

Calibration of the paediatric index of mortality in UK paediatric intensive care units

G A Pearson, J Stickley, F Shann

Abstract

Aim—To test a paediatric intensive care mortality prediction model for UK use.

Method—Prospective collection of data from consecutive admissions to five UK paediatric intensive care units (PICUs), representing a broad cross section of paediatric intensive care activity. A total of 7253 admissions were analysed using tests of the discrimination and calibration of the logistic regression equation.

Results—The model discriminated and calibrated well. The area under the ROC plot was 0.84 (95% CI 0.819 to 0.853). The standardised mortality ratio was 0.87 (95% CI 0.81 to 0.94). There was remarkable concordance in the performance of the paediatric index of mortality (PIM) within each PICU, and in the performance of the PICUs as assessed by PIM. Variation in the proportion of admissions that were ventilated or transported from another hospital did not affect the results.

Conclusion—We recommend that UK PICUs use PIM for their routine audit needs. PIM is not affected by the standard of therapy after admission to PICU, the information needed to calculate PIM is easy to collect, and the model is free.

(Arch Dis Child 2001;84:125–128)

Keywords: mortality; intensive care; prediction model

Birmingham Children's Hospital,
Steelhouse Lane,
Birmingham B4 6NH,
UK

G A Pearson
J Stickley

Royal Children's Hospital, Parkville,
3052 Victoria,
Australia
F Shann

Correspondence to:
Dr Pearson
Gale.Pearson@
Bhamchildrens.wmids.nhs.uk

Accepted 20 September
2000

Table 1 Data by hospital

Hospital	Duration of data collection (mth)	Number of admissions (no. of deaths)	Admissions per bed per year	SMR (95% CI)	Mean probability of death (SD)	% retrieved	Vent in first hour	Average length of stay, days (SD)	Median length of stay, hours (IQR)	Area under ROC plot
Alder Hey	16	1245 (85)	44.46	0.89 (0.72 to 1.06)	0.077 (0.133)	23%	78%	5.31 (10.82)	48 (24 to 144)	0.8344
Birmingham Children's	48.5	4380 (374)	77.91	0.87 (0.80 to 0.95)	0.098 (0.155)	3%	83%	3.68 (8.32)	39 (20 to 91)	0.8373
St George's	19	498 (25)	62.9	0.83 (0.53 to 1.12)	0.061 (0.124)	37%	35%	3.32 (5.48)	36 (19 to 90)	0.8571
Great Ormond St	6	341 (26)	68.2	0.88 (0.56 to 1.17)	0.087 (0.151)	4%	87%	4.08 (4.48)	48 (22 to 144)	0.8336
Guy's	12	789 (67)	46.41	0.89 (0.69 to 1.07)	0.096 (0.145)	54%	77%	3.53 (5.72)	47 (24 to 95)	0.8019

requires the use of a validated “off the shelf” model, although detailed comparisons for formal epidemiological studies are better done using a study specific model. There is currently a choice between two risk adjustment tools for routine use in children in intensive care: the paediatric risk of mortality (PRISM—currently marketed in its first revision, PRISM III),^{2,3} and the paediatric index of mortality (PIM).⁴ Neither model has been adequately evaluated in the UK. We sought to test PIM in five UK paediatric intensive care units.

Methods

The variables that contribute to PIM were collected prospectively in the PICUs at five hospitals: Alder Hey Children's Hospital in Liverpool, Birmingham Children's Hospital, and Great Ormond Street, Guy's, and St George's Hospitals in London. Data collection was by junior medical staff and validated by a named consultant collaborating in the study. Four of the five centres were recognised “lead centres” for paediatric intensive care provision and training, admitting in excess of 800 admissions per year (or more than 600 intubated admissions per year) and catering for a broad cross section of paediatric intensive care diagnostic case mix. The fifth had a smaller throughput and a greater proportion of high dependency cases. However, all were staffed by full time consultant paediatric intensive care physicians and dedicated junior staff with no commitments other than the PICU.

Admissions were classified by diagnostic groups defined in advance as respiratory, cardiac, neonatal, postoperative, accident/trauma, neurological, or miscellaneous.

