

ORIGINAL ARTICLE

Reliability of PRISM and PIM scores in paediatric intensive care

J G van Keulen, K H Polderman, R J B J Gemke

Arch Dis Child 2005;90:211–214. doi: 10.1136/adc.2003.046722

See end of article for authors' affiliations

Correspondence to:
Dr K H Polderman,
Department of Intensive
care, VU University
Medical Center, PO Box
7057, 1007 MB
Amsterdam, Netherlands;
k.polderman@tip.nl

Accepted 20 May 2004

Aims: To assess the reliability of mortality risk assessment using the Paediatric Risk of Mortality (PRISM) score and the Paediatric Index of Mortality (PIM) in daily practice.

Methods: Twenty seven physicians from eight tertiary paediatric intensive care units (PICUs) were asked to assess the severity of illness of 10 representative patients using the PRISM and PIM scores. Physicians were divided into three levels of experience: intensivists (>3 years PICU experience, n = 12), PICU fellows (6–30 months of PICU experience, n = 6), and residents (<6 months PICU experience, n = 9). This represents all large PICUs and about half of the paediatric intensivists and PICU fellows working in the Netherlands.

Results: Individual scores and predicted mortality risks for each patient varied widely. For PRISM scores the average intraclass correlation (ICC) was 0.51 (range 0.32–0.78), and the average kappa score 0.6 (range 0.28–0.87). For PIM scores the average ICC was 0.18 (range 0.08–0.46) and the average kappa score 0.53 (range 0.32–0.88). This variability occurred in both experienced and inexperienced physicians. The percentage of exact agreement ranged from 30% to 82% for PRISM scores and from 28 to 84% for PIM scores.

Conclusion: In daily practice severity of illness scoring using the PRISM and PIM risk adjustment systems is associated with wide variability. These differences could not be explained by the physician's level of experience. Reliable assessment of PRISM and PIM scores requires rigorous specific training and strict adherence to guidelines. Consequently, assessment should probably be performed by a limited number of well trained professionals.

Scoring systems such as the Paediatric Risk of Mortality (PRISM) score and Paediatric Index of Mortality (PIM) are widely used in paediatric intensive care. These are third generation scoring systems that allow assessment of the severity of illness and mortality risk adjustment in heterogeneous groups of patients in an objective manner, enabling conversion of these numbers into a numerical mortality risk based on logistic regression analysis. The purpose of their usage varies, and may include comparison of severity of illness between different treatment arms in clinical trials and comparison of quality of care between paediatric intensive care units (PICUs) using standardised (that is, severity of illness adjusted) mortality rates. Both the PRISM and PIM scoring system have been developed and carefully validated in tertiary PICUs.^{1–3} In some centres that were closely involved in developing these scoring systems, preliminary data have indicated that the degree of inter-observer reliability was acceptable.^{4–5} As was the case in these centres, patients are preferably scored by a small and dedicated number of thoroughly trained professionals (for example, research nurses). In principle, this form of organisation can be expected to have a high inter-observer agreement. However, the practical situation in numerous ICUs and PICUs throughout Europe is that severity scoring is performed by a varying number of residents, fellows, (paediatric) intensivists, paediatricians, or nurses, with varying degrees of PICU experience and training with the PRISM and PIM score. This may imply that these scoring systems will not achieve the same degree of accuracy and reliability in everyday clinical practice.

We previously showed significant degrees of inter-observer and intra-observer variability in the use of the APACHE II scoring system, both in the research setting and in everyday clinical practice.^{7–11} To our knowledge no studies have been carried out to systematically address this issue in paediatric

ICUs. We therefore decided to assess the accuracy and reliability of PRISM and PIM scoring in everyday clinical practice, stratified by level of experience.

METHODS

Physicians from eight academic paediatric ICUs (tertiary referral centres) with residency and fellowship training programmes were asked to participate in our study. Physicians were divided into three categories: residents (n = 9) with limited experience in paediatric intensive care (average 3 months, range 6 weeks–6 months); PICU fellows (n = 6, average experience 18 months, range 6–30 months), and paediatric intensivists (n = 12) with at least three years of full time PICU experience. For the Dutch situation this means that all major paediatric centres were represented, and that around 50% of all paediatric intensivists and PICU fellows participated in our study.

