ORIGINAL ARTICLE

A comparison of three scoring systems for mortality risk among retrieved intensive care patients

S M Tibby, D Taylor, M Festa, S Hanna, M Hatherill, G Jones, P Habibi, A Durward, I A Murdoch

Arch Dis Child 2002;87:421–425

See end of article for authors' affiliations

Correspondence to:
Dr S M Tibby, Paediatric
Intensive Care Unit, 9th
floor, Guy's Tower, Guy's
Hospital, London SE1 9RT,
UK;
shane.tibby@
gstt.sthames.nhs.uk

Accepted 24 June 2002

Aims: To assess the impact of two paediatric intensive care unit retrieval teams on the performance of three mortality risk scoring systems: pre-ICU PRISM, PIM, and PRISM II.

Methods: A total of 928 critically ill children retrieved for intensive care from district general hospitals in the south east of England (crude mortality 7.8%) were studied.

Results: Risk stratification was similar between the two retrieval teams for scores utilising data primarily prior to ICU admission (pre-ICU PRISM, PIM), despite differences in case mix. The fewer variables required for calculation of PIM resulted in complete data collection in 88% of patients, compared to pre-ICU PRISM (24%) and PRISM II (60%). Overall, all scoring systems discriminated well between survival and non-survival (area under receiver operating characteristic curve 0.83–0.87), with no differences between the two hospitals. There was a tendency towards better discrimination in all scores for children compared to infants and neonates, and a poor discrimination for respiratory disease using pre-ICU PRISM and PRISM II but not PIM. All showed suboptimal calibration, primarily as a consequence of mortality over prediction among the medium (10–30%) mortality risk bands.

Conclusions: PIM appears to offer advantages over the other two scores in terms of being less affected by the retrieval process and easier to collect. Recalibration of all scoring systems is needed.

Mortality risk scoring systems are integral to the provision of modern intensive care, providing a measure of performance both between and within individual intensive care units over time.¹ A valid scoring system must predict mortality accurately while adjusting for case mix and disease severity,^{1–3} but also requires data capture that is feasible in clinical practice.³

The common paediatric intensive care scores identify physical intensive care unit (ICU) admission as a crucial event, and may utilise data captured either prior or subsequent to ICU admission, or from a combination of both.⁴ A tacit assumption in these scores is that ICU care begins when the patient enters the ICU. However, the advent of mobile retrieval teams in the United Kingdom over the last decade has meant that ICU care may effectively commence prior to this. The precise point at which ICU care begins is difficult to define, and is a function of many factors including the acuity and type of disease process, experience and resources available to the referring team in the district general hospital, adequacy of telephonic communication with the referral centre, and skill of the retrieval team.

The impact of retrieval practice as such on the performance of mortality risk scoring systems is unknown. Paediatric retrievals are now commonplace in the South Thames region, undertaken primarily by teams based in two referral hospitals: Guy's and St Mary's. Since the inception of the retrieval services in the early 1990s, the total workload has increased in accordance with national recommendations,^{5,6} currently accounting for 50% and 85% of total ICU admissions to each of the above hospitals respectively. Our aim in this study was to evaluate the performance of commonly used paediatric mortality risk scoring systems among patients retrieved by these two hospitals. The secondary goals were: (a) to investigate whether differences in retrieval practice, age, and case mix affected scores' performance; and (b) to assess the clinical feasibility of data collection required for each of the scoring systems.

METHODS

Data were collected prospectively on 928 children retrieved by two teaching hospitals (hospital A, 592; hospital B, 336) in the south east of England over a 21 month period (December 1997 to September 1999). Hospital A has a 16 bed multidisciplinary ICU (including cardiac surgery); while hospital B contains an eight bed general ICU, with an interest in sepsis. The retrieval services are run separately; both are coordinated by dedicated ICU consultants and staffed by ICU fellows and nurses from the respective units. The catchment areas for both hospitals overlap; hospital A covering primarily (but not exclusively) South Thames, hospital B covering North and South Thames. The remaining retrieval service operating in the Thames region, Great Ormond Street, covers North Thames primarily, and was not included in the study.

Data were gathered as part of routine clinical management, thus the collection of every variable required to fully calculate mortality risk using each scoring system was not necessarily achieved in every patient. This strategy avoided subjecting patients to unnecessary blood tests, and served as a means of assessing the feasibility of data collection as part of standard clinical practice for each of the scoring systems. As the study utilised data collected as part of routine management, ethical approval was not sought.

