

Propiedades psicométricas de las escalas de gravedad de obstrucción bronquial en pediatría. Revisión sistemática y metaanálisis

Psychometric properties of scales used for grading the severity of bronchial obstruction in pediatrics: A systematic review and meta-analysis

Klga. Mg. Soledad Luarte-Martínez^{a,b}, Klgo. Dr. Iván Rodríguez-Núñez^{c,d}, Enf. Mg. Paula Astudillo^{b,c,e,f} y Dr. Carlos Manterola^{b,c,e,g}

- a. Departamento de Kinesiología, Universidad de Concepción, Concepción, Chile.
- b. Programa de Magíster Ciencias Médicas, Universidad de La Frontera, Chile.
- c. Programa de Doctorado en Ciencias Médicas, Universidad de La Frontera, Chile.
- d. Escuela de Kinesiología, Universidad San Sebastián, Concepción, Chile.
- e. Departamento de Cirugía, Universidad de La Frontera, Chile.
- f. Programa Doctorado en Psicología, Universidad de Girona, España.
- g. Centro de Investigaciones Biomédicas, Universidad Autónoma de Chile.

Correspondencia:
Enf. Mg. Paula Astudillo:
paula.astudillo@ufrontera.cl

Financiamiento:
Financiado por la Vicerrectoría de Investigación y Desarrollo (VRID) de la Universidad de Concepción, Proyecto N.º 215.082.050IN.

Conflicto de intereses:
Ninguno que declarar.

Recibido: 28-6-2016
Aceptado: 7-11-2016

RESUMEN

Introducción. En pediatría, es determinante establecer precozmente la gravedad de la obstrucción bronquial (GOB).

Objetivo. Evaluar las propiedades psicométricas de escalas de medición de GOB en pacientes pediátricos.

Población y método. Revisión sistemática de estudios sobre validez y confiabilidad de escalas de GOB realizados en niños menores de 3 años. Se buscó en Medline, WoS, EMBASE, SciELO, Google Scholar. Los coeficientes de correlación de cada artículo fueron metaanalizados mediante el modelo de efectos aleatorios para determinar la validez de criterio y la confiabilidad a través de promedios ponderados de los coeficientes según el tamaño de la muestra.

Resultados. Se incluyeron 9 artículos con un total de 2699 niños; 3 artículos presentaron calidad metodológica buena o excelente. Cuatro artículos determinaron la validez de criterio concurrente considerando la saturación de oxígeno; coeficiente de correlación ponderado -0,627 (IC 95%: de -0,767 a -0,431; $p < 0,001$); 2 artículos determinaron la validez de criterio convergente; coeficiente de correlación ponderado 0,809 (IC 95%: de 0,721 a 0,871; $p < 0,001$); 6 artículos determinaron la confiabilidad interobservador; coeficiente de correlación ponderado de 0,500 para kappa y 0,891 para coeficiente de correlación intraclase.

Conclusión. La evaluación de las propiedades psicométricas para apoyar la estimación del constructo GOB mostró una validez de criterio de moderada a adecuada. El porcentaje de acuerdo entre observadores respecto de la medida observada (GOB) se muestra adecuado; sin embargo, hay que considerar las debilidades presentadas en el diseño de los artículos, lo cual podría afectar a la validez interna de los resultados.

Palabras clave: reproducibilidad de resultados, enfermedades pulmonares obstructivas, fiabilidad de los resultados, validez de los resultados, escalas.

<http://dx.doi.org/10.5546/aap.2017.241>

Texto completo en inglés:

<http://dx.doi.org/10.5546/aap.2017.eng.241>

Cómo citar: Luarte-Martínez S, Rodríguez-Núñez I, Astudillo P, et al. Propiedades psicométricas de las escalas de gravedad de obstrucción bronquial en pediatría. Revisión sistemática y metaanálisis. *Arch Argent Pediatr* 2017;115(3):241-248.

INTRODUCCIÓN

La bronquiolitis aguda (BA) es una enfermedad frecuente en la infancia y llega a ser la principal causa de ingreso por infección respiratoria aguda (IRA) de vías bajas en niños menores de 2 años.¹ En Latinoamérica, las IRA constituyen la primera causa de hospitalización en pediatría (98% son secundarias a infección del tracto respiratorio inferior).^{2,3} Entre las IRA bajas, se destacan los cuadros bronquiales obstructivos, incluidos el síndrome bronquial obstructivo agudo (SBO) y la bronquiolitis.^{2,4,5} El SBO se caracteriza por obstrucción respiratoria aguda con sibilancias, por lo general, de etiología viral;^{4,6} y la bronquiolitis es el primer episodio de obstrucción en lactantes, cuyo diagnóstico es preferentemente clínico basado en la historia, signos y síntomas del paciente.⁶