DATA AND STATISTICS

The data needed to calculate PIM were collected within one hour from the time of first face to face contact between the ICU doctor and the patient, even when this occurred in another hospital—for example, when patients were transported from another institution. The method for calculating the probability of death is detailed in the appendix. In addition, various observational measures of case mix were

Table 2 Calibration of the PIM model

		Decile										Total
		1	2	3	4	5	6	7	8	9	10	
Died	Observed	4.0	7.0	8.0	9.0	33.0	30.0	45.0	58.0	113.0	270.0	577.0
	Expected	4.3	7.5	10.2	14.0	20.8	31.3	46.2	71.4	117.6	336.6	659.9
	O:E ratio	0.94	0.93	0.79	0.64	1.59	0.96	0.97	0.81	0.96	0.80	0.87
Survived	Observed	721.0	719.0	717.0	716.0	693.0	695.0	680.0	667.0	613.0	455.0	6676.0
	Expected	720.7	718.5	714.8	711.0	705.2	693.7	678.8	653.6	608.4	388.4	6593.1
	O:E ratio	1.00	1.00	1.00	1.01	0.98	1.00	1.00	1.02	1.01	1.17	1.01

recorded, such as whether the patient's intensive care admission was a result of interhospital transport and whether or not the patient was intubated during their admission. Patients who were intubated at any point can be regarded as having received a period of intensive care.

We tested the fit of PIM mortality prediction model in two ways. Discrimination was assessed using the area under the receiver operating characteristic (ROC) plot.⁵ This measure expresses how well the model distinguishes between patients who lived and those who died. An area under the ROC of 0.75 or more is considered clinically useful, and in the design of mortality prediction there is a trade off between the simplicity of the model (easier data collection and quality) and enhanced area under the ROC plot. An area under the ROC plot of 0.75 means that a randomly selected non-survivor would have a higher PIM value than a randomly selected survivor 75% of the time; it does not mean that prediction of death is correct 75% of the time. Calibration evaluates how well the model classifies patients into low, medium, and high risk categories. We evaluated this by examining a Hosmer-Lemeshow goodness of fit 4×10 table, which displays how well the model matches observed outcomes in deciles of the population ranked

by probability of death.¹ The results are also presented as observed:expected mortality ratios within standard risk categories (predicted mortality <1%, 1–4%, 5–14%, 15–29%, or 30% or more) and diagnostic categories. The overall expected death rate is the sum of the probability of death for each admission, and the ratio of observed to expected death rates is known as the standardised mortality ratio (SMR). Values less than one imply good performance, and values greater than one imply poor performance, although either may be affected by the fit of the model (see below). Confidence limits for the SMR were derived from a parametric approach.⁶

Results

We analysed a total of 7253 admissions. One centre (Great Ormond Street) did not collect data on consecutive admissions in all its intensive care units, so these were excluded. However, data from the cardiac ICU were complete for a six month period and were included. Table 1 shows the sample periods, standardised mortality ratios, number of patients, and the discrimination of the model in each hospital. The area under the ROC plot was 0.84 (95% confidence interval (CI) 0.819 to 0.853; see fig 1). Table 2 shows calibration of the PIM model. Table 3 shows the performance of the model in different diagnostic categories and in the standard severity of illness bands.

There were clinically significant differences in the proportion of patients transferred by an intensive care transport team to the receiving hospital, which were not associated with the proportion that were ventilated at admission to PICU ($r = -0.05$). There was remarkable similarity in the performance of the PICUs as assessed by the SMR. The overall SMR was 0.87 (95% CI 0.81 to 0.94) and this apparent

