

Stéphane Leteurtre  
Bruno Grandbastien  
Francis Leclerc  
Roger Parslow  
Groupe Francophone  
de Réanimation et Urgences  
Pédiatriques  
Paediatric Intensive  
Care Audit Network

## International comparison of the performance of the paediatric index of mortality (PIM) 2 score in two national data sets

Received: 5 May 2011  
Accepted: 8 April 2012  
Published online: 9 May 2012  
© Copyright jointly held by Springer and ESICM 2012

Groupe Francophone de Réanimation et Urgences Pédiatriques (GFRUP): The members of the GFRUP who participated in this study are listed in the [Appendix](#).

Paediatric Intensive Care Audit Network (PICANet): We are grateful to all the audit clerks, secretaries, nurses and doctors who support and contribute to PICANet from their own PICUs, the Paediatric Intensive Care Society, the Paediatric Intensive Care Society Study Group, the PICANet Steering Group and the PICANet Clinical Advisory Group.

These data were presented, in part, at the 20th European Society of Paediatric and Neonatal Intensive Care (ESPNIC) annual congress on 14–17 June 2009, Verona, Italy.

### Electronic supplementary material

The online version of this article (doi:[10.1007/s00134-012-2580-6](https://doi.org/10.1007/s00134-012-2580-6)) contains supplementary material, which is available to authorized users.

S. Leteurtre · F. Leclerc  
Paediatric Intensive Care Unit, Jeanne de Flandre University Hospital, 59037 Lille Cedex, France

B. Grandbastien  
Department of Epidemiology and Public Health, Calmette Hospital, 59037 Lille Cedex, France

S. Leteurtre (✉) · B. Grandbastien · F. Leclerc  
Univ Lille Nord de France, UDSL, EA2694, 1, place de Verdun, 59045 Lille Cedex, France  
e-mail: [stephane.leteurtre@chru-lille.fr](mailto:stephane.leteurtre@chru-lille.fr)  
Tel.: +33-3-20445982  
Fax: +33-3-20446133

R. Parslow  
Paediatric Epidemiology Group, Centre for Epidemiology and Biostatistics, 8.49 Worsley Building, University of Leeds, Leeds LS2 9JT, UK

**Abstract Objectives:** To test the performance of PIM2 in French-speaking (FS) paediatric intensive care units (PICUs) and its relative performance when recalibrated using data from FS and Great Britain (GB) PICUs of different size. **Methods:** Consecutive admissions to 15 FS ( $n = 5,602$ ) and 31 GB PICUs ( $n = 20,693$ ) from June 2006 to October 2007 were included. The recalibrated PIM2 were applied to PICUs of different size within the FS and GB PICUs and between the two groups. PICU size was defined using number of admissions/month. Discrimination and calibration were evaluated using the area under the ROC curve (AUC) and the goodness-of-fit test, respectively. Logistic regression, funnel plots and

standardized W scores were performed in the two groups and between different PICU sizes. **Results:** In FS PICUs, the original PIM2 had good discrimination (AUC = 0.85) and moderate calibration ( $p = 0.07$ ). The recalibrated PIM2 scores had good calibration in FS ( $p = 0.33$ ) and moderate calibration in GB ( $p = 0.06$ ). Calibration was poor when the recalibrated FS PIM2 was applied to GB ( $p = 0.02$ ) but good when the GB recalibration was applied to the FS ( $p = 0.36$ ). Using the original PIM2 coefficients, calibration was poor in large units in both groups but improved following recalibration. There were no effects of PICU size on risk-adjusted mortality in GB and a significant effect in the FS PICUs with a minimum risk-adjusted mortality at about 35 admissions/month. **Conclusion:** The PIM2 score was valid in the FS population. The recalibration based on GB data could be applied to FS PICUs. Such recalibration may facilitate comparisons between countries.

**Keywords** Paediatric intensive care · Mortality · Severity of illness index · Critical care · Child

## Introduction

Prognostic models for groups of children admitted to paediatric intensive care units (PICUs) are now standard components of the methodology used in intensive care quality assurance and research [1–4]. Indeed, they are integrated in databases of paediatric networks [5–7]. Two predictive scoring systems have been proposed for critically ill children: the pediatric risk of mortality (PRISM and PRISM III) and the paediatric index of mortality (PIM and PIM2) [8, 9]. PIM2 was initially validated in 20,787 critically ill children from 14 large PICUs in Australia, New Zealand and the UK and was validated after recalibration in the UK demonstrating satisfactory performance [10]. This score has never, however, been studied in a prospective multicentre study in France.

Prognostic scoring systems that are developed in one country require recalibration (i.e. adapted coefficients) before being used to provide risk-adjusted outcomes of PICU mortality for units within a new health care setting [10]. Local recalibration has been reported to improve the performance of adult scoring systems in individual countries or regions [11–13]. Nevertheless, some authors have suggested that recalibration can compromise the ideal of comparisons between different geographic zones [13, 14]. Thus, the impact of the recalibration between two European countries on the performance of the score, in order to enable some comparisons, has never been evaluated.

Structural and process factors might explain variations in outcomes [15, 16]. The effect of volume of patients on quality of paediatric intensive care has been a subject of debate in North America [17] as three out of four studies found higher volume units had lower mortality [15, 18–20]. A better understanding of this relationship is needed to develop effective regionalization and referral policies for critically ill children [15, 18]. The performance of a unit for its case mix can be compared with the expected performance of a reference group of providers for a similar case mix [21]. Before any comparison of performance between PICUs of different size is conducted, it is important that the risk-adjusted mortality is accurately predicted.

