

Estudio de concordancia clínica en educandos de pre y postítulo en Pediatría. Puntaje de Tal

Soledad Urzúa B.¹, Gastón Duffau T.¹,
Guillermo Zepeda. F.², Susana Sagredo J.²

Resumen

Los programas de formación médica deberían alcanzar criterios y destrezas en los educandos, que condujeran a evaluaciones concordantes ante una misma situación clínica. *Hipótesis:* Un sistema de puntaje (Tal) para calificar estado clínico en patología respiratoria, que agrupa diferentes signos, genera mayor concordancia entre observadores que cada elemento integrante (diferencia de 0,20 o mayor en el índice kappa). *Material y método:* Muestra: parejas de internos (n = 70) y médicos alumnos del Programa de Especialización en Pediatría (APE) (n = 30), evaluaron a 345 niños con patología respiratoria aguda (167 y 178 cada grupo respectivamente), según puntaje de Tal modificado. Se evaluó el sistema de puntaje, consignando además edad del paciente, género, estado nutricional y diagnósticos de ingreso. *Estudio estadístico:* Comparación de proporciones por χ^2 y de índices kappa. *Resultados:* El puntaje de Tal, por internos, mostró kappa de 0,515 y los signos individuales, de 0,080 a 0,379. En APE, 0,230 y 0,163 a 0,277, respectivamente. Una agrupación con menos categorías no mostró ventajas. Ambos grupos de examinadores sólo difirieron en el puntaje total, más concordante entre internos que médicos. Al comparar kappa para cada signo con el resultado del puntaje de Tal, se detectó cambio significativo sólo en internos y para sibilancias y cianosis, en cuatro categorías ($p < 0,05$). *Conclusión:* La concordancia encontrada es escasa, tanto para un solo signo como para el conjunto y no mejora al reducir las categorías de evaluación. La propuesta de Tal genera mejor concordancia que los signos aislados sólo en internos y para cianosis o sibilancias. Es indispensable revisar los objetivos y métodos de los programas docentes, para alcanzar entre los educandos, un nivel de acuerdo mayor en la valoración clínica de los enfermos. (**Palabras clave:** Concordancia, Puntaje de Tal). Rev. Chil. Pediatr. 73 (5); 471-477, 2002.

A study of clinical concordance in pre and postgraduate students in Paediatrics. Tal score

Medical education programmes should result in the uniform development of clinical skills and criteria leading to a concordant assessment of clinical situations. *Hypothesis:* a scoring system (Tal) for assessing the clinical status for a group of signs on physical examination will produce a higher level of concordance between observers than each individual sign (a kappa index difference of > 0.20). *Patients and Methods:* sample

1. Departamento de Pediatría y Cirugía Infantil, Facultad de Medicina, Universidad de Chile. Hospital Clínico de Niños Roberto del Río. Santiago de Chile.
2. Interno de Medicina. Facultad de Medicina, Campus Norte, Universidad de Chile.

Trabajo recibido el 31 de mayo de 2002, devuelto para corregir el 15 de julio de 2002, segunda versión el 01 de agosto de 2002, aceptado para publicación el 11 de septiembre de 2002.

size; according to previous data, with a standard error of the kappa index of 0.08, a difference within and between groups of 0.20, which was confirmed in a pilot study, estimated the sample size to be between 74 to 126 cases per group, with an alpha error of 5%. Observers were a random group of medical interns ($n = 70$) and medical doctors in the Paediatric Specialty Programme ($n = 30$). The scoring system was assessed, considering the age of the patients, gender, nutritional status and diagnosis on admission. *Statistics:* Kappa indexes, χ^2 in comparison of proportions. *Results:* 345 infants and children admitted with lower respiratory tract infections were examined, 167 by interns and 178 by medical doctors, both groups being similar at the initial examination. The Tal score in interns resulted in a kappa index of 0.515 and for individual signs from 0.080 to 0.379 and in doctors 0.230 and 0.163 to 0.277 respectively. A significant difference was observed for the Tal score in interns with a higher kappa index. In comparing concordance for each sign against the total score, there was only a significant change in the intern group, and for cyanosis and wheezing in 4 categories ($p < 0.05$). *Conclusions:* The observed concordance was rather weak for isolated signs and also for the complete scoring system of Tal. This gave a better concordance for individual signs only in interns and for cyanosis and wheezing. Teaching goals and methods to develop clinical skills should be reviewed.