The charts of 10 patients that had been admitted to a single PICU in the course of a one year period were selected for scoring. The charts were selected to reflect typical PICU patients, and were not chosen for difficulty of scoring. Relevant data from the medical charts and copies of blank data collection sheets from the PRISM and PIM scores were provided to all participating physicians. No specific additional instructions were given to those who scored the patients, and no time limit was given in order to mimic the routine working situation as closely as possible. All physicians assessed PRISM and PIM scores, which were noted on similar data collection sheets as are used for routine assessment. Mean (SD) and range of the PRISM and PIM scores were calculated for each individual patient for the

Abbreviations: ICC, intraclass correlation; ICU, intensive care unit; PICU, paediatric intensive care unit; PIM, Paediatric Index of Mortality; PRISM, Paediatric Risk of Mortality

Table 1 Overall inter-observer agreement of PRISM score based mortality risk (%) and PIM score based mortality risk (%)

	PRISM score based mortality risk*			PIM score based mortality risk†		
	Mean (SD)	EA	κ scores (95% CI)	Mean (SD)	EA	κ scores (95% CI)
Patient 1	4.7 (6.9)	30	0.57 (0.51 to 0.64)	9.7 (8.8)	28	0.37 (0.28 to 0.45)
Patient 2	33.0 (27.4)	42	0.42 (0.35 to 0.50)	2.1 (2.0)	32	0.32 (0.20 to 0.44)
Patient 3	0.6 (0.3)	69	0.68 (0.62 to 0.76)	1.3 (1.3)	64	0.60 (0.54 to 0.65)
Patient 4	2.7 (2.4)	32	0.58 (0.50 to 0.64)	7.0 (9.1)	24	0.37 (0.31 to 0.42)
Patient 5	0.5 (0.3)	52	0.61 (0.57 to 0.66)	2.7 (1.8)	64	0.56 (0.48 to 0.65)
Patient 6	1.8 (0.5)	82	0.84 (0.80 to 0.88)	4.8 (3.1)	84	0.88 (0.81 to 0.94)
Patient 7	3.2 (2.1)	33	0.28 (0.21 to 0.35)	6.8 (4.5)	40	0.36 (0.32 to 0.41)
Patient 8	17.2 (14.0)	41	0.63 (0.58 to 0.68)	15.2 (15.3)	36	0.58 (0.51 to 0.64)
Patient 9	0.6 (0.4)	78	0.87 (0.80 to 0.93)	13.3 (25.1)	36	0.61 (0.55 to 0.67)
Patient 10	2.5 (0.9)	44	0.54 (0.46 to 0.61)	4.2 (1.2)	64	0.60 (0.55 to 0.65)

EA, percentage exact agreement.

*ICC PRISM=0.51 (range 0.32–0.78); †ICC PIM=0.18 (range 0.08–0.46).

overall group of physicians and for each of the three categories of physicians, according to methods described previously.⁸ Weighted kappa scores and intraclass correlations (ICC) were also determined to assess inter-observer agreement. A deviation of ± 1 point (PRISM score) or ± 1 item (PIM score) was designated as mild disagreement, ± 2 or 3 points (items) as moderate disagreement, and > 3 points (items) as severe disagreement. Statistical analysis was performed using Student's *t* test for unpaired variables for paired groups and by analysis of variance (ANOVA). Statistical significance was accepted for $p < 0.05$. Excel (Microsoft Inc.) and SPSS 9.0 (SPSS Inc.) software was used to perform the necessary calculations.

RESULTS

Tables 1–3 present the results. Table 1 depicts overall inter-observer agreement as the percentage of exact agreement, kappa scores, and ICC for the PRISM and PIM based mortality risks, respectively. Tables 2 and 3 show the inter-observer agreement for the PRISM and PIM score derived mortality risk.

As is evident from the tables, a wide variation in both PRISM and PIM based mortality risk assessment was found, regardless of the level of expertise. Indeed, the differences between the three groups were small, with statistically significant differences observed only between residents and intensivists for the PRISM based mortality risk ($p < 0.01$), and for intensivists versus fellows and residents versus fellows for the PIM based mortality risk ($p < 0.05$).

On analysis, the most frequent problems and sources of error in determining PRISM scores were:

- The $\text{PaO}_2/\text{FiO}_2$ determination. An arterial PaO_2 and a measured FiO_2 are mandatory, so it is only possible to

include this measurement if oxygen is given through an endotracheal tube or by a blender, or when FiO_2 is measured by a “head” or “Gardner” box. In the absence of an arterial PaO_2 , the score should be missing.