The first objective of this study was to test the performance of the PIM2 score in French-speaking (FS) PICUs and the relative performance of the PIM2 score when recalibrated using data from FS countries and Great Britain (GB). The second objective was to compare the performance of large, medium and small sized units in FS and GB PICUs.

## Methods

All the 33 neonatal and PICUs (NPICUs)/PICUs and two PICUs affiliated to the Groupe Francophone de Réanimation et Urgences Pédiatriques (GFRUP) were invited to

participate on a voluntary basis: 17 agreed, and 15 (13 NPICUs/PICUs and two PICUs) provided the information requested (14 in France and one in Belgium). All consecutive children admitted to these PICUs between 21 June 2006 and 31 October 2007 were included. Children with a history of prematurity and hospitalized after birth were included. Those over 18 years and newborns who were premature (<37 weeks gestation) and admitted at birth were excluded.

Although the period of data collection varied between units, for each unit all patients admitted consecutively during the study period were included.

During the same period, 31 PICUs in GB from the PICANet database were analysed in order to perform comparative analysis. The selection process and data collection methods have been previously published [10]. Information were prospectively collected for each admission (demographic details, PIM2 score, outcome at PICU discharge (death, alive)). Children were excluded from the analysis if they were aged over 18 years, or less than 2 years and born prematurely (<37 weeks gestation).

### Legal considerations

Data collected from routine practice were used in this FS observational study. The two databases were declared to and approved by the French and English authorities (see electronic supplementary material, ESM).

### Data collection and management

Clinical data were prospectively recorded on a standardised case-report form. Previously trained physicians (one per centre) entered data into a Web-based database respecting confidentiality requirements (Epiconcept<sup>TM</sup>, Paris, France) (see ESM).

### Statistical analysis

#### *Design of analyses*

Outcome was vital status at PICU discharge. Validation of the PIM2 score was evaluated with the original published coefficients (FS-PIM2 and GB-PIM2 scores respectively). In each data set, a random sample of half of the population was selected to provide a development data set. Re-estimation of coefficients (recalibration) was conducted in this development set: individual component variables of the PIM2 score (independent variables) were entered into a logistic regression using the PICU discharge as dependent variable. These two steps were repeated 5,000 times (bootstrap procedure). Bootstrapping is a method employed to resample from the original data to create

replicate data sets, from which the variability of the quantities of interest can be assessed. The aim is to create alternative versions of data that “we might have seen” [22]. From the 5,000 coefficients obtained, the best model was selected using the mean coefficients of each independent variable. Final coefficients of the recalibrated model were applied on the data set of each country, and two models named ‘GB-Rec-PIM2’ score and ‘FS-Rec-PIM2’ score were provided (Fig. 1, ESM). A cross-recalibration was performed as follows: coefficients of the recalibrated FS model were applied to the GB data set (model named ‘GB-Rec<sub>FS</sub> coeff-PIM2’ score) and coefficients of the recalibrated GB model were applied to the FS data set (model named ‘FS-Rec<sub>GB</sub> coeff-PIM2’ score) (Fig. 1, ESM).

This analysis was stratified by PICU level of activity. To define small, medium or large units, an equal number of PICUs in each group taking into account the number of admissions per month (admissions per year) was determined from the FS data set; the defined cut-offs were less than 20 (240), 20–35 (240–420) and greater than 35 (420), close to the two cut-offs mentioned in the study by Wolfler et al. [14].

#### Calibration, discrimination and performance

**Calibration:** In order to compare observed with expected mortality and to estimate the calibration of the PIM2 score, a Hosmer–Lemeshow goodness-of-fit test was performed using the logit of the PIM2 score following logistic regression [23] (see ESM).

**Discrimination:** Area under the receiver operating characteristic curve (AUC) and standard error were calculated to estimate the discrimination of the scores (see ESM).

**Funnel plots:** Funnels plots of PICU mortality ratios by number of admissions along with corresponding 99.8 % control limits were produced. Funnel plots are a variation on control charts that identify PICUs with unexpectedly high (above upper control limit) or low (below the lower control limit) mortality (see ESM).

**Standardized W score:** The standardized W score (Ws) produced the number of excess or lack of survivors per 100 patients compared with the prediction [24, 25]. The standard error (SE) of Ws is also calculated to provide the 95 % confidence interval (CI) [24] (see ESM).

**Logistic regression:** A logistic regression was used to study size effect, with size as categorical or continuous variable (see ESM). Statistical analyses were performed using STATAv10 statistical software (StataCorp LP 2004, Texas, USA).

## Results

Among the 15 FS PICUs, no PICUs were exclusively devoted to cardiac surgery. Fourteen PICUs were devoted to medical, trauma and post-operative care (including cardiac surgery). One PICU admitted only medical patients. Among 5,651 patients, we excluded two patients older than 18 and 47 patients with incomplete data. Thus, 5,602 patients (414 died, mortality rate 7.39 %) were included. Primary category of illness on admission was congenital disease (33 %), infection (24 %), trauma (8 %), chemical injury (2 %), drug (0.1 %), cancer (3 %), diabetes (1 %), allergic immunologic disease (2 %) and other/undetermined (21 %). For GB PICUS, data are available on the PICANet website [26]. In GB PICUs 20,693 patients (1,014 died, mortality rate 4.90 %) were included. Characteristics and comparisons between the two populations are given in Table 1.