(Key words: concordance, Tal score). Rev. Chil. Pediatr. 73 (5); 471-477, 2002.

INTRODUCCIÓN

En un estudio anterior¹, se evaluó el grado de concordancia clínica entre educandos de pre y postítulo para manifestaciones del examen físico de común ocurrencia en pacientes pediátricos hospitalizados. Los resultados obtenidos mostraron un bajo nivel de acuerdo entre ambos grupos, que mejoraba en los médicos alumnos del Programa de Especialización en Pediatría (APE) al reducir el número de categorías de valoración.

Una de las metas de los programas de formación médica tanto de pre como postítulo, debería ser alcanzar criterio y habilidades semejantes en los grupos, así como una evaluación concordante de un mismo paciente o situación clínica.

Si bien el estudio previo reveló un escaso nivel de acuerdo, esto no pareció reflejarse en el resultado final de la atención del paciente, dado tal vez por las múltiples evaluaciones que de él se hacen durante la evolución y porque estas se realizan por distintas personas, disminuyendo así la probabilidad de error, dando más opción para corregir rumbo.

La estructuración de diversos hallazgos clínicos en un sistema organizado, que genera un puntaje total, como el propuesto por Tal y colaboradores², podría mejorar el resultado de una evaluación global reflejada en el puntaje obtenido y la concordancia entre distintos observadores, comparado con

la valoración de los elementos clínicos independientes.

Hipótesis: Un sistema de puntaje estructurado para calificar estado clínico en patología respiratoria, que agrupa diferentes signos, genera una evaluación global más concordante entre observadores que la valoración de los signos integrantes, individualmente (diferencia $\geq 0,20$ en el índice kappa).

MATERIAL Y MÉTODO

El estudio se desarrolló en el Departamento de Pediatría, Campus Norte, Facultad de Medicina, Universidad de Chile con internos de la carrera de Medicina y médicos APE, en formación durante tres años, correspondiendo el campo clínico al Hospital Clínico de niños "Roberto del Río".

El tamaño de la muestra se determinó efectuando una estimación del tamaño muestral empleando información obtenida anteriormente¹, considerando un error estándar de kappa de aproximadamente 0,08, buscando detectar una diferencia intergrupo del índice, de un grado (0,20) e intragrupo un valor de kappa de igual magnitud, como significativamente diferente de cero. Un nuevo estudio piloto confirmó las presunciones formuladas y arrojó una estimación de muestra (centrada en la presencia de sibilancias) de 102 casos variando de 74 a 176 niños, para error alfa de 5%.

Los grupos a estudiar se integraron de

las combinaciones posibles para conformar parejas de observadores en el nivel de internos de pediatría ($n = 70$) y médicos APE de primero, segundo y tercer año del programa ($n = 30$). Cada pareja estuvo representada en la muestra por un número similar de observaciones.

Durante los años 1999 y 2000, bajo la supervisión de uno de los investigadores, las parejas de educandos evaluaron pacientes hospitalizados según los elementos propuestos en el sistema de puntaje de Tal modificado³. Este incluye los siguientes criterios: frecuencia respiratoria, sibilancias, cianosis y retracción.

El primero de ellos es valorado con rangos de frecuencia según la edad (menor y mayor o igual a 6 meses), el resto de los signos, según presencia e intensidad, como: (-), (+), (++) y (+++). Generando resultados que varían de 0 a 12 puntos.

La información obtenida se vació a un formulario *ad-hoc* en forma independiente por cada observador de modo que cada uno ignorase la evaluación del otro. Asimismo, no conocían la hipótesis en verificación. El formulario contenía, además, edad, género, peso, talla y estado nutricional por criterio peso/edad o peso/talla según curvas del NCHS y procedimiento sugerido por el MINSAL⁴. Se consignaron, finalmente, los diagnósticos del paciente, pareja participante y fecha.

El lapso entre el examen efectuado por un observador y el siguiente, fue el menor posible y siempre inferior a 15 minutos.

El estudio de concordancia genera el índice kappa⁵, que se divide en tramos de 0 a 1 y se interpretan como: pobre (0,00 a 0,20), regular (0,21 a 0,40), buena (0,41 a 0,60), substancial (0,61 a 0,80) y casi perfecta o perfecta (0,81 a 1,00).