Uno de los factores que determinan el curso clínico de las IRA en lactantes es el establecimiento precoz de la gravedad de la obstrucción bronquial (GOB). Por este motivo, se han desarrollado diversas escalas de puntuación clínica, construidas considerando distintos dominios representativos de los signos y síntomas que caracterizan estas patologías y dan origen al constructo GOB.⁷ Entre estas, se encuentran la escala de gravedad de bronquiolitis aguda,⁸ escala de Wang (EWN),⁹ escala de dificultad respiratoria, Ministerio de Salud de Argentina (EDRAR),¹⁰ escala de dificultad respiratoria (EDR),¹¹ escala de puntuación respiratoria

del Hospital Pediátrico de Wisconsin (EHW),¹² escala de Wood,¹³ escala de Tal¹⁴ y escala de Tal modificada.¹⁵ Las propiedades de estas deben considerar íntegramente el constructo GOB y presentar asociación entre el resultado de la medición y la GOB, lo cual permite categorizar a los individuos y direccionar las estrategias terapéuticas.¹⁶

Diversos estudios han evaluado las propiedades psicométricas de las escalas de GOB y mostraron resultados discrepantes en cuanto a su validez y confiabilidad.^{10-12,14,15,17} Por lo tanto, surge la necesidad de determinar las propiedades de medición de este tipo de instrumentos de manera integrada, con el propósito de evaluar la utilidad de la estimación del constructo GOB a partir de métodos indirectos en el contexto clínico.^{7,17,18}

El objetivo de esta revisión fue determinar las propiedades psicométricas de las escalas de medición de GOB en niños menores de 3 años.

POBLACIÓN Y MÉTODO

Diseño: Revisión sistemática con metaanálisis de acuerdo con la guía PRISMA (*preferred reporting items for systematic reviews and meta-analyses*).¹⁹⁻²¹

Criterios de inclusión de los estudios: Todos los artículos dirigidos a determinar la validez y confiabilidad de escalas de GOB realizados en niños menores de 3 años ingresados a un centro hospitalario por cuadro bronquial obstructivo, sin restricción de género ni raza. Fueron considerados artículos escritos en idioma inglés, español y portugués.

Criterios de exclusión de los estudios: Se excluyeron artículos con temáticas no atingentes al estudio, artículos de revisión, discusión, así como también artículos que agruparon niños con alguna patología cardiovascular o pulmonar crónica concomitante, estudios basados en pruebas diagnósticas que presentasen amplia variabilidad de criterios de referencia, puntos de corte y reporte inconstante del área bajo la curva.

Variables de respuesta consideradas en los estudios: a) La validez de criterio concurrente: si se correlacionó la escala con algún criterio exterior ("estándar de oro"), se buscó saber si el puntaje total proporcionado por la escala de GOB se acercó o no al criterio. b) La validez de criterio convergente: si las mediciones del mismo rasgo realizadas con distintos métodos correlacionaban entre sí. Los valores de las correlaciones oscilaron entre -1 y +1; cuanto más cercano a 1 (sin tener en cuenta el signo), mayor era el grado de validez y, cuanto más cercano a 0, menor era la

validez; el signo +/- dependía de la dirección en la relación. c) La confiabilidad interobservador: si correlacionaron los puntajes obtenidos entre distintos observadores; mostraba el porcentaje de acuerdo respecto de la medida observada (GOB) y corregía el factor azar, es decir, la capacidad de la escala para producir los mismos resultados independientemente de quién la utiliza (sus valores oscilan entre 0 y 1; cuánto más próximo a +1, mayor era el grado de concordancia).

Fuentes de información e identificación de los estudios: En una primera búsqueda, fueron consideradas las bases de datos Medline, WoS, EMBASE, SciELO y Google Scholar, desde sus inicios hasta noviembre de 2015. En una segunda búsqueda, fueron consideradas las bases de datos Medline y SciELO desde noviembre de 2015 hasta junio de 2016.

Fueron utilizados términos descriptores médicos (*Medical Subjects Headings*; MeSH, por sus siglas en inglés): bronquiolitis, reproducibilidad de los resultados, estadística, bronquitis viral, enfermedades pulmonares obstructivas y validación de estudios. Se utilizaron también términos libres: obstrucción bronquial, bronquiolitis aguda, bronquitis aguda, validación, reproducibilidad, confiabilidad, fiabilidad, correlación, concordancia, escala, score, score clínico. Se aplicaron los operadores booleanos *and* y *or*, y como límites de búsqueda, "humanos", "lactantes" y "niños".

Recolección de datos: La extracción de datos de los estudios que cumplieron con los criterios de inclusión fue en planilla ad hoc realizada por dos de los investigadores, independientemente uno de otro. Se registró año e idioma de la publicación, tamaño de la muestra, edad de los participantes, escala GOB evaluada, criterio de validez, confiabilidad y variable fisiológica considerada como criterio de referencia. Las diferencias en el registro de datos fueron resueltas por consenso con una tercera persona del grupo de los investigadores.

