd).

Algoritmo de búsqueda

Para seleccionar las palabras claves de la búsqueda se utilizó la matriz con los componentes de la pregunta de investigación (PACOT) y la posterior búsqueda de términos Mesh que más se ajustaban a los componentes descritos en la matriz. Los términos de búsqueda seleccionados (Mesh) fueron “*Health Care Evaluation Mechanisms*”, “*Validation Study*”, “*Evaluation Study*” “*Reproducibility of Results*”, “*Surveys and Questionnaires*” los cuales están definidos en el Anexo 2 y el término booleano utilizado fue “AND”. Finalmente, la estrategia de búsqueda utilizada en PUBMED fue "Health Care Evaluation Mechanisms"[Mesh] AND "Validation Study" [Publication Type] AND "Evaluation Study" [Publication Type] AND “Reproducibility of Results” [Mesh] AND “Surveys and Questionnaires” [Mesh].

Tabla 1: Matriz lógica para la estrategia de búsqueda.

Componentes de la pregunta			
Población	Intervención	Comparación	Resultados
Estudios en humanos	Validación EDePAM	Instrumentos diagnósticos validados	EDePAM es válido y fiable
Transformando a estrategia de búsqueda con palabras clave controladas y texto libre			
Instrumentos diagnósticos "Health Care Evaluation Mechanisms"[Mesh]	Estudios de validación instrumentos diagnósticos Mecanismos de evaluación de "Validation Study" [Publication Type] "Evaluation Study" [Publication Type] "Reproducibility of Results" [Mesh]	Encuestas y cuestionarios Estudios médicos "Surveys and Questionnaires" [Mesh]	Validación

Crterios de inclusión y exclusión

Los criterios de inclusión de esta revisión fueron: ensayos clínicos aleatorizados, estudios transversales y longitudinales, estudios en humanos y artículos en inglés desde el 2000 en adelante, no se estableció límite de edad. Además, se realizó una búsqueda manual a través de las referencias de los estudios encontrados. Los criterios de exclusión fueron: publicaciones en congresos, estudios en animales e *in vitro*, tesis y revisiones sistemáticas, adaptaciones de cuestionarios, cuestionarios auto reportados, versiones cortas de cuestionarios, validación de guías clínicas, dispositivos diagnósticos o métodos clínicos.

Proceso de selección de artículos

Dos investigadores (PD, XM), en forma independiente, revisaron todos los resúmenes, aplicando los criterios de inclusión y exclusión. En los artículos que cumplieron con los criterios de selección, se realizó la revisión de texto completo, luego de la eliminación de duplicados y registros irrelevantes, definiendo y registrando los motivos por los cuales el estudio se excluyó. La fiabilidad inter revisor en el proceso de selección de estudios fue determinado por el Coeficiente kappa de Cohen. Este cálculo se realizó para cada una de las búsquedas en las distintas

bases de datos. Los desacuerdos fueron resueltos por un tercer investigador (FC), que sirvió como juez independiente.

Evaluación de la calidad de los artículos

Con el fin de determinar la calidad de los artículos incluidos en esta revisión se realizó una lista de verificación con 4 dominios y 16 ítems (Tabla 2). Los dominios corresponden a introducción, metodología utilizada, propiedades del cuestionario y validez y fiabilidad. Dentro del dominio de introducción se evaluó si indicaba objetivo y la definición del constructo, en la metodología, el tamaño de la muestra, definición de la población objetivo del cuestionario, el método de recolección de información para la construcción del cuestionario, la forma de administración y el formato de aplicación del cuestionario. En el dominio de propiedades del cuestionario, se encontraba la definición de dimensiones a evaluar, definición de ítems, si el contenido es unidimensional o multidimensional, la puntuación de los ítems y si correspondía a un estudio o una prueba piloto. La validez y fiabilidad del artículo se definió en el ítem de si evaluaba consistencia interna, validez, estabilidad temporal y concordancia inter observadores.

Tabla 2. Lista de verificación de calidad utilizada para evaluar la calidad de los artículos seleccionados para la revisión sistemática.

Ítem	SI	NO
Introducción		
Indica objetivo		
Definición del constructo		
Metodología		
Tamaño de la muestra		
Define población objetivo del cuestionario		
Método de recolección de información para la construcción del cuestionario		
Forma de administración del cuestionario		
Formato		
Propiedades del cuestionario		
Definición de dimensiones a evaluar		
Definición de ítems		
Contenido Uni o Multidimensional		
Puntuación de los ítems		
Estudio/Prueba Piloto		
Validez y fiabilidad		
Evalúa consistencia interna		
Evalúa Validez		
Evalúa estabilidad temporal		
Concordancia interobservadores		

Síntesis de datos

Debido al alto nivel de heterogeneidad en las variables de validación en los estudios incluidos, no fue posible realizar un metaanálisis del hallazgo. Por ello, se realizó una síntesis cualitativa de los resultados de la revisión y se elaboraron tablas que incluyen las características y los resultados de los estudios seleccionados.

Resultados

Búsqueda y selección de artículos

Como resultado de la búsqueda en las distintas bases de datos se obtuvo un total de 647 artículos. De este total, 45 artículos se eliminaron por estar duplicados, quedando 602 para la lectura del resumen y aplicación de criterios de inclusión y exclusión. Finalmente 48 artículos fueron incluidos en la revisión. En la Figura 1 se resume el proceso de búsqueda, selección e inclusión de los artículos incluidos en la revisión sistemática.

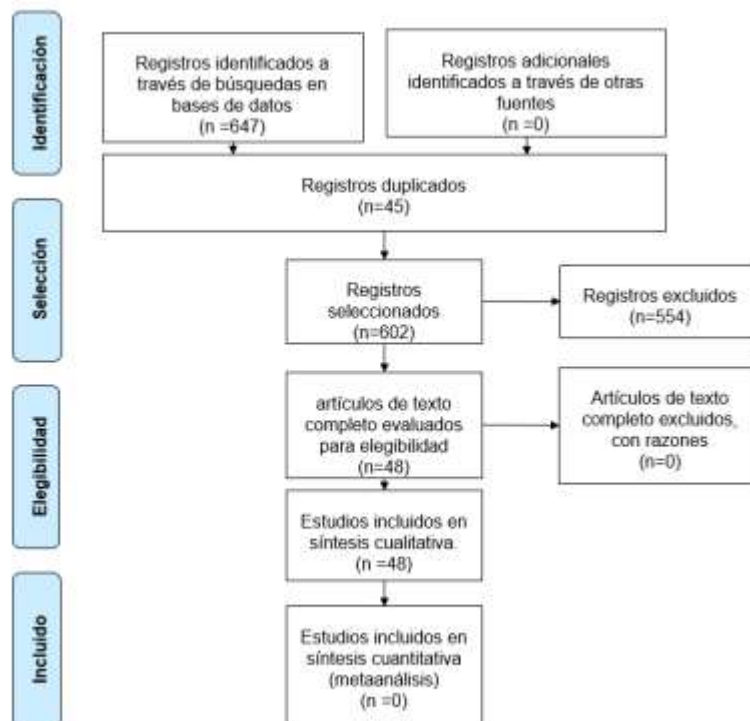


Figura 1. Flujograma del proceso de identificación y selección de los artículos para la revisión sistemática según modelo PRISMA.

Calidad de los artículos

Al aplicar la lista de verificación de calidad se pudo observar que el 75% de los artículos evaluó más de un 80% del total de ítems y el 25% restante evaluó entre el 60 y 80% de los ítems. Solo el 21% de los artículos evaluó la concordancia inter observadores, siendo el ítem más bajo seguido por “Definir la puntuación de los ítems” con un 42% y “Evaluar la estabilidad temporal” con un 58% de los artículos.

Descripción de los estudios incluidos

De los 48 estudios seleccionados, el que presenta mayor antigüedad fue del año 2004 y el más reciente de abril del 2020. El tipo de estudio se distribuyó en 2 grandes grupos, 38 artículos fueron estudios descriptivos transversales y 10 estudios longitudinales prospectivos.

La distribución de los estudios por países se encuentra distribuida con un 47% de los estudios en Europa y el porcentaje restante se divide entre América, Asia y Oceanía tal como se ve en la Figura 2.

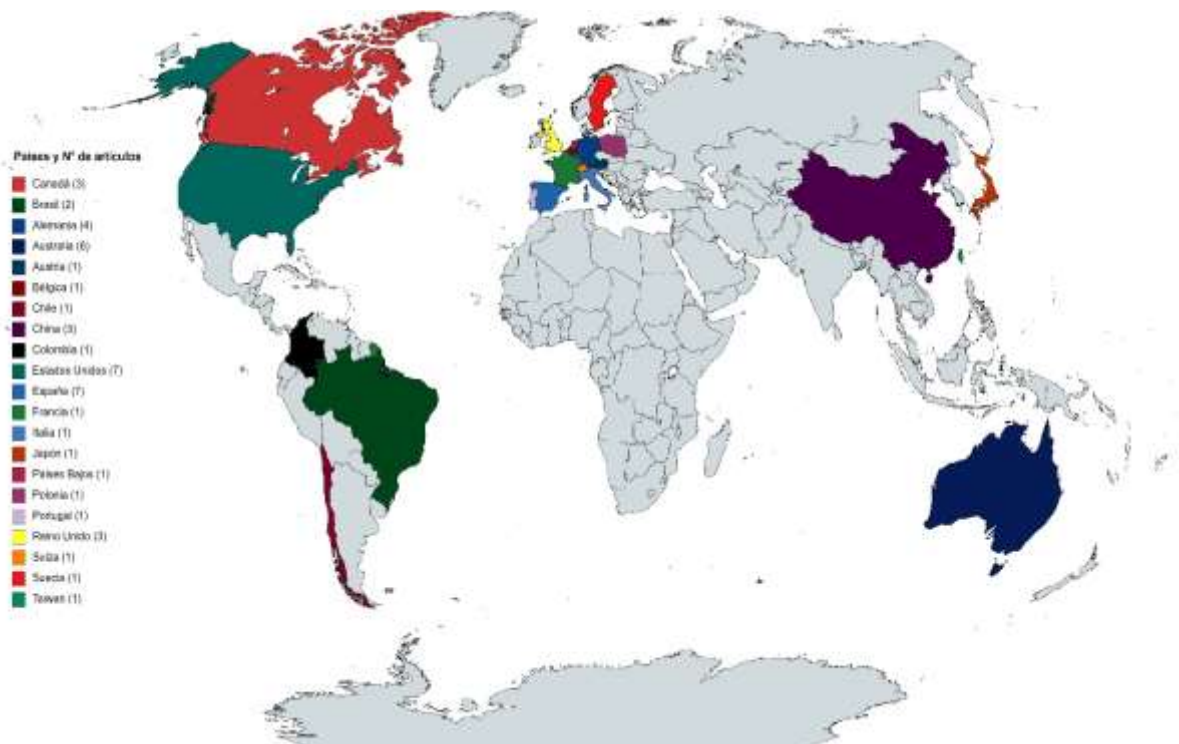


Figura 2: Distribución geográfica de los artículos.