Three scoring systems were compared (fig 1):

- (1) Pre-ICU PRISM (paediatric risk of mortality)⁷; a physiological based system incorporating 14 variables collected in-hospital over a maximum of 24 hours immediately prior to physical ICU admission. This may include data collected at the referring hospital (for example, from the accident and emergency department or the ward) as well as during the retrieval period.

Abbreviations: ICU, intensive care unit; IQR, interquartile range; PIM, paediatric index of mortality; SMR, standardised mortality ratio

Figure 1 Periods of data collection for the three mortality risk scores.

- (2) PIM (paediatric index of mortality)⁸; a point-of-care score encompassing eight variables collected from the time of first physical contact between the patient and the ICU retrieval team up until one hour after physical ICU admission.
- (3) PRISM II⁹; utilising the same variables as the pre-ICU PRISM, but covering the first 24 hours following physical ICU admission. A PRISM II score was calculated only if the patient stayed in the ICU for more than eight hours.

Details of the components of each of the scores, together with the coefficients allowing calculation of mortality risk are given elsewhere.⁷⁻⁹ Only written data were included, and were extracted from the local (referring) hospital case notes, the retrieval logs, and from the ICU case notes.

An updated score, PRISM III, has been available since 1996.¹⁰ This test utilises 17 physiological and several categorical factors, providing greater predictive accuracy. However, the coefficients remain the property of the developer, and a charge is levied for use. Because of this, the score is not widely used in the UK, and was thus not assessed in this study.

Statistical analysis

Quality of data collection was verified by an intraclass correlation coefficient of reliability >0.9 for all scoring systems on 50 randomly selected case notes.¹¹ Demographic data between hospitals were analysed using the Mann-Whitney test. Comparisons of the proportion of patients falling within mortality risk bands ($<1\%$, $1-5\%$, $5-15\%$, $15-30\%$, and $>30\%$) for each score, and for the distribution of patients within each disease category were by χ^2 analysis.

Mortality discrimination was quantified by calculation of the area under the receiver operating characteristic curve (a plot of sensitivity versus $1-\text{specificity}$ for each of the scores).¹²⁻¹⁵ Discrimination refers to the ability of the test to calculate a higher mortality probability among non-survivors

than survivors across the whole group, with acceptable discrimination represented by an area under the curve of $0.70-0.79$, and good discrimination by an area ≥ 0.80 .¹⁴

Calibration signifies how well the test predicts both mortality and survival across subcategories of risk. Here we employed the Hosmer-Lemeshow test, where acceptable calibration is evidenced by a p value ≥ 0.10 .¹⁵ This test has been criticised as being unreliable when the number of covariate patterns is less than the number of subjects,¹⁴ thus we also present the data visually as an observed versus expected mortality plot (with 95% confidence intervals) for deciles of risk.

Mortality was also standardised for case mix using the standardised mortality ratio (SMR),¹⁶ Z score,¹⁷ and standardised Z score.¹⁸ The SMR is the ratio of risk adjusted, observed mortality to the expected mortality as derived from the development set. If the 95% confidence intervals around the SMR are less than 1.0, then mortality is lower than that seen in the development set; conversely confidence intervals greater than 1 signify a higher mortality. The Z score expresses risk adjusted mortality in terms of a z distribution, with a score >2 implying good performance, and ≤ 2 poor performance (excess mortality). The standardised Z score adjusts for differences in the proportion of patients falling within different risk categories between the study cohort and the patients in the development set. For example, if the study group contains a much higher proportion of patients in a particular risk category than the development set, this may produce undue influence on the unstandardised Z score.

Statistical programmes included Microsoft Excel, Instat (Graphpad Software, San Diego, California, USA) and Analyse-it (Analyse-it Software, Leeds, UK). Unless otherwise specified, data are expressed as median and interquartile range (IQR).