Evaluación de calidad metodológica (CM) y riesgo de sesgo: La CM fue realizada por dos de los investigadores, independientemente uno de otro, a través de la lista de verificación *Consensus-based Standards for the Selection of Health Measurement* (COSMIN), que permitió determinar la CM de estudios dirigidos a la evaluación de las propiedades psicométricas de parámetros de medición en salud. Solo fue considerada la sección de la lista enfocada en evaluar estudios de confiabilidad y validez de criterio.^{22,23}

En la clasificación, se asignó excelente, buena, razonable, pobre. Se calificó de “excelente” cuando el aspecto metodológico de la calidad del estudio fue adecuado. Se calificó de “bueno” cuando la información pertinente no se informó en el artículo, pero se podía suponer que la calidad era adecuada. Se puntuó como “razonable” si existieron dudas respecto a que la CM fuera adecuada. Se calificó como “pobre” si se aportaron pruebas de que la CM no era adecuada. Las diferencias en la evaluación de CM se resolvieron por consenso con una tercera persona del grupo de los investigadores.

El riesgo de sesgo de publicación fue determinado mediante correlación entre la magnitud del valor absoluto del índice estadístico que determinaba la propiedad de medición y el tamaño de muestra, a través del cálculo del coeficiente de correlación (CC) de rangos tau de Kendall (prueba de correlación de rangos de Begg y Mazumdar). Para ello, todos los coeficientes fueron multiplicados por -1. Adicionalmente, fue construido el gráfico de embudo para los índices de validez de criterio con el fin de determinar el riesgo de sesgo de selección, en el que la línea vertical representaba la media ponderada de los coeficientes y las líneas diagonales representaban los límites (IC 95%) de la distribución esperada en ausencia de sesgo de selección.²⁴

Análisis estadístico: El análisis estadístico de los datos fue realizado en el software MedCalc, versión 15.8 (MedCalc Software bvba, Ostend, Belgium; <https://www.medcalc.org>; 2015). Se realizó estadística descriptiva mediante el cálculo de promedio y desviación estándar para variables cuantitativas y porcentaje para variables categóricas.

Como validez de criterio concurrente, se consideró la correlación bivariada (r de Pearson). Como índice de confiabilidad, fue considerado el coeficiente de correlación intraclase (CCI) y el coeficiente kappa.

El metaanálisis de los estudios que determinaron la validez de criterio fue realizado de acuerdo con el método de Hedges-Olkin,²⁵ transformando la z de Fisher de los CC. La inconsistencia fue calculada a través del estadístico I². Considerando la divergencia existente en cuanto a la CM entre artículos, el metaanálisis fue realizado basado en el modelo de efectos aleatorios. Para el metaanálisis de los estudios que determinaban la confiabilidad, fue calculado el promedio ponderado (PP) del CCI y de la estadística kappa según el tamaño

de muestra, de acuerdo con la sumatoria de los coeficientes ponderados de cada artículo (β). El cálculo de PP del CCI y la estadística kappa es la suma de los coeficientes ponderados de acuerdo con la siguiente fórmula:

$$PP = \Sigma\beta = \frac{Q_i * n_i}{\Sigma n_i}$$

Las abreviaciones:

i: representa el artículo.

Q: representa el coeficiente de confiabilidad utilizado (CCI o kappa).

n: representa el número de sujetos.

Σn_i: representa la suma de todos los “n” de los artículos que usaron Q.

Consideraciones éticas: Se realizó enmascaramiento de los autores, centros de estudio y títulos de los artículos primarios para evitar sesgos de selección y análisis.

RESULTADOS

Selección de los estudios: La búsqueda arrojó 679 artículos. De ellos, 275 fueron encontrados en la base de datos Medline; 11, en WoS; 17, en EMBASE; 11, en SciELO; y 365, en Google Scholar. Se excluyeron artículos con temática no atingente y duplicación; quedaron 30 resúmenes potencialmente relevantes y, de ellos, 19 artículos fueron excluidos porque consideraban adultos, con patología crónica concomitante sin incluir análisis por subgrupo, no definieron la escala utilizada o no cumplieron con algún criterio de selección. De los 11 artículos (textos completos), 2 fueron excluidos por incluir adultos y análisis estadístico mediante curva receiver operating characteristic (ROC) (Figura 1).

Características de los artículos: De los 9 artículos seleccionados, 7 estaban escritos en idioma inglés y 2, en español. El rango de años de publicación fue entre 1999 y 2015. El tamaño muestral encontrado en los estudios fue entre 36 y 1765 participantes, y no fue reportado en 1 artículo. En 8 de 9 artículos, se reportó el promedio de edad y, en 5 de 9 artículos, el rango de edad. El PP de la edad fue 4,2 meses (el PP de la edad máxima y mínima fue 6,3 y 1,7 meses, respectivamente).

En relación con las propiedades de medición evaluadas, 3 artículos determinaron solo la validez de criterio (n= 374; 12,9%); 4, solo la confiabilidad (n= 2417; 83,3%); y 2, ambas propiedades (n= 108; 3,7%) (Tabla 1).