In FS PICUs, the FS-PIM2 score had a good discriminatory power (AUC 0.85; 95 % CI 0.83–0.87) and a moderate calibration ( $p = 0.07$ ). The FS-Rec-PIM2 score had a good calibration ( $p = 0.33$ ) and the GB-Rec-PIM2 had a moderate calibration ( $p = 0.06$ ) (Table 2). The GB-Rec<sub>FS</sub> coeff-PIM2 score displayed lack of fit and therefore poor calibration ( $p = 0.02$ ), whereas the FS-Rec<sub>GB</sub> coeff-PIM2 score displayed good calibration ( $p = 0.36$ ) (Table 2).

In the GB data set, calibration plots showed that the GB-PIM2 tended to overestimate risk in low-risk

**Table 1** Characteristics of the population

	GB PICUs ( $n = 20,693$ )	FS PICUs ( $n = 5,602$ )	$p$ value
Data collection period, months; median (IQR)	17 (17–17)	15 (8–17)	<0.0001
Children per year per unit; median (IQR)	391 (292–633)	255 (205–458)	0.12
Age, months; median (IQR)	15.7 (2.7–80.8)	12.3 (1.5–66.6)	<0.0001
PICU mortality; $n$ (%)	1,014 (4.90)	414 (7.39)	<0.0001
Length of stay, days; median (IQR)	2 (1–5)	3 (1–6)	<0.0001
Mechanical ventilation (%)	65.5	48.7	<0.0001
Recovery post procedure (%)	36.4	21.5	<0.0001
PIM2 score (probability of death %); median (IQR)	2.1 (0.8–5.3)	1.6 (0.8–5.1)	0.0021
Elective admission (%)	37.4	22.8	<0.0001

PICU Paediatric Intensive Care Unit, IQR interquartile range, PIM2 paediatric index of mortality 2 score, GB Great Britain, FS French-speaking

**Table 2** Discrimination (AUC) and calibration (chi square, *p* value) from the FS and GB data sets

	<20 admissions/month	20–35 admissions/month	>35 admissions/month	Total
FS data set				
Number (mortality rate)	1,072 (7.93)	1,269 (8.67)	3,261 (6.72)	5,602 (7.39)
Recalibrated PIM2 score (FS-Rec-PIM2)				
AUC (95 % CI)	0.85 (0.80–0.89)	0.89 (0.86–0.92)	0.85 (0.82–0.87)	0.86 (0.84–0.88)
Chi square	13.58	6.73	9.52	11.30
<i>p</i> value	0.19	0.75	0.52	0.33
Recalibrated PIM2 with coefficients from GB data set (FS-Rec <sub>GB</sub> coeff-PIM2)				
AUC (95 % CI)	0.84 (0.79–0.88)	0.89 (0.86–0.92)	0.83 (0.81–0.86)	0.85 (0.83–0.87)
Chi square	8.83	9.63	10.98	10.96
<i>p</i> value	0.55	0.47	0.36	0.36
GB data set				
Number (mortality rate)	828 (2.42)	5,331 (4.58)	14,534 (5.16)	20,693 (4.90)
Recalibrated PIM2 score (GB-Rec-PIM2)				
AUC (95 % CI)	0.87 (0.78–0.95)	0.85 (0.83–0.88)	0.83 (0.81–0.84)	0.84 (0.82–0.85)
Chi square	6.70	12.38	9.99	17.84
<i>p</i> value	0.75	0.26	0.44	0.06
Recalibrated PIM2 with coefficients from FS data set (GB-Rec <sub>FS</sub> coeff-PIM2)				
AUC (95 % CI)	0.88 (0.80–0.95)	0.85 (0.83–0.88)	0.82 (0.80–0.83)	0.84 (0.82–0.85)
Chi square	2.90	12.42	9.97	20.66
<i>p</i> value	0.98	0.26	0.44	0.02

GB Great Britain, FS French-speaking

mortality patients by less than 1 % in GB and 4 % in FS using the original and recalibrated equations (Fig. 2a, b for GB and Fig. 2c, d for FS PICUs, in ESM).

#### PICU size

Using the original PIM2 coefficients, calibration was good in small and medium units ( $p = 0.25$  and  $p = 0.78$  in FS and  $p = 0.99$  and  $p = 0.25$  in GB respectively) and poor in large units ( $p = 0.03$  in FS and  $p = 0.001$  in GB). In both groups, recalibrated and cross-recalibrated PIM2 scores provided good discriminatory power and calibration in small, median and large units (Table 2).

Funnel plots of crude mortality ratios by number of admissions for each unit indicated that three FS and seven GB PICUs had unadjusted mortality ratios that were lower or higher than expected (Fig. 1). Funnel plots using the original PIM2 score indicated that no PICU in France and only two PICUs in GB had adjusted mortality ratios that were lower than expected (Fig. 1). Funnel plots using the recalibrated PIM2 score indicated that no PICU in FS and only one PICU in GB had an adjusted mortality ratio that was lower than expected (Fig. 1). In the GB data set, one PICU had an observed mortality ratio equal to zero (no patient died in this PICU during the study period) and so this PICU was out of the lower limit of agreement in the three analyses (Fig. 1).

Using the original PIM2 score, the standardized *Ws* scores demonstrated fewer survivors per 100 cases in small and large FS PICUs, whereas the standardized *Ws* scores demonstrated an excess of survivors per 100 cases in medium and large GB PICUs (Table 3). With the

recalibrated PIM2 scores, the standardized *Ws* scores did not demonstrate a lack or excess of survivors in small, medium and large PICUs in either group (Table 3).

