

Avaliação Clínica da Gravidade da Doença em Cuidados Intensivos Pediátricos

PAULO OOM

*Clinica Universitária de Pediatria
Hospital de Santa Maria*

Resumo

A avaliação da gravidade da doença representa uma das mais importantes habilitações clínicas. É pela ponderação de diversos factores que o médico, na sua prática diária, consegue avaliar a gravidade da doença que pretende tratar e desta forma actuar da forma que considera mais eficaz e apropriada. Esta avaliação da gravidade feita individualmente é, no entanto, subjectiva e não é de estranhar que, perante uma mesma situação, dois clínicos possam divergir na sua avaliação do prognóstico a transmitir ao doente ou aos seus familiares. Foi objectivo deste trabalho avaliar de forma objectiva o desempenho da avaliação clínica da gravidade da doença e compará-lo com o obtido pelas escalas objectivas *Pediatric Risk of Mortality* (PRISM) e *Pediatric Index of Mortality* (PIM). A avaliação subjectiva da gravidade da doença feita pelo clínico na primeira hora após a admissão da criança mostrou um desempenho bom (SMR=0,90; $\chi^2(p)=0,42$; AUC=0,93). No confronto directo com as escalas objectivas, a avaliação clínica mostrou um poder discriminativo semelhante ($\chi^2(p)=0,78$).

Palavras-Chave: Cuidados intensivos pediátricos, escalas preditivas, mortalidade, avaliação clínica, PRISM, PIM.

Summary

Clinician Mortality Estimates in Pediatric Intensive Care

Severity of illness measurement represents one of the most important clinicians qualifications. Is by weighting several factors in daily practice that physicians can assess the severity of their patients condition and act properly. This personal evaluation is, consequently subjective. This study compared the performance of subjective

mortality estimates by physicians with two objective estimates, the Paediatric Risk of Mortality (PRISM) and the Paediatric Index of Mortality (PIM). Clinical judgement performed well (SMR=0,90; $\chi^2(p)=0,42$; AUC=0,93). Healthcare providers subjective mortality predictions, PRISM and PIM discriminated equally well ($\chi^2(p)=0,78$).

Key-Words: Paediatric intensive care, scores, mortality, clinical judgement, PRISM, PIM

Introdução

A avaliação da gravidade da doença feita pelo médico por meios subjectivos (avaliação clínica) ou objectivos (monitorização, avaliação laboratorial) está presente em todos os actos médicos praticados numa unidade de cuidados intensivos pediátrica (UCIP) e dela dependem atitudes diárias como triagem ou transferência de doentes, início, intensidade ou modificação da terapêutica e mesmo a suspensão de medidas de suporte avançado de vida. Esta forma de medir a gravidade da doença (de “prognosticar”) baseia-se na experiência e conhecimentos de cada médico e é, de longe, a mais utilizada desde sempre. No entanto, ao depender da experiência do clínico é uma avaliação subjectiva, não reproduzível. A própria experiência clínica é sempre reduzida (apoiada numa base de doentes que corresponde aos casos anteriormente tratados e lembrados pelo mesmo clínico, cujo número nunca é suficientemente grande) e influenciada pelas experiências mais recentes, mais raras e mais dramáticas.^{1,2}

O desenvolvimento de escalas preditivas de mortalidade como a *Pediatric Risk of Mortality* (PRISM) e a *Pediatric Index of Mortality* (PIM) veio permitir a avaliação da gravidade da doença de forma objectiva.^{3,4} No entanto, o julgamento clínico mantém-se a forma de avaliação utilizada na quase totalidade das decisões clínicas pelo que é interessante compará-lo com a medição objectiva.

