



# Revista Colombiana de Cancerología

[www.elsevier.es/cancerologia](http://www.elsevier.es/cancerologia)



ORIGINAL

## Consistencia interna y validez de contenido del instrumento DELBI



Fabio Sierra<sup>a,\*</sup>, Esperanza Peña<sup>a</sup>, Magda Alba<sup>a</sup> y Ricardo Sánchez<sup>a,b</sup>

<sup>a</sup> Grupo de Investigación Clínica, Instituto Nacional de Cancerología, Bogotá D. C., Colombia

<sup>b</sup> Facultad de Medicina, Universidad Nacional de Colombia, Bogotá D. C., Colombia

Recibido el 10 de diciembre de 2013; aceptado el 27 de septiembre de 2014

Disponible en Internet el 19 de diciembre de 2014

### PALABRAS CLAVE

Cuestionario;  
Guía de práctica  
clínica;  
Neoplasias de la  
mama;  
Neoplasias de la  
próstata;  
Neoplasias del colon;  
Investigación sobre  
Servicios de Salud;  
Indicadores de  
Calidad de la  
Atención de Salud

### Resumen

**Objetivos:** Evaluar la consistencia interna y la validez de contenido de la escala DELBI para evaluación de guías de práctica clínica (GPC).

**Métodos:** Fueron seleccionadas 146 GPC de cáncer de mama, colorrectal y de próstata, a las que se aplicó el instrumento DELBI. Se evaluó la consistencia interna de la escala completa y por dominios utilizando el coeficiente alfa de Cronbach y se evaluó la validez de contenido de la escala por medio de un análisis factorial exploratorio (AFE) con rotación ortogonal y uno confirmatorio (AFC) a través de ecuaciones estructurales; para el AFE se determinó que la matriz fuera factorizable por medio del test de esfericidad de Bartlett y el criterio KMO y para el AFC se utilizaron matrices policóricas y métodos de estimación asintótica, teniendo en cuenta que las opciones de respuesta son ordinales.

**Resultados:** Se observó una consistencia interna buena para la escala total (0,89), la cual varió entre 0,46 y 0,74 en los dominios. El AFE mostró una estructura de cuatro factores sugiriendo una agrupación de los ítems en cuatro dominios: aplicabilidad, claridad y presentación, rigor en la elaboración y participación de los implicados. El AFC mostró un mejor ajuste de la escala organizada en cuatro dominios que en los ocho dominios de la escala original, los cuales tuvieron mejor consistencia interna.

**Conclusiones:** Se concluye que la escala tiene propiedades psicométricas aceptables y se sugiere considerar cuadro dimensiones de calidad de las GPC en la interpretación de las evaluaciones realizadas con el DELBI.

© 2013 Instituto Nacional de Cancerología. Publicado por Elsevier España, S.L.U. Todos los derechos reservados.

\* Autor para correspondencia.

Correo electrónico: [fasierra@cancer.gov.co](mailto:fasierra@cancer.gov.co) (F. Sierra).

**KEYWORDS**

Questionnaires;  
Practice guidelines;  
Breast neoplasm;  
Prostatic neoplasms;  
Colonic neoplasms;  
Health services  
research;  
Health care quality  
indicators

**Internal consistency and content validity of DELBI scale****Abstract**

**Objectives:** To assess internal consistency and content validity of the DELBI scale used to evaluate clinical practice guidelines (CPG).

**Methods:** The DELBI scale was applied to a selection of 146 CPG for breast, colorectal and prostate cancer. Internal consistency of the full scale and by domains was assessed using Cronbach's alpha coefficient, and the content validity was assessed using an Exploratory Factor Analysis (EFA) with orthogonal rotation and a Confirmatory Factor Analysis (CFA) applying structural equation modelling. Bartlett's test of sphericity and KMO test were used to determine the factorability of the intercorrelation matrix, with the ordinal nature of the item-scores as a Likert-type scale. Estimation methods were performed using polychoric correlation, and asymptotic covariance matrices.

**Results:** For the full scale, a good internal consistency was observed (0.89), which varies between 0.46 and 0.74 by domains. A four-factor structure was found: applicability, clarity and presentation, rigour of development, and stakeholder involvement. The goodness-of-fit tests of the scale organised by 4 domains was better than the 8 domains from the original scale and had a better internal consistency.

**Conclusions:** It is concluded that the scale has acceptable psychometric properties and it is suggested to consider four dimensions of quality of CPGs in the interpretation of the evaluations performed with the DELBI scale.

© 2013 Instituto Nacional de Cancerología. Published by Elsevier España, S.L.U. All rights reserved.

**Introducción**

El cáncer tiene una importante carga de la enfermedad para Colombia. De acuerdo con los datos reportados por Globocan, la incidencia es de 160,8 por 100.000 hombres al año y 144,1 por 100.000 mujeres al año, con una mortalidad de 97,5% por 100.000 hombres al año y 81,7% por 100.000 mujeres al año, para mujeres y hombres, respectivamente<sup>1</sup>. Debido a la importancia de esta enfermedad, se han propuesto herramientas que ayuden a los profesionales de salud a realizar una detección temprana así como un tratamiento integral del cáncer. Entre estas se encuentran las guías de práctica clínica (GPC) y los protocolos de manejo.