RESULTS

Patient demographics and case mix

The median patient age was 15 months (IQR 3–54), and differed statistically ($p < 0.001$) between hospitals: A, 11 months (2–49); B, 18 months (6–57). Sites of patient referral included: general paediatric ward (39%), accident and emergency department (28%), adult ICU (16%), neonatal ICU (8%), operating theatre (6%), and other (3%). Prior to referral, patients had been in the district hospital for a median time of six hours (IQR 3–18). Disease categories included respiratory (39%), sepsis (30%), neurological (14%), cardiac (5%), other (3%), trauma (3%), poisoning (2%), diabetic ketoacidosis (2%), non-septic shock (1%), and postoperative haemorrhage (usually following tonsillectomy) (1%). The distribution of patients across disease categories differed between hospitals

Table 1 Data collection profile for the three scoring systems

	Pre-ICU PRISM (n=928)	PIM (n=928)	PRISM II (n=903)
Data collection complete	24	88	60
1 variable missing	17	6	16
2 variables missing	16	6	8
3 variables missing	23	0	7
≥ 4 variables missing	20	0	9
Breakdown of missing variables (expressed as a percentage of total patients)			
Serum bilirubin	58	NA	21
Coagulation profile	55	NA	25
Serum calcium	46	NA	17
$\text{PO}_2:\text{FiO}_2$ ratio	19	12	12
Base excess	NA	6	NA

All numbers are expressed as percentages.

$\text{PO}_2:\text{FiO}_2$, ratio of partial pressure of dissolved arterial oxygen to inspired oxygen fraction; NA, not applicable.

Figure 2 Comparison of distribution of patients across mortality risk bands between hospitals for each scoring system.

($p < 0.0001$): hospital A had a greater proportion of respiratory (42 v 35%), neurological (16 v 11%), and cardiac (8 v 0%) illness, while sepsis predominated in hospital B (46 v 21%). Seventy six per cent of patients were mechanically ventilated prior to retrieval. The in-hospital crude mortality was 7.8% (72/928), with all deaths occurring in the ICU. The median time until death was 36 hours (IQR 12–120), with 30 deaths occurring less than 24 hours after admission. Among survivors the median length of ICU stay was 55 hours (IQR 27–112).

Figure 3 Semilogarithmic plot showing observed versus predicted mortality for "deciles of risk". Note poor calibration in mid-range mortality risk deciles. Error bars represent 95% confidence intervals.

Completeness of data collection

Pre-ICU PRISM and PIM data were available for all 928 patients, while PRISM II data were available for 903 (A, 570; B, 333). The inability to calculate PRISM II mortality risk for 25 patients was a result of early (<8 hours) ICU discharge ($n = 13$) or death ($n = 12$).

Complete data were obtained in a much higher proportion of patients for PIM (88%) compared to both pre-ICU PRISM (24%) and PRISM II (60%). Table 1 shows the number and type of missing variables.

Risk stratification

The two hospitals showed similar distributions of patient mortality risk for both pre-ICU PRISM ($p = 0.54$) and PIM ($p = 0.37$) but not PRISM II ($p = 0.02$), possibly reflecting differing retrieval practice (fig 2). This was further suggested by a median (IQR) stabilisation time (defined as the time spent by the retrieval team in the district general hospital) for hospital A of 1 hour and 16 minutes (55 minutes to 2 hours) and hospital B of 2 hours and 35 minutes (2 hours to 3 hours 26 minutes) ($p < 0.001$). For combined data, mortality risk distribution differed for pre-ICU PRISM, PIM, and PRISM II (all $p < 0.001$, data not shown).

Scores' performance for combined data

All scoring systems displayed a similar discrimination for predicting death (table 2). The apparent divergence in standardised mortality ratios and Z scores between PIM and PRISM II was minimised after calculation of the standardised Z score. This was predominantly caused by a differing distribution of patients across risk bands in our study compared to the original PIM derivation population. All showed suboptimal calibration; however, PRISM II was the closest to exhibiting an acceptable p value (table 2). Calibration appeared deficient in the mid-range mortality risk strata for all systems (fig 3).

Scores' performance with subgroup analysis

Table 3 presents discrimination and calibration data, examining potential differences in scores' performance between hospitals, across age ranges, diagnostic categories, and for differing stabilisation times. The categories generally showed acceptable discrimination; however, borderline values were seen for both pre-ICU PRISM and PRISM II with respiratory disease, and for pre-ICU PRISM among infants. Calibration was superior across all categories for PRISM II.