Los artículos fueron publicados en revistas científicas vigentes indexadas a bases de datos cuyos responsables como organización editora se encuentran Universidades e Institutos de Investigación en su mayoría, con una distribución por año que se puede encontrar en la Figura 3. Se obtuvo el factor de impacto de cada una de las revistas científicas para evaluar la importancia relativa de ésta y del artículo seleccionado. El promedio del factor de impacto fue de 2,953 con una mediana de 2.278.



Figura 3: Distribución de los artículos por año.

Comité de ética

Todos los artículos mencionaron la aprobación de su estudio por un comité de ética. Además, los participantes elegidos dieron su consentimiento por escrito para participar en los estudios.

Validez aparente

La validez aparente es una forma de validez de contenido que mide el grado en el que los ítems parecen medir lo que se proponen. Solo el 15% de los artículos analizados mencionó explícitamente que evaluó la validez aparente del cuestionario o instrumento diagnóstico desarrollado.

Turner et al. (2007) mencionó en primer lugar la generación de ítems mediante una revisión de la literatura para identificar variables existentes, generando una lista de ítems que se entregó a un grupo Delphi de 48 especialistas en el área pidiéndoles que comenten la lista e identifiquen otros factores, además debían clasificar cada elemento de 0 a 3 siendo cero no importante y 3 extremadamente importante (Turner et al., 2007).

McAllister & Flynn (2016), para el desarrollo de su cuestionario organizó un simposio de 50 enfermeras donde desarrolló 5 aspectos importantes del cuestionario, para posteriormente formar un grupo de 5 enfermeras que desarrollaron inicialmente 110 ítems dentro de los 5 aspectos desarrollados en el simposio. Luego realizó un panel de pares el cual redujo a 98 ítems siendo evaluado con una escala Likert de 5 pasos. El borrador final fue enviado a un grupo de 5 enfermeras mediante una plataforma online para ser examinado tanto la relevancia y legibilidad y el tiempo de finalización (McAllister & Flynn, 2016).

Cobos et al. (2014), para la selección del grupo de expertos, se basó en los años de experiencia profesional que poseen seleccionando solo dos expertos (Cobos et al., 2014). Kowalski et al. (2015), los seleccionó en base a su cargo y al área de interés, seleccionando así un total de 15 participantes incluyendo gerentes de calidad, médicos, enfermeras y representantes de autoayuda (Kowalski et al., 2015).

Biagioli et al. (2019) en el proceso de validación aparente, con el fin de verificar la relevancia y representatividad de cada ítem, así como fomentar la generación de ítems, realizó un grupo focal con 10 expertos: cinco enfermeros, dos hematólogos, dos psicólogos y un expaciente. Los comentarios se utilizaron para desarrollar un primer borrador del cuestionario (Biagioli et al., 2019).

Thomason et al. (2014), conformó dos paneles de expertos para el desarrollo de los apartados del instrumento. Un primer panel se dedicó a la identificación y selección de los ítems importantes para el cuestionario y un segundo panel, para la revisión de aquellos ítems potenciales. El primer grupo realizó una lluvia de ideas, para posteriormente discutir la viabilidad de cada variable y su contribución a los objetivos

del estudio. Al segundo grupo, se le envió por correo los elementos seleccionados para su revisión y selección, además de solicitar sugerencias de redacción y terminología. Ambos grupos fueron compuestos por 11 expertos entre los que se encontraban médicos, terapeutas ocupacionales, especialistas en enfermería clínica y otros expertos relacionados con los ítems del cuestionario(Thomason et al., 2014).

Medina-Mirapeix et al.(2015), mencionó que la redacción de los ítems se basó en los comentarios de los pacientes y se modificaron mediante un proceso de discusión entre los miembros del equipo del estudio, el formato de respuesta del equipo fue mediante el uso de una escala de frecuencia Likert de 5 puntos. El equipo estaba conformado por 7 profesionales con una experiencia media de 10 años y 94 pacientes con trastornos musculoesqueléticos los cuales fueron seleccionados al azar. De tener 30 ítems pasaron a tener 27 en el instrumento final, el ítem se mantenía si el 80% de los participantes o de los profesionales determinaba que era relevante(Medina-Mirapeix et al., 2015). Webber et al. (2014), también utilizó una escala de frecuencia Likert de 5 puntos a un panel de 15 expertos en el área mediante un proceso Delphi, pero el ítem se mantenía si al menos el 75% de los expertos estaban de acuerdo o totalmente de acuerdo que el elemento era relevante(Webber et al., 2014).

Validez de contenido

La validez de contenido evalúa de manera cualitativa si el cuestionario abarca todas las dimensiones del fenómeno que se quiere medir, ya que se considera que un instrumento es válido en su contenido si contempla todos los aspectos relacionados con el concepto que mide. Un 79% de los estudios analizados evaluó la validez de contenido del instrumento y para esto se utilizaron distintos tipos de pruebas.

Se ha descrito una evaluación preliminar de la validez de los cuestionarios, Liang, en el 2015, realizó una prueba piloto para una validez inicial de la generación de los ítems del cuestionario. Esta prueba preliminar se aplicó a una muestra piloto con pacientes de las mismas características de público objetivo y los resultados fueron analizados por un panel de expertos, agregando o eliminando ítems dependiendo

de los resultados determinados en el análisis exploratorio de los componentes principales(Liang et al., 2015).

López-Torres-Hidalgo et al. (2005) realizó una prueba piloto en 16 pacientes con el objetivo de establecer la claridad de las preguntas y las instrucciones, así como registrar el tiempo necesario para completar la prueba(López-Torres et al., 2005).

McMurray et al. (2018), para el desarrollo de su herramienta de evaluación, inicialmente realizó una revisión de la literatura generando una lista de elementos potenciales. Luego, conformó un grupo de 14 expertos, el cual “se definió como un médico con experiencia especializada en el dolor irruptivo y / o el desarrollo de instrumentos para el dolor y considerado por sus pares como un líder de opinión clave” calificando los ítems con una escala Likert de 5 puntos con un acuerdo mínimo del 75% de los expertos para mantener el ítem. Seleccionados los artículos se realizó un muestreo a pacientes en dos oportunidades para comprobar que la redacción fuera clara y legible. Para esto, se les solicitó a los participantes que, a medida que leían el cuestionario, pensarán en voz alta y registrarán esta información(McMurray et al., 2018).

McMurray et al. (2019) realizó tres pasos para la validación de contenido. El primero fue una escala de Likert de 5 pasos a un grupo de expertos, posteriormente se hizo con una escala de Likert de 7 pasos, pero no se observaron diferencias significativas. El segundo paso fue formar grupos con pacientes y cuidadores y realizar una clasificación de tarjetas, las cuales tenían impresas distintos ítems del cuestionario y se les pidió que las clasificaran en orden de importancia y justificaran su elección. Como tercer y último paso, se les pidió a los proveedores de salud que diseñaran preguntas y discutieran cual era la que más los representaba (McMurray et al., 2019).

Zhao et al. (2014) envió dos encuestas a expertos. La primera incluyó una carta de presentación, el consentimiento por escrito y un borrador del cuestionario. Estos expertos tenían una semana para enviar los resultados a los autores, los cuales analizaron las respuestas y, dos semanas después, enviaron un segundo

cuestionario con las sugerencias de los expertos. El análisis del segundo cuestionario por parte de los expertos corresponde al juicio de validez de contenido, la correlación y la intención de los ítems para finalmente responder una escala preliminar para comprobar la comprensión y la redacción (Zhao et al., 2017).

Liem et al. (2017) mencionó que un factor importante a considerar para la validación del contenido era la facilidad de entender las preguntas. En este estudio, con el fin de que el paciente comprendiera la pregunta, se escribió una breve definición de conceptos claves(Liem et al., 2017).

Validez de criterio

Dentro de la validez de criterio, se analiza tanto la validez concurrente como la validez predictiva y se las relaciona con conceptos como el poder predictivo y poder explicativo (Argibay, 2006b). El 60% de estos artículos mencionó comprobar la validez de criterio en al menos uno de sus dos aspectos.

Liang et al. (2015), para evaluar la validez concurrente, correlacionó la escala de autoeficacia para el manejo de síntomas en el cáncer de mama, que era su instrumento por validar, con la escala de autoeficacia general que es el estándar de oro utilizado hasta ahora. Para hacer esto, utilizaron el coeficiente de correlación de Pearson (Liang et al., 2015). Esta forma de evaluar la validez concurrente también fue utilizada por Webber et al. (2014) y determinaron que los valores del coeficiente de correlación $> 0,1$ equivalen a una pequeña correlación, $> 0,3$ equivalen a una correlación media y $> 0,5$ equivalen a una fuerte correlación (Webber et al., 2014).

Kowalski et al., 2015 realizó el coeficiente de correlación de Spearman para evaluar la correlación del instrumento con el estándar de oro para evaluar la validez concurrente, mientras que para evaluar la validez predictiva utilizó un modelo lineal jerárquico comparando si la puntuación de la prueba se correlacionaba positivamente con la información proporcionada a los pacientes utilizando el coeficiente de correlación interclase (Kowalski et al., 2015).