CM y riesgo de sesgo: Solo 3 presentaron una CM buena o excelente; el resto presentó una CM

razonable o pobre (Tabla 1). Respecto al riesgo de sesgo de publicación, se observó un bajo nivel de correlación entre el valor absoluto de los CC y el tamaño de muestra de los estudios. En este

sentido, el CC tau de Kendal para los estudios de validez de criterio fue -0,447 ($p=0,1415$) y, para los estudios de confiabilidad, fue -0,414 ($p=0,1734$), lo que descartó la existencia de sesgo

FIGURA 1. Diagrama de flujo de los estudios primarios

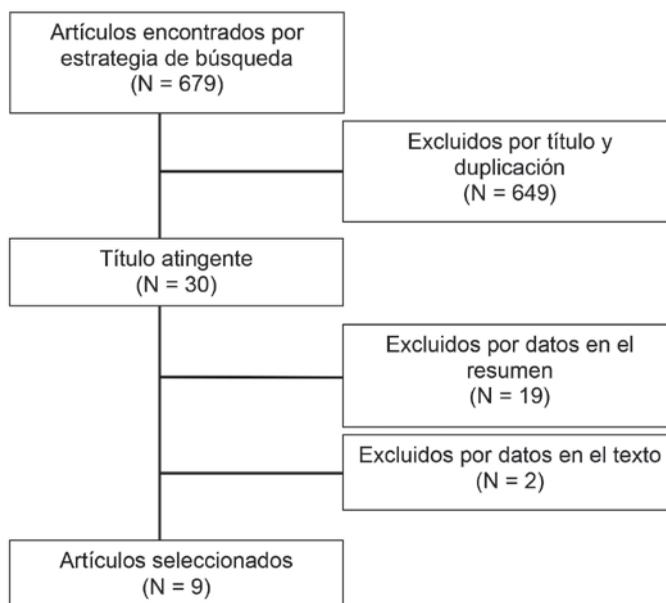


TABLA 1. Características de los artículos primarios incluidos en la revisión sistemática. N= 9

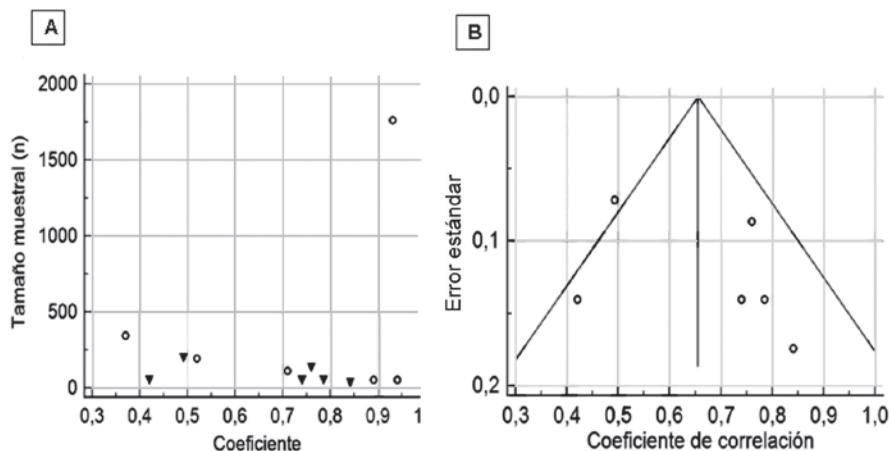
| Autor, año | N.º de casos (% hombres) | Edad (meses) | Escala | Propiedad de medición | Criterio de referencia | Estadística | CM |
|-------------------|--------------------------|--------------|--------------------------------------------------|---------------------------------|----------------------------------|---------------|-----------|
| Ho Jen Chin, 2004 | 54 (63) | 6-18 | Kristjansson Wang | Validez de criterio concurrente | Saturación de O ₂ | R de Pearson | Razonable |
| | | | | Confiabilidad interobservador | NA | CCI | Razonable |
| Pavón, 1999 | 138 (66) | 1-24 | Tal modificada | Validez de criterio concurrente | Saturación de O ₂ | R de Pearson | Razonable |
| Fernández, 2015 | 1765 (62) | < 12 m | Dificultad respiratoria | Confiabilidad interobservador | NR | CCI | Razonable |
| Camargo, 2014 | 36 (55) | < 24 m | Tal modificada | Validez de criterio convergente | Escala de Wood-Downes modificada | R de Pearson | Razonable |
| McCallum, 2012 | 112 (65) | < 24 m | Tal Tal modificada | Confiabilidad interobservador | NR | Kappa | Excelente |
| Duarte, 2012 | 54 (56) | < 24 | Wood-Downes modificada | Validez de criterio convergente | Escala de Tal | R de Spearman | Razonable |
| | | | | Confiabilidad interobservador | NA | Kappa | Razonable |
| Destino, 2012 | 195 (60) | < 12 | Hospital de Wisconsin Dificultad respiratoria | Confiabilidad interobservador | NA | CCI | Buena |
| Urzúa, 2002 | 345 (55) | 1-24 | Tal modificada | Confiabilidad interobservador | NA | Kappa | Pobre |
| Coarasa, 2010 | NR | 1-24 | Dificultad respiratoria Argentina | Validez de criterio concurrente | Saturación de O ₂ | R de Spearman | Excelente |