Las GPC consisten en conjuntos de afirmaciones (recomendaciones) para asistir a médicos y pacientes en la toma de decisiones sobre atención en salud apropiada para circunstancias específicas<sup>2-4</sup>. Se han propuesto actualizar el conocimiento de los profesionales para disminuir variabilidad en la práctica clínica y hacer más racional la prestación de los servicios de salud y así mejorar la calidad al brindar la atención de forma efectiva, consistente y eficiente<sup>5,6</sup>. La elaboración de GPC basadas en la evidencia puede desarrollarse por medio de dos metodologías: desarrollo de novo o adaptación. Se ha recomendado la adaptación de GPC existentes teniendo en cuenta que el proceso requiere el uso de recursos y que la información de las GPC elaboradas en otros países puede ser aplicable en un contexto local y así se evita la duplicación de tareas y esfuerzos<sup>5</sup>.

La adaptación de GPC consiste en el uso o la modificación de una guía producida en un contexto cultural u organizacional para aplicarla en un contexto diferente<sup>7,8</sup>. Este

proceso requiere varios pasos que son propuestos por diferentes grupos desarrolladores de GPC<sup>7,9,10</sup>, y puede implicar la búsqueda, la selección y la evaluación de la calidad de GPC existentes. La evaluación se enfoca en la calidad metodológica de la guía, por lo tanto en su validez interna, es decir, se refiere a los elementos apropiados que debieron considerarse durante su elaboración, además, de la factibilidad de sus recomendaciones, su validez externa, y la capacidad para influir en el cuidado de la salud<sup>5,11,12</sup>.

Teniendo en cuenta los elementos de calidad de una GPC, se han elaborado instrumentos que faciliten una evaluación rigurosa de las mismas. Entre estos se encuentran el Instrumento para la Evaluación de Guías de Práctica Clínica (AGREE, por sus siglas en inglés, *The Appraisal of Guidelines for Research and Evaluation*) y el Instrumento Alemán para la Apreciación Metodológica de Guías de Práctica Clínica (DELBI, por sus siglas en alemán, *Deutsches Instrument zur methodischen Leitlinien-Bewertung*). El DELBI es un instrumento genérico para asistir a los desarrolladores y usuarios de GPC (profesionales de la salud, educadores, tomadores de decisiones) en la evaluación de su calidad metodológica<sup>11</sup>, publicado en junio del 2005 y enmendado en septiembre de 2008, como resultado de la cooperación entre la Asociación de Sociedades Médicas Científicas en Alemania (AWMF, por sus siglas en alemán, *Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften*), la Agencia Alemana para Calidad en Medicina (AQuMed, por sus siglas en alemán, *Ärztliches Zentrum für Qualität in der Medizin*) y un grupo de expertos<sup>13</sup>.

El DELBI está organizado en ocho dominios, seis de los cuales corresponden a los dominios del AGREE (alcance y objetivos, participación de los grupos de interés, rigor metodológico en la elaboración, claridad y presentación,

aplicabilidad e independencia editorial)<sup>12,14</sup>; los otros dos dominios son la aplicabilidad de las guías en el sistema de salud alemán y la calidad metodológica de guías cuya elaboración siguió un proceso de adaptación de GPC existentes. Este fue el instrumento propuesto por la Guía Metodológica para la elaboración de Guías de Atención Integral en el Sistema General de Seguridad Social en Salud Colombiano<sup>5</sup>, dado que tiene más dominios y, además, tiene ayudas para la evaluación como son los criterios para seleccionar cada una de las opciones de respuesta en la escala tipo Likert<sup>11</sup>.

En el momento de realización de este trabajo no había sido reportado algún estudio que evaluara propiedades psicométricas del DELBI. El conocimiento de la validez de esta herramienta es necesario en cuanto aporta elementos adicionales para decidir cuál herramienta es la mejor para evaluar la calidad de las GPC y garantiza un proceso de adaptación riguroso.

Teniendo en cuenta que este instrumento se comporta como una escala, se consideró necesario analizar su consistencia interna y su validez de contenido, como se realizó con el instrumento AGREE<sup>15-17</sup>. La consistencia interna es una medida de confiabilidad que mide el grado de correlación entre los ítems de la escala<sup>18</sup> e indica si los ítems son o no homogéneos. La validez de contenido indica si la muestra de ítems que conforman la escala es representativa del universo de ítems que pueden medir el constructo<sup>19</sup> y la calidad de la guía. Esta se evalúa por medio de análisis factorial exploratorio o confirmatorio: en el primer caso, se determina el número de factores en que pueden agruparse los ítems; en el segundo caso, se parte de la estructura previamente determinada de la escala o de la encontrada en el análisis exploratorio y se determina si cada ítem correlaciona con el factor para el que fue propuesto<sup>20</sup>.

En este estudio se analizó la consistencia interna y la validez de contenido del instrumento DELBI, con el propósito de analizar si es una herramienta confiable en la evaluación de GPC y si la estructura dominios propuesta es adecuada o si los ítems pueden agruparse en una estructura de dominios diferente.

## Metodología

### Instrumento

El instrumento DELBI consiste en 34 ítems en escala tipo Likert con cuatro opciones de respuesta (1: “fuertemente en desacuerdo”, 2: “en desacuerdo”, 3: “de acuerdo” y 4: “fuertemente de acuerdo”), organizados en 8 dominios (alcance y objetivos; participación de los implicados; rigor en la elaboración; claridad y presentación; aplicabilidad; independencia editorial; aplicabilidad al sistema de salud local (alemán en la versión original de la escala) y rigor metodológico cuando se usan GPC existentes) (tabla 1). Este instrumento puede ser aplicado a GPC de cualquier área clínica y de cualquier área de cuidado (diagnóstico, prevención, promoción, tratamiento, etc.). En este estudio no se incluyó el dominio 8 (rigor metodológico cuando se usan GPC existentes) debido a que no todas las GPC se realizaron por medio de un proceso de adaptación.

