

Nivel de evidencia científica en la cirugía pediátrica española*

V. Ibáñez¹, V. Modesto², J. Lluna³, E. Hernández³

¹Unidad de Cirugía Pediátrica. Hospital General de Castellón. ²Sección de UCI Pediátrica. ³Servicio de Cirugía Pediátrica. Hospital Infantil «La Fe», Valencia.

RESUMEN: Introducción. El cuerpo de la literatura en cirugía pediátrica se caracteriza por un predominio de estudios observacionales y una escasez de ensayos clínicos aleatorizados que estudian principalmente terapias médicas.

Objetivo. Evaluar el nivel de evidencia científica producido en nuestro medio.

Material y métodos. Artículos publicados en la revista Cirugía Pediátrica, artículos sobre cirugía publicados en Anales Españoles de Pediatría, así como los publicados por autores españoles en Journal of Pediatric Surgery, Pediatric Surgery International y European Journal of Pediatric Surgery en el período entre 1997 y 2002. Se utilizaron las guías de lectura crítica de Sackett para evaluar la validez interna y la clasificación de niveles de evidencia del Centro de Medicina Basada en la Evidencia de la Universidad de Oxford.

Resultados. El 94,5% de los artículos sobre terapia, el 61,5% sobre pruebas diagnósticas, el 83,3% sobre pronóstico y el 52,2% sobre factores de riesgo o etiología presentaron un nivel de evidencia 4, equivalente a estudios descriptivos. Se hallaron únicamente 4 ensayos clínicos aleatorizados, todos comparando terapias médicas (antibiototerapia, analgesia y control del estrés quirúrgico).

Conclusión. Los resultados son similares a los ofrecidos por revisiones sobre revistas anglosajonas, destacando el bajo nivel de evidencia científica de la mayoría de los artículos. La difusión de guías para la comunicación de estudios descriptivos puede mejorar el nivel de evidencia de los mismos.

PALABRAS CLAVE: Medicina basada en la evidencia; Cirugía pediátrica.

EVIDENCE-BASED MEDICINE IN SPANISH PEDIATRIC SURGERY REPORTS

ABSTRACT: Introduction. Pediatric surgical literature is characterized by a high prevalence of observational studies and a paucity of randomized controlled trial, comparing mainly medical therapies.

Aims. To assess scientific evidence level produced by spanish pediatric surgeons.

Correspondencia: Dr. V. Ibáñez Pradas. Unidad de Cirugía Pediátrica. Hospital General de Castellón. Avda. Benicassim s/n, 12004 Castellón.

*Trabajo presentado en el XLII Congreso de la Sociedad Española de Cirugía Pediátrica. Valencia, 28-31 de mayo de 2003.

Recibido: Mayo 2003

Aceptado: Septiembre 2003

Material and methods. Articles reported in Cirugía Pediátrica, surgical reports in Anales Españoles de Pediatría, and reports from spanish authors published in the Journal of Pediatric Surgery, Pediatric Surgery International and European Journal of Pediatric Surgery, from 1997 to 2002. Internal validity was assessed through the critical guides from Sackett, and evidence level was classified through the Oxford Centre for Evidence Based Medicine classification.

Results. 94.5% on therapy reports, 61.5% on diagnosis, 83.3% on prognosis and 52.2% on risk factors or etiology, were classified as level 4 (descriptive studies). Only four randomized controlled trials were found, all of them comparing medical therapies (analgesia, antibiotics and surgical stress control).

Conclusions. Our results are similar to those offered by previous reviews, showing up a low level of scientific evidence in most articles. Diffusion of guides on reporting descriptive studies may improve the evidence level of our reports.

KEY WORDS: Evidence-based medicine; Pediatric surgery.

INTRODUCCIÓN

El término de medicina basada en la evidencia (MBE) fue acuñado en los años 80 en la Escuela de Medicina de la Universidad canadiense McMaster, y consiste en el proceso de búsqueda sistemática, evaluación crítica y aplicación de los hallazgos de la investigación a la toma de decisiones clínicas. Su difusión en la práctica no se produjo hasta principios de los 90 a través de los artículos publicados en la revista JAMA y progresivamente su sistemática de trabajo ha arraigado en la comunidad médica^(1, 2).