Table 3 Calibration across diagnostic categories and severity of illness bands

Diagnostic group	Bands of death probability					Total
	<1%	1% to <5%	5% to <15%	15% to <30%	30% +	
<i>Respiratory</i>						
Expected deaths	1.01	9.49	36.08	20.07	28.04	94.69
Actual deaths	1	3	16	13	17	50
n	139	385	409	97	56	1086
<i>Cardiac</i>						
Expected deaths	2.79	37.54	50.43	47.53	121.79	260.07
Actual deaths	2	44	50	38	73	207
n	359	1718	528	229	229	3063
<i>Neonatal</i>						
Expected deaths	0.12	3.16	15.63	19.97	18.69	57.57
Actual deaths	0	0	6	14	13	33
n	20	122	150	99	33	424
<i>Postoperative</i>						
Expected deaths	0.95	4.33	5.90	2.15	1.77	15.10
Actual deaths	0	3	2	1	1	7
n	189	196	72	10	5	472
<i>Accident/trauma</i>						
Expected deaths	0.34	5.85	8.05	3.67	14.62	32.54
Actual deaths	0	5	15	6	19	45
n	45	173	114	16	28	376
<i>Neurological</i>						
Expected deaths	0.73	8.01	19.74	12.11	30.49	71.09
Actual deaths	2	8	11	14	23	58
n	109	295	245	60	58	767
<i>Miscellaneous</i>						
Expected deaths	0.95	9.37	27.14	22.49	68.85	128.81
Actual deaths	2	17	46	43	69	177
n	145	390	309	106	115	1065
<i>Total</i>						
Expected deaths	6.90	77.77	162.97	127.98	284.25	659.87
Actual deaths	7	80	146	129	215	577
n	1006	3279	1827	617	524	7253

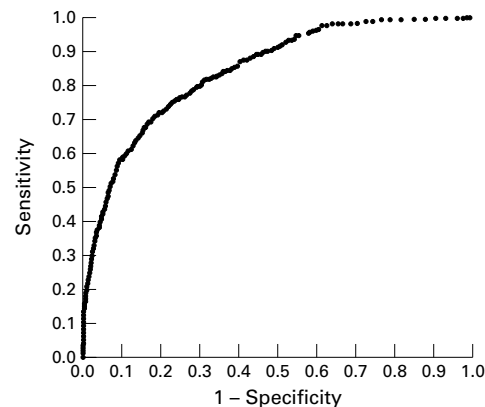


Figure 1 ROC curve.

good performance of the units appears on calibration to be systematic (table 2). As expected for a tool derived from a population of all intensive care admissions, the performance of the model was better in some diagnostic groups than in others. The most common diagnoses in the miscellaneous group were septic shock (204 children), other infection (71), liver transplant (64), neoplasm (53), acute hepatitis or liver failure (49), haemopoietic disease (48), renal failure (38), and inborn errors of metabolism (31).

Discussion

Until now, neither PIM nor PRISM has been adequately tested in UK paediatric intensive care units, and the performance of PRISM has been questioned.^{7,8} Mortality prediction models which, like PRISM, use the worst values of their predictor variables in the first 12–24 hours in PICU, have three disadvantages.⁹ Firstly, the data are difficult to collect. Secondly, they are not a good tool for comparing different intensive care units: for example, patients mismanaged in a poor quality unit will have higher scores than similar patients managed in a high quality unit, so the high mortality rate of the former might be incorrectly attributed to its having sicker patients. Thirdly, they appear to be more accurate than they really are: about 40% of deaths occur in the first 24 hours in PICU, so in these cases the score is really diagnosing death rather than predicting it.

PIM is free, and the data are collected within an hour of the time of first face to face contact between a PICU doctor and the patient. This means that the data have to be recorded by the PICU doctor rather than a data clerk, but it avoids the problem, common to all 12–24 hour scores, that the standard of care provided in a unit alters the predicted mortality rate. First contact data collection also means that PICU retrieval is, justifiably, evaluated as part of the PICU care. PIM also adjusts for the influence of premorbid states that can profoundly influence outcome.

PIM compares PICU performance to that of the PICUs that contributed to the derivation dataset. PIM was derived from data that were collected from consecutive admissions to all seven dedicated PICUs in Australia and one in the UK in 1994–95, so it is not surprising that UK PICUs in 1998–99 had 13% fewer deaths than predicted by the model. A revised version of PIM will be available soon to bring the model up to date.

Sixty per cent of the patients and 65% of the deaths in this study were from Birmingham Children's Hospital (table 1). However, this does not invalidate our finding that PIM works well: the SMR and area under the ROC plots were remarkably similar in all five PICUs (table 1).