Parra-Anguita et al. (2019), para la validación de criterio concurrente comparó el cuestionario desarrollado con un cuestionario descrito como estándar de oro. Se utilizó el coeficiente de correlación intraclase (CCI) como estadística para la estimación de la concordancia (Parra-Anguita et al., 2019).

Crane et al. (2012) creó un cuestionario para identificar las necesidades de las personas mayores con asma. El autor mencionó que actualmente no existe un estándar de oro con el cual comparar el cuestionario, por lo cual realizó una comparación de los cuestionarios validados que se centraban en constructos particulares que poseía el cuestionario a validar (Crane et al., 2012).

Thomason et al., (2014) desarrolló una herramienta para monitorear la cicatrización de úlceras. Al evaluar la validez predictiva que tendría el instrumento, correlacionaron el tiempo de cicatrización y la compararon con la puntuación del estadio de la úlcera descrita al inicio del estudio (Thomason et al., 2014).

Validez de constructo

La validez de constructo es una medida que expresa si el instrumento indica correctamente la base teórica que pretende medir.

Lockhart et al. 2013, en su cuestionario para determinar el impacto de la neoplasia intraepitelial vulvar, evaluó la validez de constructo mediante dos maneras. Primero, cada pregunta en el cuestionario se correlacionó con la puntuación total corregida para evaluar en qué medida los elementos individuales estaban midiendo un solo constructo. En segundo lugar, para evaluar si el cuestionario estaba midiendo algo distinto a lo medido por los otros cuestionarios, cada una de las preguntas se correlacionó con las puntuaciones totales de los cuestionarios analizados como control (Lockhart et al., 2013).

Parra-Anguita et al. (2019), probó la validez de constructo mediante análisis factorial exploratorio, dando como resultado que el cuestionario era una escala unidimensional (Parra-Anguita et al., 2019). Mientras que Delclos et al. (2006),

buscó asociaciones entre los ítems de su cuestionario con otras medidas en las cuales se conoce su resultado (Delclos et al., 2006).

López-Torres et al. (2005), garantizó una adecuada validez de constructo mediante un análisis cualitativo utilizando un análisis factorial con extracción de factores por el método de componentes principales y una rotación posterior, explorando dimensiones subyacentes comprobando que el contenido del cuestionario fuera concordante con el concepto teórico de la enfermedad y comparable con otros instrumentos que miden el mismo concepto (López-Torres et al., 2005).

Consistencia interna

La consistencia del instrumento es una medida para determinar si los distintos ítems producen el mismo resultado. Este paso se evaluó en el 88% de los artículos analizados.

Delclos et al. (2006) desarrolló un instrumento de encuesta sobre el asma para estudios epidemiológicos y para evaluar la consistencia interna lo realizó de dos formas. La primera fue mediante el coeficiente de Cronbach para los grupos de ítems en las secciones de asma y la segunda forma fue mediante un análisis exploratorio de factores principales (Delclos et al., 2006). Kaczmarek et al. (2014), para realizar este mismo paso, solo utilizó el coeficiente de alfa de Cronbach como análisis estadístico para investigar la consistencia interna de su cuestionario y los subconjuntos de éste (Kaczmarek et al., 2014).

Kowalski et al. 2015, realizó un análisis de ítems para examinar hasta qué punto se correlacionó cada ítem con la puntuación del instrumento total mediante la correlación ítem-total corregida. Además, calculó el alfa de Cronbach para examinar la consistencia interna de los ítems (Kowalski et al., 2015).

Fiabilidad intra observador

La fiabilidad intra observador tiene como objetivo evaluar el grado de consistencia al efectuar la medición de un observador consigo mismo (Sánchez Fernández et al.,

2005). El 56% de los artículos mencionó la evaluación de la fiabilidad intra observador.

Webber et al. (2014), al medir la estabilidad del instrumento, realizó dos evaluaciones adicionales a la inicial, una a 24 horas después de la primera evaluación y la última evaluación la realizaron con al menos una semana de diferencia. El análisis estadístico se realizó con el cálculo de coeficiente de correlación intraclass e indicaron una buena fiabilidad prueba re prueba si el resultado era superior a 0,8 (Webber et al., 2014). Por su parte, Noll et al., (2013) en su estudio realizó algo similar a Webber, pero aplicando el cuestionario a los 7 días, contando con un examinador que estaba instruido en explicar el cuestionario pero sin mencionar el procedimiento de repetición de la prueba hasta que hubiese terminado la prueba inicial. Para el análisis estadístico de la prueba y la re prueba utilizó en el coeficiente kappa no ponderado para escalas nominales clasificando los resultados según su valor. Los resultados se clasificaron en malos ($k \leq 0,2$), regulares ($0,2 < k \leq 0,4$), moderados ($0,4 < k \leq 0,6$), buenos ($0,6 < k \leq 0,8$) o muy buenos ($0,8 < k$) y para que una pregunta permaneciera incluida en el cuestionario, se requería obtener un valor mínimo de $k \geq 0.5$ (Noll et al., 2013). De manera similar, Lockhart et al., 2013 utilizó valores superiores a 0,4 de kappa para una concordancia moderada y más de 0,6 para reflejar una buena concordancia (Lockhart et al., 2013).

Lockhart et al., 2013, para evaluar la confiabilidad prueba re prueba de su cuestionario, lo envió en dos ocasiones, con un intervalo de 7 días entre cada una. Se eligió el intervalo de 7 días ya que era poco probable que hubiera cambios sustanciales en la salud de los participantes en ese tiempo, pero sería lo suficientemente largo como para minimizar que los participantes recordaran sus respuestas a la primera administración del cuestionario (Lockhart et al., 2013). Por su parte, Tong et al. (2009) validó un instrumento que evaluó la funcionalidad durante el tratamiento del cáncer nasofaríngeo, pero en la prueba re prueba utilizó un tiempo de 2 semanas para volver a aplicar el instrumento de medición (Tong et al., 2009).

Bejia et al. (2006), en su encuesta para identificar problemas de espalda en adolescentes, diseñó una prueba re prueba con 7 días de diferencia. Para ello utilizó 72 sujetos que participaron en ambas fases y para el análisis se utilizó el coeficiente de kappa con los criterios de Landis y Koch (Bejia et al., 2006).

McAllister & Flynn, (2016) utilizaron el coeficiente de correlación intraclase, correlaciones de Spearman y una prueba de rango con signo de Wilcoxon para evaluar la fiabilidad prueba re prueba del cuestionario. Los indicadores de confiabilidad aceptable fueron un coeficiente de correlación intraclase mayor que 0,75 y con un intervalo de confianza del 95%, un coeficiente de correlación de Spearman mayor a 0,61 y un resultado no significativo de una prueba de rango con signo de Wilcoxon (McAllister & Flynn, 2016).

Fiabilidad inter observador

La fiabilidad inter observador pretende evaluar la consistencia entre dos o más observadores distintos cuando evalúan una misma medida en un mismo individuo. Solo el 19% de los artículos evaluó la concordancia interobservadores, siendo el ítem más bajo evaluado en la dimensión de confiabilidad del instrumento.

Dentro de los artículos que midieron la confiabilidad interobservador, Turner et al., (2007) evaluó el índice de actividad de colitis en dos oportunidades por dos médicos en la misma visita (Turner et al., 2007). Por su parte, Eslava-Schmalbach et al. (2006) realizó esta medición con diferencia de horas y los evaluadores eran ciegos de los resultados obtenidos previamente. Para el análisis estadístico utilizó el coeficiente de concordancia de Kendall para evaluar el nivel de concordancia entre los expertos (Eslava-Schmalbach et al., 2006).

López-Torres et al. (2005), realizó una prueba con 20 pacientes en dos ocasiones distintas en un plazo inferior a una semana por dos médicos para determinar la fiabilidad entre los observadores y utilizó el coeficiente de correlación intraclase como método analítico (López-Torres et al., 2005).

Malouf et al., (2014) utilizó a tres evaluadores de forma independiente y comparó los resultados con el coeficiente de correlación intraclase, mencionando que se prefirió éste sobre los índices más clásicos, como el coeficiente de correlación de Pearson debido a que este puede producir resultados engañosos (Malouf et al., 2014).

Uezato et al. (2014), formó pares de evaluadores y cada uno de ellos evaluó al mismo paciente por separado el mismo día, posteriormente a esto se evaluó el coeficiente kappa para cada ítem definiendo una confiabilidad excelente como igual o mayor a 0,8, buena si el valor kappa era igual o mayor a 0,6, moderada igual o mayor a 0,4, regular mayor o igual a 0,2 y mala si el valor kappa es menor a 0,2 (Uezato et al., 2014).

Programas estadísticos

Los programas estadísticos están diseñados para resolver problemas en esta área, apoyando a los investigadores en su trabajo. El uso de los distintos programas o paquetes estadísticos dependerá de los datos a analizar. Así es como Kowalski et al., 2015a utilizó IBM SPSS Statistics 22.0 para el análisis descriptivo, IBM SPSS AMOS 22.0 para el modelado de ecuaciones estructurales y HLM 7 para el análisis multinivel, al igual que Raña-Rocha et al. (2020) utilizó SPSS y el paquete de psiquiatría de R. Uezato et al., (2014), para el análisis interobservador también mencionó utilizar el software R.

Parra-Anguita et al. (2019) tabularon sus datos en una hoja de cálculo antes de su análisis, para posteriormente ser ingresado al software JMetrik. Por su parte, Gau & Hung (2014) realizaron la estadística descriptiva utilizando SPSS 13.0 y el software LISREL 8.0 para determinar la estructura factorial de la escala.

Granado de la Orden et al. (2007) ingresó las encuestas recogidas en una base de datos con el programa informático ACCESS especialmente diseñada para su estudio y posteriormente utilizó el programa SPSS para el análisis estadístico. López-Torres-Hidalgo et al. (2005), utilizó el programa Win Episcopo 2.0 para realizar el análisis y grafica de sus resultados

Discusión

La lista de verificación de calidad utilizada en esta revisión nos permite, en primer lugar, cuantificar la calidad de la evidencia recogida.

