

Síntomas en la mujer con síndrome coronario agudo: desde las pruebas de validez

Symptoms in Women with Acute Coronary Syndrome: from Validity Tests

Sintomas na mulher com síndrome coronário agudo: desde as provas de validade

VIVIANA MARYCEL CÉSPEDES CUEVAS*

Resumen

El presente artículo expone los resultados de una fase investigativa psicométrica propuesta para demostrar la validez de un instrumento diseñado para medir la experiencia de los síntomas de mujeres con Síndrome coronario agudo según la percepción, evaluación y respuesta al síntoma a partir del Modelo Conceptual del Manejo del Síntoma. La muestra fue de 380 mujeres con diagnóstico confirmado, hospitalizadas en dos instituciones de la ciudad de Bogotá, Colombia. El instrumento final reportó validez de contenido a través del Modelo de Lawshe modificado, validez de escala por medio del Modelo recta de diseño, validez discriminante con la utilización de la Matriz Multirrasgo-Multimétodo y validez de constructo mediante la técnica de Análisis Factorial. Las evidencias de validez son un requisito de calidad de la medida, que posteriormente se complementan con la confiabilidad, la cual ha de ser un atributo de medición secundario y permanente. Demostrar la validez de un instrumento es fundamental no solamente para la investigación, sino también para el desarrollo de una

práctica clínica segura, como prueba de tamizaje de síntomas sospechosos de evento coronario en la mujer.

Palabras clave: validez de las pruebas, síndrome coronario agudo, salud de la mujer (fuente: DeCs, BIREME).

Abstract

This article presents the results of a psychometric investigation phase proposed to demonstrate validity of an instrument designed to measure experience of symptoms in women suffering from acute coronary syndrome according to perception, assessment and response to the symptom based on the Conceptual Model of Symptom Management. The population was 380 women with a confirmed diagnose, who were inpatients of two institutions of the city of Bogota, Colombia. The final instrument reported content validity through the modified Lawshe Model, scale validity through the model of design straight line, Multitrait-Multimethod Matrix-based discriminant validity and construct validity through the Factor Analysis technique. Validity proofs are a measurement quality requisite that is later completed with reliability, which should be a secondary and permanent measurement feature. Demonstrating validity of an instrument is essential not only for the research, but also for development of a safe clinical practice as a screening test of suspicious symptoms of coronary events in women.

Key words: Test validity, acute coronary syndrome, Women Health.

* Doctora en Enfermería, profesora asociada, Facultad de Enfermería, Universidad Nacional de Colombia. vmcespedesc@unal.edu.co, vmcespedesc@gmail.com, Bogotá, Colombia.

Resumo

Este artigo expõe os resultados de uma fase de pesquisa psicométrica proposta para demonstrar a validade de um instrumento desenhado para medir a experiência dos sintomas de mulheres com Síndrome coronário agudo conforme a percepção, avaliação e resposta ao sintoma a partir do Modelo Conceptual de Manejo do Sintoma. A amostra foi de 380 mulheres com diagnóstico confirmado, hospitalizadas em duas instituições da cidade de Bogotá, Colômbia. O instrumento final evidenciou a validade do conteúdo através do Modelo de Lawshe modificado, a validade de escala por meio do Modelo de reta de desenho, validade discriminatória com o uso da Matriz Multirrasgo - Multimétodo e validade de constructo mediante a técnica de Análise Fatorial. As evidências de validade são um requerimento de qualidade da medida que posteriormente são complementadas com a confiabilidade, deve ser também um atributo de medição secundária e permanente. Demonstrar a validade de um instrumento é fundamental não só para a pesquisa, mas também para o desenvolvimento de uma prática clínica segura, como prova do crívado de sintomas suspeitos de evento coronário na mulher.

Palavras-chave: validade das provas, síndrome coronário agudo, Saúde da Mulher.

INTRODUCCIÓN

Hablar de enfermería cardiovascular específica por género es una necesidad actual dentro de un sistema de salud que cada vez exige mayor exactitud y precisión diagnóstica. Cuando los resultados y las recomendaciones para la detección de síndrome coronario agudo (SCA) se generalizan de manera indiscriminada para hombres y mujeres, se conduce a serios problemas para atender a la mujer de forma oportuna y eficaz ante los síntomas de SCA (1), generándose errores de diagnóstico que pueden derivar en grandes inversiones para el manejo de complicaciones (2) o, en la peor de las situaciones, en la muerte prematura, que para el caso de una enfermedad prevenible resulta difícil de comprender.

Es así como la medición de los síntomas como herramienta de valoración y apreciación clínica en los diversos procesos de desviación de la salud ha sido una de las actividades medulares en enfermería (3), en otras ciencias de la salud y en las ciencias sociales (4).

El presente artículo expone el diseño de un instrumento útil para medir la experiencia de los síntomas de SCA en la mujer, haciendo énfasis en las fases requeridas para garantizar su validez, definida como *atributo*

de la medición que permite medir lo que se pretende medir. Futuras publicaciones presentarán los resultados obtenidos en las pruebas de confiabilidad, a partir de métodos analíticos e interpretativos novedosos que permiten valorar la *homogeneidad o consistencia interna y estabilidad* de los sistemas de medición en enfermería. El modelo conceptual del manejo del síntoma (SMM) (5) proporciona la estructura teórica a partir de tres elementos conceptuales: percepción, evaluación y respuesta a los síntomas.

MARCO TEÓRICO

El SMM se encuentra vinculado al Centro para el Manejo del Síntoma de la Facultad de Enfermería de la Universidad de California, San Francisco, y propone que los síntomas deben ser vistos como “experiencias subjetivas que reflejan un cambio en la función biopsicosocial de la sensación o cognición de una persona” (6, 7). El modelo contempla tres dimensiones, denominadas: “Experiencia del síntoma”, “Estrategias de manejo de los síntomas” y “Resultados”, las cuales se encuentran interrelacionadas y en permanente intercambio con el entorno.