When an individual PICU or group of PICUs applies a logistic regression model such as PIM to their data, differences between the observed and expected number of deaths may be a result of either the performance of the ICU, or the performance of the model. The

subsequent clinical interpretation is rarely objective if there are more deaths than predicted; there is a tendency to blame the fit of the model for the discrepancy. A proliferation of new mortality prediction models results, because the coefficients of the original model are often changed to compensate for poor performance. In fact, consistent differences across deciles of risk, as seen in this study, are more likely to be a result of differences in clinical performance rather than fundamental errors in the structure of the model. The reverse is largely true for non-systematic errors. It is not usually appropriate to respond to an SMR of more than 1.0 (which implies poor clinical performance) by changing the model. Furthermore one must be very cautious in interpreting small series (for example, with fewer than 20 deaths per unit).

CONCLUSION

We found that PIM provides a consistent guide to the performance of these five PICUs. Great care has to be taken to ensure that the data needed to calculate PIM are accurate; the fact that all the information is collected in the first hour simplifies this task. The model incorporates the quality of retrieval services in its assessment, and adjusts for the presence of important premorbid conditions. We recommend that PIM be used routinely as a mortality prediction model for paediatric intensive care in the UK.

The collaborators in this study, who collected and cleaned the data from their institutions, were Dr P Baines (Alder Hey), Dr A Goldman (Great Ormond Street), Dr P J Rye (St George's), and Dr I Murdoch (Guy's). Further information about PIM can be obtained in the software section of <http://pedscm.wustl.edu/>.

Appendix

PIM is calculated from information collected at the time a child is admitted to your ICU. Because PIM describes how ill the child was at the time you started intensive care, the observations to be recorded are those made at or about the time of first face to face (not telephone) contact between the patient and a doctor from your intensive care unit (or a doctor from a specialist paediatric transport team). Use the *first* value of each variable measured within the period from the time of first contact to one hour after arrival in your ICU. The first contact may be in your ICU, or your emergency department, or a ward in your own hospital, or in another hospital (e.g. on a retrieval). The pupils' reactions to light are used as an index of brain function; *do not record an abnormal finding if this is probably caused by drugs, toxins, or local injury to the eye*. If information is missing (e.g. base excess not measured), record zero (*except for systolic blood pressure, which should be recorded as 120*); PIM assumes that missing values are normal (e.g. that the base excess is 0 if it is not measured). Variables 1–7 are descriptive variables that are not needed to calculate PIM. Include *all* children admitted to your ICU.

- (1) Identifier (e.g. ICU admission number):
- (2) Age (months):
- (3) Diagnostic group (1 = Resp, 2 = CVS, 3 = Postop, 4 = Accident, 5 = Neurol, 6 = Other):
- (4) Date admitted to ICU (dd/mm/yyyy):
- (5) Days in ICU on this admission:
- (6) Endotracheal tube in situ at any time during this ICU admission (no = 0, yes = 1):
- (7) Outcome of ICU admission (discharged from ICU = 0, died in ICU = 1):
- (8) Booked admission to ICU after elective surgery; or elective admission for a procedure (e.g. insertion

- of a central line), or monitoring, or review of home ventilation (no = 0, yes = 1):
- (9) Record the number in square brackets if the condition is present (if in doubt, record 0):
 - [0] none
 - [1] cardiac arrest out of hospital
 - [2] severe combined immune deficiency
 - [3] malignancy after completion of 1st induction
 - [4] spontaneous cerebral haemorrhage from aneurysm or AV malformation
 - [5] cardiomyopathy or myocarditis
 - [6] hypoplastic left heart, <1 mth, requiring Norwood
 - [7] HIV infection
 - [8] IQ <35, worse than Down's
 - [9] a neurodegenerative disorder (progressive ongoing loss of milestones)
 - (10) Response of pupils to bright light (>3 mm and both fixed = 1, other = 0, unknown = 0):
 - (11) Base excess in arterial or capillary blood, mmol/l (unknown = 0):
 - (12) Pao₂, mm Hg (unknown = 0):
 - (13) Fio₂ at time of Pao₂ if oxygen via ETT or headbox (unknown = 0):
 - (14) Systolic blood pressure, mm Hg (unknown = 120):
 - (15) Mechanical ventilation at any time during the first hour in ICU (no = 0, yes = 1):

EXAMPLE OF HOW TO CALCULATE PIM

Child has: pupils react (fixed = no = 0), myocarditis (specified diagnosis = yes = 1), emergency admission (elective = no = 0), ventilated (= yes = 1), systolic blood pressure 40 mm Hg, base excess -16.0 mmol/l, Fio₂ 1.00, and Pao₂ 60 mm Hg.