Correspondência: Paulo Oom

Unidade de Cuidados Intensivos
Serviço de Pediatria
Hospital de Santa Maria
Av. Prof. Egas Moniz 1649-035 Lisboa

Doentes e Métodos

O estudo foi realizado na Unidade de Cuidados Intensivos de Pediatria do Hospital de Santa Maria. Foram incluídas de forma prospectiva todas as crianças internadas na unidade durante um período de 21 meses, excepto as que apresentavam algum critério para exclusão: 1 – admissão não se justificava, por se tratar de situações que não exigiam cuidados intensivos e que, por rotina, seriam tratadas em outros locais do Serviço de Pediatria, 2 – tempo de internamento inferior a 2 horas, 3 – admissão durante manobras de reanimação cardiorespiratória e não tendo sido obtidos sinais vitais estáveis durante um período de, pelo menos, 2 horas. Foram obtidos de cada criança os seguintes dados: data de nascimento, data de internamento, idade, sexo, tipo de admissão (electiva, emergência), origem (enfermaria, outro hospital, urgência), destino (enfermaria, outro hospital, alta) e data de alta. Foram consideradas em pós-operatório as crianças internadas até 24 horas após o procedimento cirúrgico. Foram igualmente colhidos, para cada criança, dados respeitantes aos procedimentos utilizados em cuidados intensivos de acordo com os critérios definidos por Pollack⁵ adoptando uma modificação das alterações propostas por Gemke.⁶ Para o cálculo das escalas preditivas de mortalidade foram seguidas as instruções dos autores originais.^{3,4} A avaliação clínica da gravidade da doença foi realizada pelo médico que recebeu a criança, decorrida uma hora após a admissão, em termos de probabilidade de morte, expressa em percentagem entre 0 e 100. A mortalidade foi entendida como mortalidade durante o internamento na UCIP.

O número de crianças a estudar foi determinado tendo em conta a necessidade de incluir pelo menos 20 falecimentos o que corresponde ao limite geralmente aceite de interesse clínico no estudo da mortalidade em cuidados intensivos pediátricos.⁷ Para uma mortalidade estimada de 3,5% foi calculada a necessidade de incluir 571 internamentos.

Este trabalho teve o parecer favorável da Comissão de Ética da Faculdade de Medicina da Universidade de Lisboa e da Comissão de Ética para a Saúde do Hospital de Santa Maria.

Todos os dados obtidos foram introduzidos no programa Excel versão 5.0 para Windows (Microsoft Corporation, Seattle, WA). Para a análise do desempenho da avaliação clínica e das escalas PRISM e PIM foram determinadas a sua calibração e poder discriminativo. A calibração foi testada pelo cálculo da razão de mortalidade padronizada (*standardized mortality rate*, SMR) e pela aplicação do teste de χ^2 *Goodness-of-fit* de Hosmer e Lemeshow.^{8,9} A discriminação foi testada pelo cálculo da área abaixo da curva ROC (*area under curve*, AUC) como proposto por Hanley e McNeil.¹⁰ As diferentes curvas ROC

foram comparadas pelo teste de χ^2 utilizando o algoritmo descrito por DeLong.¹¹ A análise estatística foi efectuada utilizando o programa Stata versão 7.0 para Windows (Stata Corporation, College Station, TX). Foram considerados significativos valores de $p < 0,05$.

Resultados

O estudo decorreu entre 1 de Fevereiro de 2001 e 31 de Outubro de 2002. Durante este período ocorreram 625 internamentos na unidade de cuidados intensivos de Pediatria. Destes, apresentavam um ou mais critérios de exclusão 45 pelo que foram analisados os dados respeitantes a 580 internamentos. As características desta população estão discriminadas na tabela I. A idade variou entre os 2 dias de vida e os 19 anos com uma mediana de 2,6 anos.

Tabela I
Características da amostra (n=580)

Sexo	Masculino	338 (58,3)
	Feminino	242 (41,7)
Admissão	Emergência	386 (66,6)
	Electiva	194 (33,4)
	Readmissões	110 (19,0)
Doentes médicos		371 (64,0)
Doentes cirúrgicos		209 (36,0)
Origem	Enfermaria	306 (52,8)
	Outro hospital	182 (31,4)
	Urgência	92 (15,8)
Destino	Enfermaria	469 (80,9)
	Outro hospital	51 (8,8)
	Alta	29 (5,0)
	Falecido	31 (5,3)

Resultados expressos em número absoluto (percentagem).