Los resultados que sustentan la presente publicación enfatizan en la primera dimensión teórica, la cual incluye a su vez la percepción individual de los síntomas típicos (definidos como presencia de dolor retro-esternal opresivo irradiado o no irradiado) y atípicos (definidos como presencia de dolor en otras partes del cuerpo, palpitations, síntomas digestivos, respiratorios, músculo-esqueléticos, vegetativos, emocionales, cognitivos y aquellos relacionados con el control de la temperatura corporal) de enfermedad coronaria, la evaluación del origen (definido como la percepción de origen cardiaco o no asociado con la patología, o bien psicossomático) y la seriedad (definida como evaluación insignificante, dañina o amenazante) de los síntomas y las acciones de respuesta de manejo (definidas como las actividades que realiza la paciente u otras personas ante los síntomas y el tiempo de demora).

Pese a la existencia de abordajes teóricos que abogan por la ocurrencia sincrónica de los síntomas y desarrollan un enfoque basado en las comunalidades existentes entre ellos (8, 9), el modelo SMM es de gran utilidad para comprender las diferentes fases que experimentan las mujeres con síntomas de enfermedad coronaria antes de solicitar ayuda profesional. La comprensión de las fases

de evaluación cognitiva del síntoma y las acciones de respuesta, es uno de los mayores atributos de la teoría.

Para la construcción del presente sistema de medición se revisaron en la literatura las escalas e instrumentos existentes, dentro de los cuales se encuentran *instrumentos para la medición del dolor*, como las escalas de descriptores verbales (vds) (10), escalas Análogas visuales (vas) (11, 12, 13) y cuestionarios del dolor (14, 15, 16, 17); *instrumentos para la medición de los síntomas de SCA*, los cuales incluyen escalas para la medición aislada de la percepción (18) y la evaluación (19, 20, 21), y cuestionarios para la medición de un síntoma como el SF-36 (22) y el Nottingham Health Profile (23).

Los anteriores instrumentos no reportan el comportamiento métrico de sus escalas, por lo que no se cuenta con elementos suficientes que garanticen la validez y la confiabilidad de los resultados obtenidos a partir de su uso. Asimismo, incluyen mediciones parcializadas de las diversas fases de interés para el estudio, con lo cual se pone en evidencia la conveniencia de desarrollar un único sistema de medición con escalas independientes, pero teóricamente relacionadas a las necesidades del fenómeno y al SMM, que permita medir lo que se pretende

medir en el más amplio sentido del concepto de *validez de pruebas*.

MATERIALES Y MÉTODOS

Los resultados psicométricos que aquí se presentan son producto de la primera fase de una investigación de tipo cuantitativa, descriptiva y exploratoria de corte transversal, la cual se desarrolló como tesis doctoral (24).

Se estudiaron 380 mujeres mayores de veinte años, con diagnóstico confirmado de síndrome coronario agudo (SCA, evidenciado por cateterismo coronario), hospitalizadas en dos instituciones de Bogotá, Colombia. Para el diseño del instrumento se realizó una exhaustiva revisión bibliográfica, con el objetivo de definir el banco inicial de reactivos de la prueba, a partir de los elementos conceptuales del SMM, el cual estuvo conformado por 76 ítems (33 para percepción, 14 para evaluación y 29 para respuesta) con respuesta dicotómica con direccionalidad la cual se amplía más adelante. Una vez consolidada la propuesta inicial, se ejecutaron los procedimientos analíticos para soportar la validez (tabla 1) así:

Tabla 1. Procedimientos analíticos

Evidencias de validez	Pruebas utilizadas
Validez de contenido	Índice de validez de contenido de Lawshe modificado
Validez escalar	Modelo lineal 20/80
Validez discriminante	Correlación punto-biserial o ítem-total
Validez de constructo	Análisis factorial con rotación varimax
Validez discriminante confirmatoria	Matriz multirrasgo-multimétodo

Fuente: (24)

RESULTADOS

La *validez de contenido* se verificó con la intención de definir las especificaciones conceptuales de la escala. Para ello un panel de cuatro expertos evaluó cada uno de los ítems en tres categorías: esencial, útil pero no esencial y no necesario, siguiendo el modelo de Lawshe (25). El análisis y la interpretación se basó en el modelo de Lawshe modificado (26), el cual presenta una propuesta de comprobación útil para la valoración de la validez de contenido de las pruebas diseñadas, adaptadas o utilizadas en enfermería. Entre las principales cualidades del modelo modificado están:

- Permite abordar paneles conformados por menos de cinco panelistas.
- Explica el modelo utilizado para obtener el punto de corte mínimo aceptable.
- Establece el número de acuerdos mínimos como una proporción constante e independiente del número de panelistas, equivalente a 0,528.

Con esta nueva expresión se puede observar que no se tiene el problema del efecto de tamaño, ya que la exigencia es la misma para todos los paneles de expertos elegidos. Para cada ítem se calculó la razón de validez de

contenido modificada (CVR') y para todo el instrumento se obtuvo el índice de validez de contenido (CVI). La expresión utilizada para el CVI fue:

$$CVI = \frac{\sum_{i=1}^M CVRi}{M}$$

Donde:

CVRi = razón de validez de contenido de los ítems aceptables con el criterio de Lawshe.

M= total de ítems aceptables de la prueba.

Como se observa en la tabla 2, los resultados obtenidos evidenciaron que el CVR para el instrumento de medición global fue de 0,97. Para cada uno de los instrumentos de medición fue de 0,94 para percepción, 0,85 para evaluación y 0,94 para respuesta. Estos valores indican que los ítems se constituyen en unidades de análisis representativas de las variables.

Tabla 2. Puntuación de contenido

Ítems	Concordancia de expertos			Índices	
	Esencial	Útil pero no esencial	No necesario	ivc $CVR' = \frac{n_e}{N}$	ivc $CVI = \frac{\sum_{i=1}^M CVRi}{M}$
Percepción					
1 al 7	III			1	31,25/33 = 0,94
8	III	I		0,75	
9	II	II		0,5	
10 y 11	III			1	
12 y 13	II	II		0,5	
14 al 33	III			1	
Evaluación					
34 al 36	III			1	12/14 = 0,85
37 y 38	II	II		0,5	
39	III	I		0,75	
40	III			1	
41 al 43	III	I		0,75	
44 al 46	III			1	
76	III			1	
Respuesta					
47	III	I		0,75	27,5/29 = 0,94
48 y 49	III			1	
50	III	I		0,75	
51 y 52	II	II		0,5	
53 al 75	III			1	
Total ivc				0,97	

Fuente: Céspedes (2009) (24).