$PIM\ logit = (2.357 \times 0) + (1.826 \times 1) + (-1.552 \times 0) + (1.342 \times 1) + (0.021 \times \text{absolute}(40 - 120)) + (0.071 \times \text{absolute}(-16.0)) + (0.415 \times 100 \times 1.00/60) - 4.873 = 1.803$.

$\text{Predicted probability of death} = e^{\logit} / (1 + e^{\logit}) = 2.7183^{1.803} / (1 + 2.7183^{1.803}) = 0.8585$, or 86%.

At <www.pedsccm.wustl.edu/clinical/pim-readme.html> there is a free program that calculates PIM.

COMMON MISTAKES IN COLLECTING PIM DATA

- (1) Do not overdiagnose the specified conditions (number 9 on the form)—if there is any doubt, do not record a specified condition. For example: do not code cerebral haemorrhage for intracerebral bleeding associated with trauma; impaired cardiac function associated with sepsis or surgery should not be coded as cardiomyopathy; Down's syndrome should not be coded as IQ <35; and a static disability should not be coded as neurodegenerative (even if it is severe) unless there is progressive ongoing loss of milestones. "Lymphoma/leukaemia after 1st induction" has been changed to "malignancy after 1st induction".
- (2) You should record the *first* value of each variable from the time of first contact up to one hour after arrival in your ICU (not the worst value).

- (3) If a variable is not measured within one hour of admission to ICU it should be coded as *missing* (for example, if the first blood gas is not done until two hours after admission, the base excess and Pao₂ should both be coded as missing). Missing data are treated as being normal when PIM is calculated.
- (4) The PIM equation is used to calculate the PIM logit. If any information is missing, that variable should add nothing to the PIM logit. For example, if the Pao₂ or the Fio₂ is missing, the value of "0.415 × 100 × Pao₂/Fio₂" should be set to zero.
- (5) Record the Fio₂ being given *at the same time* that the first Pao₂ is measured (that is, both the Fio₂ and Pao₂ that you record must relate to the same time).
- (6) Make sure that you are consistent about the Pao₂ units (all mm Hg, or all kPa), and the Fio₂ (all between 0.0 and 1.0, *not* percentages between 0 and 100).
- (7) Read very carefully the definition, "booked [prearranged] admission to ICU after *elective* surgery; or *elective* admission for a *procedure* (e.g. insertion of a central line), or *monitoring*, or *review* of home ventilation".
- (8) The pupils are only recorded as fixed if both are >3 mm, *and* both are fixed, *and* the finding is not caused by drugs or toxins or direct injury to the eye.
- (9) If systolic blood pressure is not measured in the first hour, record 120—*do not record zero*.
- (10) Randomly sample about every 20th admission to your ICU and get another person to collect the PIM data independently a second time, so that you can check the accuracy of your data.
- (11) You should include *all* admissions to your ICU, not just selected cases.

- 1 Hosmer DW, Lemeshow S. *Applied logistic regression*. New York: John Wiley & Sons, 1989.
- 2 Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med* 1988;16:1110-16.
- 3 Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med* 1996;24:743-52.
- 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 Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561-77.
- 6 Rapoport J, Teres D, Lemeshow S, et al. A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study. *Crit Care Med* 1994;22:1385-91.
- 7 Goddard JM. Pediatric risk of mortality scoring overestimates severity of illness in infants. *Crit Care Med* 1992;20:1662-5.
- 8 Pearson G, Shann F, Barry P, et al. Should paediatric intensive care be centralised? Trent versus Victoria. *Lancet* 1997;349:1213-17.
- 9 Shann F. Mortality prediction model is preferable to APACHE. *BMJ* 2000;320:714.