### Procedimiento

El proceso de validación consistió en la búsqueda y selección de GPC, aplicación del instrumento DELBI y evaluación de su consistencia interna y validez. Este trabajo hizo parte del desarrollo de GPC realizado en el Instituto Nacional de Cancerología E.S.E., que siguió los lineamientos de la Guía Metodológica para la elaboración de Guías de Atención Integral en el Sistema General de Seguridad Social en Salud Colombiano<sup>5</sup>.

Se incluyeron guías para cáncer de mama, próstata y colon-recto publicadas en inglés, español, francés o alemán. Estas guías fueron evaluadas por tres revisores independientes que fueron entrenados en la aplicación del instrumento DELBI; dos con formación en epidemiología clínica y uno con formación clínica (oncología, mastología, urología y gastroenterología). Posterior a su aplicación se calificó cada GPC de acuerdo con el algoritmo propuesto por el sistema DELBI<sup>11</sup>.

### Análisis estadístico

La consistencia interna de la escala fue analizada por medio del coeficiente alfa de Cronbach. Estos coeficientes fueron calculados con todos los ítems y retirando ítems para la escala completa y para cada uno de los dominios.

La validez de contenido fue evaluada por medio de análisis factorial exploratorio (AFE por sus siglas en español) y confirmatorio (AFC por sus siglas en español). Para el análisis factorial exploratorio, se determinó si la matriz de correlaciones entre ítems era factorizable, a partir del test de esfericidad de Bartlett y del criterio de Kaiser Meyer Olkin (KMO por sus siglas en inglés) y se empleó el método de factores principales y una rotación ortogonal para extraer los factores. Para escoger el número de factores se utilizó el criterio de valores propios mayores o iguales a 1 más proporción de la varianza explicada.

El AFC se realizó por medio de ecuaciones estructurales. Teniendo en cuenta que las variables del DELBI están construidas en escala tipo Likert y son de naturaleza ordinal, la estimación de los modelos se realizó a partir de la matriz de correlaciones policóricas y se utilizó el método de mínimos cuadrados ponderados, dado que no se puede asumir una distribución normal de los datos. Se evaluaron dos modelos, presentados en la figuras 1 y 2: uno asumiendo la estructura original de la escala y otro teniendo en cuenta los resultados del AFE. Los dominios fueron considerados como variables latentes y los ítems de la escala como variables observadas.

Previo al análisis de los modelos se determinó si estos eran identificables calculando el número máximo de parámetros libres<sup>21</sup> y considerando el número mínimo de variables observadas por cada dominio (mínimo tres para modelos de más de dos factores)<sup>22</sup>.

Asimismo, para garantizar la normalización de los modelos, dado que tienen variables latentes, se fijaron las varianzas de las variables latentes en 1. Se calculó la consistencia interna también para el segundo modelo (obtenido de acuerdo con los resultados del AFE).

Para decidir sobre la calidad del ajuste del modelo a los datos se utilizó la prueba  $\chi^2$ , la razón  $\chi^2/\text{grados de libertad}$ , el Índice de Ajuste Comparativo (CFI, por sus siglas en inglés,

**Tabla 1** Resultados del AFE de la escala DELBI

	Dominio – Ítem	Factor 1	Factor 2	Factor 3	Factor 4	Unicidad
<i>Alcance y objetivos</i>						
i1	El(los) objetivo(s) general(es) de la guía está(n) específicamente descrito(s)	0,28	0,59	0,13	0,09	0,55
i2	Lo(s) aspectos(s) clínico(s) cubierto(s) por la guía está(n) específicamente descrito(s)	0,43	-0,02	0,24	0,09	0,75
i3	Los pacientes a quienes se pretende aplicar la guía están específicamente descritos	0,09	0,55	0,30	0,19	0,57
<i>Participación de los implicados</i>						
i4	El grupo que desarrolla la guía incluye individuos de todos los grupos profesionales relevantes	0,39	0,35	0,00	0,48	0,50
i5	Se han tenido en cuenta los puntos de vista del paciente y sus preferencias	0,41	-0,03	0,03	0,59	0,48
i6	Los usuarios diana de la guía están claramente definidos	0,10	0,64	0,15	0,18	0,52
i7	La guía ha sido probada entre los usuarios diana	-0,16	0,23	0,37	-0,02	0,79
<i>Rigor en la elaboración</i>						
i8	Se han utilizado métodos sistemáticos para la búsqueda de la evidencia	0,09	0,03	0,83	-0,04	0,29
i9	Los criterios para seleccionar la evidencia se describen con claridad	0,13	0,06	0,76	0,17	0,37
i10	Los métodos utilizados para formular las recomendaciones están claramente descritos	0,28	0,32	0,37	0,09	0,38
i11	Al formular las recomendaciones han sido considerados los beneficios en salud, los efectos secundarios y los riesgos	0,69	0,00	0,24	0,10	0,45
i12	Hay una relación explícita entre cada una de las recomendaciones y las evidencias en las que se basan	0,27	0,55	-0,05	0,07	0,62
i13	La guía ha sido revisada por expertos externos antes de su publicación	0,19	0,29	0,43	0,34	0,58
i14	Se incluye un procedimiento para actualizar la guía	0,45	0,12	0,12	0,31	0,67
<i>Claridad y presentación</i>						
i15	Las recomendaciones son específicas y no son ambiguas	0,26	0,49	0,24	-0,12	0,62
i16	Las distintas opciones para el manejo de la enfermedad o condición se presentan claramente	0,68	0,22	0,04	0,16	0,46
i17	Las recomendaciones clave son fácilmente identificables	0,17	0,33	0,22	-0,15	0,79