La *validez de escala* del instrumento está relacionada con la dirección de medida de las variables en estudio, que a su vez depende de la direccionalidad y la escalaridad de los ítems.

Con respecto a la *direccionalidad de los ítems*, es importante considerar que estos se diseñaron para registrar respuestas dicotómicas, que para el caso de la presente investigación garantizó que cada ítem estableciera una medida discreta del rasgo continuo subyacente, de tal modo que el valor discreto “1” indicaría que la persona tenía “más” del rasgo medido y que además se encontraba en la dirección de crecimiento de dicho rasgo, mientras que 0 indicaría que la persona tenía “menos” del rasgo medido en la misma dirección de crecimiento del rasgo. La siguiente codificación puede ilustrar lo anterior:

- 1 (Presencia) = la respuesta del ítem va en la dirección positiva de la variable.
- 0 (Ausencia) = la respuesta del ítem va en la dirección contraria de la variable.

Así se tiene que la escala corre siempre de 0,0 a 1,0 en términos de p (proporción de respuestas correctas). Se trata de un tratamiento novedoso de la dicotomía considerada en el diseño de instrumentos en el área de la salud o de las ciencias sociales, que tradicionalmente supone la escala como aditiva, es decir, acumulando los “puntos” que va respondiendo cada persona, pudiendo ser procedentes de ítems dicotomos o en escala tipo Likert.

Continuando con la comprensión de la *dirección de escala*, se consideró la interpretación dada por Wright y Stone (27), que focalizan en la calidad del criterio teórico que sustenta la escala, con base en un conjunto de ítems previamente calibrados y aceptados con el modelo clásico de los tests (a partir de los parámetros de dificultad y de discriminación), lo cual permitió conocer objetivamente si la distribución de los ítems al interior de cada una de las dimensiones teóricas de la prueba correspondía con un criterio teórico externo.

Para el presente estudio se consideró que una variable cuenta con validez de escala cuando se satisfacen los siguientes criterios expuestos por Tristán y Vidal (28):

1. La variable se define en relación con un rasgo, en una dimensión cartesiana positiva; es decir, se

puede representar en un eje que va de “poco” a “mucho” de dicho rasgo.

2. Los ítems que miden la variable pueden ser ubicados en dicha dimensión en una posición correspondiente a la media de respuestas positivas de los sujetos.
3. El orden que guardan los ítems en el eje cartesiano tiene sentido respecto a la variable medida, es decir, un ítem de mayor medida indica “más” del atributo medido.
4. Los ítems quedan centrados en una posición media del rasgo medido.
5. Los ítems se distribuyen en todo el rango de medidas, cubriendo las posibles medidas del atributo en todos los sujetos en estudio.
6. Los ítems se distribuyen uniformemente en todo el rango de medidas.

Si estos atributos se cumplen se puede encontrar una variable que define una escala, sobre la cual se puede ordenar a los sujetos, dictaminar niveles o posiciones específicas para tomar decisiones sobre ellos y clasificarlos para alguna intervención.

Para revisar las cualidades de validez de escala descritas, se dispuso de un modelo denominado “Recta de diseño” (o Test Design Line, TDL), que consistió en representar en el plano “ p vs. número de ítem” los ítems que integran cada dimensión del instrumento, ordenándolos de menos a más con respecto a p (media de respuestas positivas en cada ítem). Si se contaba con una escala perfecta, todos los ítems deberían quedar alineados en una recta definida de “ p ” inferior = 0,2 a “ p ” superior = 0,8 (valores teóricos sugeridos).

Para comparar las posiciones reales de los ítems con las posiciones teóricas se calculó la diferencia absoluta media (DAM), la cual no debe ser mayor al 5 % (DAM = 0,05). A continuación se relaciona el cálculo de la DAM:

$$DAM = \frac{\sum_{i=1}^M |p_i - p_e|}{M}$$

Donde:

M = número de ítems

p_i = media de respuestas positivas en el ítem i

$$p_e = p_{inf} + (p_{sup} - p_{inf}) \frac{i - 1}{M - 1}, \text{ media esperada en}$$

la recta que va de p_{inf} a p_{sup} .

Cuando los ítems del cuestionario se alejaban significativamente de una distribución rectilínea, se consideraba dudosa la escalaridad de la variable. La tabla 3 presenta las puntuaciones de validez de escala obtenidas por dimensión.

Tabla 3. Puntuación de validez de escala

Dimensión 1: Percepción		Dimensión 2: Evaluación		Dimensión 3: Respuesta	
P (proporción de respuesta en el ítem)		P (proporción de respuesta en el ítem)		P (proporción de respuesta en el ítem)	
P inferior	P superior	P inferior	P superior	P inferior	P superior
26	87	8	54	0	80
Promedio	56	Promedio	31	Promedio	40
Diferencia absoluta	4,57	Diferencia absoluta	1,69	Diferencia absoluta	7,90

Fuente: (24).

La *variable percepción*, que va de típica a atípica, presentó ajuste a una recta de diseño 26-87, centrada en $p = 56$ del atributo medido. La $D_{MA} = 4,57$ confirmó que el rasgo se presenta más cerca del extremo “percepción atípica”.

La *variable evaluación*, que va de inadecuada a adecuada, presentó ajuste a una recta de diseño 8-54, centrada en $p = 31$ del atributo medido. La $D_{MA} = 1,69$ confirmó que el rasgo se presenta en el extremo “evaluación inadecuada”.

La *variable respuesta*, que va de inadecuada a adecuada, presentó ajuste a una recta de diseño 0-80, centrada en $p = 40$ del atributo medido. La $D_{MA} = 7,90$ confirmó que el rasgo se presenta en el extremo “respuesta adecuada”.

Como se mostró, las dimensiones de percepción, evaluación y respuesta se comportaron como escalares. Sin embargo, la dimensión de percepción fue la más cercana a un diseño centrado, mientras que la dimensión de evaluación quedó corta, al no haber ítems con más del 60 % de respuestas positivas, lo cual implicó una nueva revisión por expertos y por evidencia soportada en revisión de literatura. Al no encontrarse nuevos ítems para completar la medición, se consideró que la presente escala es hasta el momento la mejor propuesta diseñada para medir esta dimensión.

