Intensive Care Med (2013) 39:942–950
DOI 10.1007/s00134-013-2857-4

PEDIATRIC ORIGINAL

Idse H. E. Visser
Jan A. Hazelzet
Marcel J. I. J. Albers
Carin W. M. Verlaat
Karin Hogenbirk
Job B. van Woensel
Marc van Heerde
Dick A. van Waardenburg
Nicolaas J. G. Jansen
Ewout W. Steyerberg

Mortality prediction models for pediatric intensive care: comparison of overall and subgroup specific performance

Received: 16 April 2012
Accepted: 11 November 2012
Published online: 22 February 2013
© Springer-Verlag Berlin Heidelberg and ESICM 2013

Electronic supplementary material
The online version of this article (doi:10.1007/s00134-013-2857-4) contains supplementary material, which is available to authorized users.

I. H. E. Visser (✉)
Department of Pediatrics, Erasmus MC, Sophia Children's Hospital, Room Sp-1537, P.O. Box 2060, 3000 CB Rotterdam, The Netherlands
e-mail: i.visser@erasmusmc.nl
Tel.: +31-63-6075775
Fax: +31-10-7036796

J. A. Hazelzet
Division of Pediatric Intensive Care, Department of Pediatrics, Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands

M. J. I. J. Albers
Pediatric Intensive Care Unit, Beatrix Children's Hospital, University Medical Center Groningen, Groningen, The Netherlands

C. W. M. Verlaat
Department of Pediatric Intensive Care, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands

K. Hogenbirk
Pediatric Intensive Care Unit, Leiden University Medical Center, Leiden, The Netherlands

J. B. van Woensel
Pediatric Intensive Care Unit, Academic Medical Center, Emma Children's Hospital, Amsterdam, The Netherlands

M. van Heerde
Department of Pediatric Intensive Care, VU University Medical Center, Amsterdam, The Netherlands

D. A. van Waardenburg
Pediatric Intensive Care Unit, Maastricht University Medical Center, Maastricht, The Netherlands

N. J. G. Jansen
Department of Pediatric Intensive Care, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands

E. W. Steyerberg
Department of Public Health, Center for Medical Decision Making, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

Abstract *Aim:* To validate paediatric index of mortality (PIM) and pediatric risk of mortality (PRISM) models within the overall population as well as in specific subgroups in pediatric intensive care units (PICUs). *Methods:* Variants of PIM and PRISM prediction models were compared with respect to calibration (agreement between predicted risks and observed mortality) and discrimination (area under the receiver operating characteristic curve, AUC). We considered performance in the overall study population and in subgroups, defined by diagnoses, age and urgency at admission, and length of stay (LoS) at the PICU. We analyzed data from consecutive patients younger than 16 years admitted to the eight PICUs in the Netherlands between February 2006 and October 2009. Patients referred to another ICU or deceased within 2 h after admission were excluded. *Results:* A total of 12,040 admissions were included, with 412 deaths. Variants of PIM2 were best calibrated. All models discriminated well, also in patients <28 days of age (neonates), with overall higher AUC for PRISM variants (PIM = 0.83, PIM2 = 0.85, PIM2-ANZ06 = 0.86, PIM2-ANZ08 = 0.85, PRISM = 0.88, PRISM3-24 = 0.90). Best discrimination for PRISM3-24 was confirmed in 13 out of 14 subgroup categories. After recalibration PRISM3-24 predicted accurately in most (12 out of 14) categories. Discrimination was poorer for all models (AUC < 0.73) after LoS of >6 days at the PICU. *Conclusion:* All models discriminated well, also in most subgroups including neonates, but had difficulties predicting mortality for patients >6 days at the PICU. In a western European setting both the PIM2(-ANZ06) or a recalibrated version of PRISM3-24 are suited for overall individualized risk prediction.

Keywords Benchmarking · Outcome/quality assessment · Critical illness/mortality · Intensive care units pediatric · Risk adjustment · Severity of illness index · Validation studies · Registries · Diagnosis-related groups

Introduction

In the 1980s the first general mortality prediction models for the pediatric intensive care were developed to standardize mortality between pediatric intensive care units (PICUs) and make outcome assessment between PICUs more objective [1, 2]. Nowadays mortality prediction models are well known in the PICU setting and used in many studies and registries all over the world [3]. Mortality prediction models are essential to adjust for case-mix and severity of illness, and to describe characteristics of diverse subgroup populations [4–7].

Although mortality rates in pediatric intensive care have roughly halved in more economically developed countries [1, 2, 8–11], mortality is still an important outcome for benchmarking PICUs [7, 12, 13].

Two important families of general mortality prediction models are available for PICUs: the paediatric index of mortality (PIM) and the pediatric risk of mortality (PRISM). Both models have meanwhile been updated to PIM2 and PRISM3, respectively [2, 9–11]. The PIM models use data collected at first contact with a pediatric intensivist up to the first hour. The PRISM is available as severity of illness score (noted here as PRISMscore and PRISM3score) and as a predicted risk of mortality based on the most extreme value recorded in the first 12 h (PRISM3-12) or in the first 24 h on the PICU (PRISM and PRISM3-24). PIM2 has recently been recalibrated in the Australian and New Zealand PICUs into the PIM2-ANZ06 and PIM2-ANZ08 for local benchmarking [13, 14].