Puesto que el objetivo final es la aplicación de la mejor evidencia científica disponible para resolver los problemas individuales de los pacientes, una parte fundamental de la práctica de la MBE es la valoración crítica de los artículos, clasificando la calidad de la evidencia que aportan y, en consecuencia, la importancia de las recomendaciones que de ellos se derivan. Distintas revisiones sobre revistas anglosajonas en nuestra especialidad han hallado que el actual cuerpo de la literatura se caracteriza por la escasez de artículos con al-

Tabla I Clasificación de la evidencia según la Agencia de Evaluación de Tecnologías

1. Metaanálisis de ensayos controlados y aleatorizados.
2. Ensayo controlado y aleatorizado de muestra grande.
3. Ensayo controlado y aleatorizado con muestra pequeña.
4. Ensayo prospectivo controlado no aleatorizado (controles coincidentes).
5. Ensayo prospectivo controlado no aleatorizado (controles históricos).
6. Estudio de cohortes.
7. Estudio de casos y controles.
8. Series clínicas no controladas. Estudios descriptivos. Comités de expertos.
9. Casos únicos.

to grado de evidencia y un predominio de casos clínicos y series de datos, que constituyen el menor nivel de evidencia disponible⁽³⁻⁵⁾.

El objetivo de este trabajo es evaluar el grado de evidencia producido en nuestro ámbito, teniendo en cuenta la pregunta de estudio a la que intentan responder los autores.

MATERIAL Y MÉTODOS

Revisamos los artículos originales publicados entre los años 1997 y 2002 en las revistas de cirugía pediátrica, los firmados por autores españoles en las revistas *Journal of Pediatric Surgery*, *European Journal of Pediatric Surgery* y *Pediatric Surgery International*, y los artículos publicados en *Anales Españoles de Pediatría* seleccionados mediante los términos *surgery* o *surgical*, excluyéndose de la revisión los casos clínicos y los editoriales. Se revisó el original de los publicados en *Cirugía Pediátrica*, y el resumen proporcionado por el repertorio bibliográfico MEDLINE para las restantes. En caso de que no estuviera disponible o su contenido no permitiera una clasificación adecuada del trabajo se acudió a la publicación original. Para la clasificación del artículo se utilizó la información proporcionada por los autores en el apartado de material y métodos, y resultados.

Los trabajos se clasificaron de modo orientativo atendiendo a su diseño según la clasificación de la Agencia de Evaluación de Tecnologías del Servicio Catalán de Salud (AET)⁽⁶⁾ (Tabla I). Para establecer el grado de evidencia se utilizó la clasificación de los niveles de evidencia del centro para la MBE de la Universidad de Oxford⁽⁷⁾ (Tabla II), según el estudio valorara la efectividad de una terapia, la utilidad de una prueba diagnóstica, el pronóstico de una enfermedad o etiología-factores de riesgo.

Para todos los estudios con diseño superior a una serie de casos se valoró para la asignación de su grado de evidencia la validez interna, según las guías de lectura crítica de Sackett⁽⁷⁾ (Tabla III). En caso de que no se ajustaran a dichas guías (la de terapia está diseñada para ensayos clínicos) se valoró que los autores hubieran hecho un intento por controlar los posibles sesgos y factores de confusión (estratificación de resultados por variables relevantes o análisis multivariante).

RESULTADOS

En el período de estudio se han publicado un total de 236 artículos en las cinco revistas analizadas: 178 en *Cirugía Pediátrica*, 23 en *Anales* y 35 en las revistas internacionales. De forma global, las publicaciones más frecuentes versaron sobre cirugía general y digestivo, urología y cirugía experimental (Tabla IV). En las revistas de difusión internacional los trabajos experimentales constituyeron el tema de publicación más frecuente. Los trabajos sobre cirugía experimental se excluyeron del análisis de la evidencia, puesto que sus conclusiones no son extrapolables a la clínica.

La clasificación de los trabajos según la AET se muestra en la tabla V, con un claro predominio de las series de casos y trabajos descriptivos, que junto con los cuatro casos clínicos publicados como originales, suponen el 76,5% de la producción total.