Asimismo, la D_{MA} de la variable respuesta superó el 5 % recomendado, por lo que se sospechó la presencia

de ítems no escalares. En la revisión de la subdimensión “Tiempos de demora”, que integra la variable respuesta, se identificó que los ítems del 58 al 75 son de tipo contextual y no se refieren a una escala, por lo cual se hizo necesario su retiro.

La nueva línea de tendencia para la variable respuesta presentó el mejor ajuste en un diseño 10-80, con un coeficiente de determinación de $R = 0,92$, lo cual garantizó que las tres variables de estudio contaran con una escala de medición independiente.

Por otra parte, el estudio de las tres variables en una sola escala, habiendo hecho el retiro de los ítems contextuales y obteniendo un instrumento total conformado por 58 ítems, se ajustó al modelo de recta de diseño 20-80, con un coeficiente de determinación de $R = 0,98$, el cual se trató de un ajuste óptimo que permitió afirmar que se dispone de una escala de medición para la experiencia del síntoma con sus tres componentes (percepción, evaluación, respuesta).

El análisis de validez de escala de la prueba, con ayuda de la recta de diseño, permitió conocer objetivamente que la distribución de los ítems al interior de cada una de las dimensiones teóricas de la prueba corresponde con un criterio teórico externo de medición, tanto en las tres escalas independientes como en la escala total.

La *validez discriminante* se enfocó en identificar la variable que mide el instrumento, en contraposición a las otras variables que pudieran estar presentes, pero

que corresponden a otros rasgos o constructos. A diferencia de la validez de contenido, que es obtenida por la opinión de jueces, y de la validez de criterio, que se enfoca en buscar la semejanza de la variable medida con otros instrumentos, la validez discriminante pretende determinar grupos de ítems que miden rasgos diferentes, teniendo consistencia dentro del grupo pero que no se correlacionan con los rasgos que miden otros grupos de ítems. El presente estudio determinó la validez discriminante a partir del análisis de la correlación ítem-total (correlación punto-biserial), en tres casos posibles:

- Correlación positiva: cuando hay homogeneidad positiva en los ítems.
- Correlación negativa: cuando hay homogeneidad negativa en los ítems.
- Correlaciones bajas (inferiores a 0,196): cuando hay heterogeneidad en los ítems, lo cual se trata de una agrupación dudosa.

De cada uno de los anteriores conjuntos de ítems se consideró una posible variable, si se cumplió con las siguientes condiciones:

- La variable cuenta con cuatro ítems por lo menos.
- Los ítems correlacionan positivamente con el conjunto (superior a 0,196).
- El conjunto produce valores altos del coeficiente de consistencia interna alfa, conforme a los valores de referencia del modelo denominado máxima validez-máxima confiabilidad (29, 30), el cual considera curvas de valores óptimos de alfa de Cronbach para pruebas que cumplen la validez de diseño en recta 20-80 (dependiendo de la varianza de respuestas de los sujetos).

Como se puede apreciar en las anteriores condiciones, el coeficiente de consistencia interna alfa fue valorado con un método novedoso que precisa la confiabilidad que una prueba requiere conforme a la validez escalar determinada. Este modelo dictamina para el valor de alfa un punto de corte justificado que no depende de la longitud de la prueba, lo cual sería de valiosa utilidad para evitar la redundancia métrica en el diseño de instrumentos de las ciencias de la salud y, específicamente, en las pruebas utilizadas en enfermería, en las que se requieren sistemas estandarizados de ágil aplicación. Con la utilización del modelo denominado máxima

validez-máxima confiabilidad se confirma que la confiabilidad, la homogeneidad o la consistencia interna, es un atributo secundario de toda prueba que debe medirse de manera *permanente* en los procesos de investigación, que no requiere protocolos de estudio específicos para su cálculo y que, incluso si no hay garantías de validez, sería innecesaria su determinación.

Como el foco central del presente artículo es la comprensión de la validez del instrumento diseñado, próximas publicaciones ampliarán la explicación de la confiabilidad como método complementario para asegurar la calidad en el diseño de pruebas.

Continuando con el proceso anteriormente descrito, a partir de los datos disponibles se obtuvieron las correlaciones ítem-total y se clasificaron en los tres grupos citados. Los resultados de los subgrupos obtenidos se detallan en la tabla 4.

Como se puede verificar, para la dimensión de *percepción* se identificó un subgrupo dentro del cual se aceptaron los ítems 3 y 23, a pesar de poseer una correlación punto-biserial inferior a 0,19, debido a que se encontraban en una zona de duda próxima al punto de corte establecido y su retiro no afectaba la puntuación de consistencia interna alfa obtenida.

El ítem 31, el cual presentó correlación ítem-total positiva pero menor que 0,19, y el ítem 2, el cual presentó correlación ítem-total negativa, no conformaron agrupación debido a que dentro del presente análisis las agrupaciones de menos de 4 o 3 reactivos no se consideraron concluyentes.

Los demás componentes de la agrupación identificada cumplieron con todas las condiciones para ser considerada una variable. Es de gran interés dentro de la consolidación de dicha agrupación el significado que recobra la percepción mixta de los síntomas (típica y atípica).