CM: calidad metodológica. NR: no reporta. NA: no aplica. CCI: coeficiente de correlación intraclase. O₂: oxígeno.

de publicación en los estudios (Figura 2. A). Por su parte, el gráfico de embudo reveló que la mayoría de los estudios de validez se ubicaron dentro del límite de confianza para la inexistencia de sesgo de selección. Solo un artículo relacionado con validez de criterio concurrente de las escalas GOB se ubicó fuera del límite de confianza (Figura 2. B).

Escalas GOB identificadas: 9 fueron identificadas: escala de Kristjansson (EK) (n = 54; 1,9%),²⁶ EWN²⁶ (n = 54; 1,9%), escala de Tal²⁷ (n= 112; 3,9%), escalas de Tal modificadas por McCallum²⁸ (ETMc) (n= 112; 3,9%), escala de Tal modificada por Pavón¹⁵ (ETMP) (n= 138; 5,1), EDR^{11,12} (n = 1960; 67,6%), escala de Wood-

FIGURA 2. Riesgo de sesgo de los estudios



2. A: Riesgo de sesgo de publicación mediante la correlación entre los índices de validez y confiabilidad con el número de sujetos de estudios. Los triángulos representan los estudios que determinan la validez de criterio y los círculos, los estudios que determinan la confiabilidad de las escalas de gravedad de obstrucción bronquial.

2. B: Riesgo de sesgo de selección de los estudios de validez mediante el gráfico de embudo.

TABLA 2. Características y propiedades psicométricas de las escalas identificadas

| Escala | Ítems evaluados | Rango de puntaje | Validez (r) | Confiabilidad |
|--------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|---------------|---------------------|
| Kristjansson | FR, retracciones, sonido pulmonar, color de piel, condición general | 0-8 | -0,75 | CCI: 0,89 |
| Wang | FR, sibilancias, retracciones, condición general | 0-12 | -0,42 | CCI: 0,99 |
| Tal (original) | FR, sibilancias, cianosis, uso de musculatura accesoria | 0-12 | NR | Kappa: 0,72 |
| Dificultad respiratoria | Sibilancias, retracciones | 0-17 | NR | CCI: 0,93 CCI: 0,30 |
| Wood-Downes modificada | Saturación de O ₂ , murmullo pulmonar, sibilancias, uso de musculatura accesoria, función cerebral | 0-10 | 0,76 | Kappa: 0,89 |
| Hospital de Wisconsin | Sonidos respiratorios, disnea, retracciones, FR, FC, necesidad de O ₂ , apariencia, capacidad para toser secreciones, radiografía de tórax, sonidos pulmonares, estado quirúrgico | 0-24 | NR | CCI: 0,73 |
| Tal modificada por McCallum | FR, sibilancias, saturación de O ₂ , uso de musculatura accesoria | 0-12 | NR | Kappa: 0,70 |
| Tal modificada por Pavón | FR según edad, sibilancias, uso de musculatura accesoria, cianosis | 0-12 | -0,76 0,84 | Kappa: 0,37 |
| Dificultad respiratoria de Argentina | FR, FC, sibilancias, retracción costal | 0-12 | -0,492 | NR |

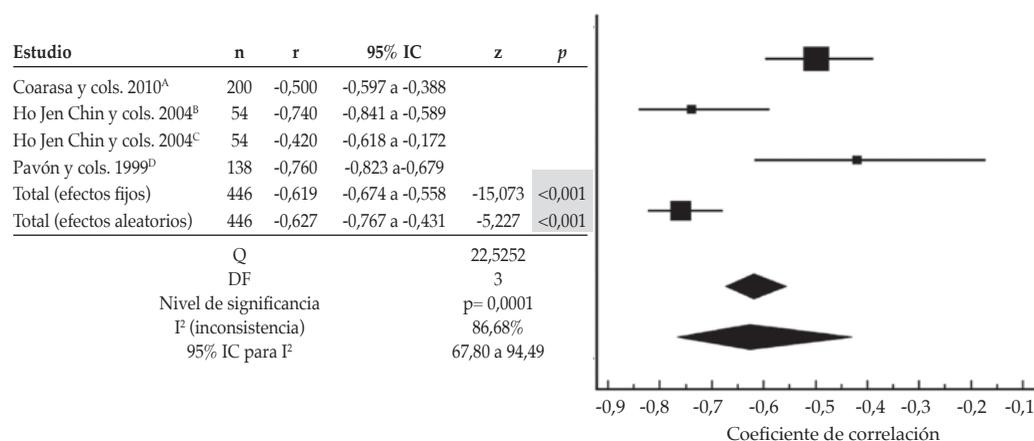
FR: frecuencia respiratoria. FC: frecuencia cardíaca. NR: no reportado. CCI: coeficiente de correlación intraclase. O₂: oxígeno. r: correlación. Coeficientes con signo negativo son referidos a validez de criterio concurrente y coeficientes de correlación positivos, a validez de criterio convergente reportada.