La clasificación del nivel de evidencia se realizó según la pregunta a la que respondiera el artículo (Tabla VI), si bien 52 trabajos no fueron clasificados en ninguno de estos grupos, bien porque se trataba de series descriptivas que trataban distintos aspectos de una patología (síntomas, diagnóstico y tratamiento), bien porque su tema era claramente distinto, como resultados de encuestas o análisis bibliométrico. Tras esta asignación, el número de casos asignados a la evidencia grado 4, que es la correspondiente a las series de casos, pasó de 156 según la clasificación de la AET a 181, puesto que en 25 casos se consideró que a pesar del diseño, la evidencia proporcionada no era superior a la de un estudio descriptivo. Sólo en dos estudios de los 42 clasificados como cohortes o caso-control, se utilizó análisis multivariante (regresión logística) para el control de factores de confusión, siendo más frecuente la estratificación por algunas de las variables consideradas de confusión.

Los trabajos sobre la eficacia de una terapia o tratamiento fueron los que presentaron mayor proporción de estudios con evidencia grado 4 (94,5%). En este grupo, la máxima evidencia la proporcionaron cuatro ensayos clínicos aleatorizados y un estudio de cohortes. Todos estudiaban terapias médicas: antibioterapia en la apendicitis, analgesia (un ensayo clínico y un estudio prospectivo controlado no aleatori-

Tabla II Niveles de evidencia según el objetivo de estudio

Nivel	Terapia/Prevención Etiología/Efecto de una exposición	Pronóstico	Diagnóstico
1a	RS de ECR	RS de estudios de cohortes de inicio. Guías de decisión clínica validadas en distintas poblaciones	RS de estudios de nivel 1. Guías de decisión clínica con estudios 1b de distintos centros
1b	ECR	Cohorte de inicio	Cohorte de validación
1c	Todos o ninguno*	Series de casos todos-ninguno	SpPins y SnNouts**
2a	RS de estudios de cohortes	RS de cohortes retrospectivas	RS con estudios diagnósticos de nivel > 2
2b	Estudio de cohorte	Cohorte retrospectiva	Cohorte exploratoria
2c	Investigación de resultados en salud. Estudios ecológicos	Investigación de resultados en salud	
3a	RS de estudios caso-control		RS de estudios de grado \geq 3b
3b	Estudio caso/control		Estudio sin aplicación consistente del patrón de referencia
4	Series de casos y estudios de cohortes y caso-control de baja calidad	Series de casos y estudios de cohortes y caso-control de baja calidad	Estudio caso-control. Patrón de referencia de baja calidad o aplicado de forma no-dependiente
5	Opinión de expertos	Opinión de expertos	Opinión de expertos

*RS, revisión sistemática; ECR, ensayo clínico randomizado; *Resultados dramáticos como el que produjo la introducción de la penicilina. **SpPin, hallazgo diagnóstico cuya especificidad es tan alta que un resultado positivo confirma el diagnóstico. SnNout, hallazgo diagnóstico cuya sensibilidad es tan alta que un resultado negativo descarta el diagnóstico.*

Grados de recomendación: A, Consistente con estudios de nivel 1; B, Consistente con niveles 2 o 3. Extrapolaciones de nivel 1; C, Nivel 4. Extrapolaciones de estudios de nivel 2 o 3; D, Nivel 5. Estudios inconsistentes o inconclusivos de cualquier nivel.

Tabla III Valoración de la validez interna según las guías de lectura crítica de Sackett