Para la dimensión de *evaluación* se identificó un subgrupo para medir el origen y la seriedad de los síntomas, respectivamente. El ítem 40 fue aceptado en la agrupación de origen del síntoma, a pesar de poseer una correlación punto-biserial inferior a 0,19, debido a que se encontraba en una zona dudosa próxima al punto de corte establecido y su retiro no afectaba la puntuación de consistencia interna alfa obtenida. Los demás componentes de las agrupaciones identificadas cumplieron con todas las condiciones para ser consideradas como

Tabla 4. Agrupaciones según análisis discriminante

Agrupación	Ítems	Ítem-Total	Significado teórico		
N.º 1 Percepción Alpha Teórica: 0,76 Experimental: 0,80	1 Dolor o molestia retroesternal	.24	Percepción Mixta: típica y atípica		
	3 Dolor o molestia en el pecho	.12			
	4 Dolor o molestia en la espalda	.38			
	5 Dolor o molestia en el cuello	.31			
	6 Dolor o molestia en los brazos	.40			
	7 Dolor o molestia en la mandíbula	.22			
	8 Dolor o molestia en la cabeza	.38			
	9 Dolor o molestia debajo del seno	.32			
	10 Dolor o molestia en el estómago	.31			
	11 Palpitaciones	.36			
	12 Pérdida del apetito	.26			
	13 Indigestión	.38			
	14 Respiración entrecortada	.42			
	15 Sensación de ahogo	.40			
	16 Tos	.19			
	17 Fatiga inusual	.38			
	18 Incapacidad para moverse	.32			
	19 Entumecimiento	.25			
	20 Cansancio	.42			
	21 Desaliento	.39			
	22 Deseos de vomitar	.34			
	23 Vómito	.19			
	24 Mareo	.39			
	25 Desmayo	.27			
	26 Confusión	.27			
	27 Miedo inusual	.27			
	28 Depresión	.35			
	29 Angustia	.30			
	30 Sudoración	.37			
	32 Escalofrío	.30			
	33 Sensación de calor	.22			
	N.º 1 Evaluación Alpha Teórica: 0,35 Experimental: 0,67	34 Problema cardíaco		.53	Origen mixto: cardíaco y extracardiaco
		35 Problema del estómago		.29	
36 Problema muscular		.21			
37 Gripe o resfriado		.34			
38 Problema dental		.25			
39 Problema de respiración		.33			
40 Problema cerebral		.17			
41 Estrés o sobrecarga		.35			
42 Preocupaciones personales		.41			
43 Fatiga	.40				
N.º 2 Evaluación Alpha Teórica: 0,35 Experimental: 0,94	44 Insignificante	.95	Seriedad mixta: insignificante y amenazante		
	45 Dañino	.94			
	46 Amenazante	.95			
	76 Sensación de muerte	.65			
N.º 1 Respuesta Alpha Teórica: 0,35 Experimental: 0,67	47 Ignorarlo	.35	Acciones de respuesta mixtas: inadecuadas y adecuadas		
	49 Tomar remedios caseros	.30			
	51 Tomar aguja	.44			
	52 Respirar profundamente	.49			
	53 Esperar a que pasara	.49			
	55 Llamar al médico	.22			
	56 Llamar a la ambulancia	.20			
57 Ir al servicio de urgencia	.14				

Fuente: (24).

variables. Es de gran interés dentro de la consolidación de dichas agrupaciones, que se encontraron evaluaciones mixtas desde el origen y la seriedad de los síntomas.

Para la dimensión de *respuesta* se identificó un subgrupo dentro del cual se aceptó el ítem 57, a pesar de poseer una correlación punto-biserial inferior a 0,19, debido a que se encontraba en una zona de duda próxima al punto de corte establecido y su retiro no afectaba la puntuación de consistencia interna alfa obtenida. Los demás ítems de esta agrupación cumplieron con todos los criterios establecidos para ser considerada como variable. El significado teórico de los hallazgos evidencia una dimensión de respuesta caracterizada por acciones mixtas para el manejo de los síntomas (adecuadas e inadecuadas).

Las cuatro agrupaciones obtenidas desde el análisis de validez discriminante permitió establecer grupos de ítems que miden rasgos diferentes, con alta consistencia *dentro* del grupo, pero sin correlación *entre* los grupos, lo cual apunta no solamente a determinar una escala cada vez más precisa, sino a derivar hallazgos de mayor interpretabilidad a la experiencia del síntoma de la mujer con SCA.

Pasando ahora a la comprensión de la *validez de constructo*, la cual permitió confirmar la existencia de un criterio teórico externo de soporte de la escala, se utilizó el análisis factorial con el método de componentes principales, incluyendo la rotación varimax.

Fueron aceptadas las cargas factoriales superiores a $\pm 0,35$ (31). El procedimiento utilizado para aceptar ítems dentro de cada factor, en aquellos casos en que un ítem se encontraba cargado en dos o más factores, consistió en aceptar el ítem dentro del factor donde presentara una mayor carga factorial (más cercana a 1 o -1) y una diferencia matemática de carga factorial superior a 0,01, en relación con los demás factores.

El análisis factorial se forzó al número de agrupaciones identificadas por análisis de correlación ítem-total, con el propósito de hacer más viable la determinación de las agrupaciones de cada una de las dimensiones.

Adicionalmente al resultado del análisis factorial, cada uno de los factores encontrados fue revisado con los criterios para determinar si se trataba de una posible variable, empleados en el análisis de validez discriminante. Se decidió emplear de nuevo estos criterios para poder comparar las dos aproximaciones en la definición

de las variables dentro de las dimensiones. La tabla 5 detalla las agrupaciones obtenidas por análisis factorial.

Para la dimensión de *percepción* la varianza explicada con tres factores fue del 28,75 %, siendo evidente la correspondencia entre cada uno de los factores identificados y el referente teórico que lo sustenta, especialmente enfocado en la presencia de percepción atípica de síntomas de SCA, lo cual refleja una adecuada validez de constructo.

Para la dimensión de *evaluación* la varianza explicada con tres factores fue del 57,13 %, evidenciándose ajuste entre los factores identificados y el sustento teórico que los soporta, específicamente en lo relacionado con evaluación mixta de origen (cardíaco y extracardíaco) y seriedad (insignificante y dañino), lo cual refleja una adecuada validez de constructo.

Para la dimensión de *respuesta* la varianza explicada con dos factores fue del 37,88 %, haciéndose evidente el ajuste entre los factores identificados y el referente teórico, principalmente relacionado con las acciones de respuesta inadecuadas, lo cual evidencia adecuada validez de constructo.