Con este objeto, se consideró adecuado establecer una gradación de calidad en una escala de puntos, de forma que habría una puntuación mínima de 40 puntos. Valores inferiores indicaban que los resultados del estudio no deberían ser tomados en consideración. Si la puntuación alcanzada se encontraba entre 40 y 60, se acordó que el estudio no era suficientemente informativo, por lo que sus conclusiones deberían ser consideradas con cautela. Así mismo, una puntuación igual o superior a los 60 puntos, sería garantía suficiente para que los resultados de la evaluación se considerasen relevantes. En este trabajo, el 75% de los artículos evaluó más de un 80% del total de ítems y el 25% restante evaluó entre el 60 y 80% de los ítems por lo que la información recabada cuenta con un alto nivel de calidad (Ver Anexo 3).

La revisión identificó diferencias entre los artículos en relación con la metodología de validación. Solo el artículo de Eslava-Schmalbach et al., en el 2006 mencionó la validación y confiabilidad de todos los aspectos de su instrumento. Esto no significa que posea una mejor validación sobre los demás estudios, pero sí es un factor que se debe considerar a la hora de utilizar ese instrumento ya que utilizó diferentes estrategias de validación.

La validación de contenido fue el aspecto más estudiado en los artículos, debido a que se busca comprobar que el instrumento mida correctamente todos los aspectos necesarios a evaluar, en contraste con la validez aparente, que fue el aspecto menos mencionado explícitamente en los artículos. Sin embargo, al momento de la lectura detallada de éstos, aquellos que hacían referencia a validez de contenido, también mencionaban muchos aspectos de validación aparente o no hacían distinción de ellos. Es por esto que la validación de contenido presentaba valores mayores a otros tipos de validación, pero no quiere decir que la validez aparente no fue un aspecto estudiado.

Según la Asociación Estadounidense de Investigación Educativa, la validación es el proceso de desarrollar un argumento sólido sobre cómo interpretar los puntajes en una prueba y la relevancia de la prueba para su uso propuesto. Debido a la amplia gama de tipos de pruebas, los tipos de evidencia que mejor apoyan la validez de una prueba propuesta también varían ampliamente. Por lo tanto, se recomienda que la validación se aclare mediante el desarrollo de un “conjunto de proposiciones que apoyen la interpretación propuesta para el propósito particular de la prueba” (Backhoff et al., 2006). Es así como el EDePAM debe seguir un conjunto de pasos para la confección y validación del instrumento.

Primera Fase: Elaboración

Para la confección del EDePAM fue necesario tener claridad de los objetivos y de las teorías subyacentes que fundamentaban la elaboración de este instrumento definiendo si el instrumento tenía como objetivo medir solo un atributo o más de uno siendo considerado unidimensional o multidimensional, respectivamente.

Es aquí donde se debe dar la redacción del primer set de ítems, sin importar el tipo de validación que se utilizó. Todos estos artículos de esta revisión sistemática, hicieron referencia a que es necesario, en primer lugar, una búsqueda bibliográfica con el objetivo de encontrar si existen encuestas previas, estudios relacionados, técnicas similares que puedan guiar en la elaboración y/o traducción de los instrumentos diagnósticos. La búsqueda de esta información debe ser realizada por más de un investigador con el fin de reducir el sesgo, se menciona la creación de grupos de investigadores, los cuales deben ser expertos en el área, para así mejorar la selección de los artículos. La selección del grupo de expertos para la confección del instrumento es ambiguo y dependiente de los títulos de los expertos, su función o cargo (Kowalski et al., 2015), por los años de experiencia profesional que poseen (Cobos et al., 2014) o pueden ser seleccionados de forma azarosa y voluntaria (McAllister & Flynn, 2016). El número de expertos es un aspecto que se debe considerar a la hora de reclutarlos, ya que debe considerar los abandonos y las no respuestas. Sin embargo, la selección de este grupo de expertos depende del tipo de instrumento que se quiera validar, porque un instrumento que mide aspectos

cualitativos del paciente debe incorporarlos para evaluar la relevancia y comprensión de cada ítem (Storch et al., 2011).

Segunda fase: Validación por expertos

La búsqueda realizada debe dar como resultado una lista con ítems potenciales con los temas que se quieren abarcar. Se pueden realizar entrevista de grupo focales para explorar la experiencia de los pacientes, las preguntas abiertas se utilizan para acomodar el conocimiento, las experiencias y las necesidades de los participantes, generando así un primer acercamiento al instrumento diagnóstico. Este debe ser sometido a un proceso de juicio por un segundo comité de expertos, distinto al primero en lo posible.

El EDePAM construyó cada una de sus dimensiones en base a la discusión que generó un panel de expertos siguiendo metodología eDELPHI. Éste consistió en tres rondas de discusión donde en cada una, los expertos entregaban sus respuestas de forma anónima donde podían calificar y corregir sus respuestas iniciales alcanzando un consenso final. A pesar de que el método de confección de la encuesta haya sido realizado mediante la extracción de preguntas ya validadas siempre se debe validar el instrumento final.

Como proceso de elaboración del instrumento diagnóstico es necesario determinar si la implementación es factible, para esto se debe evaluar la claridad de las instrucciones del cuestionario, el tiempo y la facilidad para completar el cuestionario, redacción adecuada a la población en estudio y determinar cualquier dificultad asociada con el cumplimiento de los cuestionarios. A los expertos se les pide la revisión de la redacción, terminología y validación aparente del instrumento (Thomason et al., 2014), generando así un cuestionario a validar. Al momento de ordenar las preguntas, éstas deben seguir un orden lógico iniciando con las más sencillas, siguiendo una secuencia incremental y reservando para el final los aspectos con mayor carga emocional, siguiendo agrupaciones temáticas y evitando el efecto de contaminación de unas preguntas sobre otras, logrando una conversación fluida. El lenguaje y los términos empleados deben ser pensados en el usuario final que utilizará este instrumento. Con el fin de que la respuesta por

parte del entrevistado sea acorde a la pregunta y que éste haya entendido aquellos términos propios del área y no de conocimiento general, siempre se debe considerar el uso de un glosario con términos específicos bajo la pregunta para ser explicado al entrevistado y ayudar a la comprensión.

Fase 3: Prueba piloto

Una vez se hayan reducido los ítems y se tenga una versión preliminar se debe realizar una prueba piloto en pacientes con las mismas características de la población final. Esto tiene la finalidad de evaluar si el contenido abarcado en el instrumento y la claridad de redacción es la adecuada, si el instrumento tiene la capacidad de explicarse por sí solo y si su aplicación final no requiere de la presencia del investigador si ese fue el propósito de su creación.

La validez de contenido se debe comprobar a través de una escala Likert, realizada por el panel de expertos. Algunos autores sugieren que estas escalas deben tener cuatro opciones de respuesta, añadiendo la opción “sin opinión” o “no sé” (Nadler et al., 2015) o bien, escalas de siete opciones para evitar una excesiva interpolación, tal como sugiere Finstad (2010), sin embargo no se encontraron diferencias significativas entre una escala de 5 o 7 opciones (McMurray et al., 2019).

Fase 4: Validez del instrumento

La validez concurrente consiste en determinar en qué medida nuestro instrumento se correlaciona con una medida reconocida previamente o estándar de oro. El método descrito en las publicaciones para realizar esta correlación fue el coeficiente de correlación de Spearman. La prueba de Spearman tiene la ventaja de ser muy sencilla de calcular, y es una medida estadística de asociación lineal que cuantifica la correlación de datos no paramétricos (Dagnino, 2014). Para estimar este coeficiente se deben considerar una variable en la cual se asigna un 1 al valor más pequeño, 2 al siguiente valor que le sigue y así sucesivamente hasta llegar al valor más alto. Luego se repite el procedimiento para la otra variable. Finalmente estos datos son ingresados a la fórmula de Spearman y obtenemos un valor comprendido entre -1 y 1 y cuya interpretación del coeficiente de Spearman concuerda en valores próximos a 1 indican una correlación fuerte y positiva, valores próximos a -1 indican

una correlación fuerte y negativa y valores próximos a cero indican que no hay correlación lineal entre nuestro instrumento a validar con el estándar de oro (Ortega et al., 2009).

El EDePAM es un instrumento diagnóstico único, que se puede comparar a una ficha clínica dental, sin embargo, ninguna de estas está basada en categorías de riesgo como lo hace el EDePAM, por lo cual no se contempla el estudio de la validación concurrente en este instrumento. Sin embargo, podría predecir los cambios en el estado de salud bucal de los pacientes y la evolución de la enfermedad por lo que la validez predictiva sí es factible. El método para la validación predictiva es propio para cada instrumento ya que no todos miden el mismo aspecto y en relación con el EDePAM se puede analizar en función a la categorización de riesgo global o específica para cada ítem. Como requisito es necesario realizar el examen en dos oportunidades como mínimo, y en lo posible realizar un seguimiento en el tiempo para evidenciar pequeños cambios. La primera aplicación del examen se utiliza para diagnosticar y categorizar al paciente, mientras que la segunda medición se debe realizar con un desfase de tiempo en el que existan cambios en la evolución de la enfermedad, para comparar estos con la ficha inicial y evaluar si la categorización de riesgo descrita al inicio del estudio predijo la evolución del paciente.

Para realizar la validación de constructo, el análisis factorial exploratorio fue la medida más utilizada. Este análisis estadístico pretende identificar el número y composición de los factores comunes o indicadores que miden el constructo del cuestionario. Seguidamente, estos indicadores son analizados para buscar patrones de relación entre ellos, agrupándolos para definir finalmente el constructo y su dimensionalidad, si es uni o multidimensional (Perez & Urbano, 2014).

Fase 5: Fiabilidad del instrumento

Para evaluar la consistencia interna del EDePAM se debe utilizar el coeficiente de Cronbach, ya que este coeficiente fue el más ampliamente descrito en la literatura para evaluar la consistencia de instrumentos diagnósticos. Este coeficiente es el promedio de las correlaciones entre los ítems del instrumento lo que se interpreta

como la medida en la cual algún constructo, concepto o factor medido está presente en cada ítem. Generalmente, un grupo de ítems que explora un factor común muestra un elevado valor de alfa de Cronbach. Una manera de mejorar el valor del coeficiente de consistencia interna cuando las escalas tienen varios dominios es estratificarla en subescalas dependiendo del contenido (Celina & Campos, 2016).