Odds ratios of PICUs size (small PICUs as reference) were not significant in both groups after adjustment for original and recalibrated PIM2 scores respectively (Table 3).

In the English data neither linear nor quadratic models showed any significant effect. In contrast with the French data a linear model showed no significant effect, but a quadratic model did (overall  $\chi^2 = 8.6$ , 2 *df*,  $p = 0.014$ ) (Fig. 3a, b, ESM). The parameters indicated a minimum risk-adjusted mortality at about 35 admissions/month in the FS PICUs (Fig. 3a, ESM).

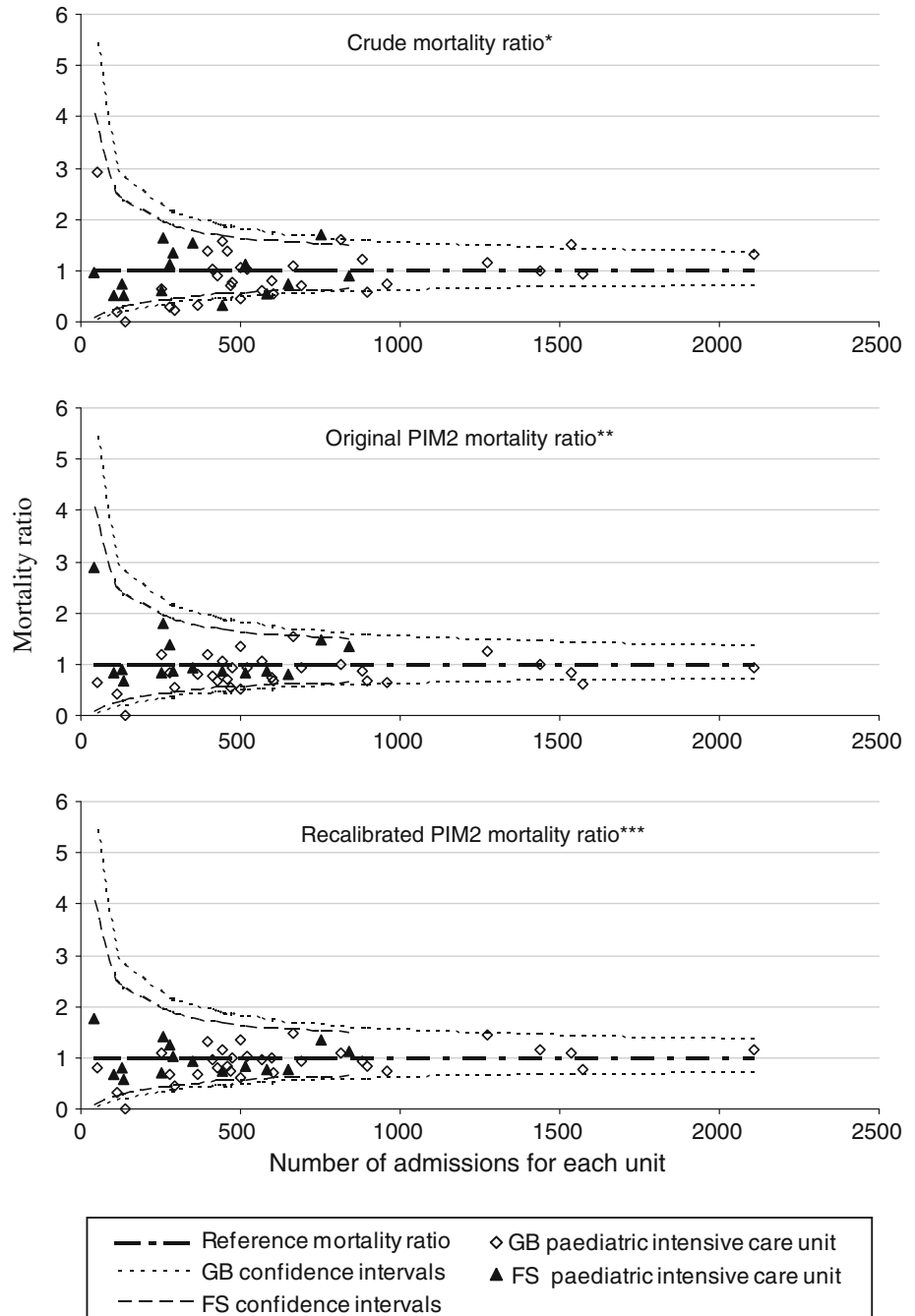
## Discussion

First, this study has shown that the PIM2 score was valid in the FS population and that the recalibration based on GB data could be applied to FS PICUs. Second, for the recalibrated PIM2 score, the volume of patients showed no effect in GB but did in the FS PICUs, with a minimum risk-adjusted mortality at about 35 admissions per month.

The PIM2 score was chosen because it is free and is very simple to collect. Moreover, several studies have showed that the PIM2 score had a good performance in different countries with [3, 10, 27] or without [14, 28] recalibration. In our study, and considering only the FS data set, the performance of the PIM2 score was good with and without recalibration.