Finalmente, como estrategia de corroboración de la presencia de validez de constructo y discriminante, se utilizaron los valores de confiabilidad de las agrupaciones para la construcción de la matriz multirrasgo-multimétodo propuesta por Donald T. Campbell y Donald W. Fiske (32, 33, 34, 35), la cual exige altas correlaciones intradimensión como evidencia de validez de constructo, y bajas correlaciones interdimensión cuando los rasgos son asumidos como diferentes y existe validez discriminante. Los resultados obtenidos se presentan en la tabla 6:

Nótese las correlaciones de Pearson altas (superiores a 0,196) al interior de cada componente teórico (zonas por debajo de la diagonal en negro claro), las cuales son evidencia de validez de constructo. Asimismo, se muestran correlaciones de Pearson bajas (inferiores a 0,196) entre componentes (zonas por debajo de la diagonal en negro oscuro). Los colores negros debajo de la diagonal clarifican para el primer caso (con negrilla) la validez de constructo y para el segundo caso (sin negrilla) la validez discriminante lo cual es indicativo de componentes teóricos (rasgos o atributos) distintos y, por lo tanto, demuestra la presencia de validez discriminante.

Las correlaciones al interior del componente teórico de evaluación poseen valores de correlación inferiores a

Tabla 5. Agrupaciones según análisis factorial

Factor	Ítems	Carga	Significado teórico		
N.º 1 Varianza explicada 15,31 % N.º 2 Varianza explicada 6,94 % N.º 3 Varianza explicada 6,49 %	1 Dolor o molestia retroesternal	.35	Percepción mixta: típica y atípica Percepción atípica: digestiva y vegetativa Percepción atípica: emocional		
	4 Dolor o molestia en la espalda	.52			
	5 Dolor o molestia en el cuello	.48			
	6 Dolor o molestia en la mandíbula inferior	.55			
	8 Dolor o molestia en la cabeza	.41			
	9 Dolor o molestia debajo del seno	.37			
	11 Palpitaciones	.38			
	14 Respiración entrecortada	.61			
	15 Sensación de ahogo	.63			
	17 Fatiga inusual	.57			
	20 Cansancio	.57			
	21 Desaliento	.56			
	30 Sudoración	.43			
	12 Pérdida del apetito	.43			
	13 Indigestión	.48			
	22 Deseos de vomitar	.66			
	23 Vómito	.64			
	24 Mareo	.45			
	25 Desmayo	.56			
	26 Confusión	.49			
	32 Escalofrío	.44			
	3 Dolor o molestia en cualquier parte del pecho	.35			
	18 Incapacidad para moverse	.37			
	27 Miedo inusual	.64			
	28 Depresión	.58			
	29 Angustia	.71			
	N.º 1 Varianza explicada 30 % N.º 2 Varianza explicada 14,77 % N.º 3 Varianza explicada 12,35 %	44 Insignificante		.97	Evaluación de seriedad mixta Evaluación de origen extracardiaco-emo- cional Evaluación de origen extracardiaco mixto
		45 Dañino		.97	
		46 Amenazante		.97	
76 Sensación de muerte		.74			
41 Estrés o sobrecarga de trabajo		.87			
42 Preocupaciones personales		.88			
43 Fatiga		.68			
34 Problema cardiaco		.50			
35 Problema del estómago		.57			
36 Problema muscular		.39			
37 Gripe o resfriado		.65			
N.º 1 Varianza explicada 24,30 % N.º 2 Varianza explicada 13,57 %	38 Problema dental	.57	Respuestas inadecua- das Respuestas mixtas		
	39 Problema de respiración	.57			
	40 Problema cerebral	.45			
	51 Tomar agua	.82			
	52 Respirar profundamente	.91			
	53 Esperar a que pasara	.87			
	47 Ignorarlo	.59			
	49 Tomar remedios caseros	.52			
50 Tomar medicinas alternativas	.47				
55 Llamar al médico	.37				
56 Llamar una ambulancia	.42				
57 Ir al servicio de urgencia	.67				

Fuente: (24).

Tabla 6. Matriz multirrasgo-multimétodo de la experiencia del síntoma

Percepción				Evaluación				Respuesta		
0,81										
0,8229	0,76									
0,7041	0,312	0,69								
0,5407	0,5335	0,2233	0,61							
0,1739	0,0827	0,0038	0,0924	0,67						
0,1479	0,1093	0,1153	0,091	0,7517*	0,77					
0,1299	0,1339	0,0038	0,1912	-0,079	-0,121	0,63				
0,1371	0,1146	0,0242	0,1808	-0,129	-0,165	0,9128*	0,94			
0,2008	0,2563	0,0648	0,1321	0,1641	0,1734	-0,272	-0,316	0,65		
0,1773	0,2366	-0,007	0,1907	0,1118	0,105	-0,158	-0,182	0,8365	0,85	
0,0848	0,1592	-0,066	0,1761	0,082	0,0313	-0,145	-0,162	0,7017	0,8343	0,49

* P valor = 0,05.

Fuente: (24).

0,196, así como valores de correlación negativos, pero también reporta correlaciones significativas * p valor= 0,05. Por lo anterior no es posible establecer juicios sobre la validez de constructo en esta dimensión teórica.

DISCUSIÓN

En la literatura se dispone de abundantes sistemas de medición de la percepción de los síntomas, sin evidencia de los mínimos parámetros de estabilidad psicométrica necesaria que haga meritoria su utilización en procesos de investigación o en la práctica, con el agravante de encontrarse soportados en los estándares masculinos.

Por otra parte, la literatura no dispone de propuestas métricas para el estudio de la evaluación y la respuesta a los síntomas, dejándose de lado como algo no medible o reduciéndose a la cuantificación de los tiempos de demora únicamente.

Contar con un sistema que sustente la validez de medida de la experiencia del síntoma desde la percepción, la evaluación y la respuesta a los síntomas del sca en la mujer a partir de varios modelos, partiendo desde la apreciación de jueces hasta aproximaciones matemáticas de mayor complejidad, es la principal contribución de la presente investigación.

Un aporte relevante se trata de la implementación de un concepto bastante difundido en la investigación metodológica, el cual hace referencia a que *todo instru-*

mento válido puede ser confiable, pero no todo instrumento confiable es válido.