Tabla 1 (continuación)

	Dominio – Ítem	Factor 1	Factor 2	Factor 3	Factor 4	Unicidad
i18	La guía se apoya con herramientas para su aplicación	0,49	0,16	0,09	0,30	0,64
<i>Aplicabilidad</i>						
i19	Se han discutido las barreras organizativas potenciales a la hora de aplicar las recomendaciones	0,55	0,17	0,14	0,03	0,65
i20	Han sido considerados los costos potenciales de la aplicación de las recomendaciones	0,56	0,17	0,23	0,05	0,60
i21	La guía ofrece una relación de criterios clave con el fin de realizar monitorización y/o auditoría	0,54	0,30	0,14	0,22	0,55
<i>Independencia editorial</i>						
i22	La guía es editorialmente independiente de la entidad financiadora	0,09	0,08	0,15	0,69	0,49
i23	Se han registrado los conflictos de intereses de los miembros del grupo de desarrollo	0,02	0,16	0,12	0,52	0,69
<i>Aplicabilidad al sistema de salud local</i>						
i24	Hay recomendaciones para medidas de prevención, diagnóstico, terapéuticas y de rehabilitación en diferentes áreas de la atención	-0,06	0,63	0,14	0,00	0,58
i25	Hay información sobre qué medidas parecen ser inadecuadas, redundantes u obsoletas	0,36	0,20	0,38	-0,16	0,66
i26	La información clínica de la guía está organizada de tal forma que el proceso de toma de decisiones clínicas es sistemático y fácilmente comprensible	0,18	0,58	0,00	0,13	0,62
i27	Una estrategia/concepto para el fácil acceso y disseminación de la guía es presentada	0,21	-0,01	0,36	-0,24	0,77
i28	Un concepto para la implementación de la guía es descrito	0,42	0,35	0,07	0,39	0,55
i29	La guía es complementada con una descripción de los métodos utilizados	0,13	0,31	0,63	0,20	0,45

*Comparative Fit Index*), el Índice de Tucker-Lewis (TLI, por sus siglas en inglés, *Tucker-Lewis Index*), el Índice de Aproximación de la Raíz de Cuadrados Medios del Error (RMSEA, por sus siglas en inglés, *Root Mean Square Error of Approximation*) y el Índice de la Raíz de los Cuadrados Medios del Residuo (SRMR, por sus siglas en inglés, *Standardized Root*

*Mean Squared Residuals*). Valor p de la prueba  $\chi^2$ /grados de libertad mayor a 0,05, razón  $\chi^2$ /grados de libertad menor a 3, valores de CFI y TLI mayores a 0,95 y valores de RMSEA y SRMR menores a 0,05, indican buen ajuste del modelo<sup>20-24</sup>.

Los análisis fueron realizados en el paquete estadístico Stata 12® (consistencia interna y AFE) y en el paquete