Os doentes cirúrgicos representam as crianças admitidas provenientes do bloco operatório

O número de procedimentos intensivos realizados em cada doente variou entre 0 e 15 com uma mediana de 2. O tempo de internamento variou entre 1 e 113 dias, com uma mediana de 2 dias. Faleceram 31 crianças correspondendo a 5,3% dos internamentos.

A escala PRISM mostrou uma boa calibração com uma razão de mortalidade padronizada de 0,90 com intervalo de confiança de 95% (IC 95%) entre 0,61 e 1,27. O teste de Hosmer-Lemeshow mostrou um valor de $\chi^2_{(8)}=6,23$ ($p=0,62$). O poder discriminativo da escala PRISM, avaliado através da determinação da área sob a curva ROC, foi

de 0,91 (IC 95% 0,86 - 0,96).

A escala PIM mostrou uma boa calibração com SMR de 1,03 (IC 95% 0,70 - 1,46) e o teste de Hosmer-Lemeshow um valor de $\chi^2_{(8)}=9,51$ ($p=0,30$). A avaliação do poder discriminativo mostrou AUC de 0,92 (IC 95% 0,88 - 0,96).

A estimativa da probabilidade de morte feita pelo clínico ao fim da primeira hora de internamento foi re-avaliada em 552 doentes (95,2% da população do estudo). As estimativas clínicas da probabilidade de morte variaram entre 0 e 1 com uma mediana de 0,01 e os percentis 25 e 75 de, respectivamente 0 e 0,05. A distribuição de valores está representada na figura 1.

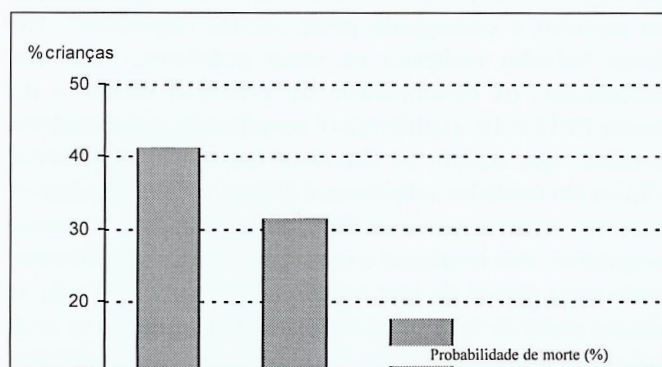


Fig. 1 - Distribuição das estimativas de probabilidade de morte de acordo com a avaliação clínica

A avaliação clínica mostrou igualmente uma boa calibração com SMR de 0,90 (IC 95% 0,61 - 1,28) e um teste de Hosmer-Lemeshow com um valor de $\chi^2_{(7)}=7,09$

Tabela II

Poder discriminativo da avaliação clínica em diferentes subgrupos de crianças

	n (%)	AUC	IC 95%
Grupos etários			
< 1 ano	172 (31,2)	0,93	0,85 - 1,00
≥ 1 < 3 anos	121 (21,9)	0,93	0,87 - 0,99
≥ 3 < 6 anos	72 (13,0)	0,96	0,90 - 1,00
≥ 6 < 12 anos	93 (16,8)	1,00	1,00 - 1,00
≥ 12 anos	94 (17,0)	0,85	0,71 - 0,99
Doentes médicos			
	354 (64,1)	0,90	0,84 - 0,95
Tempo internamento			
≤ 1 dia	56 (10,1)	1,00	1,00 - 1,00
> 1 dia	496 (89,9)	0,91	0,87 - 0,96
Número de procedimentos			
> 1	301 (54,5)	0,89	0,83 - 0,95

AUC - área sob a curva ROC; IC 95%-intervalos de confiança de 95%; Doentes médicos refere-se aos doentes que não foram admitidos no pós-operatório. A avaliação dos doentes cirúrgicos ou com procedimentos ≤ 1 não foi possível por não terem ocorrido mortes nestes subgrupos.