Con la anterior premisa se dio una revisión a la propuesta de validez de contenido de Lawshe modificada por Tristán, la cual permitió emitir un juicio con base en un nivel de aceptabilidad del *ivc*, calculado en 0,58 por métodos de comprobación e independiente del número de panelistas. Con ello se puede resolver la dificultad detectada en los procesos investigativos que deben adaptar escalas o demostrar parámetros mínimos de contenido para la utilización de pruebas en contextos diferentes, en relación con la conformación de los paneles y la determinación de los puntos mínimos de aceptabilidad, que para la propuesta original de Lawshe se plantean cinco panelistas para un valor mínimo de *cvr* de 1, que corresponde a total acuerdo (lo cual es bastante exigente cuando se trabaja con un grupo de expertos) y cuarenta panelistas (lo cual es a veces inexistente en nuestro contexto) para un valor mínimo de *cvr* de 0,29, pudiéndose notar que no hay propuesta para menos de cinco panelistas.

Asimismo, la implementación del concepto de validez escalar permitió desarrollar una propuesta que pasó de ser una colección de preguntas a una verdadera escala con tres subescalas independientes que plantean direccionalidad, dependiendo de la cantidad de atributos de percepción, evaluación y respuesta que posean las mujeres con sca.

Este importante desarrollo escalar favoreció corroborar la validez discriminante y de constructo de los

rasgos medidos, los cuales derivan no solamente en un sistema de medición válido, sino en resultados de alta interpretabilidad para la mujer que experimenta los síntomas de SCA, que se demarcan en escenarios de percepción atípica, evaluación de un origen extracardiaco e insignificante y acciones de respuesta inadecuadas.

La versión derivada en el presente estudio es de gran utilidad no solo para la investigación, sino que a futuro podrá ser utilizada en la práctica disciplinar de enfermería y de las ciencias de la salud en general, gracias a las pruebas cortas derivadas, las cuales serían fáciles de aplicar en escenarios reales, especialmente en los sistemas de clasificación-Triage.

Finalmente y como elemento parcialmente presentado en la presente publicación, los métodos que soportaron la validez favorecieron la implementación de sistemas novedosos para dictaminar la confiabilidad del instrumento como parte de la medición de la calidad de pruebas, como el modelo de máxima validez-máxima confiabilidad utilizado, el cual seguramente será de gran utilidad en la medición de los fenómenos de enfermería, que por su naturaleza subjetiva requiere un manejo especial, principalmente al establecer los puntos de corte. Lo anterior se puede ejemplificar claramente con el tan utilizado parámetro de 0,7 para determinar la aceptabilidad del valor de alpha de Crombach de un reactivo dentro de una prueba, que como se demuestra en la presente investigación, podría tratarse de redundancia métrica (aumento de longitud de prueba para incrementar la confiabilidad) en pruebas que demuestran una óptima validez escalar.

Los parámetros de validez obtenidos en el diseño del instrumento derivado en la presente investigación constituyen la principal garantía para su utilización, seguido del cálculo de la confiabilidad necesaria para la validez obtenida, lo que abre el camino al proceso de estandarización de la prueba en mediciones repetidas.

CONCLUSIONES

La presente investigación exploró varias evidencias de validez, en particular:

Validez de contenido. Se realizó en función de un análisis de jueces expertos. En este caso se empleó el modelo de cuantificación de la validez de contenido a partir de los supuestos planteados en el modelo de Lawshe modifica-

do, encontrándose un índice de validez de contenido para el instrumento global de 0,97 y comprobándose que cada ítem alcanzó y superó el criterio de garantía métrica. Este hallazgo indica que los ítems se constituyen en unidades de análisis representativas de los elementos incluidos para la percepción, la evaluación y la respuesta a los síntomas medidos.

Validez de escala. El instrumento propuesto debía medir un rasgo latente asociado a la experiencia del síntoma en tres componentes específicos: percepción, evaluación y respuesta. Tratándose de un instrumento de medición y no de un simple conjunto de preguntas inconexas, se esperó que el “puntaje” que recibía una persona correspondiera con una medida dentro de las tres variables específicas. Se planteó la necesidad de verificar que el instrumento disponía de una escala subyacente, por lo cual se requirió probar la validez de escala. En este caso se empleó el concepto de validez de escala, revisándose la distribución de ítems, el rango y el ajuste al modelo denominado recta de diseño 20-80, el cual aportó la evidencia de esta validez, encontrándose que los tres componentes de percepción, evaluación y respuesta correspondían a escalas reales de medición de manera independiente y total. Como medida correctiva del componente de respuesta analizado por contenido exacto, al no optar por el comportamiento escalar, los ítems asociados al ambiente social y los tiempos de demora ameritaron ser retirados del instrumento. Sin embargo, su valor como reporte fue reubicado y utilizado para análisis desde el contexto.

Validez discriminante: los ítems debían medir cada uno de los componentes específicos de percepción, evaluación y respuesta y no otros; es decir, debía haber una clara distinción entre los ítems del primer componente y los de los demás componentes. Para esta verificación se ejecutó el análisis de validez discriminante, con base en una exploración exhaustiva de los ítems, utilizando la correlación punto-biserial y el modelo de máxima validez-máxima confiabilidad, los cuales permitieron comprobar que cada uno de los componentes estudiados eran diferentes entre sí.

Validez de constructo: el instrumento debía demostrar que la medición de los ítems correspondía a un referente teórico externo, para lo cual se empleó el análisis factorial exploratorio, utilizando el modelo de extracción de valores característicos y rotación varimax. Este análisis

se complementó con la aplicación del modelo matemático de análisis multirasgo-multimétodo, demostrándose que el conjunto de ítems que conformaba los componentes de percepción, evaluación y respuesta correspondieron con un elemento teórico externo en particular, evidenciándose así la validez de constructo del instrumento creado mediante dos estrategias metodológicas diferentes.

La metodología seguida en este trabajo es de aplicación sencilla, pero muy potente y rigurosa para el diseño de pruebas.