Eficacia de un tratamiento	Principal	¿La asignación de sujetos ha sido aleatoria? ¿Han sido seguidos todos? ¿Estrategia de análisis por intención de tratar?
	Secundarias	¿Se ha utilizado alguna técnica de enmascaramiento? ¿Los grupos eran comparables al inicio? ¿Los grupos han sido seguidos y tratados en la misma forma (excepto el tratamiento en estudio)?
Etiología Efecto perjudicial de una exposición	Principal	¿Se han identificado claramente grupos comparativos similares respecto a importantes determinantes de los resultados excepto a la exposición? ¿Exposición y resultados se han medido de la misma manera en los grupos que se comparan? ¿El seguimiento ha sido lo suficientemente largo y completo?
	Secundarias	¿La secuencia temporal es la correcta? ¿Existe un gradiente dosis-respuesta?
Prueba diagnóstica	Principal	¿Se ha realizado una comparación a ciegas e independiente con una prueba de referencia? ¿La muestra estudiada incluyó el espectro apropiado de pacientes a los que la prueba será aplicada en la práctica?
	Secundarias	¿Los resultados de la prueba diagnóstica influyeron en la decisión de realizar la prueba de referencia? ¿Se describen los métodos de realización de la prueba diagnóstica con el suficiente detalle para permitir su reproducibilidad?
Pronóstico	Primarias	¿Se ha estudiado una muestra representativa y bien definida, formada por pacientes en un mismo momento del curso de la enfermedad? ¿El seguimiento ha sido lo suficientemente largo y completo?
	Secundarias	¿Se han utilizado criterios objetivos y no sesgados para medir los desenlaces? ¿Se ha ajustado por los factores pronósticos importantes?

Tabla IV Temas de publicación

	Revista					Total
	<i>Cir Pediatr</i>	<i>An Esp Pediatr</i>	<i>J Pediatr Surg</i>	<i>Pediatr Surg Int</i>	<i>Eur J Pediatr Surg</i>	
Cir. gral. y digestivo	53	9	4	4	4	74
	29,8%	39,1%	26,7%	40,0%	40,0%	31,4%
Urología	36	3	2			41
	20,2%	13,0%	13,3%			17,4%
Experimental	16		7	5	3	31
	9,0%		46,7%	50,0%	30,0%	13,1%
Tórax-tráquea	16	3	1	1	2	23
	9,0%	13,0%	6,7%	10,0%	20,0%	9,7%
Oncología	14	4	1			19
	7,9%	17,4%	6,7%			8,1%
Maxilofacial	11					11
	6,2%					4,7%
Plástica-reparadora	8					8
	4,5%					3,4%
Trauma y ortopedia	2					2
	1,1%					0,8%
Otros	22	4			1	27
	12,4%	17,4%			10,0%	11,4%
Total	178	23	15	10	10	236
	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

Tabla V Resultado según la clasificación de la AET

	N	%
Metaanálisis de ensayos controlados y aleatorizados	1	0,4
Ensayo controlado y aleatorizado. Muestra grande	1	0,4
Ensayo controlado y aleatorizado. Muestra pequeña	3	1,4
Ensayo prospectivo controlado no aleatorizado (controles coincidentes)	1	0,4
Ensayo prospectivo controlado no aleatorizado (controles históricos)	-	-
Estudio de cohortes	36	17,5
Estudios de casos y controles	6	2,9
Series clínicas no controladas. Estudios descriptivos. Comités de expertos	153	74,6
Casos únicos	4	1,9
Total	205	100

zados, realizados por anestelistas) y control del estrés quirúrgico (dos ensayos clínicos realizados por psicólogos). En ningún ensayo clínico se especificaba el sistema utilizado para la aleatorización y en uno de ellos no había ningún dato que permitiera al lector comprobar que los grupos de estudio eran comparables, como afirmaban sus autores.

En el caso de los estudios sobre pruebas diagnósticas y sobre factores de riesgo/etiología, estudios que no requieren de

ensayos aleatorizados para obtener buenas evidencias, el nivel de evidencia fue superior al nivel 4 en 38,5% y en el 47,8%, respectivamente; mientras que en estudios sobre pronóstico de una enfermedad este porcentaje volvió a caer hasta el 16,7%.

DISCUSIÓN

Los resultados de nuestra revisión muestran que cerca de un 90% de los artículos poseen un bajo nivel de evidencia (nivel 4), así como una escasez de ensayos clínicos aleatorizados (ECA), que trataron sobre terapias médicas. Estos hallazgos son similares a los de series más amplias como la de Hardin⁽³⁾, que encuentra que el 98% de las publicaciones en las principales revistas de habla inglesa son trabajos observacionales y sólo el 0,3 eran ECA, tratando principalmente temas médicos como la antibioterapia en la apendicitis o tratamientos quimioterápicos en pacientes oncológicos. Moss⁽⁴⁾ encuentra 134 ECA en el período 1966-1996, lo que supone un 0,17%, siendo las principales áreas de estudio la analgesia postoperatoria (49%), la antibioterapia y la ECMO.