Downes modificada^{28,29} (EWDM) (n= 54; 1,9%), EHW¹² (n= 195; 6,7%) y la EDRAR¹⁰ (n = 200; 6,9%). En la *Tabla 2*, se pueden observar las escalas identificadas, sus características metodológicas y propiedades psicométricas.

Validez de criterio concurrente y convergente: 4 estudios (n= 392; 13,5%) determinaron la validez de criterio concurrente,^{10,15,27,28} todos consideraron la saturación de oxígeno (O₂) como criterio de referencia. Un estudio²⁶ determinó la validez de criterio concurrente de dos escalas, las cuales

fueron consideradas por separado en el análisis. Considerando las divergencias en la CM entre artículos, el modelo de efectos aleatorios arrojó un CC ponderado de -0,627 (IC 95%: de -0,767 a -0,431; p < 0,001) (*Figura 3*). Además, dos estudios (n= 90; 3,1%) determinaron la validez de criterio convergente, utilizando como criterio de referencia la escala de Tal³⁰ y la EWDM.²⁹ En ellos, el modelo de efectos aleatorios arrojó un CC ponderado de 0,809 (IC 95%: de 0,721 a 0,871; p < 0,001) (*Figura 4*).

FIGURA 3. Validez de criterio concurrente de las escalas de obstrucción



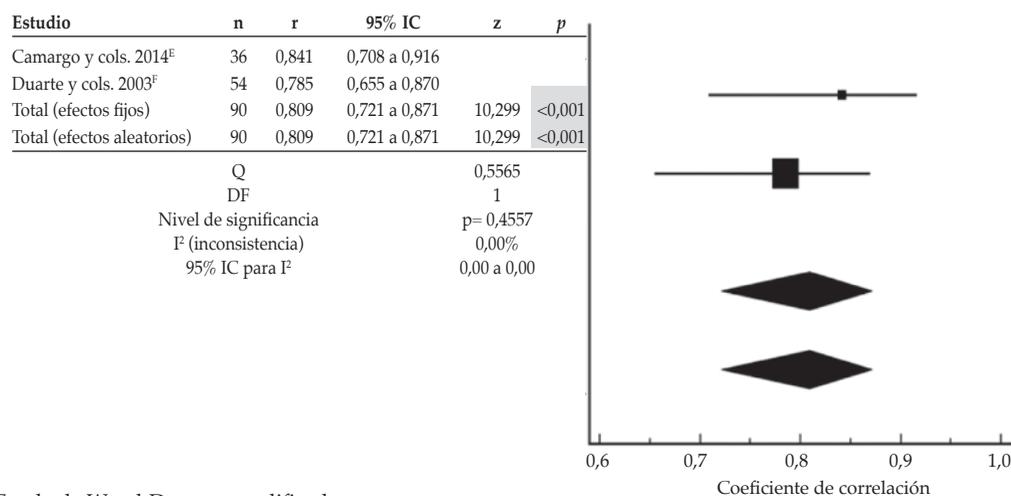
^A Escala de dificultad respiratoria, Ministerio de Argentina (EDRAR).

^B Escala de Kristjansson (EK).

^C Escala de Wang (EW).

^D Escala de Tal modificada.

FIGURA 4. Validez de criterio convergente de las escalas de obstrucción



^E Escala de Wood-Downes modificada.

^F Escala de Tal.

Confiabilidad interobservador y CCI: 6 artículos determinaron la confiabilidad interobservador de las escalas de GOB. Tres consideraron la estadística kappa^{27,28,30} (n= 511; 17,6%) y tres consideraron el CCI como índice de confiabilidad^{11,12,26} (n= 2015; 69,5%). En tres artículos, fue determinada la confiabilidad interobservador de dos escalas,^{12,27,28} en los cuales se consideró la media ponderada del rendimiento de ambas como índice de confiabilidad; los coeficientes ponderados calculados fueron 0,500 para kappa y 0,891 para CCI (Tabla 3).