($p=0,42$). A calibração da avaliação clínica manteve-se quando os foram avaliados diferentes subgrupos como grupo etário ($\chi^2_{(8)}=11,90$ $p=0,16$), tempo de internamento ($\chi^2_{(7)}=8,80$ $p=0,27$) e número de procedimentos intensivos realizados ($\chi^2_{(6)}=5,86$ $p=0,44$).

A avaliação do poder discriminativo mostrou AUC de 0,93 (IC 95% 0,89 - 0,97). A capacidade discriminativa manteve-se quando foram analisados diferentes grupos de crianças (tabela II).

A comparação da discriminação obtida pelas diferentes escalas ($n=552$) mostrou não existirem diferenças significativas entre elas todas revelando AUC superiores a 0,90 ($\chi^2_{(2)}=0,50$ $p=0,78$) conforme se observa na tabela III e figura 2.

Tabela III

Poder discriminativo das diferentes escalas de mortalidade ($n=552$)

Escala	AUC	IC 95%
PRISM	0,91	0,86 - 0,96
PIM	0,92	0,88 - 0,96
Clínica	0,93	0,89 - 0,97

PRISM-Pediatric risk of mortality; PIM-Pediatric index of mortality; Clínica-avaliação clínica; AUC-área sob a curva ROC. IC 95%-intervalos de confiança de 95%. $\chi^2_{(2)}=0,50$ $p=0,78$

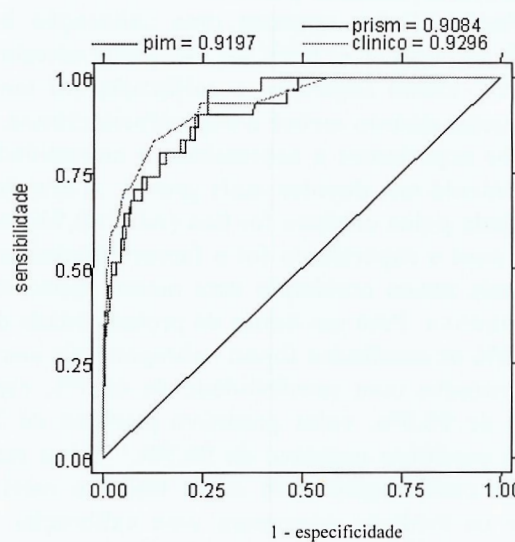


Fig. 2 - Comparação do poder discriminativo das diferentes escalas de mortalidade

Discussão

A avaliação da gravidade da doença representa um exercício diário para o médico que trabalha em cuidados intensivos pediátricos. Esta avaliação pressupõe a colheita

e integração de múltiplas variáveis, clínicas e laboratoriais, a partir das quais o clínico elabora uma opinião pessoal sobre a gravidade da doença de cada doente em particular. Esta forma de avaliação faz parte e confunde-se com a própria actividade médica e é, desde sempre, a mais utilizada.

Neste trabalho, estudámos o desempenho da avaliação clínica subjectiva da gravidade, expressa em estimativa da probabilidade de morte. No total, a avaliação clínica previu a morte de 34,5 crianças num total de 31 falecidos originando uma SMR de 0,90 e um valor de p no teste de Hosmer-Lemeshow de 0,42. Esta calibração manteve-se ao longo de diferentes intervalos de risco, grupos etários, tempos de internamento e número de procedimentos intensivos realizados. A avaliação do poder discriminativo mostrou AUC de 0,93 (IC 95% 0,89 - 0,97). Para um limiar de probabilidade de morte de 50% a avaliação clínica mostrou-se pouco sensível (48%) mas muito específica (99%) com um valor preditivo positivo de 75% e valor preditivo negativo de 97%.