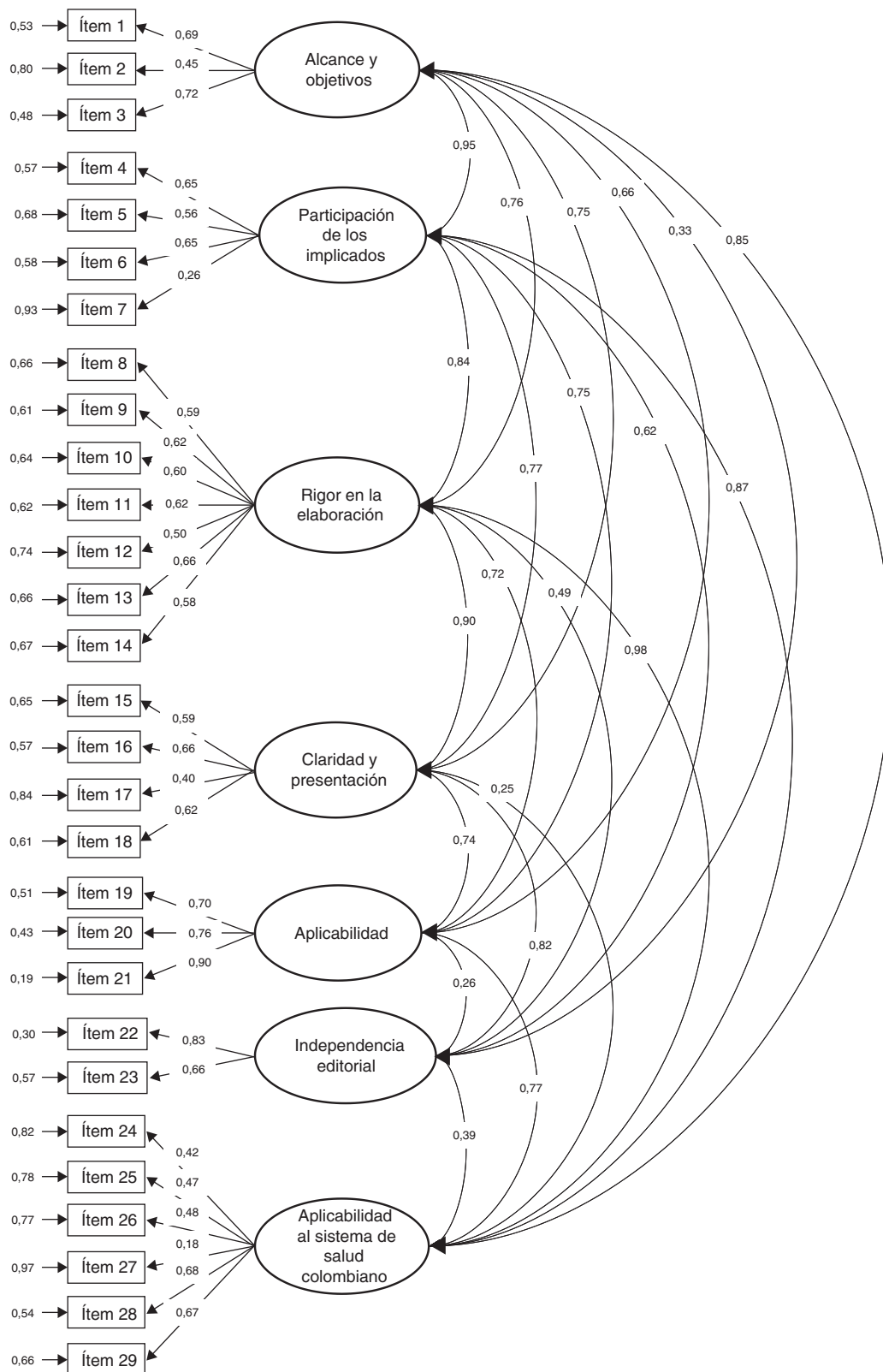


Figura 1 Modelo de Análisis Factorial Confirmatorio para la estructura original de la escala DELBI.

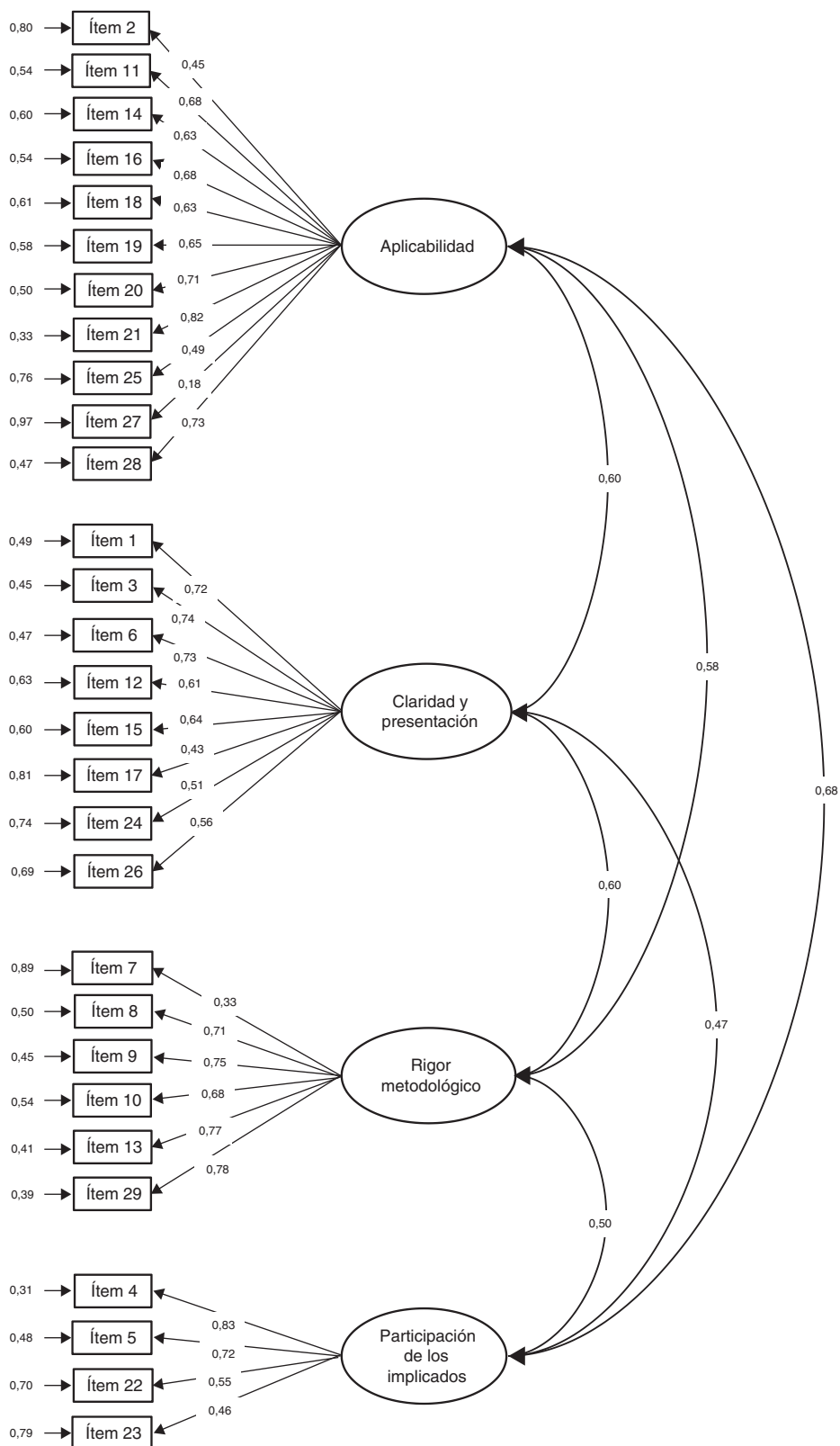


Figura 2 Modelo de Análisis Factorial Confirmatorio para la estructura de cuatro dominios de la escala DELBI.

**Tabla 2** Consistencia interna de la escala DELBI

Dominio	Alfa de Cronbach sin retirar ítems
Toda la escala	0,89
Alcance y objetivos	0,54
Participación de los implicados	0,46
Rigor en la elaboración	0,72
Claridad y presentación	0,57
Aplicabilidad	0,74
Independencia editorial	0,66
Aplicabilidad al sistema de salud local	0,62

estadístico R utilizando las librerías psych<sup>25</sup> y lavaan<sup>26,27</sup> (para el AFC).