Para evaluar la estabilidad del EDePAM se debe realizar el método de prueba y re prueba, que es el mecanismo más ampliamente utilizado para este tipo de exámenes. Prieto et al., (2000) en su trabajo menciona que, para cuantificar la fiabilidad de las mediciones asociadas a las variables cuantitativas continuas, el índice estadístico que se debe utilizar es el coeficiente de correlación intraclase (Prieto et al., 2000). Este método menciona la implementación del instrumento mínimo en dos oportunidades con una diferencia de tiempo que no permita a los usuarios recordar sus respuestas, pero sin que existan cambios en el estado de salud significativos y que sean aplicados en condiciones idénticas (Lockhart et al., 2013). Sin embargo, la manera correcta de abordar el análisis depende de la naturaleza de los datos y, si éstos son de tipo continuo, es habitual la utilización de estimadores del coeficiente de correlación intraclase, mientras que cuando se trata de datos de tipo categórico, el estadístico más empleado es el índice kappa (López et al., 2001)

La estrategia de informar a los pacientes que se realizaría una segunda prueba al final de ser aplicada y que el tiempo de diferencia entre estas sea de una semana mostró resultados positivos en cuanto a disminuir el efecto de arrastre de las respuestas de la primera prueba a la segunda (Noll et al., 2013). En relación con la categorización de riesgo del paciente, éste no varía de una semana a otra por lo que no se esperaría cambios en los resultados al realizar una nueva prueba transcurrido este lapso.

El propósito del análisis mediante la prueba y re prueba es asegurar que la construcción subyacente sea estable y que los resultados obtenidos en una ocasión sean similares a los obtenidos en otra ocasión. Esto se logra mediante el cálculo de coeficientes de correlación intraclase (Webber et al., 2014). Este cálculo contempla

las diferencias o concordancia entre los ítems asignando valores que oscilan entre 0 y 1 siendo este último equivalente a la concordancia absoluta. La interpretación se ha consensuado en que si el resultado es menor a 0.40 es una concordancia baja, si está entre 0.41 y 0.75 es una concordancia regular y si este valor es sobre los 0.75 es una concordancia muy buena (Enderlein, 2007)

Un instrumento es preciso cuando sus resultados son consistentes cuando se aplica más de una vez al mismo individuo por distintos examinadores bajo las mismas circunstancias. Al hablar de confiabilidad inter observador los resultados arrojaron que es una característica poco evaluada por los investigadores para sus cuestionarios. Este depende de la cantidad de investigadores calificados para aplicar la encuesta (que mínimo deben ser dos) y que sea aplicada en las mismas condiciones, ya que si esto no se cumple, no es posible cumplir con este tipo de fiabilidad. Se debe tener en cuenta que los examinadores deben estar calibrados y capacitados para evitar el error del examinador. Los valores obtenidos deben ser analizados mediante el coeficiente kappa. Este coeficiente relaciona el acuerdo que exhiben los observadores (Posadas Yábar et al., 2013),

La secuencia de desarrollo y validación de los instrumentos diagnósticos queda ejemplificada en la Figura 4.

Al momento de desarrollar los análisis estadísticos los investigadores mencionaron la utilización de softwares para la simplificación y reducción del trabajo, mencionando diferentes programas computacionales dentro del que se destaca el paquete estadístico para ciencias sociales o conocido por sus siglas en inglés SPSS en sus distintas versiones. Este es un programa de pago que, si bien ofrece una prueba gratuita, el pago de la suscripción es un factor a considerar dentro del presupuesto de la investigación. Es así como también se mencionaron otros software de pago como es Microsoft Office, Stata, LISREL y SAS. Sin embargo, se mencionaron la utilización de software de código abierto dentro de los que se destacan el proyecto R para estadística computacional, Win Episcope y la utilización de hojas de cálculos como Google sheets u OpenOffice.

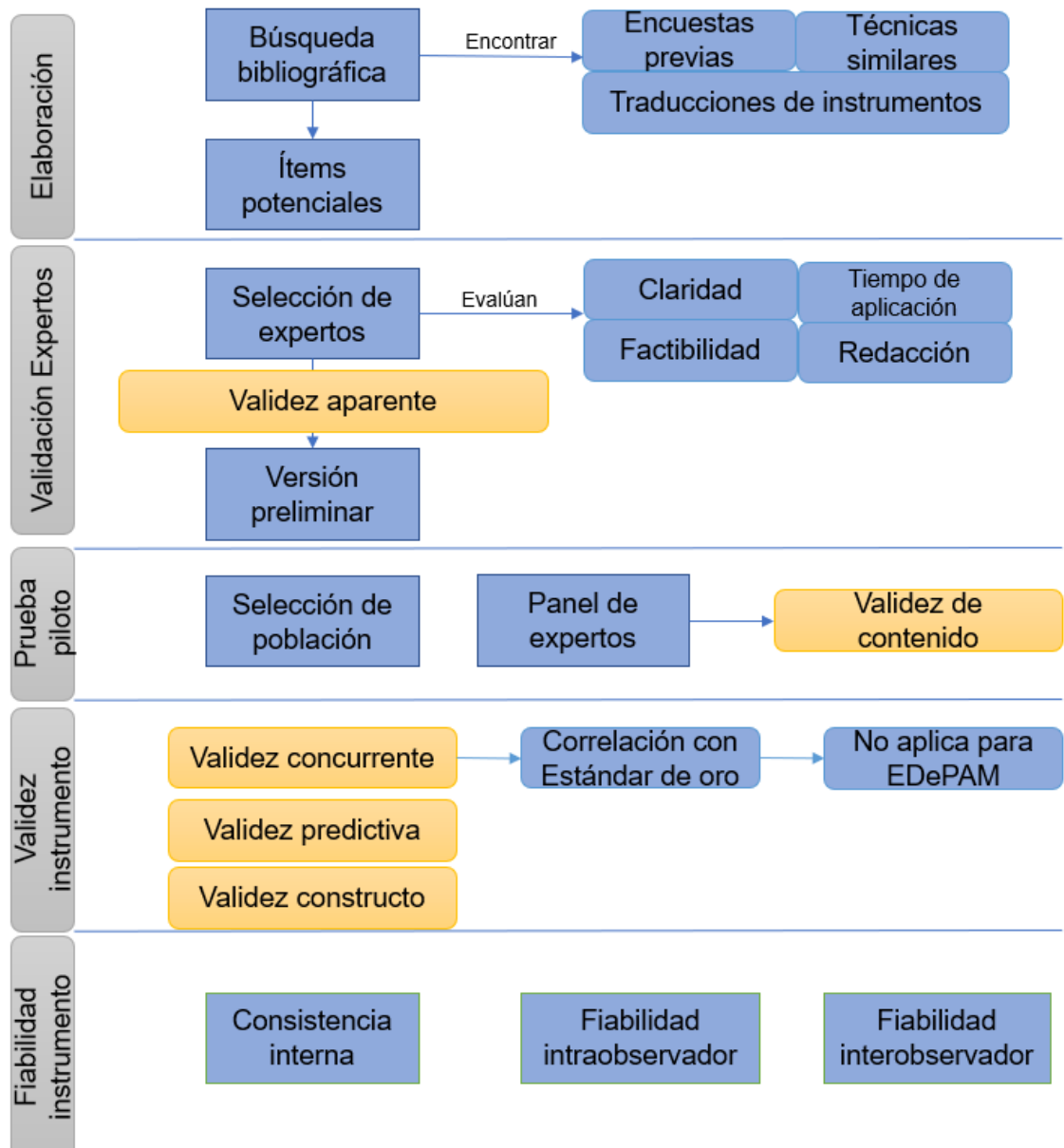


Figura 4. Secuencia y validación de instrumentos diagnósticos.

Limitaciones de la revisión

Las metodologías de validación y fiabilidad en los instrumentos en varias oportunidades estaban mencionadas de forma parcial o no claramente explicadas, muchas de ellas no mencionaban los criterios de selección de los expertos o el tamaño de la muestra.

Conclusión

La salud oral tiene un alto impacto en la calidad de vida de las personas mayores, con diversas consecuencias sobre la salud en general, asociándose a la mayoría de las patologías crónicas y con deterioro cognitivo, con altos costos económicos y sociales. Medidas de intervención focalizadas en la población de alto riesgo y la vigilancia e implementación de actividades preventivas en las poblaciones de menor riesgo permiten un uso eficiente de los recursos.

Los cuestionarios pueden ser muy útiles como elementos de apoyo al diagnóstico clínico e instrumentos para medir la gravedad de la sintomatología, por ello, es imprescindible que estén correctamente adaptadas al entorno cultural en que serán utilizadas y dispongan de unas propiedades psicométricas consistentes. El hecho de que estas encuestas tengan con frecuencia un excesivo número de ítems y, especialmente, que estén planteados con múltiples posibilidades de respuesta, dificulta su aplicación y justifica los intentos para su simplificación y validación.

Los instrumentos diagnósticos utilizados en esta revisión miden diferentes características o aspectos, por lo que la validación de un instrumento puede no ser aplicado para otro. El EDePAM es un instrumento diagnóstico único en su área y la metodología de validación de otros instrumentos debe ser utilizada con criterio.

El EDePAM es un instrumento diagnóstico basado en la categorización del paciente según su condición bucodental, ofreciendo opciones de tratamiento que contempla una fase preventiva, una fase de rehabilitación y fase de seguimiento, procesos que cambiarán el nivel de riesgo del paciente y, por lo tanto, contribuirán a mejorar su calidad de vida. Pero este instrumento se debe validar siguiendo las recomendaciones descritas en esta revisión sistemática para su aplicación en las personas mayores y así otorgar la mejor calidad de atención en base a la mayor evidencia disponible abarcando la cantidad de contenido adecuado y suficiente, cumpliendo con el objetivo que fue construido.