Although PIM and PRISM models have been externally validated in different settings, they have not been compared simultaneously on specific subgroup populations, fully different in time and place from the development setting [15–17]. The aim of this study was to determine the predictive performance of (variants of) PIM and PRISM models in direct comparison to each other within the overall population as well as in specific subgroups.

Methods

Data-setting

We analyzed data from consecutive patients admitted to the eight Dutch PICUs, from February 2006 up to October 2007. After this first period, six centers carried on voluntarily with the study for different periods up to October 2009. Each center collected data for all consecutive admissions up to the last day of the month ending their inclusion period on all models; PIM, PIM2, PIM2-ANZ06, PIM2-ANZ08, PRISM, and PRISM3-24. Data collection for this study was performed as part of the pediatric intensive care evaluation (PICE), a national PICU registry where routinely measured values on admissions to the PICUs is collected for national benchmarking and the maintenance of a clinical database on patient characteristics and outcome [7]. Data quality was assessed following standardized procedures including local audit site visits with re-extraction on approximately 40 randomly chosen admissions at each site.

For direct comparison patients were selected if they fulfilled the study criteria from both PIM and PRISM models [2, 9–11]. Therefore patients were excluded who were 16 years of age and older, transferred to another ICU and those dying within 2 h in the PICU after admission or not achieving stable vital signs for at least 2 h within the first 24 h. Details on data collection process and quality are presented in Electronic Supplementary Material, (ESM) Tables 4 and 5. Mortality in the PICU was the end point for statistical analysis. Missing values were accepted as normal and hence not contributing any additional risk to the models. The mortality risk for the proprietary PRISM3-24 was provided by the Virtual PICU Systems (VPS LLC, Los Angeles, CA, United States of America) (https://portal.myvps.org).

External validation

The external validity of the PIM and PRISM models was compared directly on the same patients by calibration and discrimination in the total study population as well as in subpopulations [18, 19].

Calibration

The overall prediction, or calibration-in-the-large, was analyzed by the ratio between observed and the expected mortality and expressed as a standardized mortality ratio (SMR) with a 95 % confidence interval based on a Poisson distribution [20]. Overall prediction of observed mortality was considered accurate when the SMR was not significantly below or above 1. The calibration of the models was further assessed graphically. We plotted

observed outcomes by deciles of risk, and compared observed to predicted mortality with a nonparametric curve. When the plotted curve is a straight diagonal line (slope 1, intercept 0), the predicted mortality matches observed mortality. The larger the deviation from this ideal diagonal line, the less accurate the calibration [18, 21].

Discrimination

The discriminative ability of each model was analyzed by calculating the area under the receiver operating characteristic (ROC) curve (AUC) [22]. With an AUC of 0.5 the model does not discriminate any better than chance and with an AUC of 1.0 the model discriminates perfectly. Statistically significant differences in AUC were determined by taking 5,000 bootstraps in all subgroups comparing the AUC between two models at a time in each bootstrap [18].

Subgroup analysis

We analyzed subgroups of patients with respect to diagnoses, age and urgency at admission to the PICU, and according to length of stay (LoS) at the PICU. Subgroups with mutually exclusive categories were chosen for their relevance to clinical and validation studies and also considering their sample size [9, 11, 15, 17, 23, 24]. The LoS was categorized with consideration of the distribution at the median (3 days) and the mean (6 days) LoS.

To compare the mortality prediction between models on their SMR more directly within subgroups, the models were customized to the total study population by a logistic recalibration procedure. A logistic regression model was refitted with the linear predictor of a prediction model as the single covariable. This procedure was followed for each of the prediction models considered. This made that predicted mortality equaled the observed mortality in the total population (SMR = 1), and that the average predictive effects were calibrated to the PICE setting (regression slope = 1). The discriminative ability remained the same since the relative weights of risk factors in the models were not altered [18, 21].

Statistical analyses were conducted using SPSS PASW version 17.0.2 (11-03-2009) [SPSS inc. Chicago USA] and R version 2.12.1 (16-12-2010) [The R Foundation for Statistical Computing, Vienna, Austria]. Statistical significance was defined as $p < 0.05$.

Results

Data

In the total study period, 13,642 consecutive admissions were registered, with 484 deaths. The combined criteria for PIM and PRISM models led to the exclusion of 946 admissions (51 deaths): 564 patients were 16 years of age and older on admission, 352 were transferred to another ICU, 14 patients died within 2 h at the PICU, and 16 patients did not achieve stable vital signs for at least 2 h within the first 24 h. Besides these exclusion criteria one center (601 admissions, 19 deaths) had to be excluded because data quality was insufficient. A further 55 admissions (two deaths) from six different centers were excluded from analysis because of poor data quality. Eventually, 12,040 admissions with 412 deaths were analyzed, originating from seven Dutch centers and admitted from February 2006 up to October 2009 (Fig. 1 and ESM Tables 2, 4).

Overall performance

Calibration

The PIM2 and PIM2-ANZ06 had good overall prediction of mortality, with an SMR not significantly below or above 1 (Table 1). PIM, PRISM and PRISM3-24 predicted mortality significantly higher than observed (SMR < 1) and the most recent regionally updated model, PIM2-ANZ08, predicted mortality significantly lower (SMR > 1). Most patients had a low risk of mortality: 80 % had a predicted risk below 10 % in all models.