## Resultados

Fueron seleccionadas 146 GPC: 52 (35,4%) para cáncer de mama, 14 (9,5%) para cáncer de próstata y 81 (55,1%) para cáncer de colon-recto.

La consistencia interna de la escala completa y por dominios se presenta en la [tabla 2](#). Esta es alta para la toda la escala y mejora al retirar ítems indicando que las preguntas del instrumento miden el mismo constructo en forma confiable. Sin embargo, cuando se analiza por dominios se observan menores coeficientes. Los dominios con menor consistencia mejoran el coeficiente al retirar ítems. *Alcance y objetivos* incrementa a 0,72 retirando el ítem 2: "Los aspectos clínicos cubiertos por la guía están específicamente descritos" y *Participación de los implicados* mejora a 0,60 retirando el ítem 7: "La guía ha sido probada entre los usuarios diana".

En el AFE se encontró una matriz factorizable, pues el resultado del test de esfericidad de Bartlett fue  $\chi^2_{406}=1.759,458$ ,  $p=0,000$  y el del KMO fue 0,799. Se encontró una solución de 4 factores, como se observa en la [tabla 1](#), que explica el 80,2% de la varianza (24%, 22,2%, 20% y 14%, respectivamente); los valores propios de estos factores fueron: 7,18, 1,99, 1,63 y 1,24. En esta tabla se muestran también los valores de unicidad para cada ítem y se observa que los ítems 2 y 7, cuyo retiro mejora la consistencia interna, tienen valores altos: 0,75 y 0,79.

Puede observarse que la estructura de ocho dominios de la escala original no se conserva. Considerando los aspectos propuestos para la evaluación de calidad de las GPC, la estructura de factores encontrada y el contenido de los ítems que cargan en esos factores, se sugieren los siguientes nombres para los cuatro dominios: *aplicabilidad*, *claridad y presentación*, *rigor en la elaboración* y *participación de implicados*.

Las estimaciones de los parámetros de los modelos de AFC se presentan en las [figuras 1 y 2](#). Todos los coeficientes de senda entre los ítems y sus dominios fueron estadísticamente diferentes de 0; excepto el ítem 27 en el modelo 2. En ambos modelos los ítems 7 y 27 tienen las menores cargas; al cambiar el ítem 7 al dominio de *rigor metodológico* su carga mejora, mientras que para el ítem 27 disminuye y deja de ser estadísticamente significativa. Las correlaciones entre dominios fueron estadísticamente significativas y altas; *independencia editorial* tuvo las correlaciones más bajas, excepto con *participación de los implicados*. En la [tabla 3](#) se presentan los estadísticos de ajuste para cada uno de los modelos analizados.

Puede observarse un mejor ajuste del modelo que agrupa los ítems de la escala en 4 dominios, dado que son mejores los valores de todos los índices, indicando que la estructura de la escala sugerida en el AFC es más adecuada para el instrumento y por lo tanto que la calidad metodológica de las GPC se puede analizar por medio de esos 4 aspectos. Al analizar la consistencia interna se encontraron mejores coeficientes para este modelo: 0,81 (aplicabilidad), 0,77 (claridad y presentación), 0,78 (rigor en la elaboración) y 0,71 (participación de los implicados).

## Discusión y conclusiones

Dada la importancia que ha ganado la calidad de la atención en salud, la disminución de la variabilidad en la práctica clínica y la necesidad de realizar un uso más racional de los recursos en los sistemas de salud, se ha presentado una producción creciente de GPC basadas en la evidencia. Teniendo en cuenta que la realización de una GPC implica uso de recursos y tiempo, la adaptación de las ya existentes es una forma que ayuda a evitar la duplicación de esfuerzos<sup>5</sup>. Es precisamente en este contexto en que el instrumento DELBI aporta una evaluación de calidad de esas guías, siendo un paso fundamental en el proceso. Este es un instrumento clave para el desarrollo de GPC dado que es necesario garantizar su confiabilidad y validez, que hasta la fecha de realización de este estudio no ha sido analizada y reportada.

Teniendo en cuenta lo anterior, en este estudio se analizó la consistencia interna y validez de contenido del DELBI utilizando las GPC incluidas en el desarrollo de guías para el sistema de salud colombiano. Los resultados encontrados sugieren una buena confiabilidad por consistencia interna (0,89) y una aceptable validez de contenido (CFI>0,9 y RMSEA=0,067), que puede mejorarse si los ítems de la escala se organizan en cuatro dominios. Estos resultados son similares a los encontrados para el instrumento AGREE, empleado también en la evaluación de GPC, y que comparte los primeros seis dominios con el DELBI. El grupo desarrollador del AGREE, empleando 100 GPC, calculó la consistencia

**Tabla 3** Estadísticos de ajuste para los modelos de AFC

Modelo	X2	gl	Valor p	X2/gl	CFI	TLI	RMSEA	SRMR
Uno	813,159	356	0,000	2,28	0,910	0,897	0,094	0,117
Dos	615,672	371	0,000	1,66	0,952	0,947	0,067	0,104



interna a partir del puntaje promedio de cada ítem (calificado por varios evaluadores) y encontró coeficientes entre 0,64 y 0,88<sup>15</sup>; aunque su estructura de dominios, obtenida por medio de componentes principales, conserva seis dominios, el ítem 7 ("la guía ha sido probada entre los usuarios diana") tiene también una carga baja. En el estudio de MacDermid et al., 2005<sup>28</sup>, con 55 GPC de terapia física, también fue encontrada una solución de cuatro factores, en la que también se agrupan los ítems de *independencia editorial* con ítems de *participación de los implicados*, e ítems de *aplicabilidad* con ítems de *claridad y presentación* y con un ítem de *alcance y objetivos*.