DISCUSSION

Risk adjusted mortality remains the commonest benchmark for neonatal,¹⁹ paediatric,⁴ and adult¹ ICU performance. However, the validity of a scoring system for calculation of mortality risk depends on several factors. The model must be

Table 2 Performance of the three scoring systems for combined hospital data

	Pre-ICU PRISM (n=928)	PIM (n=928)	PRISM II (n=903)
Median percentage risk (IQR)	6.1 (2.9–17.5)	8.3 (4.3–15.5)	3.3 (1.4–10.1)
Mean percentage risk (SD)	14.4 (20.7)	13.6 (16.2)	9.4 (15.3)
Area under ROC curve (95% CI)	0.83 (0.78 to 0.89)	0.87 (0.82 to 0.92)	0.86 (0.81 to 0.92)
Standardised mortality ratio (95% CI)	0.54 (0.42 to 0.67)	0.57 (0.44 to 0.70)	0.71 (0.53 to 0.89)
Z score	7.08	5.85	3.31
Standardised Z score	7.05	2.69	2.44
Hosmer-Lemeshow χ^2	59.4	38.9	14.2
Hosmer-Lemeshow p	<0.0001	<0.0001	0.08

ROC, receiver operating characteristic; CI, confidence interval.

Table 3 Performance of the three scoring systems for subgroups

	n	Area under ROC curve (95% CI)			Hosmer-Lemeshow p		
		Pre-ICU PRISM	PIM	PRISM II	Pre-ICU PRISM	PIM	PRISM II
Hospital							
A	592	0.83 (0.77 to 0.90)	0.87 (0.81 to 0.92)	0.87 (0.81 to 0.93)	<0.005	0.013	0.18
B	336	0.86 (0.75 to 0.96)	0.87 (0.78 to 0.97)	0.84 (0.70 to 0.97)	<0.001	<0.005	0.34
Age (months)							
<1	92	0.80 (0.66 to 0.94)	0.79 (0.62 to 0.96)	0.85 (0.72 to 0.98)	0.02	0.11	0.27
1–12	338	0.71 (0.58 to 0.84)	0.79 (0.68 to 0.90)	0.84 (0.75 to 0.94)	<0.001	0.02	0.06
>12	498	0.89 (0.83 to 0.95)	0.92 (0.87 to 0.97)	0.88 (0.80 to 0.95)	<0.001	0.002	0.20
Stabilisation times							
<2 hours	528	0.84 (0.78 to 0.92)	0.86 (0.80 to 0.92)	0.86 (0.79 to 0.94)	<0.001	0.03	0.09
>2 hours	400	0.85 (0.75 to 0.95)	0.85 (0.75 to 0.98)	0.85 (0.72 to 0.97)	<0.001	<0.01	0.10
Diagnosis							
Respiratory	362	0.67 (0.54 to 0.80)	0.79 (0.69 to 0.90)	0.72 (0.56 to 0.87)	0.02	0.007	0.001
Sepsis	277	0.90 (0.85 to 0.96)	0.89 (0.83 to 0.95)	0.94 (0.90 to 0.98)	<0.001	0.05	0.05
Neurological	131	0.97 (0.94 to 1.00)	0.89 (0.70 to 1.00)	0.90 (0.73 to 1.00)	0.03	0.08	0.20
Other	158	0.80 (0.68 to 0.93)	0.84 (0.74 to 0.95)	0.85 (0.77 to 0.94)	<0.001	0.08	0.38

For n<300 the Hosmer-Lemeshow calculations were performed using quintiles, rather than deciles of risk.
ROC, receiver operating characteristic; CI, confidence interval.

accurate,^{1–20,21} it obliges data capture that is feasible in clinical practice,³ and should be updated to reflect changes and advances in ICU care.²² The accuracy of the common paediatric ICU scores is well established; however, all were developed from data collected in units where paediatric retrieval is not common. It has been shown that retrieval teams alter many of the physiological variables used in the calculation of mortality risk (particularly so for variables collected after ICU admission),²³ and this may have a knock-on effect on the performance of the score itself. Our aim in this study was thus to evaluate the effect of the retrieval process on the validity of three common scores, as defined above.