- Assessing the Glasgow Coma Score (GCS) when sedatives have been given. It is only permitted to assess GCS before administration of sedatives or after their effect has worn off.
- In some observers there were difficulties in determining the appropriate time window for PRISM calculation. This should include the worst values in the first 24 hours after PICU admission; patients admitted for less than 2 hours should be excluded from scoring.^{1–3}

The most frequent problems in determining PIM scores were:

- PIM scores should be determined by taking the first value (instead of inappropriately using the worst value) during the first hour after starting intensive care treatment. This treatment may start outside the PICU, for example, in the emergency room, by a retrieval service or a trauma team, so that (in contrast to PRISM scores) in some cases values obtained before PICU admission can be used to assess PIM scores.
- Misinterpretation of underlying conditions. For example, cardiac failure during sepsis was inappropriately scored as cardiomyopathy, and muscular dystrophy was inappropriately counted as a neurodegenerative illness.
- Use of an FiO_2 value measured at a different moment than PaO_2 .
- Adequate definition of a booked admission to the ICU: this should be a *pre-arranged* admission, for example after *elective* surgery.

Table 2 PRISM score based mortality risk; inter-observer agreement by physician category

	Residents* (n=9)			Fellows† (n=6)			Intensivists‡ (n=12)		
	Mean (SD)	EA	κ scores (95% CI)	Mean (SD)	EA	κ scores (95% CI)	Mean (SD)	EA	κ scores (95% CI)
Patient 1	2.2 (1.3)	22	0.42 (0.31 to 0.58)	9.6 (13)	0	0.43 (0.27 to 0.68)	4.8 (2.5)	50	0.71 (0.64 to 0.78)
Patient 2	17.0 (24)	0	0.21 (0.11 to 0.34)	27.0 (24)	50	0.57 (0.38 to 0.76)	47.8 (25.1)	50	0.54 (0.42 to 0.72)
Patient 3	0.6 (0.3)	56	0.84 (0.62 to 0.98)	0.5 (0.2)	83	0.84 (0.66 to 0.98)	0.7 (0.2)	66	0.82 (0.72 to 0.91)
Patient 4	1.9 (0.1)	44	0.54 (0.37 to 0.64)	4.5 (4.4)	33	0.39 (0.20 to 0.58)	2.4 (1.2)	50	0.41 (0.30 to 0.51)
Patient 5	0.4 (0.1)	66	0.74 (0.58 to 0.89)	0.4 (0.1)	50	0.62 (0.49 to 0.78)	0.7 (0.4)	78	0.71 (0.64 to 0.77)
Patient 6	2.1 (0.8)	44	0.53 (0.41 to 0.68)	1.8 (0.0)	100	1.00 (0.92 to 1.0)	1.7 (0.3)	89	0.92 (0.83 to 0.96)
Patient 7	4.5 (2.9)	33	0.38 (0.24 to 0.53)	2.5 (1.4)	66	0.58 (0.45 to 0.70)	2.4 (0.9)	44	0.42 (0.32 to 0.51)
Patient 8	20.0 (16.4)	22	0.34 (0.26 to 0.41)	18.1 (15)	33	0.42 (0.29 to 0.55)	17.2 (8.0)	56	0.63 (0.55 to 0.72)
Patient 9	0.8 (0.6)	66	0.59 (0.46 to 0.70)	0.5 (0.0)	83	0.92 (0.80 to 0.99)	0.5 (0.1)	75	0.88 (0.81 to 0.65)
Patient 10	2.1 (0.7)	33	0.52 (0.41 to 0.64)	2.3 (0.7)	50	0.66 (0.51 to 0.80)	2.9 (1.1)	58	0.73 (0.66 to 0.80)

EA, percentage exact agreement.

*ICC residents=0.24 (range 0.07–0.57); †ICC fellows=0.40 (range 0.16–0.73); ‡ICC intensivists=0.73 (range 0.57–0.91).