No primeiro trabalho sobre a avaliação clínica da gravidade da doença, realizado em adultos, foi igualmente observada uma tendência dos clínicos para sobrestimar a mortalidade originando uma SMR de 0,67 mas infelizmente não foi estudada a calibração pelo teste de Hosmer-Lemeshow. A avaliação da capacidade discriminativa revelou uma AUC=0,89.¹² No primeiro trabalho pediátrico sobre o mesmo tema a avaliação clínica mostrou uma calibração boa com SMR de 1,09 e o teste de Hosmer-Lemeshow com $p=0,11$. Como esperado, a calibração foi tanto mais deficiente quanto menor a experiência clínica, com os menos experientes a sobrestimar a mortalidade, principalmente nos doentes mais graves. A discriminação revelada pelos clínicos foi boa (AUC=0,93) mas também aqui a experiência foi o factor determinante com os mais novos revelando uma menor capacidade discriminativa. Para um limiar de probabilidade de morte de 50% os resultados foram sobreponíveis aos por nós encontrados com sensibilidade de 48,7%, especificidade de 98,9%, valor preditivo positivo de 73,1% e valor preditivo negativo de 96,9%.¹³ Estes resultados foram confirmados num outro trabalho mais recente onde os médicos mostraram uma calibração e poder discriminativo bons e proporcionais à sua experiência clínica em cuidados intensivos pediátricos com os menos experientes sobrestimando a mortalidade.¹⁴ No período neonatal os resultados foram semelhantes.¹⁵ De uma forma geral estes resultados reforçam a ideia de que a avaliação subjectiva da gravidade, expressa como estimativa de probabilidade de morte é um instrumento bem calibrado e com suficiente capacidade discriminativa para utilização diária em cuidados

intensivos pediátricos. A sua utilização para comparar o desempenho entre diferentes unidades será, no entanto, sempre abusiva, dada a subjectividade que contém.

Neste trabalho comparámos igualmente o desempenho obtido pela avaliação clínica em relação ao conseguido pelas escalas objectivas (PRISM e PIM). Nos três casos a calibração foi boa, traduzida por valores de SMR próximos da unidade e o teste de Hosmer-Lemeshow mostrando sempre valores de p superiores a 0,05. Nos trabalhos realizados em adultos a calibração conseguida pela avaliação clínica é geralmente suplantada pela conseguida com as escalas objectivas devido, em grande parte ao exagero na predição da mortalidade.^{12,16-18} No entanto, nos extremos de gravidade, a avaliação clínica tem uma calibração idêntica ou superior à conseguida pelas escalas objectivas.¹² No único trabalho realizado na idade pediátrica, onde são comparados os desempenhos da avaliação clínica e da escala PRISM III a calibração é semelhante sendo também evidente que aquela se deteriora quando a experiência clínica em cuidados intensivos é menor.¹⁴ Estes resultados parecem mostrar que o médico, principalmente o menos experiente, tem tendência a considerar que os seus doentes estão mais graves do que o estão na realidade. Ou seja, o clínico sente-se mais tranquilo (perante si próprio, os seus colegas e o doente ou a sua família) se falhar mais na previsão da morte do que na previsão da sobrevivência.

No nosso estudo a análise comparativa da capacidade discriminativa não mostrou diferenças significativas entre a avaliação clínica e as escalas objectivas ($\chi^2_{(2)}=0,50$ $p=0,78$). Este dado está de acordo com o de outros trabalhos realizados em adultos e crianças onde a discriminação conseguida pela avaliação clínica é idêntica ou superior à das escalas objectivas.^{12,14-18} Este facto pode ser explicado pelo facto de a avaliação clínica utilizar outros indicadores de gravidade, não incluídos nos modelos preditivos, como o tipo de medidas de suporte de vida utilizadas, a agressividade da terapêutica administrada ou a intensidade da monitorização exigida por cada criança. Porventura mais importante, o clínico inclui na sua análise a evolução das diferentes variáveis ao longo do período de avaliação, ao contrário das escalas que utilizam o primeiro ou o pior valor de cada variável obtido num intervalo de tempo determinado.