The three scoring systems evaluated in the current study utilise data from overlapping time periods (fig 1). Pre-ICU PRISM includes the “worst” data prior to physical ICU admission (including retrieval data), while PRISM II is recorded for the 24 hour interval following admission. PIM captures data from point of first retrieval team contact until one hour post-ICU admission. However, rather than capture the “worst” data from this time period, PIM utilises the first measurement of a particular variable captured after contact with the retrieval team, and therefore, like the pre-ICU PRISM, PIM tends to reflect what is happening to the patient prior to arrival of the retrieval team. Not surprisingly, this resulted in a differing pattern of risk stratification between the scores, with PRISM II showing a lower overall predicted mortality risk than pre-ICU PRISM and PIM (fig 2, table 2). This was presumably a result of increased “normalisation” of physiological and laboratory variables occurring during the retrieval process before physical ICU admission. Indeed, this retrieval effect is further highlighted by comparison of the two hospitals. Both hospitals show a similar mortality risk distribution (implying comparable disease severity) for pre-ICU PRISM and PIM. However,

the reduction in mortality risk seen with PRISM II is accentuated in hospital B (fig 2), presumably as a consequence of the longer time spent stabilising the patients in the district general hospital.

Despite these differences there was no disparity in mortality discrimination between the scores (area under the curve 0.83–0.87). However, all scores showed suboptimal calibration (table 2) because of an over prediction of mortality in the mid-range risk groups (10–30%, fig 3). This was reflected in a lower SMR and higher Z score for all three models. However, it is worth noting that the performance of PIM approximated that of PRISM II after standardisation of the Z score. Whether this “over performance” represents a general improvement in ICU care with time, given that PRISM II was developed from data collected prior to 1988, pre-ICU PRISM prior to 1995, and PIM from data collected in 1994–96, or rather that this represents an additional positive impact of retrieval per se remains to be seen. However, recent data suggest that both factors may be important. Pearson *et al* have published PIM derived data from five UK ICUs covering the period 1998 to 1999.²⁴ Eighty four per cent of the 7253 patients studied were not retrieved. The overall SMR was 0.87 (95% confidence interval 0.81 to 0.94), which was remarkably similar to the largest participating unit, Birmingham Children’s Hospital, where only 3% of total admissions were retrieved. Our data show a further reduction in SMR to 0.57 (0.44–0.70), which may suggest a retrieval effect, and/or differences in case mix; however, this cannot be proven without a direct comparison between retrieved and non-retrieved patients of similar case mix from the same units. We were unable to examine this in our study, owing to the small number of outside admissions who were not retrieved.

We also performed subgroup analysis, examining the performance of the scores according to hospital, age, stabilisation time, and disease category (table 3). Although definitive conclusions from subgroup analysis are limited owing to small sample size, several patterns did emerge. Differences in retrieval practice did not appear important, either between the two hospitals or in terms of stabilisation times. All scores provided better discrimination in children compared to infants and neonates; however, PIM and PRISM II were still acceptable across all age groups. There was a tendency towards poor discrimination with respiratory disease for both pre-ICU PRISM and PRISM II (the lower limits of the 95% confidence intervals for the area under the ROC curve being 0.54 and 0.56 respectively). Calibration data for all subgroups reflected the combined data, namely PRISM II > PIM > pre-ICU PRISM.

Sensibility is another aspect of a score's validity, which encompasses the feasibility of data collection required to calculate the score.³ Here PIM proved superior, with an 88% completion rate. Debate exists as to the impact of missing data on a score's performance; certainly one inference is that missing data is often normal, and thus does not contribute significantly to a risk score.²⁵ Nonetheless, one must question the clinical validity of a score such as pre-ICU PRISM where complete data collection occurs in only 24% of patients, and 20% were missing four or more variables.

Although a score does not provide a risk assessment for individual patients, it does permit categorisation into a particular risk category, which may allow for targeting of novel or high risk therapies towards the sickest patient groups. A significant proportion of paediatric mortality occurs soon after ICU admission,²⁶ thus a score such as PIM that allows early identification of high risk patients has greater usefulness. Indeed this has been a criticism levelled at PRISM II, in that it may diagnose rather than predict death.⁸ This is consistent with our results, where a PRISM score could not be calculated on 12 patients who died less than eight hours after ICU admission (17% of all deaths); furthermore, 30 deaths (42%) occurred within 24 hours. Also, as highlighted earlier, PIM and pre-ICU PRISM are less affected by retrieval team practice, and thus may provide a better early snapshot of disease severity.