Para el estudio estadístico de los resultados se recurrió a los programas Epidat 2.0⁶ y WinEpiscopes 1.0⁷ obteniéndose kappa con su error estándar, intervalo de confianza del 95% (IC 95%) y el error estándar de kappa = 0. El cálculo del índice expresa el valor de éste corregido por azar aunque no ponderado. La comparación intergrupo siguió el procedimiento sugerido por Cohen⁸. χ^2 en comparación de proporciones. En este estudio se aceptó un error alfa de 5%.

RESULTADOS

Ingresaron a la investigación 345 niños

que constituyeron 3 grupos: uno de 167 casos que fueron evaluados por internos de medicina, un segundo grupo constituido por 120 pacientes evaluados por médicos APE en su primer año y por último 58 casos evaluados por médicos APE de segundo o tercer año del programa.

La distribución de los enfermos por edad, género, estado nutricional y diagnósticos de ingreso en cada grupo se muestra en las tablas 1 y 2. La mayor parte de los niños presentaba síndrome bronquial obstructivo asociado o no a bronconeumonía. Los tres grupos resultaron comparables en los aspectos mencionados.

Por no detectar diferencias significativas entre los grupos de médicos APE en las demás variables estudiadas, se presentan en conjunto.

En la tabla 3, se observa el índice kappa en internos y médicos APE, mostrando en general niveles bajos de concordancia ($\leq$ regular), con valores incluso negativos en el intervalo de confianza del 95% (IC 95%). La mejor concordancia se obtuvo en relación al puntaje total en el conjunto de internos ($k = 0,515$) y en segundo lugar el signo clínico frecuencia respiratoria ($k = 0,379$), en el mismo grupo. Para el resto de los elementos, los valores del índice fueron menores de 0,325.

En la comparación de internos con médicos APE, la única diferencia significativa se produjo entre puntajes Tal, con mayor concordancia entre los primeros. En ese grupo, por otra parte, se encontró kappa para puntaje global mejor que para signos individuales frente a cianosis y sibilancias.

Los índices de concordancia encontrados intragrupo e intergrupo para cada rubro analizado, son diferentes de cero para la mayoría de ellos, exceptuando cianosis en internos y médicos.

La situación de bajo acuerdo detectada, podría atribuirse en parte al número de categorías de cada signo clínico. Para verificar tal planteamiento se establecieron sólo 2 niveles para sibilancias, retracción y cianosis (presente o ausente) y 3 para frecuencia respiratoria (normal, aumento leve-moderado e intenso). El puntaje total se agrupó en 2 categorías: cero o leve (0-5) y moderado o intenso (6-12).

La tabla 4 muestra que la concordancia parece mejorar, pero no significativamente, destacando que alcanza un nivel de "buena"

Tabla 1. Distribución por edad, género y estado nutricional de 345 niños examinados por internos de Medicina y médicos APE

Género	Internos		Grupo examinado por: Médicos			
	n	%	APE I		APE II y III	
			n	%	n	%
Masculino	89	53,3	75	62,5	27	46,6
Femenino	78	46,7	45	37,5	31	53,4
Total	167	100,0	120	100,0	58	100,0

 χ^2 , 2 gl. = 4,58. p = 0,11**Edad (meses)**

	n	%	n	%	n	%
< 3	32	19,2	25	20,8	13	22,5
3-5	37	22,2	32	26,7	14	24,1
6-11	34	20,4	31	25,8	14	24,1
12-23	41	24,5	21	17,5	14	24,1
≥ 24	23	13,7	11	9,2	3	5,2
Total	167	100,0	120	100,0	58	100,0

 χ^2 , 8 gl. = 6,90. p = 0,54**Estado Nutricional**

	n	%	n	%	n	%
Obesidad	6	3,6	7	5,8	1	1,7
Sobrepeso	19	11,4	15	12,5	8	13,8
Normal	91	54,5	63	52,5	30	51,7
Riesgo de desnutrición	30	17,9	21	17,5	11	19,0
Desnutrición	21	12,6	14	11,7	8	13,8
Total	167	100,0	120	100,0	58	100,0

 χ^2 , 8 gl. = 2,30, p = 0,97

APE I, II, III = Alumnos del Programa de Especialización en Pediatría de primer, segundo y tercer año, respectivamente.