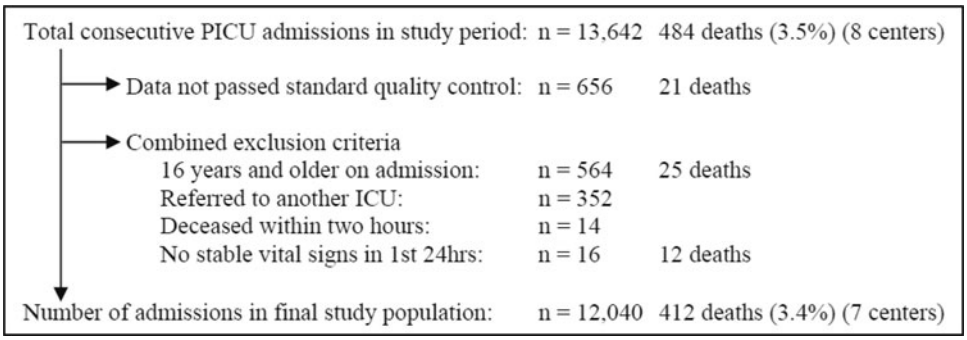

Fig. 1 Flowchart exclusion study population

Table 1 Overall performance original models by calibration (SMR) and discrimination (AUC)

Model	PIM1	PIM2	PIM2-ANZ06	PIM2-ANZ08	PRISM	PRISM3-24
SMR[a]	0.81*	0.95	1.05	1.12	0.52*	0.87*
(95 %CI)	(0.73;0.89)	(0.86;1.05)	(0.95;1.15)	(1.02;1.24)	(0.47;0.57)	(0.79;096)
AUC[b]	$0.83_{<1,2,3}$	$0.85_{<1,2,4}$	$0.86^{>3,4}_{<1}$	$0.85^{>3}_{<1,2}$	$0.88^{>2}_{<1}$	$0.90^{>1}$
(95 %CI)	(0.81;0.85)	(0.83;0.87)	(0.85;0.88)	(0.83;0.87)	(0.86;0.90)	(0.89;0.92)

Significant differences ($p < 0.05$) between models marked by $_{<}$ $^{>}$ signs; for example: $^{>2}_{<1}$ means the discrimination is significantly ($p < 0.05$) better than a model marked "$_{<2}$", and is significantly less than a model marked "$^{>1}$"

[a] Standardised mortality ratio (SMR) with 95 % Poisson confidence interval between *brackets* , values marked *asterisk* are significantly below or above SMR = 1

[b] AUC (c-statistic) with 95 % nonparametric confidence interval between brackets, differences in discrimination (AUC) are calculated by 5,000 bootstrap samples each between two models at a time