El ensayo clínico aleatorizado representa el patrón oro de la investigación clínica, y constituye el máximo nivel de evidencia científica, sólo superada por el metaanálisis de ensayos clínicos. Sin embargo, en el campo de la cirugía la experiencia personal sigue ejerciendo una fuerte influencia y la presencia de este tipo de estudios es escasa. Entre las dificultades

Tabla VI Clasificación de las publicaciones según la pregunta de estudio

	Pregunta de estudio						Total
	No procede	Terapia	Prueba diagnóstica	Pronóstico	Factor de riesgo/etiología	Revisión sistemática	
Oxfor 1b		4 (3,7%)	5 (38,5%)	1 (16,7%)			10 (4,9%)
2a					1 (4,3%)		1 (0,5%)
2b		1 (9%)			5 (21,7%)		6 (2,9%)
3b					5 (21,7%)		5 (2,5%)
4	52 (100,0%)	103 (94,5%)	8 (61,5%)	5 (83,3%)	12 (52,2%)	1 (100,0%)	181 (88,7%)
5		1 (0,9%)					1 (0,5%)
Total	52 (100,0%)	109 (100,0%)	13 (100,0%)	6 (100,0%)	23 (100,0%)	1 (100,0%)	204 (100,0%)

existentes para su realización están las preferencias personales de los cirujanos por unas u otras técnicas, la dificultad para realizar enmascaramientos doble ciego con técnicas quirúrgicas o comparaciones con placebo, y sesgos como el denominado del último caso (es poco probable que un cirujano utilice una nueva técnica si tras probarla el paciente presenta una complicación grave)^(8,9). Si además nos centramos en la cirugía pediátrica, a estos problemas tendríamos que añadir la baja incidencia de muchos de los procesos de interés, por lo que una sola institución, trabajando de forma aislada no dispondría de los suficientes pacientes para desarrollar un ensayo clínico de la potencia adecuada. Una medida indirecta de las especiales dificultades para el desarrollo de este tipo de estudios en nuestro campo es que el número de artículos con niveles de evidencia superiores a 4 mayor en aquellos trabajos que estudiaban pruebas diagnósticas o factores de riesgo, estudios que no requieren de ensayos clínicos aleatorizados.

Algunos autores consideran que las series de casos representan la única evidencia clínica para la mayoría de las intervenciones en cirugía pediátrica⁽⁵⁾, hecho que no es del todo cierto puesto que al menos dos estudios^(10,11) han hallado que el porcentaje de intervenciones sustentadas por ECA es del 11% en un hospital regional y llega al 26% en un centro terciario aunque, como contrapartida, el 83% de esos ensayos se realizaron con adultos. Organizaciones como la Colaboración Cochrane recomiendan la omisión de las series de casos en las revisiones sobre evidencia de una terapia, lo cual es correcto siempre y cuando la evidencia disponible incluya una proporción aceptable de datos prospectivos de alta calidad, lo que puede ser problemático en nuestro campo⁽⁵⁾. Se ha estimado que el 38% de las investigaciones en cirugía colorrectal de adultos podrían ser evaluadas prospectivamente mediante investigación clínica, cifra que posiblemente sea inferior en la cirugía pediátrica⁽³⁾, por lo que parece que los estudios de series de casos van a seguir jugando un papel importante en la generación de evidencia en nuestra especialidad.

Las alternativas son, por una parte, el desarrollo de estudios multicéntricos para contrarrestar así la baja incidencia de muchos de los procesos de interés y, por otra, la aplica-

ción de los principios del metaanálisis a las series de casos⁽¹²⁾, aunque la utilidad de esta opción estará limitada por la calidad de los datos analizados. Como ejemplo, y aunque en nuestra serie no hemos realizado un análisis de la metodología empleada, resulta reseñable que sólo 2 entre los 42 artículos etiquetados como estudios de cohortes o caso-control incluyeran un análisis multivariante (regresión logística) para el control de la confusión, siendo lo más frecuente la estratificación por variables no siempre relevantes, y el análisis bivariante. Un ejemplo más en este sentido es el análisis que realizó González de Dios⁽¹³⁾ sobre los artículos publicados en Anales Españoles de Pediatría durante el período 1994-2000, donde ninguno de los artículos sobre cirugía utilizó técnicas estadísticas de complejidad superior al análisis bivariable.

