External Validation of Mortality Prediction Models for the Dutch Pediatric Intensive Care Evaluation

Idse Visser MA MSc, on behalf of the **Dutch PICE study group***
* Pediatric Intensivists from: **AMC, VUmc, LUMC, UMCG, MUMC, UMCN, UMCU, ErasmusMC**

J.Hazelzet, C.Verlaat, M.Albers, D.v.Waardenburg, N.v.Dam, J.v.Woensel, R.Houmes, M.v.Heerde, K.Jansen: pediatric-intensivists from the Pediatric Intensive Care Units in the 8 University Hospitals in the Netherlands: Amsterdam, Groningen, Leiden, Maastricht, Nijmegen, Rotterdam, Utrecht. Emma Children's Hospital AMC; VU university medical center; Beatrix Children's Hospital, University Medical Center Groningen; Leiden University Medical Center; Maastricht University Medical Center; Radboud University Nijmegen Medical Center, Erasmus MC Sophia Children's Hospital; Utrecht MC Wilhelmina Children's Hospital.

OBJECTIVE

This study is part of the Dutch Pediatric Intensive Care Evaluation (PICE), the nationwide registry of admissions to the Pediatric Intensive Care Units (PICUs) in the Netherlands since 2003.
One of the aims of the PICE is to benchmark the units on PICU mortality. In this benchmark mortality prediction models are used to adjust for case-mix differences between PICU populations.
The objective of this study is to determine which general mortality prediction model is most appropriate for use in the pediatric intensive care setting in the Netherlands.

METHODS

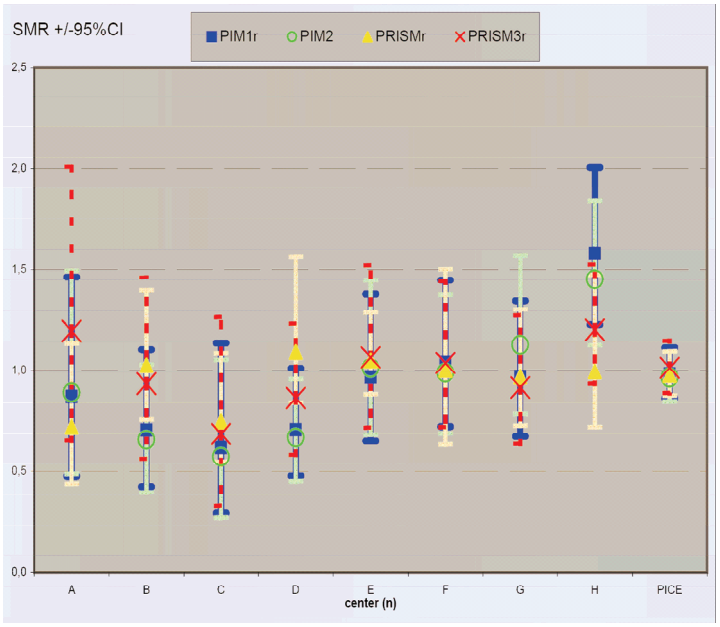
For the pediatric intensive care two validated general mortality prediction models exist with both updated versions: PIM1, PIM2, PRISM and PRISM3. The updated versions were implemented into the PICE registry in the beginning of 2006.
The models were examined in the Dutch PICE setting comparing them across units and (ANZPIC)diagnoses on overall performance using the standardized mortality ratio (SMR), discrimination by calculating the area under the ROC curve (AUC) and calibration by comparing observed-predicted ratio of deaths among deciles of risk with the Hosmer-Lemeshow goodness of fit test (H-statistic: $\chi 2$;8df) and by plotting observed mortality to expected mortality over ten intervals of risk.

RESULTS

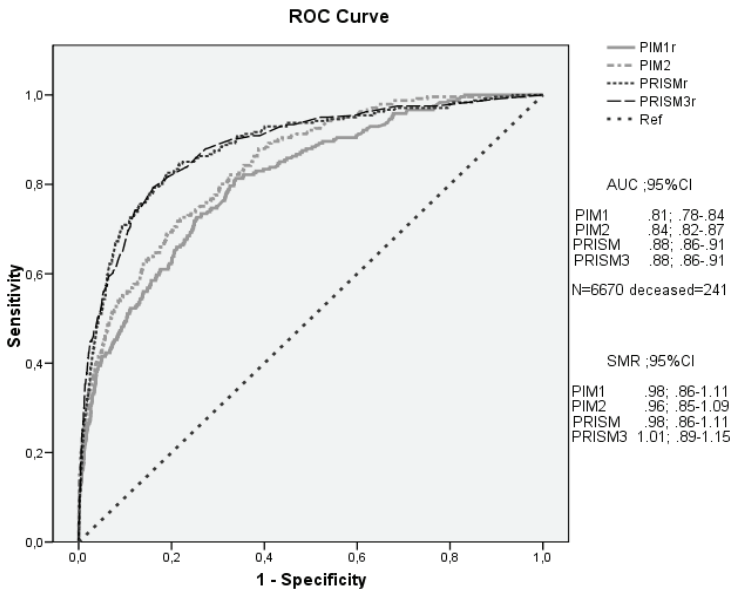
Validated admissions between March 2006 and August 2007 to all PICUs in the Netherlands were analyzed.
Adult patients (11) and patients who died within two hours on the PICU (10) or already dead on arrival (4) were excluded. A total of 6911 out of 6936 admissions were included for final analysis.
The discrimination of the models was good to excellent, with best discriminating powers for PRISM models: AUC=0.88 [95%CI: 0.86-0.91].
Only the original PRISM showed significant overprediction of mortality in all units: SMR < 1.
The PIM1 and PRISM were poorly calibrated for the Dutch setting: H-statistics $\chi 2 \geq 35$ ($p<0.001$). After simple recalibration of the models to the Dutch setting calibration was acceptable, H-statistics with $\chi 2 \leq 18$, and overall SMRs were not significantly different from one.

CONCLUSION

Except for the PIM2, the models needed recalibration to the Dutch setting. After simple recalibration overprediction of mortality decreased with good to excellent discriminating powers.
The PRISM was favored because of excellent discrimination, overall performance and ability for trend analyses including earlier years of the PICE registry. Next to PRISM, PIM2 was chosen because of acceptable calibration and discrimination and international comparison with other national PICU registries.



SMR by center according to models PIM2 and (recalibrated) PIM1, PRISM & PRISM3



Discrimination, Calibration & Overall performance of (recalibrated) models