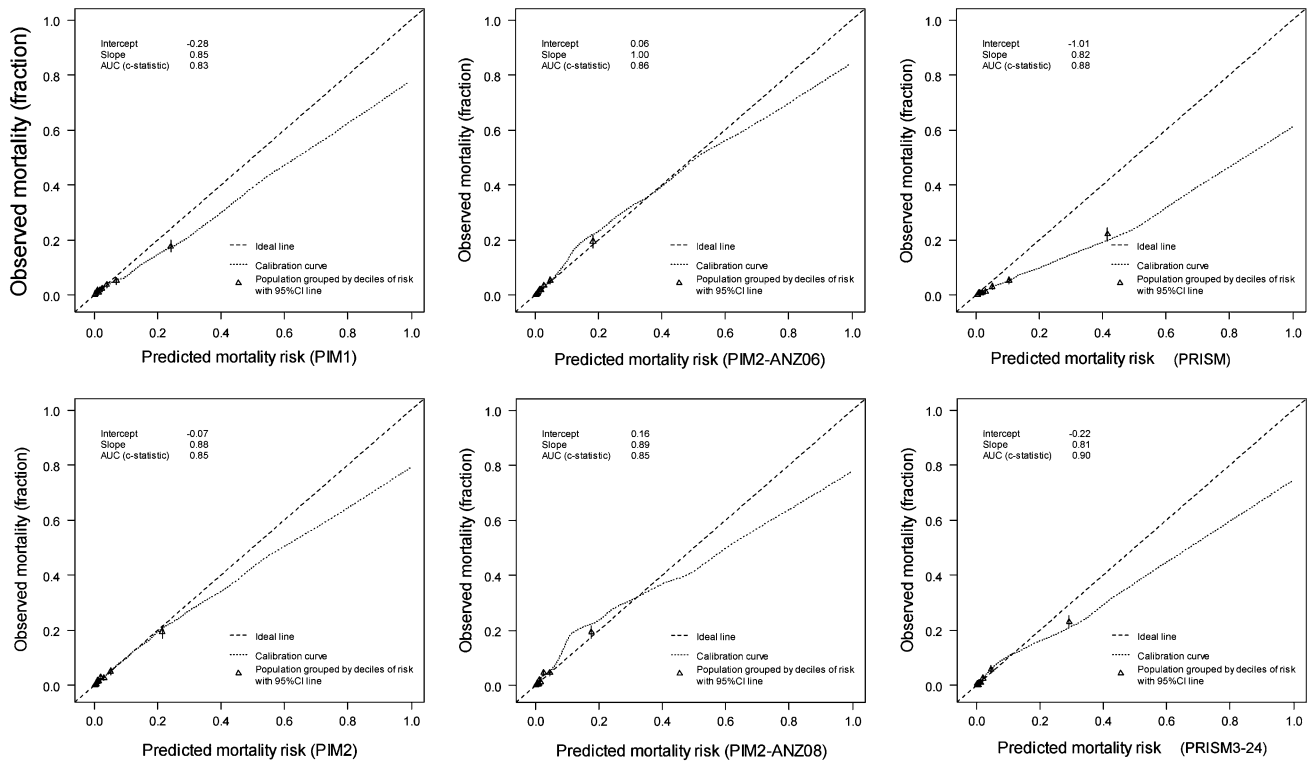

Fig. 2 Calibration plots (original) models

Mortality in the highest risk deciles was lower than predicted in all models. The PIM2-ANZ06 was best calibrated overall with a slope of 1.00 and intercept 0.06. The PIM2-ANZ08 did not calibrate as well as the PIM2-ANZ06, especially in patients with a predicted risk below 0.2 or a predicted risk between 0.4 and 0.6 (Fig. 2).

Discrimination

All mortality prediction models had good discrimination, with the PRISM3-24 (AUC = 0.90) performing significantly better than all other models (Table 1). Discrimination ranged between models from PIM (AUC = 0.83), PIM2 (AUC = 0.85), PIM2-ANZ06 (AUC = 0.86), PIM2-ANZ08 (AUC = 0.85), PRISM (AUC = 0.88) to PRISM3-24 (AUC = 0.90).

Subgroup performance

For assessment of subgroup performance we focused on the two most promising models, PIM2-ANZ06 and PRISM3-24, because they performed best on calibration and discrimination. After recalibration to the average mortality in the overall study population the PRISM3-24