DISCUSIÓN

Respecto a la evidencia de las propiedades psicométricas para apoyar la utilidad de estos métodos indirectos para la estimación del constructo GOB, podríamos decir que la validez de criterio se mostró de moderada a adecuada y el porcentaje de acuerdo entre observadores respecto de la medida observada (GOB) fue adecuado. Estos resultados observados en el análisis de validez de criterio concurrente concordaron con estudios realizados previamente bajo el enfoque de pruebas diagnósticas. McCallum y cols.,²⁸ observaron un rendimiento diagnóstico moderado de la escala de Tal (AUC = 0,69), considerando la saturación periférica de O₂ como estándar de referencia. Por su parte, Destino y cols.,¹² reportaron una sensibilidad y especificidad de 65% de la EHW, así como también una curva ROC calculada de 0,68, rendimiento similar al observado en nuestro estudio. Sin embargo, el estándar de referencia

empleado fue el requerimiento de hospitalización según la gravedad del cuadro. Por su parte, Puebla y cols.,³¹ establecieron una sensibilidad y especificidad de 77% y 88%, respectivamente, de la escala de Tal modificada, considerando como estándar de referencia la impresión clínica del médico residente.

Un criterio de referencia ampliamente utilizado para evaluar la validez de criterio concurrente de estas escalas ha sido la saturación de O₂. No obstante, Pavón y cols.¹⁵ observaron que, de los dominios existentes en la escala de Tal modificada, la cianosis tenía el menor nivel de correlación con la saturación periférica (r = -0,38). Por el contrario, estudios dirigidos a la determinación de la consistencia interna de la escala han registrado valores aceptables del alfa de Cronbach²⁷ (cianosis= 0,75 y saturación periférica= 0,72). En relación con la validez de criterio convergente, ha sido evaluada mediante la correlación entre los puntajes de dos escalas, una de ellas se escoge como estándar de referencia debido a la alta calidad de sus propiedades psicométricas establecidas en estudios precedentes.^{28,29}

Entre las limitaciones del estudio, la mayoría de los estudios considerados en esta revisión presentaron una CM razonable, principalmente, debido a debilidades en el diseño metodológico y en el proceso de ejecución, lo cual podría afectar a la validez interna de las conclusiones de este estudio. Además, se verificó un alto porcentaje de heterogeneidad en el metaanálisis de validez de criterio concurrente, posiblemente, debido a los variados criterios de referencia utilizados. No podemos descartar la existencia de sesgo de selección en los estudios que determinan esta propiedad de medición, lo cual concuerda con lo observado en el gráfico de embudo, en el que un artículo (de validez de criterio concurrente) se ubicó fuera del límite de confianza. Además, en la mayoría de los estudios, los evaluadores fueron capacitados en la aplicación de las escalas, por lo tanto, es probable que exista sobreestimación en el cálculo de los coeficientes de validez de criterio.

El hecho de que la GOB sea detectada de manera adecuada con varias metodologías diferentes es un indicador de la existencia real de ese rasgo; sin embargo, la CM de los estudios deben apoyar la validez de dichas conclusiones. Por tanto, para evaluar las propiedades de estos instrumentos de medición, son necesarios futuros estudios con mejor CM. ■

TABLA 3. Promedio ponderado de los coeficientes de confiabilidad de las escalas identificadas en los artículos primarios. N= 6

| Autor, año | N | Kappa | β |
|-------------------|------|-------|---------|
| McCallum, 2012 | 112 | 0,71* | 0,156 |
| Urzúa, 2002 | 345 | 0,37 | 0,250 |
| Duarte, 2012 | 54 | 0,89 | 0,094 |
| Total ponderado | 511 | | 0,500 |
| Autor, año | N | CCI | β |
| Ho Jen Chin, 2004 | 54 | 0,94* | 0,025 |
| Destino, 2012 | 196 | 0,52* | 0,051 |
| Fernández, 2015 | 1765 | 0,93 | 0,815 |
| Total ponderado | 2015 | | 0,891 |

β : coeficientes ponderados. N: tamaño de muestra.

CCI: coeficiente de correlación intraclass.

* El coeficiente de confiabilidad corresponde al promedio ponderado de ambas escalas.