La estructura de factores encontrada en el AFE, indica que las dimensiones de calidad de las GPC pueden reducirse a cuatro, las cuales abarcan aspectos de validez externa e interna (estos cuatro factores explican el 80,2% de la varianza). En este sentido los primeros dos factores encontrados son la *aplicabilidad* y la *claridad y presentación* de las GPC, que dentro de un proceso de adaptación ayudan a establecer si estas pueden utilizarse en un nuevo contexto, por lo tanto, este resultado está a favor de la validez y utilidad del instrumento. Los dos siguientes factores son *rigor metodológico* y *participación de los implicados*, los cuales garantizan que el instrumento está bien construido y que en su elaboración están representadas las distintas áreas clínicas y pacientes y que no existe un conflicto de interés de sus miembros.

Este estudio es un primer acercamiento a las propiedades de medición del instrumento, el cual, hasta ahora empieza a usarse, como puede observarse en las GPC para el sistema de salud colombiano publicadas en el 2013<sup>29,30</sup> y en los estudios de Alba et al., 2013<sup>31</sup>, Frese et al., 2012<sup>32</sup> y Wiegele et al., 2011<sup>33</sup>. Por esta razón y teniendo en cuenta los resultados obtenidos, se requieren más investigaciones sobre el instrumento en las que se consideren los asuntos mencionados a continuación. Además de la validez de contenido, se puede realizar una validez de constructo, al comparar los resultados del DELBI con los del AGREE, actualmente más utilizado.

Dado que uno de los aspectos que diferencian al DELBI del AGREE, además del número de dominios es la escala de respuesta, se requiere analizar si el pobre ajuste del modelo de la escala original se debe a que las opciones son cuatro y a que existe una guía que indica cuando escoger cada una de ellas. En el AGREE, el evaluador manifiesta su acuerdo o desacuerdo con el criterio que evalúa, en una escala de siete puntos, pero no se sugieren aspectos que debe considerarse para escoger cada uno de esos puntos. Es posible, que la estructura más directiva del DELBI influya en la validez del instrumento. Esta hipótesis puede ser evaluada por medio de análisis de teoría de respuesta al ítem usando modelos de respuesta graduada o modelos de Rasch, los cuales permiten hacer un diagnóstico de las opciones de respuesta<sup>34-36</sup>.

Entre las limitaciones de este estudio se encuentra el tamaño de muestra y la heterogeneidad de las GPC empleadas. Los modelos de AFC requieren muestras grandes, más aún cuando los ítems incluidos son de naturaleza ordinal<sup>22,37</sup>. Teniendo en cuenta este tamaño de muestra pequeño se utilizó como método de estimación el método de mínimos cuadrados ponderados. Sin embargo, es posible que el pobre ajuste del modelo uno (estructura original de la escala) se relacione con este tamaño. En relación con

la heterogeneidad, las GPC empleadas fueron oncológicas. Un análisis que incluya otro tipo de condiciones clínicas puede aportar una mayor variabilidad en los dominios evaluados y por lo tanto permitir análisis más robustos.

Se puede concluir que el instrumento DELBI tiene buena consistencia interna y que es válido para evaluar la calidad de GPC. Dada la estructura de factores encontrada se sugiere realizar una interpretación de los resultados de acuerdo con las cuatro dimensiones propuestas en el presente estudio: *aplicabilidad*, *claridad y presentación*, *rigor en la elaboración y participación de los implicados*. Así mismo, debe analizarse cuidadosamente el resultado del ítem 27, dada su pobre carga factorial en cada uno de los modelos analizados.

## Financiación

Departamento Nacional de Ciencia y Tecnología Colciencias Instituto Nacional de Cancerología E.S.E. en el marco del Contrato 283-2010, Convocatoria 500-2009 para el desarrollo del proyecto 2101-500-25587

## Responsabilidades éticas

**Protección de personas y animales.** Los autores declaran que para esta investigación no se han realizado experimentos en seres humanos ni en animales.

**Confidencialidad de los datos.** Los autores declaran que en este artículo no aparecen datos de pacientes.

**Derecho a la privacidad y consentimiento informado.** Los autores declaran que en este artículo no aparecen datos de pacientes.

## Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

## Bibliografía

1. International Agency for Research on Cancer (IARC). GLOBOCAN 2008 [internet]. 2008 [citado 2010 jun 6]. Disponible en: <http://globocan.iarc.fr/>.
2. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt G. Users' guides to the medical literature. VIII. How to use clinical practice guidelines. A. Are the recommendations valid? The Evidence-Based Medicine Working Group. *JAMA*. 1995;274(7):570-4.
3. Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J. Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ*. 1999;318(7182):527-30.
4. Lohr K, Field M. A provisional instrument for assessing clinical practice guidelines. En: Field M, Lohr K, editores. *Guidelines for clinical practice From development to use*. Washington, D.C: National Academy Press; 1992.
5. Ministerio de la Protección Social, Colciencias, Centro de Estudios e Investigación en Salud de la Fundación Santa Fe de Bogotá, Escuela de Salud Pública de la Universidad de Harvard. *Guía Metodológica para la elaboración de Guías Atención Integral en el Sistema General de Seguridad Social en Salud Colombiano*. Bogotá; 2010.