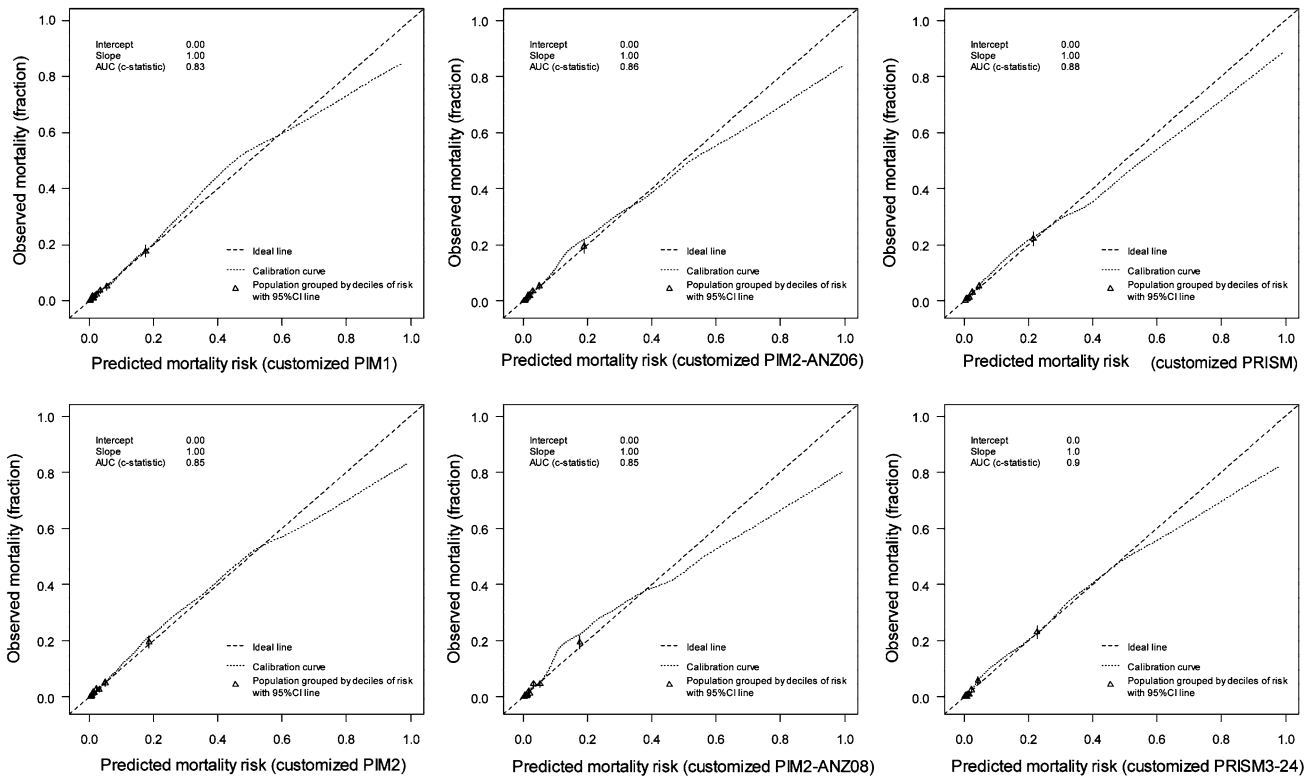

Fig. 3 Calibration plots recalibrated ('customized') models

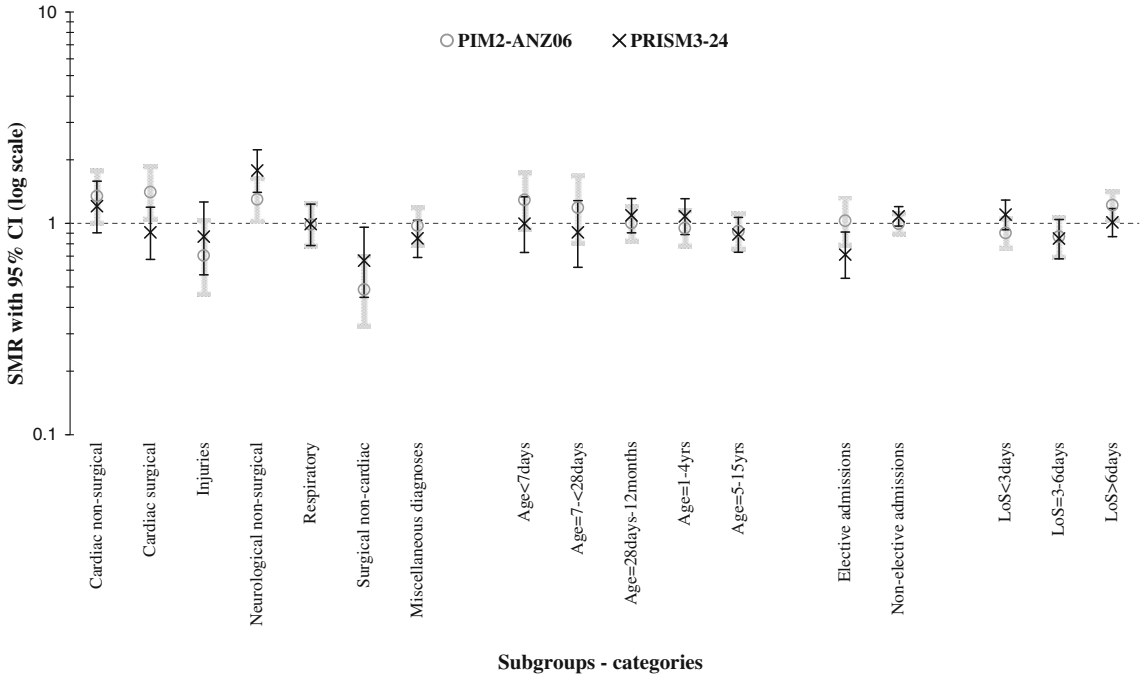

Fig. 4 Calibration (SMR) by recalibrated PIM2-ANZ06 and PRISM3-24 within subgroups