Financiamiento

La revisión sistemática se encuentra adscrita al proyecto Examen Dental Preventivo Adulto Mayor (EDePAM) con Control del Riesgo y de las Complicaciones de las Enfermedades Crónicas No Transmisibles: Mejorando la calidad de vida del adulto mayor chileno financiado por el concurso idea FONDEF/CONICYT 2018 con relación a los proyectos aprobados y beneficios otorgados a la línea de financiamiento del adulto mayor. El folio del proyecto es ID18I10034 y fue otorgado a la Universidad de Chile como institución principal.

Bibliografía

Antonio Pérez-Gil, J., Moscoso, S. C., & Moreno Rodríguez, R. (2000). Validez de constructo: el uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. In *Psicothema* (Vol. 12, Issue 2).

Argibay, J. C. (2006). Técnicas psicométricas. Cuestiones de validez y confiabilidad. *Subjetividad y Procesos Cognitivos*.

Backhoff, E., Aguilar-Villalobos, J., & Larrazolo-Reyna, N. (2006). Metodología para la validación de contenidos de exámenes normativos. *Revista Mexicana de Psicología*, 23(1), 79–86. <http://www.redalyc.org/pdf/2430/243020646010.pdf>

Bejia, I., Abid, N., BenSalem, K., Touzi, M., & Bergaoui, N. (2006). Reproducibility of a low back pain questionnaire in Tunisian adolescents. *Clinical Rheumatology*, 25(5), 715–720. <https://doi.org/10.1007/s10067-005-0150-9>

Biagioli, V., M., P., O., A., M.C., T., A., P., F., M., M.R., M., S., S., E., D. G., E., P., M., C., S., C., P., M., L., O., S., S., M., C., L., C., F., B., R., A., ... G., G. (2019). Development and initial validation of a questionnaire to assess patients' perception of protective isolation following haematopoietic stem cell transplantation. *European Journal of Cancer Care*, 28(2), e12955. <https://doi.org/10.1111/ecc.12955>

Celina, H., & Campos, A. (2016). Aproximación al uso del coeficiente alfa de Cronbach. *Revista Colombiana de Psiquiatría*, 34(4), 572–580.

<http://www.redalyc.org/pdf/806/80634409.pdf>
<http://www.redalyc.org/pdf/806/80650839004.pdf>

Cobos, J., J., T., N., S., & F., B. (2014). Development and validation of the scale to assess satisfaction with medications for addiction treatment-methadone for heroin addiction (SASMAT-METHER). *Drug and Alcohol Dependence*, 142, 79–85. <https://doi.org/10.1016/j.drugalcdep.2014.05.024>

Crane, M. A., Douglass, J. A., Goeman, D. P., Cousens, N. E., & Jenkins, C. R. (2012). Development and validation of the Patient Asthma Concerns Tool (PACT) to identify the needs of older people with asthma. *Respiratory Medicine*, 106(11), 1501–1508. <https://doi.org/10.1016/j.rmed.2012.07.004>

Dagnino, J. (2014). Correlación. *Revista Chilena de Anestesia*, 43(2), 150–153.

Delclos, G. ., A.A., A., L., A., A., C., D., L., C., L., T., S., E., S., L.W., W., F.G., B., & J.M., A. (2006). Validation of an asthma questionnaire for use in healthcare workers. *Occupational and Environmental Medicine*, 63(3), 173–179. <https://doi.org/10.1136/oem.2005.021634>

Enderlein, G. (2007). Fleiss, J. L.: The Design and Analysis of Clinical Experiments. Wiley, New York - Chichester - Brisbane - Toronto - Singapore 1986, 432 S., £38.35. *Biometrical Journal*, 30(3), 304–304. <https://doi.org/10.1002/bimj.4710300308>

Eslava-Schmalbach, J., Gaitán-Duarte, H., & Gómez-Restrepo, C. (2006). [A scale for measuring the quality of post-anaesthetic recovery from the user's point of view]. *Revista de salud publica (Bogota, Colombia)*, 8(1), 52–62. <https://doi.org/10.1590/s0124-00642006000100005>

Finstad, K. (2010). Response interpolation and scale sensitivity: evidence against 5-point scales. *Undefined*.

Gau, B.-S., & Hung, C.-C. (2014). The self-efficacy scale for preschool teachers regarding asthma care: instrument development and validation. *The Journal of*

School Health, 84(2), 91–98. <https://doi.org/10.1111/josh.12122>

Gómez-Benito, J., & Hidalgo, M. D. (2005). La validez de los tests, escalas y cuestionarios. *La Sociología En Sus Escenarios*, 8(12), 1–14. <http://aprendeenlinea.udea.edu.co/revistas/index.php/ceo/article/view/1750/1370>

Granado de la Orden, S., C.R., R., M.D.C.O., L., A.C., G., D.V., E., & P.R., P. (2007). Design and validation of a questionnaire to evaluate satisfaction of patients of the hospital outpatient clinics in Madrid, Spain, 2006. *Revista Espanola de Salud Publica*, 81(6), 637–645.

Jara, G., Baeza, M., & Mayo, G. (2019). *Situación atención dental de adultos mayores en Chile*.

Kaczmarek, W., Staszczuk, P., Chlebosz, M., & Kaczmarczyk, J. (2014). Hip-specific quality of life survey for patients with hip osteoarthritis - development and preliminary validity assessment. *Polish Orthopedics and Traumatology*, 79, 10–18.

Kowalski, C., Lee, S.-Y. D., Schmidt, A., Wesselmann, S., Wirtz, M. A., Pfaff, H., & Ernstmann, N. (2015a). The health literate health care organization 10 item questionnaire (HLHO-10): development and validation. *BMC Health Services Research*, 15, 47. <https://doi.org/10.1186/s12913-015-0707-5>

Kowalski, C., Lee, S. Y. D., Schmidt, A., Wesselmann, S., Wirtz, M. A., Pfaff, H., & Ernstmann, N. (2015b). The health literate health care organization 10 item questionnaire (HLHO-10): Development and validation. *BMC Health Services Research*, 15(1). <https://doi.org/10.1186/s12913-015-0707-5>

Liang, S.-Y., Wu, W.-W., Kuo, C.-Y., & Lu, Y.-Y. (2015). Development and Preliminary Evaluation of Psychometric Properties of Symptom-Management Self-Efficacy Scale for Breast Cancer Related to Chemotherapy. *Asian Nursing Research*, 9(4), 312–317. <https://doi.org/10.1016/j.anr.2015.09.001>

Liang, S., Wu, W., Kuo, C. Y., & Lu, Y. Y. (2015). Development and Preliminary

Evaluation of Psychometric Properties of Symptom-Management Self-Efficacy Scale for Breast Cancer Related to Chemotherapy. *Asian Nursing Research*, 9(4), 312–317. <https://doi.org/10.1016/j.anr.2015.09.001>

Liem, A., Newcombe, P. A., & Pohlman, A. (2017). Evaluation of complementary-alternative medicine (CAM) questionnaire development for Indonesian clinical psychologists: A pilot study. *Complementary Therapies in Medicine*, 33, 14–19. <https://doi.org/10.1016/j.ctim.2017.05.003>

Lockhart, J., Gray, N. M., & Cruickshank, M. E. (2013). The development and evaluation of a questionnaire to assess the impact of vulval intraepithelial neoplasia: a questionnaire study. *BJOG: An International Journal of Obstetrics and Gynaecology*, 120(9), 1133–1142. <https://doi.org/10.1111/1471-0528.12229>

López-Torres, J. D., Galdón-Blesa, M. P., Fernández-Olano, C., Escobar-Rabadán, F., Montoya-Fernández, J., Boix-Gras, C., Montes-Lozano, M. J., Blas-Hernández, L. V., Cremades-Romero, M. L., Del Campo-Del Campo, J. M., & Arnalich-Fernández, F. (2005). [Design and validation of a questionnaire for the detection of major depression in elderly patients]. *Gaceta sanitaria*, 19(2), 103–112. <https://doi.org/10.1157/13074364>

López, A., Galparsoro, D. U., & Fernández, P. (2001). Medidas de concordancia : el índice de Kappa. *Cad Aten Primaria*, 2–6.

Luján-Tangarife, J. A., & Cardona-Arias, J. A. (2015). Construcción y validación de escalas de medición en salud: Revisión de propiedades psicométricas. *Archivos de Medicina*, 11(3), 1–10. <https://doi.org/10.3823/1251>

Malouf, T., Langdon, R., & Taylor, A. (2014). The insight interview: a new tool for measuring deficits in awareness after traumatic brain injury. *Brain Injury*, 28(12), 1523–1541. <https://doi.org/10.3109/02699052.2014.922700>

McAllister, M., & Flynn, T. (2016). The Capabilities of Nurse Educators (CONE) questionnaire: Development and evaluation. *Nurse Education Today*, 39, 122–127.

<https://doi.org/10.1016/j.nedt.2016.01.022>

McMurray, J., H., M., A., G., J., E., & P., S. (2019). Building a Rehabilitative Care Measurement Instrument to Improve the Patient Experience. *Archives of Physical Medicine and Rehabilitation*, 100(1), 39–44.

<https://doi.org/10.1016/j.apmr.2018.05.022>

McMurray, J., McNeil, H., Lafortune, C., Black, S., Prorok, J., & Stolee, P. (2018). Building a Rehabilitative Care Measurement Instrument to Improve the Patient Experience. *Archives of Physical Medicine and Rehabilitation*, 97(1), 39–44.

<https://doi.org/10.1016/j.apmr.2015.08.408>

Medina-Mirapeix, F., del Baño-Aledo, M. E., Martínez-Payá, J. J., Lillo-Navarro, M. C., & Escolar-Reina, P. (2015). Development and validity of the questionnaire of patients' experiences in postacute outpatient physical therapy settings. *Physical Therapy*, 95(5), 767–777. <https://doi.org/10.2522/ptj.20140041>

Nadler, J. T., Weston, R., & Voyles, E. C. (2015). Stuck in the middle: The use and interpretation of mid-points in items on questionnaires. *Journal of General Psychology*, 142(2), 71–89. <https://doi.org/10.1080/00221309.2014.994590>

Noll, M., Tarragô Candotti, C., Vieira, A., & Fagundes Loss, J. (2013). Back pain and body posture evaluation instrument (BackPEI): development, content validation and reproducibility. *International Journal of Public Health*, 58(4), 565–572. <https://doi.org/10.1007/s00038-012-0434-1>

Ortega, R. M. M., Pendás, L. C. T., Ortega, M. M., Abreu, A. P., & Cánovas, A. M. (2009). El coeficiente de correlacion de los rangos de spearman caracterizacion. *Revista Habanera de Ciencias Medicas*, 8(2).