Tabla 2. Diagnósticos de ingreso en 345 niños hospitalizados por patología respiratoria

Diagnóstico	Internos	Grupo examinado por: Médicos	
		APE I	APE II-III
Bronconeumonía c/ SBO	126	94	48
Bronconeumonía s/ SBO	20	12	4
SBO	13	12	5
Neumonía	8	2	1
Total	167	120	58

APE I, II, III = Médicos alumnos del Programa de Especialización en Pediatría, de primero, segundo y tercer año, respectivamente. SBO = Síndrome bronquial obstructivo. χ^2 , 6 gl = 4,4, p = 0,62

Tabla 3. Concordancia clínica en internos de Medicina y médicos APE. Puntaje de Tal y según cada componente

Signo Clínico (4 categorías)	Grupo	Índice kappa	IC 95%	I v/s APE p
F. Respiratoria	I	0,379	0,268 a 0,490	0,06
	M	0,257	0,152 a 0,361	
Sibilancias	I	0,235#	0,109 a 0,361	0,31
	M	0,277	0,166 a 0,387	
Cianosis	I	0,080*	-0,068 a 0,229	0,31
	M	0,163	-0,145 a 0,472	
Retracción	I	0,325	0,211 a 0,438	0,08
	M	0,214	0,110 a 0,319	
Puntaje de Tal*	I	0,515#*	0,254 a 0,776	0,03
	M	0,230	0,060 a 0,400	

I = Internos. Examinaron 167 niños.

M = Médicos APE. Examinaron 178 niños.

* = 3 categorías (0-5 puntos: Leve, 6-8: Moderado, 9-12: Grave).

#* = $p < 0,05$.

Tabla 4. Concordancia clínica de internos de Medicina y médicos APE. Puntaje de Tal y componentes agrupados en 2 ó 3 categorías

Signo Clínicos	Grupo	Índice kappa	IC 95%	I v/s APE p
F. Respiratoria*	I	0,484	0,350 a 0,618	0,07
	M	0,339	0,196 a 0,479	
Sibilancias**	I	0,317	0,171 a 0,462	0,19
	M	0,402	0,268 a 0,537	
Cianosis**	I	0,165•	-0,137 a 0,468	0,48
	M	0,156	-0,156 a 0,468	
Retracción**	I	0,391	0,247 a 0,534	0,36
	M	0,427	0,279 a 0,573	
Puntaje de Tal**	I	0,514•	0,250 a 0,776	0,09
	M	0,292	0,103 a 0,479	

I = Internos. Examinaron 167 niños.

M = Médicos APE. Examinaron 178 niños.

* = Tres categorías (Normal, aumento leve-moderado, intenso)

** = Dos categorías. (En puntaje de Tal, 0-5 puntos: Cero o leve y 6-12: Moderado o intenso. En los restantes: Presente o ausente).

• = $p < 0,05$.

para frecuencia respiratoria en el grupo de internos y para retracción en médicos APE. Nuevamente, el puntaje global indica en internos, mejor concordancia para cianosis pero no para sibilancias.

DISCUSIÓN

Las posibles consecuencias que puede acarrear una reducida concordancia entre observadores queda aún más notoriamente expuesta si consideramos una reciente investigación entre patólogos entrenados al informar muestras cervicouterinas potencialmente neoplásicas: Su acuerdo alcanzó kappa de sólo 0,46 para citología en monocapa, equivalente a la encontrada para biopsias⁹.

En estudios publicados desde 1966^{10,11}, se ha abordado el problema de generar sistemas de puntaje que permitan la evaluación de pacientes con patología respiratoria, de manera de unificar criterios en relación a definir la intensidad del compromiso clínico y posteriormente, el tratamiento que le corresponda. La propuesta de Tal^{2,3}, es un sistema de evaluación clínica, que agrupa varios signos, destinados a establecer la gravedad de un cuadro respiratorio y en consecuencia, permite instaurar pautas de tratamiento según el puntaje detectado.

Se ha estimado que el agrupar distintos signos clínicos para definir "gravedad" de un paciente debería lograr un mayor grado de concordancia entre observadores, que el análisis de un solo signo en forma aislada. La interpretación del hallazgo de niveles altos o bajos de concordancia fue discutido en otra parte¹.

En el presente estudio el grado de acuerdo encontrado es escaso, tanto para la evaluación de un signo como para el conjunto de éstos, sólo mejorando significativamente con el sistema de puntaje en internos y comparando el hallazgo ante cianosis y sibilancias. Es destacable, además, que la concordancia entre médicos no pareció mejorar con el tiempo de práctica clínica, dentro del margen estudiado. Incluso, no se observaron diferencias importantes con el grupo de alumnos de pre grado en los rubros individuales. Del mismo modo no mejoró la concordancia al reducir el número de categorías en cada elemento clínico.