In summary, we suggest that PIM represents the most suitable mortality risk score for retrieved paediatric ICU patients. The data required for calculation of this score are easy to collect, non-proprietary, and because the data are collected at "point-of-care", risk stratification does not appear affected by retrieval practice. Mortality risk can be calculated at an early stage after ICU admission. However, recalibration is needed to reflect advances in ICU care; the authors of PIM have stated that an updated version will soon be available.¹⁴

Authors' affiliations

S M Tibby, D Taylor, S Hanna, M Hatherill, G Jones, A Durward, I A Murdoch, Department of Paediatric Intensive Care, Guy's Hospital, London, UK
P Habibi, M Festa, Department of Paediatric Intensive Care, St Mary's Hospital, London, UK

REFERENCES

- 1 Lemeshow S, Le Gall JR. Modelling the severity of illness of ICU patients. A systems update. *JAMA* 1994;**272**:1049–55.
- 2 Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Intensive Care Med* 1995;**21**:770–6.
- 3 Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997;**277**:488–94.
- 4 Marcin JP, Pollack MM. Review of the methodologies and applications of scoring systems in neonatal and pediatric intensive care. *Pediatr Crit Care Med* 2000;**1**:20–7.
- 5 British Paediatric Association. *The care of critically ill children: report of a multidisciplinary working party on intensive care*. London: British Paediatric Association, 1993.
- 6 Department of Health. Paediatric intensive care: "a framework for the future". Report from the national coordinating group on paediatric intensive care to the chief executive of the NHS executive. London: Department of Health, 1997.
- 7 Kanter RK, Edge WE, Caldwell CR, et al. Pediatric mortality probability estimated from pre-ICU severity of illness. *Pediatrics* 1997;**99**:59–63.
- 8 Shann F, Pearson G, Slater A, et al. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Med* 1997;**23**:201–7.
- 9 Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med* 1988;**16**:1110–16.
- 10 Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated Pediatric Risk of Mortality score. *Crit Care Med* 1996;**24**:743–52.
- 11 Fleiss JL. *The design and analysis of clinical experiments*. New York: John Wiley and Sons, 1986:2–3.
- 12 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29–36.
- 13 Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983;**148**:839–43.
- 14 Shann F. Are we doing a good job: PRISM, PIM and all that. *Intensive Care Med* 2002;**28**:105–7.
- 15 Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;**115**:92–106.
- 16 Bland M. *An introduction to medical statistics*, 2nd edn. Oxford: Oxford Medical Publications, 1995:293–6.
- 17 Flora JD. A method for comparing survival of burn patients to a standard survival curve. *J Trauma* 1978;**18**:701–5.
- 18 Hollis S, Yates DW, Woodford M, et al. Standardized comparison of performance indicators in trauma: a new approach to case-mix variation. *J Trauma* 1995;**38**:763–6.
- 19 The international neonatal network. The CRIB (Clinical Risk Index for Babies) score: a tool for assessing initial neonatal risk comparing performance of neonatal intensive care units. *Lancet* 1993;**342**:193–8.
- 20 Harrell FE Jr, Lee KL, Matchar DB, et al. Regression models for prognostic prediction: advantages, problems and suggested solutions. *Cancer Treat Rep* 1985;**69**:1071–7.
- 21 Harrell FE Jr, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–87.
- 22 Randolph AG. Paediatric index of mortality (PIM): do we need another paediatric mortality prediction score? *Intensive Care Med* 1997;**23**:141–2.
- 23 Britto J, Nadel S, Maconochie I, et al. Morbidity and severity of illness during interhospital transfer: impact of a specialised paediatric retrieval team. *BMJ* 1995;**311**:836–9.
- 24 Pearson GA, Stickley J, Shann F. Calibration of the paediatric index of mortality in UK paediatric intensive care units. *Arch Dis Child* 2001;**84**:125–8.
- 25 Pollack MM, Patel KM, Ruttimann U, et al. Frequency of variable measurement in 16 pediatric intensive care units: influence on accuracy and potential for bias in severity of illness assessment. *Crit Care Med* 1996;**24**:74–7.
- 26 Proulx F, Gauthier M, Nadeau D, et al. Timing and predictors of death in pediatric patients with multiple organ system failure. *Crit Care Med* 1994;**22**:1025–31.