Table 3 PIM Score based mortality risk; inter-observer agreement by physician category

	Residents* (n=9)			Fellows† (n=6)			Intensivists‡ (n=12)		
	Mean (SD)	EA	κ scores (95% CI)	Mean (SD)	EA	κ scores (95% CI)	Mean (SD)	EA	κ scores (95% CI)
Patient 1	11.8 (13.7)	15	0.27 (0.22 to 0.39)	7.1 (2.0)	33	0.42 (0.29 to 0.55)	9.6 (5.9)	44	0.48 (0.38 to 0.64)
Patient 2	16.2 (10.7)	56	0.55 (0.41 to 0.68)	39.8 (34.1)	0	0.31 (0.20 to 0.42)	14.2 (6.7)	50	0.51 (0.39 to 0.64)
Patient 3	12.6 (11.1)	66	0.54 (0.42 to 0.67)	20.2 (19.0)	50	0.70 (0.55 to 0.95)	1.0 (0.9)	80	0.74 (0.68 to 0.80)
Patient 4	5.1 (3.4)	22	0.40 (0.36 to 0.48)	13.8 (18.4)	33	0.30 (0.18 to 0.44)	5.1 (2.2)	60	0.56 (0.46 to 0.65)
Patient 5	3.8 (1.9)	44	0.51 (0.39 to 0.66)	2.2 (1.5)	50	0.48 (0.32 to 0.65)	2.3 (1.2)	60	0.62 (0.57 to 0.70)
Patient 6	6.1 (5.3)	33	0.72 (0.59 to 0.87)	4.5 (0.3)	66	0.76 (0.54 to 0.95)	4.2 (0.9)	80	0.82 (0.77 to 0.88)
Patient 7	9.0 (4.4)	33	0.38 (0.30 to 0.48)	6.5 (5.6)	33	0.36 (0.22 to 0.53)	5.3 (3.6)	30	0.34 (0.25 to 0.44)
Patient 8	22.5 (19.3)	0	0.55 (0.46 to 0.67)	19.3 (18.5)	0	0.41 (0.20 to 0.62)	7.8 (5.6)	30	0.63 (0.54 to 0.72)
Patient 9	7.3 (4.2)	44	0.67 (0.59 to 0.78)	12.1 (16.9)	33	0.62 (0.50 to 0.74)	18.5 (35.8)	40	0.54 (0.43 to 0.65)
Patient 10	4.1 (13.4)	56	0.61 (0.50 to 0.73)	4.8 (13.3)	66	0.58 (0.41 to 0.73)	4.3 (0.6)	70	0.71 (0.64 to 0.80)

EA, percentage exact agreement.

*ICC residents=0.33 (range 0.14–0.65); †ICC fellows=0.33 (range 0.10–0.69); ‡ICC=intensivists 0.16 (0.13–0.39).

- Pupils can only be scored as fixed when the diameter of *both* pupils is >3 mm *and* when there is no reaction to light. In addition, the possibility of a medication effect or a peripheral cause of absent or abnormal pupil reactions as a direct result of injury to the eyes should be taken into account. This was not always correctly interpreted.

DISCUSSION

The results of our study show that substantial inter-observer variability exists in both PRISM and PIM scoring. We had expected to find some variability on the basis of our experience in previous studies dealing with the use of the APACHE II score.^{7–11} In these studies inter-observer variability had ranged from 15% to 30%. However, we were quite surprised by the degree of inconsistency in the use of the PIM and PRISM scores. Kappa scores ranged from 0.28 to 0.87 in PRISM scores, and from 0.32 to 0.88 in PIM scores. For the PRISM score the average ICC was 0.51 (range 0.32–0.78) and the average kappa score 0.6 (range 0.28–0.87); for the PIM score the average ICC was 0.18 (range 0.08–0.46) and the average kappa score 0.53 (range 0.32–0.88). Inter-observer agreement tended to be better in less severely ill patients. In this category, the a priori chance of disagreement between observers is lower due to the fact that there are less abnormal variables.

Our results contrast with a report by Pollack and colleagues,⁶ who studied the effects of variation in frequency of variable measurements on PRISM scores and found these effects to be relatively small.⁶ This study did not deal with the issue of inter-observer reliability of the measured variables themselves, which might subsequently affect the scores. Thus the degree of variability may even have been underestimated in our study. In addition, errors may have occurred in different areas of assessment which subsequently “cancelled each other out”; for example, if one observer erroneously attributed points for chronic diseases while another erroneously attributed points for low blood pressure, these two observers will agree on the total number of points and predicted mortality; seemingly they are in concurrence, while in reality their agreement is poor.