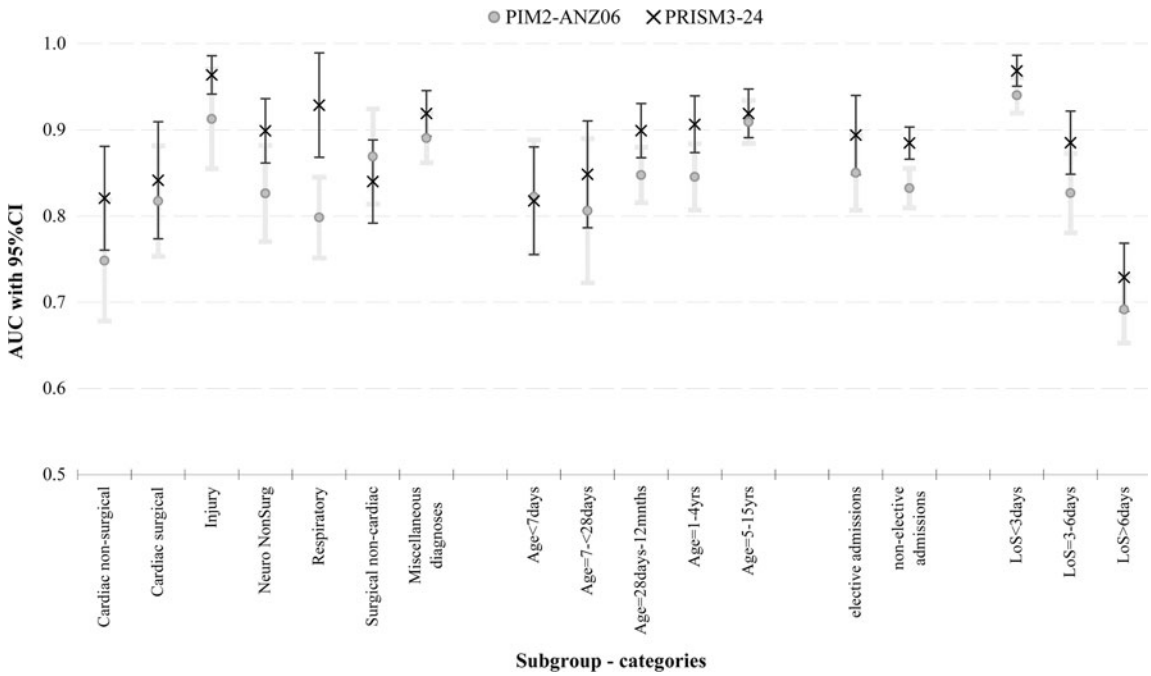

Fig. 5 Discrimination (AUC) by PIM2-ANZ06 and PRISM3-24 within subgroups

had a slightly better calibration plot than the PIM2-ANZ06 (Fig. 3). Further analysis within the subgroups revealed only better calibration for the PIM2-ANZ06 within the urgency subgroup (Fig. 4). The ranking between the models on discrimination was largely the same as in the overall population. The PRISM3-24 had higher AUC than the PIM2-ANZ06 in all categories, and this was statistically significant in half of them (Fig. 5). Details on calibration and discrimination for all models within the subgroups are shown in ESM Table 3.

Performance of recalibrated models on diagnoses, age and urgency

In cardiac patients, both surgical and nonsurgical, mortality was higher (SMR > 1) than predicted by PIM2-ANZ06. The actual mortality in neurological nonsurgical patients was higher (SMR > 1) than any model predicted, and in surgical noncardiac patients mortality was lower (SMR < 1) than predicted (Fig. 4). Both PIM2-ANZ06 and PRISM3-24 predicted mortality accurately in all age groups, but the PIM2-ANZ06, as all other PIM models, showed a tendency to predict lower mortality than observed in the younger population. In the elective admissions mortality was lower than predicted (SMR < 1) by the PRISM3-24 (Fig. 4 and ESM Table 3).

Predictive performance according to LoS in PICU

The longer the stay at the PICU, the poorer the models performed. All models were well calibrated for patients discharged within the first 6 days from the PICU, but with a longer stay good prediction became more difficult. Observed mortality was significantly higher than predicted (SMR > 1) by the PIM2-ANZ06 in patients longer than 6 days in PICU. The discriminative ability of all models was very good (AUC ≥ 0.92) for short LoS (<3 days), with PRISM models showing significantly higher AUC than the PIM models. Discrimination of all models declined with a longer stay in the PICU, where only the PRISM3-24 had an AUC above 0.7 (AUC = 0.73) for a stay of more than 6 days (Fig. 5 and ESM Table 3).

Discussion

The PIM and PRISM models performed well in a national PICU registry on their general and subgroup specific predictive capacity. The PIM2(-ANZ06) was best calibrated to the overall population. The PRISM3-24 had significantly better discrimination than all other models and, after recalibration to our study population, was best calibrated over all subgroups.

PIM and PRISM are general mortality prediction models that do not aim at one specific subgroup of pediatric intensive care. Good to excellent performance was found in different subgroups in the original validation settings [9, 11]. Some researchers have questioned the differences in performance of some of these models in specific subgroups [15, 17, 25–29]. As in our study, the main problem was found in the calibration and not in the discrimination by these models. Within the overall study population, general over- or under-prediction was easily overcome by logistic recalibration, but this did not solve miscalibration within all subgroups (Table 1 and Fig. 4). In only three of the seven categories of the diagnoses subgroup (injuries, respiratory and miscellaneous diagnoses) did all recalibrated models predict mortality correctly (ESM Table 3). The lower mortality in surgical noncardiac patients and higher mortality than predicted in neurological nonsurgical patients is in line with other studies [15, 17]. All models had poorer discrimination in younger patients. It was still quite reasonable (AUC $\geq$ 0.79) for patients <28 days of age (neonates) or even for neonates <7 days of age (ESM Table 3) [15, 30]. Only the PIM2-ANZ06 was well calibrated for both elective and nonelective admissions, which is a risk factor in the PIM models. The PRISM3-24 had better discrimination, but predicted mortality in elective admissions relatively too high. This suggests that severity of illness for elective admissions may need to be adjusted for mortality with an additional risk factor on urgency.

More extensive updating may be required, with re-estimation of the specific predictors within the models, to overcome the problem with subgroup-specific calibration. This was done for Australia and New-Zealand in the regionally updated PIM2-ANZ06 and PIM2-ANZ08 [13, 14, 18]. However, such model revision also affects the discrimination by the models, and hence the external validity [18, 31]. In our study, we found the most recent updated version of PIM2, the PIM2-ANZ08 (Table 1 and Fig. 3), was less well calibrated than the PIM2-ANZ06 in all subgroups and discriminated less in some categories (ESM Table 3). Nevertheless, both regional updated PIM2 versions showed external validity in our study population. A recent study into the external applicability of other re-estimated PIM2 versions concluded the local updated version could also be adapted outside the region of updating [32].

We included the original PIM and PRISM models into our analysis, as they are still used in different studies [5, 6]. It has been advised not to use the PRISM any longer, but we found that it performed overall still remarkably well after logistic recalibration with exception of (young) age groups and respiratory diagnoses (ESM Table 3) [33, 34]. The PRISM3-12 was not included as the data collection effort was too high, but a previous large validation study in the UK found both versions of PRISM3 to have quite similar overall performance [16]. Besides the PRISM mortality prediction models we analyzed the PRISM severity of illness scores and found they discriminated quite well on mortality also in subgroups, but their calibration was not as good as their risk prediction versions.