REFERENCIAS

- (1) Demsey SJ, Dracup K, Moser DK. Women's decisions to seek medical care for acute myocardial infarction. *Heart Lung*. 1995; 24(6): 444-456.
- (2) Hauser ST. Physician-patient relationships. *Social Contexts of Health, illness, and patient care*. Cambridge: Cambridge University Press; 1987. p. 104-140.
- (3) Johnson JI. Nurse practitioner-patient discourse: uncovering the voice of nursing in primary care practice. *Scholarly inquiry for nursing practice: an International Journal*. 1993; 7(3): 143-157.
- (4) Wright BD, Stone MH. Validity in Rasch measurement. *Research memorandum 54. MESA*. Chicago: University of Chicago; 1988. p. 12.
- (5) Dodd MJ. The pathogenesis and characterization of oral mucositis associated with cancer therapy. In *innovative approaches to oral mucositis and cancer pain: linking mechanisms to management*. *Oncology Nursing Forum*. 2004; 31(4): 5-11.
- (6) University of California. A model for symptom management. San Francisco School of Nursing Symptom Management Faculty Group. *J Nurs Scholarship*. 1994; 26 (4): 272-276.
- (7) Dodd M, Janson S, Facione N, Faucett J, Froelicher E, Humphreys J, et al. Advancing the science of symptom management. *J Advance Nurs*. 2001; 33(5): 668-676.
- (8) Lenz ER, et al. The middle-range theory of unpleasant symptoms: An update. *Advances in Nursing Science*. 1997; 19(3): 14-27.
- (9) Lenz ER, et al. Collaborative development of middle-range nursing theories: Toward a theory of unpleasant symptoms. *Advances in Nursing Science*. 1995; 17(3): 1-13.
- (10) Keele KD. The pain chart. *Lancet*. 1948; 3(2): 6-8.
- (11) Revill SI, Robinson JO, Rosen N, et al. The reliability of a linear analogue for evaluating pain. *Anaesthesia*. 1976; 31(9): 1191-1198.
- (12) Liu WH, Aitkenhead AR. Comparison of contemporaneous and retrospective assessment of postoperative pain using the visual analogue scale. *Br J Anaesth*. 1991; 67(6): 768-771.
- (13) Frank M, Olsen S. Instruments for clinical health-care research. *Measuring Pain*. 2a Ed. Jones and Bartlett Publishers; 1997. p. 528-561.
- (14) Melzack R. The short-form McGill Pain Questionnaire. *Journal of Pain Symptom Management*. 1987; 30(2): 191-197.
- (15) Dudgeon D, Raubertas RF, Rosenthal SN. The Short-form McGill Pain Questionnaire in chronic cancer pain. *Journal of Pain Symptom Management*. 1993; 8(4): 191-195.
- (16) McGuire DB, Altomonte V, Peterson DE, et al. Patterns of mucositis and pain in patients receiving preparative chemotherapy and bone marrow transplantation. *Oncology Nursing Forum*. 1993; 20(10): 1493-1502.
- (17) Swanston M, Abraham C, Nacrae WA, et al. Pain assessment with interactive computer animation. *Review of Pain*. 1993; 53(3): 347-351.
- (18) Zerwic JJ. Symptoms of acute myocardial infarction: Expectations of a community sample. *Heart Lung: Journal of Acute & Critical Care*. 1998, 27(2): 75-81.
- (19) Champion VL. Instrument development for health belief model constructs. *Advances in Nursing Science*. 1984; 6(3): 72-85.
- (20) Porter LA. Cognitive appraisal of psychological stress, appraisal contexts, and life disruption in patients having percutaneous transluminal coronary angioplasty. Rochester: The University of Rochester; 1992.
- (21) Delunas LR, Potempa K. Adaptation after treatment for Heart disease: preliminary examination within a stress appraisal context. *Heart Lung*. 1999; 28(3): 186-191.
- (22) Vilagut G, Ferrer M, Rajmil L, et al. El cuestionario de salud SF-36 español: una década de experiencia y nuevos desarrollos. *Gac Sanit*. 2005; 19(2): 135-50.
- (23) Wann-Hansson C, Rahm Hallberg I, Risberg B, et al. A comparison of the Nottingham Health Profile and Short Form 36 Health Survey in patients with chronic lower limb ischaemia in a longitudinal perspective. *Health and quality of life outcomes*. 2004; 2(9): 5-11.
- (24) Céspedes V. Modelo Conceptual del Manejo del síntoma: clasificación por percepción, evaluación y respuesta de mujeres con síndrome coronario agudo; originada por la construcción de un instrumento validado en Bogotá, Colombia [Tesis doctoral]. Universidad Nacional de Colombia; 2009.
- (25) Lawshe CH. Quantitative Approach to Content Validity. *Personnel Psychology*. 1975; 28(4): 563-575.

- (26) Tristán LA. Modificación al modelo de Lawshe para el dictamen cuantitativo de la validez de contenido de un instrumento objetivo. *Avances en Medición*. 2008; Vol. 6: 35-44.
- (27) Wright BD, Stone MH. *Making measures*. Chicago (USA): The Phaneron Press; 2004. p. 35-39.
- (28) Tristán LA, Vidal UR. Linear model to assess the scale's validity of a test. Chicago: AERA Annual Meeting; 2007.
- (29) Tristán LA. Contribución al estudio del error de medida, (Parte 1). Notas sobre evaluación criterial. N.º 5. Ceneval y familia de programas KALT; 1999. p. 12.
- (30) Tristán LA. Diseño de pruebas, (Parte 4). México: Instituto de Evaluación e Ingeniería Avanzada; 1999-2007. p. 13-17.
- (31) Batista JM, Coenders G, Alonso J. Análisis factorial confirmatorio. Disponible en: <http://www.elsevier.es/sites/default/files/elsevier/pdf/2/2v122nSupl.1a13057542pdf001.pdf>. Consultado mayo de 2011.
- (32) Campbell DT, Fiske CD. Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*. 1959; 56(2): 81-105.
- (33) Wright BD, Stone MH. *Making measures*. Chicago (USA): The Phaneron Press; 2004. p. 35-39.
- (34) Rodríguez P, Cifuentes M, Moreno E, Sanjuán P. Matriz Multirasgo Multimétodo. Disponible en: <http://redalyc.uaemex.mx/redalyc/pdf/727/72797112.pdf>. Consultado mayo de 2011.
- (35) Martínez JA, Martínez L. Matriz Multirasgo Multimétodo. Disponible en: <http://repositorio.bib.upct.es:8080/jspui/handle/10317/1600>. Consultado mayo de 2011.