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.

Keywords: mortality; intensive care; prediction model

Table 1 Data by hospital

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Duration of data collection (mth)</th>
<th>Number of admissions (no. of deaths)</th>
<th>Admissions per bed per year</th>
<th>SMR (95% CI)</th>
<th>Mean probability of death (SD)</th>
<th>% retrieved</th>
<th>Vent in first hour</th>
<th>Average length of stay, days (SD)</th>
<th>Median length of stay, hours (IQR)</th>
<th>Area under ROC plot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alder Hey</td>
<td>12</td>
<td>1245 (85)</td>
<td>44.46</td>
<td>0.89 (0.72 to 1.06)</td>
<td>0.077 (0.133)</td>
<td>37%</td>
<td>78%</td>
<td>5.31 (10.82)</td>
<td>48 (24 to 144)</td>
<td>0.8344</td>
</tr>
<tr>
<td>Birmingham Children’s</td>
<td>48.5</td>
<td>4380 (374)</td>
<td>77.91</td>
<td>0.87 (0.80 to 0.95)</td>
<td>0.098 (0.155)</td>
<td>35%</td>
<td>84%</td>
<td>3.68 (8.32)</td>
<td>39 (20 to 91)</td>
<td>0.8373</td>
</tr>
<tr>
<td>St George’s</td>
<td>19</td>
<td>498 (25)</td>
<td>62.9</td>
<td>0.83 (0.53 to 1.12)</td>
<td>0.063 (0.124)</td>
<td>37%</td>
<td>35%</td>
<td>3.32 (5.48)</td>
<td>36 (19 to 90)</td>
<td>0.8371</td>
</tr>
<tr>
<td>Great Ormond St</td>
<td>6</td>
<td>341 (26)</td>
<td>68.2</td>
<td>0.88 (0.56 to 1.17)</td>
<td>0.087 (0.151)</td>
<td>4%</td>
<td>87%</td>
<td>4.08 (4.48)</td>
<td>48 (22 to 144)</td>
<td>0.8336</td>
</tr>
<tr>
<td>Guy’s</td>
<td>12</td>
<td>789 (67)</td>
<td>46.41</td>
<td>0.89 (0.69 to 1.07)</td>
<td>0.096 (0.145)</td>
<td>54%</td>
<td>77%</td>
<td>3.53 (5.72)</td>
<td>47 (24 to 95)</td>
<td>0.8019</td>
</tr>
</tbody>
</table>

Those wishing to audit paediatric intensive care units (PICUs) have to adjust or control their results for variations in “case mix”. In its broadest sense “case mix” describes all the confounding factors that could introduce bias, or the chances of a particular outcome in a comparison between different units or periods of time. The most significant element of case mix is the severity of the illness leading to PICU admission. Traditionally this is mathematically represented as a mortality risk (odds of dying), or the related “probability of death” (POD), as determined by a mortality prediction model derived by logistic regression1 from a large database of consecutive representative admissions. The process of routine audit 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/truma, 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

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

Table 3 Calibration across diagnostic categories and severity of illness bands

<table>
<thead>
<tr>
<th>Diagnostic group</th>
<th>Bands of death probability</th>
<th>&lt;1%</th>
<th>1% to 5%</th>
<th>5% to 15%</th>
<th>15% to 30%</th>
<th>30% +</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory</td>
<td></td>
<td>1.01</td>
<td>9.49</td>
<td>36.08</td>
<td>20.07</td>
<td>28.04</td>
<td>94.69</td>
</tr>
<tr>
<td>Actuadrach</td>
<td>n</td>
<td>139</td>
<td>385</td>
<td>409</td>
<td>97</td>
<td>56</td>
<td>1086</td>
</tr>
<tr>
<td>Cardiac</td>
<td></td>
<td>2.79</td>
<td>37.54</td>
<td>50.43</td>
<td>47.53</td>
<td>121.79</td>
<td>260.07</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>359</td>
<td>1718</td>
<td>528</td>
<td>229</td>
<td>229</td>
<td>3063</td>
</tr>
<tr>
<td>Neonatal</td>
<td></td>
<td>0.12</td>
<td>1.16</td>
<td>15.63</td>
<td>19.97</td>
<td>18.69</td>
<td>57.57</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>20</td>
<td>122</td>
<td>150</td>
<td>99</td>
<td>33</td>
<td>424</td>
</tr>
<tr>
<td>Postoperative</td>
<td></td>
<td>0.95</td>
<td>4.33</td>
<td>5.90</td>
<td>2.15</td>
<td>1.77</td>
<td>15.10</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>189</td>
<td>196</td>
<td>72</td>
<td>10</td>
<td>5</td>
<td>472</td>
</tr>
<tr>
<td>Accident/trauma</td>
<td></td>
<td>0.34</td>
<td>5.85</td>
<td>8.05</td>
<td>3.67</td>
<td>14.62</td>
<td>32.54</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>45</td>
<td>175</td>
<td>114</td>
<td>16</td>
<td>28</td>
<td>376</td>
</tr>
<tr>
<td>Neurological</td>
<td></td>
<td>0.73</td>
<td>8.01</td>
<td>19.74</td>
<td>12.11</td>
<td>30.49</td>
<td>71.09</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>109</td>
<td>295</td>
<td>245</td>
<td>60</td>
<td>58</td>
<td>767</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
<td>0.95</td>
<td>9.37</td>
<td>27.14</td>
<td>22.49</td>
<td>68.85</td>
<td>128.81</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>145</td>
<td>390</td>
<td>309</td>
<td>106</td>
<td>115</td>
<td>1065</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>6.90</td>
<td>77.77</td>
<td>162.97</td>
<td>127.98</td>
<td>284.25</td>
<td>659.87</td>
</tr>
<tr>
<td>Actual deaths</td>
<td>n</td>
<td>1006</td>
<td>3279</td>
<td>1827</td>
<td>617</td>
<td>524</td>
<td>7253</td>
</tr>
</tbody>
</table>

by probability of death. 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% 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.

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
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), neoplasms (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 performance of the units to 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), neoplasms (53), acute hepatitis or liver failure (49), haemopoietic disease (48), renal failure (38), and inborn errors of metabolism (31).

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://pedsccm.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]were 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 absolute(−16.0)) + (0.415 × 1.342 × 1.803. 

(11)Systolic blood pressure, mm Hg (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 = 0):

(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 (unknown = 0):

PIM logit = (2.557 × 0) + (1.826 × 0) + (−1.552 × 0) + (1.342 × 1) + (0.021 × absolute(40 − 120)) + (0.071 × absolute(−16.0)) + (0.415 × 100 × 1.00/60) − 4.873 = 1.803.

Predicted probability of death = e^X / ( 1 + e^X ) = 2.7183e^x

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 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.

9 Shann F. Mortality prediction model is preferable to APACHE. BMJ 2000;320:714.