Because of diversity in case mix, structure, organization, staffing and management between different countries

**Fig. 1** Funnels plots of GB and FS PICUs using crude mortality ratio, original PIM2 score mortality ratio and recalibrated PIM2 score mortality ratio. The mortality ratio then is plotted on the y-axis against the number of admissions to the PICU on the x-axis. To satisfy the condition that if the distribution of the mortality ratios is random there exists an ~5 % chance of a unit falling outside the control limits, then the upper and lower control limits must represent not 95 % confidence intervals but 99.7 % (confidence intervals around a mortality ratio of 1 by number of admissions [41]. \*Crude mortality ratio: observed PICU mortality/whole population mortality. \*\*Original PIM2 mortality ratio: observed PICU mortality/original PIM2 expected mortality (respectively FS PIM2 score and GB PIM2 score). \*\*\*Recalibrated PIM2 mortality ratio: observed PICU mortality/recalibrated PIM2 expected mortality (respectively FS-Rec-PIM2 score and GB-Rec-PIM2 score). *PIM* paediatric index of mortality, *GB* Great Britain, *FS* French speaking



[8, 29], region-specific or country-specific equations have been proposed in order to compare ICUs on a similar level [30]. Usually, discrimination and calibration of prognostic scoring systems are good in the population used for the development and for the internal validation. However, in almost all cases, when these scoring systems are applied to a new population, calibration deteriorates although discrimination hardly changes [13]. Recalibration had a large impact on the performance of the models, improving in particular its calibration [31]. We have used

the same recalibrated PIM2 score to make the comparison between these two different countries. Conversely, “there is a ‘risk’ to risk adjustment” because different models will not always agree on the identity of outlier performing institutions and on ranking the institutions [31]. The PIM2 score was initially developed from a study in Australia, New Zealand and UK, without recalibration between these three countries [9]. Furthermore, cross comparisons between different countries with recalibrated severity scoring systems had never been evaluated. In our study,



**Table 3** Standardized Ws scores and odds ratios with original and recalibrated PIM2 scores in FS and GB PICUs

	PICU			
	Small	Medium	Large	All
Standardized Ws score, value (95 % CI) <sup>a</sup>				
Original PIM2 score				
FS-PIM2	-1.11 (-2.20; -0.03)	+0.06 (-0.96; +1.08)	-1.21 (-1.79; -0.63)	-0.91 (-1.37; -0.46)
GB-PIM2	+0.54 (-0.37; +1.44)	+0.47 (+0.004; +0.93)	+0.50 (+0.20; +0.80)	+0.49 (+0.25; +0.74)
Recalibrated PIM2 score				
FS-Rec-PIM2	-0.14 (-1.39; +1.10)	+0.74 (-0.40; +1.89)	-0.34 (-1.00; +0.33)	-0.06 (-0.58; +0.46)
GB-Rec-PIM2	+0.76 (-0.25; +1.77)	+0.13 (-0.34; +0.61)	-0.15 (-0.45; +0.15)	-0.04 (-0.29; +0.21)
Odds ratio (95 % CI) <sup>b</sup>				
Original PIM2 score				
FS-PIM2	1	0.83 (0.59; 1.18)	1.04 (0.78; 1.40)	0.76 (0.70; 0.83)
GB-PIM2	1	1.33 (0.79; 2.24)	1.36 (0.82; 2.26)	0.82 (0.78; 0.86)
Recalibrated PIM2 score				
FS-Rec-PIM2	1	0.82 (0.58; 1.17)	1.05 (0.78; 1.42)	1.00 (0.91; 1.07)
GB-Rec-PIM2	1	1.46 (0.87; 2.45)	1.59 (0.96; 2.62)	1.00 (0.94; 1.05)

<sup>a</sup> If the interval contains zero, then Ws is not significantly different from zero, and hence the performance is not significantly different from that in the prediction database. If Ws is positive, then the actual number of survivors is greater than expected based on the reference database

<sup>b</sup> Odds ratios of scores on mortality according to the PICUs size (small PICUs as reference) and odds ratios of scores on mortality for all the PICUs without reference

the recalibrated PIM2 score in the GB data set could be applied to the FS data set, but not vice versa. Hypotheses to explain this result might be the larger population of the GB data set and the inclusion of all GB PICUs which cover a wider range of individuals or risks. Recalibration based on the GB data set and applied to FS and GB PICUs data sets have shown that case mixes from two countries can be considered as case mixes from two regions (states) in the same country. Indeed, differences in case mix between PICUs in the same country have been previously reported. Outcome description of 20 PICUs in the USA showed a significant variation among centres in mortality rate (from 1.4 to 21.3 %) or mean PRISM III score (from 2.42 to 11.18) [32]. Moreover, the annual report of the Australian and New Zealand Paediatric Intensive Care Registry in 2009 described mortality rates from 1 to 9 % and mechanical ventilation percentages from 7.5 to 100 % (three centres with no ventilated patients excluded) [33].

Definitions of the size of units vary between paediatric studies: small and large units were determined by the population median of six or fewer beds and more than six beds [17, 19], institutions' volume as medium (mean 45.6 patients/month), low (mean 26.8 patients/month) and high (mean 72.6 patients/month) [20]. In another study, cut-offs were not defined but there was a slight increase in mortality rates among PICUs with very high annual admission volumes (about 1,500 admissions per year) [18]. In the study by Wolfler et al., all PICUs were paediatric (mean number of children per year per unit 181) and most PICUs were small (4–6 beds) and had no more than 200–400 admissions per year [14]. Using a comparable definition of size, we did not observe any

differences between centres on discrimination and calibration. In our study, the funnel plots for the recalibrated PIM2 indicated that the adjusted mortality rate for all units in France and the UK was consistent. These results were observed with the PIM, PIM2, PRISM and PRISM III scores in 26 PICUs in the UK [10]. The relation between SMR and FS PICU volume suggested a reversed J-shaped relationship. Such a relationship was previously observed in data from North America PICUs, with a lowest theoretical threshold of 1,250 annual admissions, more than observed in our study (420 admissions per year) [18]. A recent study in Australia and New Zealand PICUs used a modified plot of risk-adjusted mortality ratio versus unit mean length of stay: two units were designated as inefficient and one unit was considered to be effective at the expense of high resource use [34].