REFERENCIAS

- González de Dios J, Ochoa Sangrador C. Conferencia de Consenso sobre bronquiolitis aguda (I): metodología y recomendaciones. *An Pediatr (Barc)* 2010;72(3):221.e1-33.
- OMS. Medidas de Control de Infecciones en la Atención Sanitaria de Pacientes con Enfermedades Respiratorias Agudas en Entornos Comunitarios. Organización Mundial de la Salud. 2009. [Acceso: 8 de noviembre de 2016]. Disponible en: http://www.who.int/csr/resources/publications/WHO_HSE_GAR_BDP_2009_1/es/.
- Chile. Ministerio de Salud. Guía Clínica AUGE Infección Respiratoria Aguda Baja de Manejo Ambulatorio en menores de 5 años. Santiago: Minsal, 2013. [Acceso: 8 de noviembre de 2016]. Disponible en: <http://www.minsal.cl/portal/url/item/7220fcd4341244a9e04001011f0113b9.pdf>.
- American Academy of Pediatrics Subcommittee on Diagnosis and Management of Bronchiolitis. Diagnosis and management of bronchiolitis. *Pediatrics* 2006;118(4):1774-93.
- Ralston SL, Lieberthal AS, Meissner HC, Alverson BK, et al. Clinical practice guideline: the diagnosis, management, and prevention of bronchiolitis. *Pediatrics* 2014;134(5):e1474-502.
- McConnochie KM. Bronchiolitis. What's in the name? *Am J Dis Child* 1983;137(1):11-3.
- Bekhof J, Reimink R, Brand PL. Systematic review: insufficient validation of clinical scores for the assessment of acute dyspnoea in wheezing children. *Paediatr Respir Rev* 2014;15(1):98-112.
- Ramos Fernández JM, Córdón Martínez A, Galindo Zavala R, Urda Cardona A. Validación de una escala clínica de severidad de la bronquiolitis aguda. *An Pediatr (Barc)* 2014;81(1):3-8.
- Postiaux G, Louis J, Labasse HC, Gerroldt J, et al. Evaluation of an alternative chest physiotherapy method in infants with respiratory syncytial virus bronchiolitis. *Respir Care* 2011;56(7):989-94.
- Coarasa A, Giugno H, Cutri A, Loto Y, et al. Validación de una herramienta de predicción clínica simple para la evaluación de la gravedad en niños con síndrome bronquial obstructivo. *Arch Argent Pediatr* 2010;108(2):116-23.
- Fernández RM, Plint AC, Terwee CB, Sampaio C, et al. Validity of bronchiolitis outcome measures. *Pediatrics* 2015;135(6):e1399-408.
- Destino L, Weisgerber MC, Soung P, Bakalinski D, et al. Validity of respiratory scores in bronchiolitis. *Hosp Pediatr* 2012;2(4):202-9.
- Martín-Torres F, Rodríguez-Núñez A, Martín-Sánchez JM. Heliox Therapy in Infants With Acute Bronchiolitis. *Pediatrics* 2009;109(1):68-73.
- Tal A, Bavilski C, Yohai D, Bearman JE, et al. Dexamethasone and salbutamol in the treatment of acute wheezing in infants. *Pediatrics* 1983;71(1):13-8.
- Pavon D, Castro-Rodríguez JA, Rubilar L, Girardi G. Relation between pulse oximetry and clinical score in children with acute wheezing less than 24 months of age. *Pediatr Pulmonol* 1999;27(6):423-7.
- McDowell I, Newell C. The Theoretical and technical Foundations of Health Measurement. In: *Measuring Health: A Guide to Rating Scales and Questionnaires*. 2nd ed. New York: Oxford University Press; 1996. Págs.10-46.
- Van Miert C, Abbott J, Verheoff F, Lane S, et al. Development and validation of the Liverpool infant bronchiolitis severity score: a research protocol. *J Adv Nurs* 2014;70(10):2353-62.
- Liu LL, Gallaher MM, Davis RL, Rutter CM, et al. Use of a respiratory clinical score among different providers. *Pediatr Pulmonol* 2004;37(3):243-8.
- Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010;8(5):336-41.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009;151(4):W65-94.
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700.
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63(7):737-45.
- Terwee CB, Mokkink LB, Knol DL, Ostelo RW, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21(4):651-7.
- Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002.
- Rodríguez I, Zambrano L, Manterola C. Validez de criterio de las escalas de medición de esfuerzo percibido en niños sanos: una revisión sistemática y metaanálisis. *Arch Argent Pediatr* 2016;114(2):120-8.
- Chin HJ, Seng QB. Reliability and validity of the respiratory score in the assessment of acute bronchiolitis. *Malays J Med Sci* 2004;11(2):34-40.
- McCallum GB, Morris PS, Wilson CC, Versteegh LA, et al. Severity scoring systems: are they internally valid, reliable and predictive of oxygen use in children with acute bronchiolitis? *Pediatr Pulmonol* 2013;48(8):797-803.
- Duarte-Dorado DM, Madero-Orostegui DS, Rodríguez-Martínez CE, Nino G. Validation of a scale to assess the severity of bronchiolitis in a population of hospitalized infants. *J Asthma* 2013;50(10):1056-61.
- Camargo Crespo C. Validación de una escala de severidad en bronquiolitis viral aguda en una población de lactantes atendidos en el hospital de la Misericordia. Bogotá: Universidad Nacional de Colombia; 2014. [Acceso: 8 de noviembre de 2016]. Disponible en: <http://www.bdigital.unal.edu.co/12343/1/5599350.2014.pdf>.
- Urzúa S, Duffau G, Zepeda G, Sagredo S. Estudio de concordancia clínica en educandos de pre y postítulo en Pediatría: Puntaje de Tal. *Rev Chil Pediatr* 2002;73(5):471-7.
- Puebla Molina S, Bustos L, Valenzuela M, Hidalgo M, et al. La escala de Tal como test diagnóstico y el diagnóstico clínico como gold standard en el síndrome bronquial obstructivo del lactante. *Rev Pediatr Aten Primaria* 2008;10(37):45-53.