Oviedo, H. C., & Campo-Arias, A. (2005). *Metodología de investigación y lectura crítica de estudios Aproximación al uso del coeficiente alfa de Cronbach Title: An Approach to the Use of Cronbach's Alfa.*

Parra-Anguita, L., Sánchez-García, I., Del Pino-Casado, R., & Pancorbo-Hidalgo, P. (2019). Measuring knowledge of Alzheimer's: development and psychometric testing of the UJA Alzheimer's Care Scale. *BMC Geriatrics*, 19(1), 63. <https://doi.org/10.1186/s12877-019-1086-2>

Peña-Casanova, A. S.-M. J. (2006). *Test-retest and interrater reliability of Barcelona Test*.

Perez, F., & Urbano, L.-S. (2014). *Vista de El análisis factorial exploratorio de los ítems: algunas consideraciones adicionales*. <https://revistas.um.es/analesps/article/view/analesps.30.3.199991/165451>

Posadas Yábar, D., Pérez Muñante, M., Sosa Valle, H., & Ticse Aguirre, R. (2013). INVESTIGACIÓN ORIGINAL / ORIGINAL RESEARCH Variabilidad intra e inter examinador de la medición del índice tobillo braquial palpatorio utilizando un método estandarizado Inter and intra observer variability of the ankle-brachial index measured by palpation. *Rev Med Hered*, 24, 199–203. <http://www.scielo.org.pe/pdf/rmh/v24n3/v24n3ao3.pdf>

Prieto, L., Lamarca, R., & Casado, A. I. (2000). La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclass., 43(March), 1–9. <https://doi.org/10.1037//0033-2909.126.1.78>

Quezada Peña, C., Muñoz Ávila, D., Cueto Urbina, A., & Barraza Sáez, A. (2013). Reforma de Salud en Chile: Evaluación Garantía Explícita en Salud Oral en Población de 60 Años. *International Journal of Odontostomatology*, 7(2), 319–326. <https://doi.org/10.4067/s0718-381x2013000200024>

Raña-Rocha, R., López-de-Ullibarri, I., Movilla-Fernández, M.-J., & Carvajal, C. C. (2020). Validation of a questionnaire of knowledge and attitudes about the subcutaneous venous reservoir in nursing TT - Validação de um questionário de conhecimentos e atitudes sobre o cateter venoso totalmente implantado em Enfermagem TT - Validación de un cu. *Revista Latino-Americana de Enfermagem*,

28. <https://doi.org/10.1590/1518-8345.3255.3250>

Sánchez Fernández, P., Aguilar de Armas, I., Fuentelsaz Gallego, C., Teresa Moreno Casbas, M., & Hidalgo García, R. (2005). Fiabilidad de los instrumentos de medición en ciencias de la salud. *Enfermería Clínica*, 15(4), 227–236. [https://doi.org/10.1016/S1130-8621\(05\)71116-5](https://doi.org/10.1016/S1130-8621(05)71116-5)

Storch, E. A., Muroff, J., Lewin, A. B., Geller, D., Ross, A., McCarthy, K., Morgan, J., Murphy, T. K., Frost, R., & Steketee, G. (2011). Development and preliminary psychometric evaluation of the Children's Saving Inventory. *Child Psychiatry and Human Development*, 42(2), 166–182. <https://doi.org/10.1007/s10578-010-0207-0>

Thomason, S. S., Luther, S. L., Powell-Cope, G. M., Harrow, J. J., & Palacios, P. (2014). Validity and reliability of a pressure ulcer monitoring tool for persons with spinal cord impairment. *The Journal of Spinal Cord Medicine*, 37(3), 317–327. <https://doi.org/10.1179/2045772313Y.0000000163>

Tong, M. C. F., Lo, P. S. Y., Wong, K. H., Yeung, R. M. W., van Hasselt, C. A., Eremenco, S., & Cella, D. (2009). Development and validation of the functional assessment of cancer therapy nasopharyngeal cancer subscale. *Head & Neck*, 31(6), 738–747. <https://doi.org/10.1002/hed.21023>

Tristán-López, A. (2008). *MODIFICACIÓN AL MODELO DE LAWSHE PARA EL DICTAMEN CUANTITATIVO DE LA VALIDEZ DE CONTENIDO DE UN INSTRUMENTO OBJETIVO* (Vol. 6).

Turner, D., Otley, A. R., Mack, D., Hyams, J., de Bruijne, J., Uusoue, K., Walters, T. D., Zachos, M., Mamula, P., Beaton, D. E., Steinhart, A. H., & Griffiths, A. M. (2007). Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. *Gastroenterology*, 133(2), 423–432. <https://doi.org/10.1053/j.gastro.2007.05.029>

Uezato, A., Toyofuku, A., Umezaki, Y., Watanabe, M., Toriihara, A., Tomita, M., Yamamoto, N., Kurumaji, A., & Nishikawa, T. (2014). Oral Dysesthesia Rating Scale:

a tool for assessing psychosomatic symptoms in oral regions. *BMC Psychiatry*, 14, 1696. <https://doi.org/10.1186/s12888-014-0359-8>

Urrutia, G., & Bonfill, X. (2010). Declaración PRISMA. In *Medicina Clínica* (Vol. 135, Issue 11, pp. 507–511). http://es.cochrane.org/sites/es.cochrane.org/files/public/uploads/PRISMA_Spanish.pdf

Webber, K., Davies, A. N., Zeppetella, G., & Cowie, M. R. (2014). Development and validation of the breakthrough pain assessment tool (BAT) in cancer patients. *Journal of Pain and Symptom Management*, 48(4), 619–631. <https://doi.org/10.1016/j.jpainsymman.2013.10.026>

Zhao, Q., Yang, L., Zuo, Q., Zhu, X., Zhang, X., Wu, Y., Yang, L., Gao, W., & Li, M. (2014). Instrument development and validation of the stroke pre-hospital delay behavior intention scale in a Chinese urban population. *Health and Quality of Life Outcomes*, 12, 170. <https://doi.org/10.1186/s12955-014-0170-8>

Zhao, Y., Chen, R., Wang, Y., Qing, C., Wang, W., & Yang, Y. (2017). In Vitro and in Vivo Efficacy Studies of Lavender *angustifolia* Essential Oil and Its Active Constituents on the Proliferation of Human Prostate Cancer. *Integrative Cancer Therapies*, 16(2), 215–226. <https://doi.org/10.1177/1534735416645408>

Zulay, F., & Rayo, P. (2018). *Validez aparente, de contenido y confiabilidad Inicial del Instrumento CICAA para medir la comunicación entre el profesional de salud y el paciente.*

Anexos

Anexo 1: Lista de verificación PRISMA

Sección / tema	#	Elementos de lista de verificación	Reportado en la página #
Título			
Título	1	Identifique el informe como una revisión sistemática, metaanálisis o ambos.	
RESUMEN			
Resumen estructurado	2	Proporcione un resumen estructurado que incluya, según corresponda: antecedentes; objetivos; fuentes de datos; criterios de elegibilidad estudiados, participantes e intervenciones; estudio de evaluación y métodos de síntesis; resultados; limitaciones; conclusiones e implicaciones de los hallazgos clave; Número de registro de revisión sistemática.	
INTRODUCCION			
Razón fundamental	3	Describa los fundamentos de la revisión en el contexto de lo que ya se conoce.	
Objetivos	4	Proporcione una declaración explícita de las preguntas que se abordan con referencia a los participantes, las intervenciones, las comparaciones, los resultados y el diseño del estudio (PICOS).	
METODOS			
Protocolo y registro	5	Indique si existe un protocolo de revisión, si es así dónde se puede acceder (por ejemplo, dirección web) y, si está disponible, proporcione información de registro, incluido el número de registro.	
Criterio de elegibilidad	6	Especifique las características del estudio (p. Ej., PICOS, duración del seguimiento) y las características del informe (p. Ej., Años considerados, idioma, estado de publicación) utilizados como criterios de elegibilidad, dando una justificación.	
Fuentes de información	7	Describe todas las fuentes de información (por ejemplo, bases de datos con fechas de cobertura, contacto con los autores del estudio para identificar estudios adicionales) en la búsqueda y la fecha de la última búsqueda.	
Búsqueda	8	Presenta una estrategia de búsqueda electrónica completa para al menos una base de datos, incluidos los límites utilizados, de modo que pueda repetirse.	
Selección de estudio	9	Indique el proceso para seleccionar estudios (es decir, detección, elegibilidad, incluido en la revisión sistemática y, si corresponde, incluido en el metaanálisis).	
proceso de recolección de datos	10	Describa el método de extracción de datos de los informes (por ejemplo, formularios piloto, de forma independiente, por duplicado) y cualquier proceso para obtener y confirmar datos de	