#### Strengths and limitations

The two databases were of different size. It is possible that the different sample sizes between the two countries explain this result. The GB sample was larger than the FS sample and the probability that the GB sample included patients similar to those of the FS sample is more important than the opposite scenario. Thus, the result of the recalibration based on GB data applied to FS PICUs could be better. Nevertheless, the FS data set included more than 5,000 patients and, thus, the power of the statistical analysis seems acceptable. The case mix was different between these two countries. Infant mortality rate is lower in France than in the UK (3.6/1,000 versus 4.7/1,000) [35], but in the present study, crude mortality

was higher in the FS database. Such differences have previously been observed: crude mortality in PICUs was lower in Australia (4.1%) than in the UK (8.2 %) [9], whereas the infant mortality rate (4.4/1,000 versus 4.7/1,000 respectively) is about the same in these two countries [35]. In the study by Brady et al. optimized models were assessed by random allocation of PICUs (stratified by annual admission number) into development and validation samples in a 2:1 ratio [10]. In our study, a randomization by patient and a bootstrap procedure were applied on half of each data set to decrease the potential inflation of the new predictive equation applied to the entire data set. The bootstrap methods can be used to assess the strength of evidence that an identified variable is an important predictor [36]. For the Hosmer–Lemeshow goodness of fit test, we used degrees of freedom equal to the number of groups [37]. The use of degrees of freedom equal to number of groups minus 2 gave the same results (data not shown) except for the original FS-PIM 2 and GB-Rec-PIM2 scores which were not calibrated ( $p = 0.03$  and  $p = 0.02$  respectively). The relationship between volume and outcome is influenced by few variables, such as length of stay, nursing workload and number of premature infants cared for in NPICU, which were not taken into account in our analysis [34, 38]. Finally, in the FS data set only the voluntary PICUs were included, whereas Glance [39] has proposed that all ICUs participate in a national ICU-outcome database as was the case for the GB data set. Universal participation would eliminate the possibility of selection bias, in contrast to voluntary participation that may result in a non-representative group of ICUs [39].

## Conclusion

This study has demonstrated that PIM2 was valid in the FS PICUs. Also, the same recalibration could be applied to two different countries. Calibration of scoring systems should be reassessed periodically to ensure their continued validity. Nowadays, severity scoring systems have to be applied to monitor outcome and, thus, improve the quality of paediatric intensive care networks [40]. If a European paediatric intensive care registry is established, the results of this study support the possibility that a single, appropriately calibrated, risk adjustment model could be used for quality improvement and research within Europe and for comparative analyses between European countries.

**Acknowledgments** The authors thank Phil McShane (Paediatric Epidemiology Group, Centre for Epidemiology and Biostatistics, University of Leeds, Leeds, UK) for his help in the statistical analysis. French database: Grant from the French ministry of Health. UK database: PICANet is funded by the National Clinical

Audit and Patient Outcomes Programme via Healthcare Quality Improvement Partnership (HQIP), Health Commission Wales Specialised Services, NHS Lothian/National Service Division NHS Scotland, the Royal Belfast Hospital for Sick Children, Our Lady's Children's Hospital, Crumlin, Children's University Hospital, Temple Street and The Harley Street Clinic, London.

**Conflicts of interest** None.

## Appendix

### *Members of the GFRUP who participated*

D. Biarent (Bruxelles, Belgium), T. Blanc (Rouen, France), S. Cantagrel (Tours, France), R. Cremer (Lille, France), S. Dauger (Paris, France), P. Desprez (Strasbourg, France), M. Dobrzynski (Brest, France), G. Emériaud (Grenoble, France), P. Hubert (Paris, France), C. Milesi (Montpellier, France), S. Renolleau (Paris, France), M. Roque-Gineste (Toulouse, France), O. Noizet-Yverneau (Reims, France), D. Stamm (Lyon, France), I. Wroblewski (Besançon, France).

### *Participating French-speaking hospitals*

Data were collected in the French NPICU/PICUs of the following centres: American Memorial Hospital, CHU de Reims, France; Hôpital Charles Nicolle, CHU de Rouen, France; Hôpital Arnaud de Villeneuve, CHU de Montpellier, France; Hôpital Clocheville, CHU de Tours, France; Hôpital des enfants, CHU de Toulouse, France; Hôpital Hautepierre, CHU de Strasbourg, France; Hôpital Jeanne de Flandre, CHU de Lille, France; Hôpital Femme Mère-enfant, Hospices Civils de Lyon, France; Hôpital Morvan, CHU de Brest, France; Hôpital Necker, Assistance Publique-Hôpitaux de Paris, France; Hôpital Nord, CHU de Grenoble, France; Hôpital Reine Fabiola, Bruxelles, Belgium; Hôpital Robert Debré, Assistance Publique-Hôpitaux de Paris, France; Hôpital Saint Pierre, CHU de Besançon, France; Hôpital Trousseau, Assistance Publique-Hôpitaux de Paris, France.