La difusión de guías para la comunicación de estudios descriptivos podría mejorar la calidad de estos trabajos, al igual que la difusión del informe CONSORT (Consolidated of Reporting Trials)⁽¹⁴⁾, adoptado por editores, revisores y autores, lo tuvo sobre la calidad de los ensayos clínicos publicados con posterioridad⁽¹⁵⁾. Según Rangel⁽⁵⁾, hasta que la colaboración entre grupos de estudio permita una mayor aplicación de estudios prospectivos de alta calidad, el desarrollo de una estrategia para mejorar la calidad y la interpretación de los datos procedentes de series de casos produciría el mayor impacto de una medida en el estado de la MBE en la cirugía pediátrica.

Por último, aunque los resultados de nuestra revisión coinciden en gran medida con otras series al mostrar un predominio de trabajos descriptivos con baja evidencia y una escasez de ECA que estudian en su mayoría terapias médicas, deben considerarse las limitaciones del mismo. En primer lugar, la posible falta de exhaustividad, principalmente con artículos procedentes de Anales Españoles de Pediatría que no fueran identificados mediante la estrategia de búsqueda, y, en segundo lugar, la gran heterogeneidad de los artículos evaluados; por lo que es posible que en algunos temas muy específicos, la valoración que hayamos hecho de las variables de ajuste no haya sido la adecuada. A pesar de ello, consideramos que puede ofrecer una aproximación real a las características de nuestra actual producción científica.

BIBLIOGRAFÍA

1. Guerra Romero L. La medicina basada en la evidencia: un intento de acercar la ciencia al arte de la práctica clínica. *Med Clin (Barc)* 1996;**107**:377-382.
2. Evidence-based medicine working group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA* 1992;**268**:2420-2425.
3. Hardin WD, Stylianos S, Lally KP. Evidence-based practice in Pediatric Surgery. *J Pediatr Surg* 1999;**34**:908-913.
4. Moss RL, Henry MCW, Dimmitt RA, Rangel S, Geraghty N, Skarsgard ED. The role of prospective randomized clinical trials in pediatric surgery: State of the art? *J Pediatr Surg* 2001;**36**:1182-1186.
5. Rangel SJ, Moss RL. The need for critical analysis of case-series reporting in pediatric surgery. *Sem Pediatr Surg* 2002;**11**:184-189.
6. Jovell AJ, Navarro MD. Evaluación de la evidencia científica. *Med Clin (Barc)* 1995;**105**:740-743.
7. <http://www.cebm-net>.
8. Blakely ML, Lally KP. Using evidence based medicine principles to improve quality of patient care in pediatric surgery. *Sem Pediatr Surg* 2002;**11**:42-45.
9. Kosloske AM. Epidemiology as the search for truth. *Sem Pediatr Surg* 2002;**11**:162-167.
10. Kenny SE, Shankar KR, Rintala R, Lamont GL, Lloyd DA. Evidence-based surgery: interventions in a regional paediatric surgical unit. *Arch Dis Child* 1997;**76**:50-53.
11. Baraldini V, Spitz L, Pierro A. Evidence-based operations in paediatric surgery. *Pediatr Surg Int* 1998;**13**:331-335.
12. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D y cols. Meta-analysis of observational studies in epidemiology. A proposal for reporting. *JAMA* 2000;**283**:2008-2012.
13. González de Dios J. Valoración del nivel de calidad de la evidencia científica de Anales Españoles de Pediatría. *An Esp Pediatr* 2001;**54**:380-389.
14. Begg C, Cho M, Eastwood S y cols. Improving the quality of reporting of randomized controlled trials: The CONSORT statement. *JAMA* 1996;**276**:637-639.
15. Moss RL. The CONSORT statement: Progress in clinical research in pediatric surgery. *J Pediatr Surg* 2001;**36**:1739-1742.