6. Lozano JM, Cuervo LG. Desarrollo de guías de práctica clínica. En: Ruiz AM, Luis, editores. *Epidemiología clínica: investigación clínica aplicada*. Bogotá: Editorial Médica Internacional; 2004. p. 373–94.
7. The ADAPTE Collaboration. Resource toolkit for guideline adaptation, versión 1.0 [Internet]. 2007. Disponible en: <http://www.adapte.org/>
8. Fervers B, Burgers JS, Haugh MC, Latreille J, Mlika-Cabanne N, Paquet L, et al. Adaptation of clinical guidelines: literature review and proposition for a framework and procedure. *Int J Qual Health Care*. 2006;18(3):167–76.
9. Haute Autorité de Santé. Méthode et processus d'adaptation des recommandations pour la pratique clinique existantes. Guide Méthodologique, HAS, 2007 [citado 2013 oct 10]. Disponible en: [www.has-sante.fr](http://www.has-sante.fr)
10. Graham ID, Harrison MB. Evaluation and adaptation of clinical practice guidelines. *Evid Based Nurs*. 2005;8(3):68–72.
11. German Instrument for Methodological Guideline Appraisal. Deutsches Instrument zur methodischen Leitlinien-Bewertung (DELBI). Version 2005/2006. Disponible en: [www.english.delbi.de](http://www.english.delbi.de)
12. Agree collaboration. Introduction to AGREE II [internet]. 2013 [citado 2013 oct 10]. Disponible en: <http://www.agreetrust.org>
13. von Troschel J, Selbmann HK, Encke A. Health services research and guidelines-from the Association of the Scientific Medical Societies' (AWMF) perspective. *Z Arztl Fortbild Qualitatssich*. 2006;100(8):597–602.
14. Dans AL, Dans LF. Appraising a tool for guideline appraisal (the AGREE II instrument). *J Clin Epidemiol*. 2010;63(12):1281–2.
15. Collaboration A. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care*. 2003;12(1):18–23.
16. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 1: performance, usefulness and areas for improvement. *CMAJ*. 2010;182(10):1045–52.
17. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. Development of the AGREE II, part 2: assessment of validity of items and tools to support application. *CMAJ*. 2010;182(10):E472–8.
18. DeVellis RF. Scale development: theory and applications, viii, 2nd ed. Thousand Oaks, Calif: Sage Publications, Inc; 2003. p. 171.
19. Streiner DL, Norman GR. Health measurement scales: A practical guide to their development and use, xii, 3rd ed. Oxford; New York: Oxford University Press; 2003. p. 283.
20. Batista-Foguet JM, Coenders G, Alonso J. Análisis factorial confirmatorio. Su utilidad en la validación de cuestionarios relacionados con la salud. *Med Clin (Barc)*. 2004;122 Supl 1:21–7.
21. StataCorp. Stata Structural Equation Modeling Reference Manual Release 12. College Station, Texas: StataCorp; 2011. 294 p.
22. Manzano A, Zamora S. Sistema de ecuaciones estructurales: una herramienta de investigación. Cuaderno técnico 4. México: Centro Nacional de Evaluación para la Educación Superior (CENEVAL); 2009.
23. Diamantopoulos A, Siguaw JA. Introducing Lisrel: a guide for the uninitiated, xii. London: SAGE; 2000. p. 171.
24. Hu L-t, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999;6(1):1–55.
25. Revelle W. Psych: Procedures for Personality and Psychological Research. R package version 1.0-91 [internet]. 2010. Disponible en: <http://personality-project.org/r>, <http://personality-project.org/r/psych.manual.pdf>
26. Rosseel Y. Lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software*. 2012;48(2):1–36.
27. Rosseel Y, Oberski D, Byrnes J, Vanbrabant L, Savalei V, Merkle E, et al. Package 'lavaan' [internet]. 2012 [citado 2012 sep 15]. Disponible en: <http://cran.r-project.org/web/packages/lavaan/lavaan.pdf>
28. MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Serv Res*. 2005;5(1):18.
29. Ministerio de Protección Social. Guía de Práctica Clínica para la detección temprana, tratamiento integral, seguimiento y rehabilitación del cáncer de mama Bogotá [internet]. 2013 [citado 2013 octubre 10]. Disponible en: <http://gpc.minsalud.gov.co/guias/Pages/Gu%C3%ADa-de-c%C3%A1ncer-de-mama.aspx>
30. Peña Torres E, Alba M, Sierra F. P334 Challenges And Advantages Adapting Recommendations: Gpc Breast Cancer. *BMJ Qual Saf*. 2013;22 Suppl 1:86.
31. Alba LH, Murillo R, Castillo JS, INC Gedgdctd. Counseling interventions for smoking cessation: systematic review. *Salud Publica Mex*. 2013;55(2):196–206.
32. Frese T, Franke M, Keyser M, Rurik I, Sandholzer H. Primary care guidelines for geriatric assessment. A structured, comparative analysis. *Z Gerontol Geriatr*. 2012;45(3):224–9.
33. Wiegele C, Glattacker M, Güllich M, Helbing G, Jäckel WH. The methodological quality of rehabilitation guidelines. *Gesundheitswesen*. 2011;73(3):162–8.
34. Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002;3(1):85–106.
35. Linacre JM. Users's guide to Winsteps. Chicago: Mesa Press; 2005.
36. Linacre M. Winsteps Rasch Tutorial 3 [internet]. 2012 [citado 2012 sep 13]. Disponible en: <http://www.winsteps.com/a/winsteps-tutorial-3.pdf>
37. Jöreskog K. Structural equation modeling with ordinal variables using LISREL [internet]. 2005 [citado 2012 Jun 06]. Disponible en: <http://www.ssicentral.com/lisrel/advancedtopics.html>