### *Participating NHS trusts and hospitals*

Addenbrooke's NHS Trust: Addenbrooke's Hospital  
 Bart's and the London NHS Trust: Bart's and The Royal London  
 Birmingham Children's Hospital NHS Trust: Diana, Princess of Wales Children's Hospital  
 Brighton & Sussex University Hospitals NHS Trust: Royal Alexandra Hospital for Sick Children  
 Cardiff and Vale NHS Trust: University Hospital of Wales  
 Central Manchester and Manchester Children's University Hospitals NHS Trust: Royal Manchester Children's Hospital

Children's University Hospital, Temple Street, Dublin  
 Great Ormond Street Hospital for Children NHS Trust: Great Ormond Street Hospital for Sick Children  
 Guy's & St. Thomas' Hospital NHS Trust: Evelina Children's Hospital  
 Hull & East Yorkshire Hospitals NHS Trust: Hull Royal Infirmary  
 King's College Hospital NHS Trust: King's College Hospital  
 Leeds Teaching Hospitals NHS Trust: Leeds General Infirmary, St. James University Hospital  
 Our Lady's Children's Hospital, Crumlin, Dublin  
 Newcastle upon Tyne Hospitals NHS Trust: Great Northern Children's Hospital (formerly Newcastle General Hospital and the Royal Victoria Infirmary). Freeman Hospital  
 Oxford Radcliffe Hospitals NHS Trust: John Radcliffe Hospital  
 Queen's Medical Centre Nottingham University NHS Trust: Queen's Medical Centre  
 Royal Brompton and Harefield NHS Trust: Royal Brompton Hospital  
 Royal Liverpool Children's NHS Trust: Alder Hey Hospital  
 Royal Group of Hospitals and Dental Hospital HSS Trust, Belfast Hospital for Sick Children  
 Sheffield Children's NHS Trust: Sheffield Children's Hospital  
 Southampton University Hospitals NHS: Trust Southampton General Hospital  
 South Tees Hospitals NHS Trust: James Cook University Hospital  
 St. George's Healthcare NHS Trust: St. George's Hospital  
 St. Mary's NHS Trust: St. Mary's Hospital  
 The Harley Street Clinic, London  
 The Lewisham Hospitals NHS Trust: University Hospital, Lewisham (to July 2010)  
 University Hospitals Bristol NHS Foundation Trust: Bristol Royal Hospital for Children  
 University Hospitals of Leicester NHS Trust: Leicester Glenfield Hospital, Leicester Royal Infirmary  
 University Hospital of North Staffordshire NHS Trust: Stoke on Trent City General

## References

- Lacroix J, Cotting J (2005) Severity of illness and organ dysfunction scoring in children. *Pediatr Crit Care Med* 6:S126–S134
- Martinot A, Leteurtre S, Grandbastien B, Duhamel A, Leclerc F (1997) Characteristics of patients and use of resource in French pediatric intensive care units. *Arch Pediatr* 4:730–736
- Slater A, Shann F (2004) The suitability of the pediatric index of mortality (PIM), PIM2, the pediatric risk of mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. *Pediatr Crit Care Med* 5:447–454
- Wunsch H, Rowan KM, Angus DC (2007) International comparisons in critical care: a necessity and challenge. *Curr Opin Crit Care* 13:725–731
- LaRovere JM, Jeffries HE, Sachdeva RC, Rice TB, Wetzel RC, Cooper DS, Bird GL, Ghanayem NS, Checchia PA, Chang AC, Wessel DL (2008) Databases for assessing the outcomes of the treatment of patients with congenital and paediatric cardiac disease: the perspective of critical care. *Cardiol Young* 18(Suppl 2):130–136
- Parslow RC, Tasker RC, Draper ES, Parry GJ, Jones S, Chater T, Thiru K, McKinney PA (2009) Epidemiology of critically ill children in England and Wales: incidence, mortality, deprivation and ethnicity. *Arch Dis Child* 94:210–215
- Slater A, Shann F, McEniery J (2003) The ANZPIC registry diagnostic codes: a system for coding reasons for admitting children to intensive care. *Intensive Care Med* 29:271–277
- Pollack MM, Patel KM, Ruttimann UE (1996) PRISM III: an updated pediatric risk of mortality score. *Crit Care Med* 24:743–752
- Slater A, Shann F, Pearson G (2003) PIM2: a revised version of the paediatric index of mortality. *Intensive Care Med* 29:278–285
- Brady AR, Harrison D, Black S, Jones S, Rowan K, Pearson G, Ratcliffe J, Parry GJ (2006) Assessment and optimization of mortality prediction tools for admissions to pediatric intensive care in the United Kingdom. *Pediatrics* 117:e733–e742
- Chase JG, Shaw GM (2012) How standard is the “S” in SMR? *Intensive Care Med* 38:1–3
- Minne L, Eslami S, de Keizer N, de Jonge E, de Rooij SE, Abu-Hanna A (2012) Effect of changes over time in the performance of a customized SAPS-II model on the quality of care assessment. *Intensive Care Med* 38:40–46
- Vincent JL, Moreno R (2010) Clinical review: scoring systems in the critically ill. *Crit Care* 14:207
- Wolfler A, Silvani P, Musicco M, Salvo I (2007) Pediatric index of mortality 2 score in Italy: a multicenter, prospective, observational study. *Intensive Care Med* 33:1407–1413
- Tilford JM, Simpson PM, Green JW, Lensing S, Fiser DH (2000) Volume-outcome relationships in pediatric intensive care units. *Pediatrics* 106:289–294
- Cogo PE, Poole D, Codazzi D, Boniotti C, Capretta A, Langer M, Luciani D, Rossi C, Bertolini G (2010) Outcome of children admitted to adult intensive care units in Italy between 2003 and 2007. *Intensive Care Med* 36:1403–1409
- Pollack MM (2007) Pediatric intensive care quality factors. *J Trauma* 63:S143–S145
- Marcin JP, Song J, Leigh JP (2005) The impact of pediatric intensive care unit volume on mortality: a hierarchical instrumental variable analysis. *Pediatr Crit Care Med* 6:136–141