Los factores que podrían explicar el bajo nivel de acuerdo y la diferencia encontrada

a favor del conjunto de internos incluyen: a) se reconocen distintos centros formadores de los profesionales médicos. El grupo de internos provenía de una misma escuela y habían recibido la misma enseñanza en semiología pediátrica. El grupo de médicos APE estuvo constituido por profesionales procedentes de distintas universidades o lugares de trabajo, b) en el programa de formación de especialistas, hay instancias de uniformidad en diferentes áreas pero no así para unificar criterios semiológicos, c) en la forma espontánea de evaluar el signo frecuencia respiratoria, se constató que los observadores cuantificaban de manera distinta el número de respiraciones. Es así como algunos lo hacían en 10, 15, 20, 30 y 60 segundos, lo que pudo significar una diferencia suficiente para cambiar de categoría a un mismo paciente.

A pesar de la baja concordancia encontrada, los resultados finales en el manejo de pacientes son generalmente favorables, lo que podría explicarse porque existe supervisión de actividades, el examen de los pacientes se repite varias veces durante el día y por distintos médicos, dando opción a cambiar la calificación del enfermo y por lo tanto, las decisiones que de esta derivan, o bien porque el grado de acuerdo entre los observadores, sin corrección por efecto del azar, no es bajo, lo que contribuye a generar resultados aparentemente mejores.

En conclusión podemos decir que el nivel de concordancia encontrado fue, en general, bajo y no alcanzó 0,52 o más, en ningún rubro. Los índices fueron diferentes de cero, excepto para cianosis. En el puntaje de Tal, obtuvieron mejor acuerdo internos que médicos APE para valoración de los signos en 4 categorías. Individualmente, estos signos no generaron diferencias entre ambos grupos de observadores. El puntaje de Tal contribuye a una mejor concordancia que los signos integrantes sólo en internos, para cianosis (4 y 2 categorías) y sibilancias (sólo 4 categorías).

Es necesario incluir como objetivo de los programas docentes, el alcanzar entre los educandos, un nivel de acuerdo mayor en la valoración clínica de los enfermos.

REFERENCIAS

1. Duffau GT, Navarrete LR, Fernández CC: Estudio de concordancia clínica en educandos de pre

- y postítulo en Pediatría. *Rev Chil Pediatr* 2000; 7: 340-6.
2. *Tal A, Bavilski Ch, Yohai D, Bearman J, Gorodischer R, Moses S*: Dexamethasone and salbutamol in the treatment of acute wheezing in infants. *Pediatrics* 1983; 71: 13-8.
 3. *Fielbaum CO*: (Coordinador General). Consenso Nacional para el manejo del síndrome bronquial obstructivo del lactante. *Pediatría al Día* 1998; 14: 107-16.
 4. República de Chile. Ministerio de Salud. Comisión Coordinadora del PNAC. Gráficas de evaluación antropométrica para niños(as) menores de 6 años. Santiago: Minsal, 1993.
 5. *Armitage P, Berry G*: Estadística para la investigación biomédica. 3ª Edición, Harcourt Brace S.A., Barcelona, España 1997: 419-24.
 6. Universidad de Santiago, España. Programa HDA de OPS/OMS. *Epidat* 2.0. 1997.
 7. *Blas N, Ortega C, Frankena K, Noordhuizen J*: Win Episcopa 1.0. University Wageningen. 1995.
 8. *Cohen J*: A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37-46.
 9. *Stoler M, Schiffman M*: Interobserver reproducibility of cervical cytologic and histologic interpretations. *JAMA* 2001; 285: 1500-5.
 10. *Dabbous I, Tkachyk J, Stamm S*: A double blind study on the effects of corticosteroids in the treatment of bronchiolitis. *Pediatrics* 1966; 37: 477-84.
 11. *Bierman C, Pierson W*: The pharmacologic management of status asthmaticus in children. *Pediatrics* 1974; 54: 245-7.

AVISO A LOS AUTORES

La Revista Chilena de Pediatría puede ser visitada a texto completo en la página web: www.scielo.cl en un aporte de Conicyt a las publicaciones científicas nacionales.