Inter-observer variability in PIM scoring appeared higher than with PRISM scoring, despite the fact that the PIM score comprises fewer variables. There are several potential explanations for this observation. Firstly, all variables of the PIM score are entered directly into the logit equation from which the mortality risks are calculated, meaning that differences in the final risk are directly proportionate to differences in the underlying variable. In addition some variables of the PIM score carry a relatively high weight, hence an error in one of these variables will lead to a relatively large change in the overall score and in predicted

mortality (for example, an erroneous “specified diagnosis”). In PRISM score based mortality risk calculation, all underlying variables of acute physiology are added to a single summary score, which does not reflect that more points that are attributed to some underlying variables might have been cancelled out by fewer points that were attributed to other variables. Only this total number of points (the PRISM score) is entered in the regression equation (with age and operative status) to calculate mortality risk, and consequently underlying variation in PRISM score variables might be obscured.

Poor agreement in scoring was found in intensivists, PICU fellows, and residents. The differences between the groups were unexpectedly small, with experienced physicians performing only marginally better than inexperienced ones. In this respect it should be realised that the number of years of experience in the PICU and/or the age of the physicians does not necessarily imply dedicated training and regular checking of the appropriateness of the application of scoring systems.

Our findings have significant implications. PRISM and PIM scores are widely used in clinical trials to compare severity of illness, and for benchmarking and comparison of quality of care between different PICUs. Our observations show that inter-observer variability in both scores is of a magnitude that makes it questionable whether they can, and should, be used for these purposes when collected in this way. The results of scoring system assessments, especially when used in research protocols and as benchmarks for quality of care, should be unequivocal; this applies especially when they are used not just by health care providers but also by hospital managers, insurance companies, and other financial decision makers, as well as by patient interest organisations. At the very least a rigorous training programme is needed; ideally, patients should probably be scored by a small number of dedicated, well trained, and regularly audited staff members.

Our study has a number of potential limitations. Patients were scored mostly by physicians who were not specifically trained for scoring, and who worked in different hospitals. The patient information was presented in the form of copies of all relevant parts of the medical and nursing charts from one medical centre; the lack of familiarity with charts from another hospital might have caused some bias against physicians working in other centres with different charts. The situation in daily clinical practice in the participating centres was as follows: in five participating tertiary PICUs, patients are normally scored by intensivists and fellows only; in two centres, residents and nurses also participate in scoring; and in one centre two dedicated students do the scoring. Training and auditing are not formally regulated in any of the eight centres. Although we have not performed a formal study to assess way the severity of illness scores are

What is already known on this topic

- Severity of illness scoring is used for various purposes in (paediatric) intensive care
- Reliability has been assessed only in centres closely involved in their development
- For the APACHE score, used in adult ICUs, substantial inter-observer variation in routine assessment has been reported

collected in most European PICUs, informal inquiries indicate that circumstances similar to the one in the Netherlands exist in many other European countries. Therefore, it is highly likely that our results reflect the average situation in Europe, although further studies will be required to fully address this issue.

The situation may be very different in centres where scores are assessed exclusively by a restricted number of dedicated staff members (Frank Shann, personal communication). This method of score collection can be expected to have less inter-observer variability, although we are not aware of published studies in which this issue has been addressed. Scores are collected in this way in many centres in the United States and Australia, and in some centres in Europe. If this method is indeed superior, the problem of score collection could in theory be solved by allowing two or three specially trained and frequently monitored staff to score patients. However, this should first be shown in future studies.

Thus, although limiting score collection to a few dedicated individuals should probably be our long term goal, at this moment we urgently need to improve the performance of those currently assessing PIM and PRISM scores. The most frequent problems in scoring should not be too difficult to resolve by adequate training and guidelines. These include using the appropriate time windows, using the appropriate values (worst in the first 24 hours and first value in the first hour, respectively), appropriate assessment of GCS and pupils during use of sedatives, avoiding misinterpretation of underlying conditions, and inadequate interpretation of the use of oxygen. Methods through which such a training programme and strict guidelines can be implemented have been described previously, in studies where a marked decrease in inter-observer variability was achieved after a training programme was implemented.⁹ With such a programme, severity of illness scoring could still be performed by a relatively large and varying number of (rotating) physicians, or by a small number of well trained and dedicated (for example, nursing) staff members. The results of our study suggest that the latter option may be preferable, but it remains to be determined whether this can be achieved in all PICUs that use these scoring systems.