19. Pollack MM, Cuerdon TT, Patel KM, Ruttimann UE, Getson PR, Levettown M (1994) Impact of quality-of-care factors on pediatric intensive care unit mortality. *JAMA* 272:941–946
20. Ruttimann UE, Patel KM, Pollack MM (2000) Relevance of diagnostic diversity and patient volumes for quality and length of stay in pediatric intensive care units. *Pediatr Crit Care Med* 1:133–139
21. Shahian DM, Normand SL (2008) Comparison of “risk-adjusted” hospital outcomes. *Circulation* 117:1955–1963
22. Cevenini G, Barbini P (2010) A bootstrap approach for assessing the uncertainty of outcome probabilities when using a scoring system. *BMC Med Inform Decis Mak* 10:45
23. Lemeshow S, Hosmer DW Jr (1982) A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 115:92–106
24. Hollis S, Yates DW, Woodford M, Foster P (1995) Standardized comparison of performance indicators in trauma: a new approach to case-mix variation. *J Trauma* 38:763–766
25. Glance LG, Osler T, Shinozaki T (2000) Effect of varying the case mix on the standardized mortality ratio and W statistic: a simulation study. *Chest* 117:1112–1117
26. Ashley A, Draper E, Fleming T, Lamming C, McKinney P, McShane P, Parslow R, Skinner S, Thiru K (2009) National Report of the Paediatric Intensive Care Audit Network 2006–2008. Leeds. [http://www.picanet.org.uk/Documents/General/Annual%20report\\_2009/PICANet%20National%20Report%202006%20-%202008\\_new.pdf](http://www.picanet.org.uk/Documents/General/Annual%20report_2009/PICANet%20National%20Report%202006%20-%202008_new.pdf). Accessed 12 Nov 2011
27. Eulmesekian PG, Perez A, Mincec PG, Ferrero H (2007) Validation of pediatric index of mortality 2 (PIM2) in a single pediatric intensive care unit of Argentina. *Pediatr Crit Care Med* 8:54–57
28. Imamura T, Nakagawa S, Goldman RD, Fujiwara T (2012) Validation of pediatric index of mortality 2 (PIM2) in a single pediatric intensive care unit in Japan. *Intensive Care Med* 38:649–654
29. Leteurtre S, Martinot A, Duhamel A, Proulx F, Grandbastien B, Cotting J, Gottesman R, Joffe A, Pfenninger J, Hubert P, Lacroix J, Leclerc F (2003) Validation of the paediatric logistic organ dysfunction (PELOD) score: prospective, observational, multicentre study. *Lancet* 362:192–197
30. Metnitz B, Schaden E, Moreno R, Le Gall JR, Bauer P, Metnitz PG (2009) Austrian validation and customization of the SAPS 3 admission score. *Intensive Care Med* 35:616–622
31. Bakhshi-Raiez F, Peek N, Bosman RJ, de Jonge E, de Keizer NF (2007) The impact of different prognostic models and their customization on institutional comparison of intensive care units. *Crit Care Med* 35:2553–2560
32. Gregory CJ, Nasrollahzadeh F, Dharmar M, Parsapour K, Marciniak JP (2008) Comparison of critically ill and injured children transferred from referring hospitals versus in-house admissions. *Pediatrics* 121:e906–e911
33. Alexander J, Tregear S, Slater A (2010) Report of the Australian and New Zealand Paediatric Intensive Care Registry 2009. Available at <http://www.anzics.com.au/core/anzpicr?start=2>. Accessed 10 Sept 2011
34. Straney LD, Clements A, Alexander J, Slater A (2010) Measuring efficiency in Australian and New Zealand paediatric intensive care units. *Intensive Care Med* 36:1410–1416
35. Haub C (2010) 2010 world population data sheet. Population Reference Bureau, Washington, DC. [http://www.prb.org/pdf10/10wpds\\_eng.pdf](http://www.prb.org/pdf10/10wpds_eng.pdf). Accessed 4 May 2011
36. Beyene J, Atenafu EG, Hamid JS, To T, Sung L (2009) Determining relative importance of variables in developing and validating predictive models. *BMC Med Res Methodol* 9:64
37. Murphy-Filkins R, Teres D, Lemeshow S, Hosmer DW (1996) Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. *Crit Care Med* 24:1968–1973
38. Metnitz B, Metnitz PG, Bauer P, Valentin A (2009) Patient volume affects outcome in critically ill patients. *Wien Klin Wochenschr* 121:34–40
39. Glance LG, Szalados JE (2002) Benchmarking in critical care: the road ahead. *Chest* 121:326–328
40. Macrae D (2009) Conducting clinical trials in pediatrics. *Crit Care Med* 37:S136–S139
41. Spiegelhalter D (2002) Funnel plots for institutional comparison. *Qual Saf Health Care* 11:390–391