		los investigadores.	
Elementos de datos	11	Enumere y defina todas las variables para las que se buscaron datos (por ejemplo, PICOS, fuentes de financiamiento) y cualquier suposición y simplificación realizada.	
Riesgo de sesgo en estudios individuales.	12	Describa los métodos utilizados para evaluar el riesgo de sesgo de los estudios individuales (incluida la especificación de si esto se hizo a nivel de estudio o de resultado) y cómo se utilizará esta información en cualquier síntesis de datos.	
Resumen de medidas	13	Indique las principales medidas de resumen (por ejemplo, razón de riesgo, diferencia de medias).	
Síntesis de resultados.	14	Describa los métodos para manejar los datos y combinar los resultados de los estudios, si se realizan, incluidas las medidas de consistencia para cada metaanálisis.	
Riesgo de sesgo entre los estudios.	15	Especifique cualquier evaluación del riesgo de sesgo que pueda afectar la evidencia acumulativa (por ejemplo, sesgo de publicación, informe selectivo dentro de los estudios).	
Análisis adicionales	16	Describa los métodos de análisis adicionales (p. Ej., Análisis de sensibilidad o de subgrupos, metarregresión), si se realizó, indicando cuáles fueron especificados previamente.	
RESULTS			
Selección de estudio	17	Proporcione un número de estudios seleccionados, evaluados para la elegibilidad e incluidos en la revisión, con razones de exclusiones en cada etapa, idealmente con un diagrama de flujo.	
Características del estudio	18	Para cada estudio, presente las características para las cuales se extrajeron los datos (por ejemplo, tamaño del estudio, PICOS, período de seguimiento) y proporcione las citas.	
Riesgo de sesgo en los estudios.	19	Presente los datos sobre el riesgo de sesgo de cada estudio y, si está disponible, cualquier evaluación del nivel de resultado (ver ítem 12).	
Resultados de estudios individuales	20	Para todos los resultados considerados (beneficios o daños), presente, para cada estudio: (a) datos resumidos simples para cada grupo de intervención (b) estimaciones del efecto e intervalos de confianza, idealmente con un diagrama de bosque.	
Síntesis de resultados.	21	Presente los resultados de cada metaanálisis realizado, incluidos los intervalos de confianza y las medidas de consistencia.	
Riesgo de sesgo entre los estudios.	22	Presente los resultados de cualquier evaluación del riesgo de sesgo entre los estudios (ver el ítem 15).	
Análisis adicional	23	Proporcione resultados de análisis adicionales, si se realizan (p. Ej., Análisis de sensibilidad o de subgrupos, metarregresión [ver ítem 16]).	
DISCUSION			
Resumen de evidencia	24	Resumir los principales hallazgos, incluida la fuerza de la evidencia para cada resultado principal; considere su relevancia para grupos clave (por ejemplo, proveedores de atención médica, usuarios y formuladores de políticas).	

Limitaciones	25	Discuta las limitaciones a nivel de estudio y de resultado (por ejemplo, riesgo de sesgo) y a nivel de revisión (por ejemplo, recuperación incompleta de investigación identificada, sesgo de informe).	
Conclusiones	26	Proporcione una interpretación general de los resultados en el contexto de otra evidencia, y las implicaciones para futuras investigaciones.	
Financiamiento			
financiamiento	27	Describa las fuentes de financiamiento para la revisión sistemática y otro tipo de apoyo (por ejemplo, suministro de datos); papel de los financiadores para la revisión sistemática.	

Anexo 2: Descripción de los términos MeSH utilizados

Health Care Evaluation Mechanisms	Validation Study	Evaluation Study	Reproducibility of Results	Surveys and Questionnaires
<p>Métodos y técnicas utilizados en la evaluación de la calidad de la atención médica, su planificación y prestación.</p> <p>Año introducido: 1998</p>	<p>Trabajos consistentes en investigaciones que utilizan procesos mediante los cuales se establece la confiabilidad y relevancia de un procedimiento para un propósito específico.</p> <p>Año introducido: 2008</p>	<p>Trabajos consistentes en estudios que determinan la efectividad o utilidad de los procesos, el personal y el equipo.</p> <p>Año introducido: 2008</p>	<p>La reproducibilidad estadística de las mediciones (a menudo en un contexto clínico), incluida la prueba de instrumentación o técnicas para obtener resultados reproducibles. El concepto incluye la reproducibilidad de las mediciones fisiológicas, que pueden usarse para desarrollar reglas para evaluar la probabilidad o el pronóstico, o la respuesta a un estímulo; reproducibilidad de ocurrencia de una condición; y reproducibilidad de resultados experimentales.</p> <p>Año introducido: 1989</p>	<p>Colecciones de datos obtenidos de sujetos voluntarios. La información generalmente toma la forma de respuestas a preguntas o sugerencias.</p> <p>Año introducido: 2016 (1973)</p>

Anexo 3: Lista de verificación de calidad de los artículos

Nombre del artículo	Introducción	Metodología	Definición del cuestionario	Validez y fiabilidad	Porcentaje de ítems evaluados
A scale for measuring the quality of post-anaesthetic recovery from the user's point of view	2	5	5	4	100,0
ABILHAND-Kids: a measure of manual ability in children with cerebral palsy	2	5	3	3	81,3
Back pain and body posture evaluation instrument (BackPEI)	2	5	4	4	93,8
Building a Rehabilitative Care Measurement Instrument to Improve the Patient Experience	2	4	4	0	62,5
Conception and Content Validation of a Questionnaire Relating to the Potential Need for Information of Visually Impaired Persons with Regard to Services and Contact Person	2	5	4	1	75,0
Design and validation of a questionnaire for the detection of major depression in elderly patients	2	4	5	4	93,8
Design and validation of a questionnaire to evaluate satisfaction of patients of the hospital o. Revista Española de Salud Pública, 8+(6), 637–645.	2	5	5	2	87,5
Development and initial validation of a questionnaire to assess patients' perception of protective isolation following haematopoietic stem cell transplantation	2	5	4	3	87,5
Development and Preliminary Evaluation of Psychometric Properties of Symptom-Management Self-Efficacy Scale for Breast Cancer Related to Chemotherapy.	2	5	4	3	87,5
Development and psychometric evaluation of the Psychiatric Inpatients Comfort Scale (PICS). Enfermería Clínica	2	4	5	2	81,3
Development and Validation of a Disease-Specific Questionnaire for Basal Joint Arthritis	2	4	2	3	68,8
Development and validation of a generic questionnaire for the implementation of complex medical interventions	2	5	5	3	93,8

Development and validation of the breakthrough pain assessment tool (BAT) in cancer patients	2	5	5	3	93,8
Development and validation of the BRIGHTLIGHT Survey, a patient-reported experience measure for young people with cancer	2	5	4	1	75,0
Development and validation of the functional assessment of cancer therapy nasopharyngeal cancer subscale	2	5	5	3	93,8
Development and validation of the Patient Asthma Concerns Tool (PACT) to identify the needs of older people with asthma.	2	3	4	3	75,0
Development and validity of the questionnaire of patients' experiences in postacute outpatient physical therapy settings	2	4	4	3	81,3
Development of a novel location-based assessment of sensory symptoms in cancer patients: preliminary reliability and validity assessment	1	5	4	2	75,0
Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study	2	5	4	3	87,5
Evaluation of a scale to assess physician-patient relations	2	4	4	2	75,0
Evaluation of complementary-alternative medicine (CAM) questionnaire development for Indonesian clinical psychologists: A pilot study	2	4	5	2	81,3
Evaluation of the Symptom Representation Questionnaire (SRQ) for assessing cancer-related symptoms	2	5	4	3	87,5
EVENDOL, a new behavioral pain scale for children ages - to 7 years in the emergency department: design and validation	2	5	4	3	87,5
Hip-specific quality of life survey for patients with hip osteoarthritis - development and preliminary validity assessment	2	5	3	2	75,0
Instrument development and validation of the stroke pre-hospital delay behavior intention scale in a Chinese	2	5	5	3	93,8

urban population. Health and Quality of Life Outcomes					
Measuring knowledge of Alzheimer's development and psychometric testing of the UJA Alzheimer's Care Scale.	2	5	4	3	87,5
Measuring patient-centredness, the neglected outcome in fertility care: A random multicentre validation study	2	5	4	2	81,3
Oral Dysesthesia Rating Scale: a tool for assessing psychosomatic symptoms in oral regions	2	5	4	2	81,3
Professional Care Team Burden (PCTB) scale - reliability, validity and factor analysis	2	3	5	2	75,0
Psychometric testing of the clinical nurse leader staff satisfaction instrument. Journal of Nursing Measurement	2	5	5	2	87,5
Psychometric testing of the Perception of Pregnancy Risk Questionnaire.	2	5	3	3	81,3
Questionnaire on body awareness of postural habits in young people: construction and validation TT - Questionário de percepção corporal dos hábitos posturais de jovens: construção e validação	2	5	4	3	87,5
Sleep quality, the neglected outcome variable in clinical studies focusing on locomotor system; A construct validation study	2	5	4	3	87,5
The Capabilities of Nurse Educators (CONE) questionnaire: Development and evaluation	2	5	4	3	87,5
The depression in old age scale (DIA-S)]. Zeitschrift fur Gerontologie und Geriatrie	2	5	4	2	81,3
The development and evaluation of a questionnaire to assess the impact of vulvar intraepithelial neoplasia: a questionnaire study. BJOG	2	3	4	4	81,3
The development and evaluation of a questionnaire to assess the impact of vulvar intraepithelial neoplasia: a questionnaire study. BJOG	2	5	5	2	87,5
The insight interview a new tool for measuring deficits in	2	4	5	3	87,5

awareness after traumatic brain injury					
The self-efficacy scale for preschool teachers regarding asthma care: instrument development and validation	2	5	4	2	81,3
The supportive care needs survey for partners and caregivers of cancer survivors: development and psychometric evaluation	2	5	3	3	81,3
Validation of a knowledge questionnaire for patients with cirrhosis of the liver	1	5	3	2	68,8
Validation of a questionnaire of knowledge and attitudes about the subcutaneous venous reservoir in nursing	2	5	4	3	87,5
Validity and reliability of a pressure ulcer monitoring tool for persons with spinal cord impairment	1	5	5	3	87,5
The Vulvovaginal Symptoms Questionnaire: a questionnaire for measuring vulvovaginal symptoms in postmenopausal women	2	5	4	4	93,8
Development and initial validation of the Stress of Conscience Questionnaire.	2	4	3	2	68,8
Development and validation of the scale to assess satisfaction with medications for addiction treatment-methadone for heroin addiction (SASMAT-METHER).	2	2	5	4	81,3
The EORTC Quality of Life Questionnaire for patients with colorectal cancer: EORTC QLQ-CR29 Validation Study for Spanish patients	2	4	4	1	68,8
Validation of an asthma questionnaire for use in healthcare workers. Occupational and Environmental Medicine	1	5	4	3	81,3