We divided outcome according to LoS at the PICU and showed discrimination decreased sharply for all models with a longer stay at the PICU. Only the PRISM3-24 was able to discriminate reasonably (AUC = 0.73) when stay in PICU lasted more than 6 days. Other factors than risk factors from the first admission day might be influencing later mortality in PICU [35]. Future model updates will need to address predicting outcome in these prolonged admissions.

Strengths of our study include that we externally validated all mortality prediction models simultaneously within a large cohort from the Dutch PICUs [19]. The staff was trained in guidelines on the different models before data collection started, and rules were strictly followed. Nevertheless in a large cohort study there are some limitations. Missing values are inevitable in observational studies, especially if they have to be performed within the first hour of contact with the patient. The differences between models in number of available values and values out of range (Fig. 1 and ESM Table 4) can be explained by the number of variables and collection period, on which topic some debate has taken place [26, 36].

The downside of directly comparing models with different exclusion criteria was the exclusion of part of the total PICU population that otherwise could have been included in one of the models (Fig. 1). Nevertheless, different exclusion policies had hardly any effect on ranking between the models in direct comparison. Details on data quality and model performance depending on different population selections are available in electronic supplementary material (ESM Table 5).

Another limitation of the study is the sample size. A sample with around 100 events (deaths) and 100 nonevents (survivors) is a minimum for sound validation studies [24, 37]. So far, only a few external validation studies in pediatric intensive care have been performed with such a sufficiently large sample size for the number of events [15–17, 29]. Even though our study cohort was rather large (12,040 admissions), some clinically interesting diagnostic groups, such as patients with central nervous infections or sepsis, could not be analyzed separately because of small numbers in mortality. Such detailed studies will remain a problem that is difficult to overcome by individual centers and smaller countries. Collaborations on international PICU registries would here be helpful.

Conclusion

We consider the freely available PIM2(-ANZ06) and proprietary PRISM3-24 the best prediction models for Western European PICU registries, because of their good overall calibration and discrimination, and good performance in previous European validation studies. Both models discriminated well overall and in almost all subgroup categories, including patients <28 days of age and recalibrated PRISM3-24 was the most stable predictor between subgroups. None of the tested models here can be recommended for mortality prediction in patients that stay longer than 6 days at the PICU. Because of different performance in subgroups, models should be applied with caution for risk adjustment in subpopulations.

Acknowledgments We like to thank Hester Lingsma and Yvonne Vergouwe from the Department of Medical Decision Making, Erasmus Medical Center, Rotterdam, The Netherlands, for R scripts to bootstraps and calibration plots, and Douwe R van der Heide and Leo Bakker, pediatric intensive care nurses for data quality control during local site visits.

References

1. Yeh TS, Pollack MM, Ruttimann UE, Holbrook PR, Fields AI (1984) Validation of a physiologic stability index for use in critically ill infants and children. Pediatr Res 18(5):445–451
2. Pollack MM, Ruttimann UE, Getson PR (1988) Pediatric risk of mortality (PRISM) score. Crit Care Med 16(11):1110–1116
3. Examples of several countries with PICU registries, Australia and New-Zealand: ANZPIC Registry http://www.anzics.com.au/anzpicr. Accessed 8 Feb 2013; Netherlands: Dutch PICE registry http://www.pice.nl; Portugal: REUNIR-CIP http://newdbserver.med.up.pt/web.care/reunir/index.html; Italy: TIPNet http://tipnet.cineca.it; United Kingdom: PICANet http://www.picanet.org.uk; United States of America: the Virtual PICU Systems LLC (VPS) https://portal.myvps.org/
4. Visser DY, Jansen NJ, Ijland MM, de Koning TJ, van Hasselt PM (2011) Intracranial bleeding due to vitamin K deficiency: advantages of using a pediatric intensive care registry. Intensive Care Med 37(6):1014–1020. doi:10.1007/s00134-011-2175-7
5. Peeters B, Jansen NJ, Bollen CW, van Vught AJ, van der Heide D, Albers MJ (2010) Off-hours admission and mortality in two pediatric intensive care units without 24-h in-house senior staff attendance. Intensive Care Med 36(11):1923–1927. doi: 10.1007/s00134-010-2020-4
6. Latour JM, van Goudoever JB, Duivenvoorden HJ, van Dam NA, Dullaart E, Albers MJ, Verlaat CW, van Vught EM, van Heerde M, Hazelzet JA (2009) Perceptions of parents on satisfaction with care in the pediatric intensive care unit: the EMPATHIC study. Intensive Care Med 35(6):1082–1089. doi: 10.1007/s00134-009-1491-7
7. Visser IHE, PICE-studygroup (2008) Dutch pediatric intensive care evaluation. PICE Report 2003–2005. ISBN 9789078022039. http://www.pice.nl/eng/gegevens-eng.htm. Accessed 10 Apr 2012
8. Gemke RJ, Bonsel GJ (1995) Comparative assessment of pediatric intensive care: a national multicenter study. Pediatric Intensive Care Assessment of Outcome (PICASSO) study group. Crit Care Med 23(2):238–245
9. Pollack MM, Patel KM, Ruttimann UE (1996) PRISM III: an updated pediatric risk of mortality score. Crit Care Med 24(5):743–752
10. Shann F, Pearson G, Slater A, Wilkinson K (1997) Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. Intensive Care Med 23(2):201–207
11. Slater A, Shann F, Pearson G (2003) PIM2: a revised version of the paediatric index of mortality. Intensive Care Med 29(2):278–285
12. PICANet (2007) Paediatric intensive care audit network national report 2004–2006 (published June 2007): Universities of Leeds and Leicester. ISBN 9780853162643. http://www.picanet.org.uk/documentation.html. Accessed 6 Oct 2011
13. Alexander J, Tregea S, Slater A (2008) Report of the Australian and New Zealand paediatric intensive care registry 2006. Australian and New Zealand Intensive Care Society (ANZICS). ISBN 1876980303. http://www.anzics.com.au/anzpicr. Accessed 8 Feb 2013
14. Alexander J, Tregea S, Slater A (2010) Report of the Australian and New Zealand paediatric intensive care registry 2008. Australian and New Zealand Intensive Care Society (ANZICS). ISBN 1876980699. http://www.anzics.com.au/anzpicr. Accessed 8 Feb 2013
15. 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(8):1407–1413. doi:10.1007/s00134-007-0694-z
16. 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(4):e733–e742
17. 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(5):447–454
18. Steyerberg EW (2009) Clinical prediction models. A practical approach to development, validation, and updating. Springer, New York
19. Collins GS, Moons KG (2012) Comparing risk prediction models. BMJ 344:e3186
20. Liddell FD (1984) Simple exact analysis of the standardised mortality ratio. J Epidemiol Community Health 38(1):85–88
21. Cox DR (1958) Two further applications of a model for binary regression. Biometrika 45(3/4):562–565
22. Hanley JAMB (1982) The meaning and use of the area under receiver operating characteristics (ROC) curve. Radiology 143:29–36
23. 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(2):271–277

24. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD (2005) Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. J Clin Epidemiol 58(5):475–483. doi: 10.1016/j.jclinepi.2004.06.017
25. Thukral A, Lodha R, Irshad M, Arora NK (2006) Performance of pediatric risk of mortality (PRISM), pediatric index of mortality (PIM), and PIM2 in a pediatric intensive care unit in a developing country. Pediatr Crit Care Med 7(4):356–361. doi: 10.1097/01.PCC.0000227105.20897.89
26. Thorburn K, Baines P (2002) Predicting mortality or recognizing death in meningococcal disease? Pediatric risk of mortality (PRISM) and pediatric index of mortality (PIM). Crit Care Med 30(7):1675–1676
27. Mildh L, Pettila V, Sairanen H, Rautiainen P (2007) Predictive value of paediatric risk of mortality score and risk adjustment for congenital heart surgery score after paediatric open-heart surgery. Interact Cardiovasc Thorac Surg 6(5):628–631. doi: 10.1510/icvts.2007.157917
28. Leteurtre S, Leclerc F, Martinot A, Cremer R, Fourier C, Sadik A, Grandbastien B (2001) Can generic scores (pediatric risk of mortality and pediatric index of mortality) replace specific scores in predicting the outcome of presumed meningococcal septic shock in children? Crit Care Med 29(6):1239–1246
29. Czaja AS, Scanlon MC, Kuhn EM, Jeffries HE (2010) Performance of the pediatric index of mortality 2 for pediatric cardiac surgery patients. Pediatr Crit Care Med. doi: 10.1097/PCC.0b013e3181e89694
30. Wang C, Xie G, Cheng B, Du L, Shi L, Tan L, Shu Q, Fang X (2009) Performance of pediatric risk of mortality, pediatric index of mortality and PIM2 in term Chinese neonates. J Trop Pediatr. doi: 10.1093/tropej/fmp124
31. Steyerberg EW, Borsboom GJ, van Houwelingen HC, Eijkemans MJ, Habbema JD (2004) Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. Stat Med 23(16):2567–2586. doi: 10.1002/sim.1844
32. Leteurtre S, Grandbastien B, Leclerc F, Parslow R (2012) International comparison of the performance of the paediatric index of mortality (PIM) 2 score in two national data sets. Intensive Care Med. doi: 10.1007/s00134-012-2580-6
33. Gorman M, Sachdeva R, Rice T, Wetzel R (2009) Support for PICUs and the PRISM scoring algorithms. Pediatr Crit Care Med 10(1):143–144. doi: 10.1097/PCC.0b013e3181937bf4
34. Tilford JM, Roberson PK, Lensing S, Fiser DH (1998) Differences in pediatric ICU mortality risk over time. Crit Care Med 26(10):1737–1743
35. Kramer AA, Zimmerman JE (2010) A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. BMC Med Inform Decis Mak 10:27. doi:10.1186/1472-6947-10-27
36. Randolph AG (1997) Paediatric index of mortality (PIM): do we need another paediatric mortality prediction score? Intensive Care Med 23(2):141–142
37. Steyerberg EW, Bleeker SE, Moll HA, Grobbee DE, Moons KG (2003) Internal and external validation of predictive models: a simulation study of bias and precision in small samples. J Clin Epidemiol 56(5):441–447 (pii: S0895435603000477)