Com este trabalho, não só confirmámos os dados existentes na idade pediátrica como, pela primeira vez, demonstrámos que a calibração e discriminação clínicas se mantêm mesmo quando a estimativa de probabilidade de morte é feita até uma hora após a admissão. O nosso trabalho representa igualmente a primeira análise comparativa entre o desempenho da avaliação clínica e o conseguido pela escala PIM numa mesma população de crianças internadas em cuidados intensivos.

Agradecimentos

O autor agradece ao Prof. Doutor J. Gomes-Pedro, Prof. Doutor A Gouveia de Oliveira e Dra. Manuela Correia a revisão crítica deste trabalho. Da mesma forma agradece à Dra Filipa Negreiro o cuidado posto na análise estatística dos resultados.

Bibliografia

1. Cowen J, Kelley M: Errors and bias in using predictive scoring systems. *Crit Care Clin* 1994; 10:53-72
2. Watts C, Knaus W: The case for using objective scoring systems to predict intensive care unit outcome. *Crit Care Clin* 1994; 10: 73-89
3. Pollack M, Ruttimann U, Getson P: Pediatric Risk of Mortality (PRISM) score. *Crit Care Med* 1988; 16: 1110-6
4. Shann F, Pearson G, Slater A, Wilkinson K: Paediatric Index of Mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Med* 1997; 23: 201-7
5. Pollack M, Katz R, Ruttimann U, Getson P: Improving the outcome and efficiency of intensive care: the impact of an intensivist. *Crit Care Med* 1988; 16: 11-17
6. Gemke R, Bonsel G, Vugth A: Effectiveness and efficiency of a Dutch pediatric intensive care unit: Validity and application of the Pediatric Risk of Mortality score. *Crit Care Med* 1994; 22: 1477-84
7. Pearson G, Stickley J, Shann F: Calibration of the paediatric index of mortality in UK paediatric intensive care units. *Arch Dis Child* 2001; 84: 125-8
8. Lemeshow S, Hosmer D: A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982; 115: 92-106
9. Hosmer D, Lemeshow S. *Applied Logistic Regression*. New York: Wiley and Sons, 1989
10. Hanley J, McNeil B: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; 143: 29-36
11. DeLong E, DeLong D, Clarke-Pearson D: Comparing the areas under two or more correlated receiver operating curves. A nonparametric approach. *Biometrics* 1988; 44: 837-45
12. McClish D, Powell S: How well can physicians estimate mortality in a medical intensive care unit? *Med Decis Making* 1989; 9: 125-32
13. Marcin J, Pollack M, Patel K, Sprague B, Ruttimann U: Prognostication and certainty in the pediatric intensive care unit. *Pediatrics* 1999; 104: 868-73
14. Marcin J, Pollack M, Patel K, Ruttimann U: Combining physician's subjective and physiology-based objective mortality risk predictions. *Crit Care Med* 2000; 28: 2984-90
15. Stevens S, Richardson D, Gray J, Goldmann D, McCornick M: Estimating neonatal mortality risk: an analysis of clinician's judgments. *Pediatrics* 1994; 93: 945-50
16. Perkins H, Jonsen A, Epstein W: Providers as predictors: using outcome predictions in intensive care. *Crit Care Med* 1986; 14: 105-10
17. Kruse J, Thill Baharozian M, Carlson R: Comparison of clinical assessment with Apache II for predicting mortality risk in patients admitted to a medical intensive care unit. *JAMA* 1988; 260: 1739-42
18. Poses R, Bekes C, Winkler R, Scott E, Copare F: Are two (inexperienced) heads better than one (experienced) head? *Arch Intern Med* 1990; 150: 1874-8