In conclusion, we observed wide inter-observer variability in assessment of PRISM and PIM scores in routine clinical practice. These scoring systems are widely used in clinical trials and as instruments to assess and compare quality of care between different PICUs. In our opinion, based on these results and in the current situation, the results of these score assessments should be interpreted with great caution. Reliability of the scores may be higher if scoring is performed by a limited number of dedicated individuals, and may be

What this study adds

- In this multicentre study, wide inter-observer variation in PRISM and PIM based mortality risk was found in routine clinical practice
- Inter-observer variability could not be explained by the physicians' levels of qualification
- Education, training, and guidelines for score assessment are needed, and perhaps severity of illness scoring in PICUs should be performed only by a limited number of well trained professionals

improved by strict guidelines and regular training. Further studies are required to address these issues.

ACKNOWLEDGEMENTS

The authors thank all participating physicians from the Dutch PICUs (Academic Medical Centre Amsterdam; University Medical Centre, Utrecht; Leiden University Medical Centre; Academic Hospital Groningen; Academic Hospital Maastricht; Erasmus University Medical Centre, Rotterdam; VU University Medical Centre) without whose most appreciated contributions this study would not have been possible. We also thank Dr E de Lange, Department of Statistics and Epidemiology, VU University Medical Center, for help with the statistical analysis, and Prof. AJ van Vught, Wilhelmina Children's Hospital and University Medical Center Utrecht, for critically reading the manuscript.

Authors' affiliations

J G van Keulen, R J B J Gemke, Department of Pediatrics, VU University Medical Center, Amsterdam, Netherlands

K H Polderman, Department of Intensive Care, VU University Medical Center, Amsterdam, Netherlands

Competing interests: none declared

REFERENCES

- 1 **Pollack MM**, Ruttimann UE, Getson PR. Pediatric risk of mortality score (PRISM) score. *Crit Care Med* 1988;**16**:1110–16.
- 2 **Shann F**, Pearson G, Slater A, *et al*. Pediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Med* 1997;**23**:201–7.
- 3 **Pollack MM**, Patel KM, Ruttimann UE. The pediatric risk of mortality III—acute physiology score (PRISM III-APS): a method of assessing physiologic instability for pediatric intensive care unit patients. *J Pediatr* 1997;**131**:575–81.
- 4 **Gemke RJB**, Bonsel GJ. The Pediatric Intensive Care Assessment of Outcome (PICASSO) study group. Comparative assessment of pediatric intensive care: a national multicenter study. *Crit Care Med* 1995;**23**:238–45.
- 5 **Slater A**, Shann F, Gearson G. PIM2: a revised version of the Pediatric Index of Mortality. *Intensive Care Med* 2003;**29**:278–85.
- 6 **Pollack MM**, Patel KM, Ruttimann UE, *et al*. Frequency of variable measurement in 16 pediatric intensive care units: influence on accuracy and potential for bias in severity of illness measurement. *Crit Care Med* 1996;**24**:74–7.
- 7 **Polderman KH**, Thijs LG, Girbes ARJ. Interobserver variability in the use of APACHE II scores [letter]. *Lancet* 1999;**353**:380.
- 8 **Polderman KH**, Girbes AR, Thijs LG, *et al*. Accuracy and reliability of APACHE II scoring in 2 intensive care units. Problems and pitfalls in the use of APACHE II and suggestions for improvement. *Anaesthesia* 2002;**56**:47–50.
- 9 **Polderman KH**, Jorna EMF, Girbes ARJ. Inter-observer variability in APACHE II scoring: effect of strict guidelines and training. *Intensive Care Med* 2001;**27**:1365–9.
- 10 **Polderman KH**, Christiaans HM, Wester JP, *et al*. Intraobserver variability in APACHE II scoring. *Intensive Care Med* 2001;**27**:1550–2.
- 11 **Polderman KH**, Girbes ARJ, Thijs LG, *et al*. Accuracy and reliability of Apache II scoring in two intensive care units. Problems and pitfalls in the use of APACHE II and suggestions for improvement. *Anaesthesia* 2001;**56**:47–50.